Masters Program in **Geospatial Technologies**

DEEP LEARNING-BASED DOCUMENTATION OF CONFLICT DESTROYED BUILDINGS IN MYANMAR WITH VHR SATELLITE IMAGES

Nyi Nyi Nyan Lin

Dissertation submitted in partial fulfillment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School
Universitat Jaume I
Westfälische Wilhelms-Universität Münster

DEEP LEARNING-BASED DOCUMENTATION OF CONFLICT DESTROYED BUILDINGS IN MYANMAR WITH VHR SATELLITE IMAGES

Dissertation supervised by:

Prof. Dr. Roberto Henriques, PhD

NOVA Information Management School
Lisboa, Portugal

Co-supervised by:

Prof. Dr. Joaquín Huerta Guijarro, PhD

Universitat Jaume I
Castelló, Spain

Prof. Dr. Christian Knoth, PhD

IFGI, Universität Münster
Münster, Germany

February, 2024

STATEMENT OF ORIGINALITY

I declare that the work described in this document is my own and not from someone else. All the assistance I have received from other people is duly acknowledged and all the sources (published or unpublished) are referenced.

This work has not been previously evaluated or submitted to NOVA Information Management School or elsewhere.

Lisbon, February 26, 2024.

Nyi Nyi Nyan Lin

The signed original has been archived by the NOVA IMS services.

This thesis is dedicated to the brave and loving hearts in Myanmar and across the globe—those valiant souls who battle tirelessly for their freedom, the rights of others, and the pursuit of justice.

Acknowledgements

My heartfelt thanks are extended to my supervisors, Dr. Roberto Henriques, Dr. Christian Knoth, and Dr. Joaquín Huerta Guijarro, for their unwavering support and invaluable advice, which were instrumental in bringing this thesis to fruition. I am especially grateful to Dr. Marco Painho for his generosity, attentive guidance, and nurturing care throughout my Master's journey, significantly contributing to my thesis's development. I also wish to acknowledge NOVA Information Management School for providing the necessary resources, facilities, and an enriching academic atmosphere. Gratitude is extended to Planet's Education and Research Program for their support which contributed to this thesis work. I extend my heartfelt thanks to the faculty, educators, and administrative personnel at NOVA IMS and IFGI, WWU for their unwavering support. My gratitude also goes to Alexandre Neto, whose guidance in software development and implementation was invaluable for this thesis. I am also indebted to Thu Ya Aung and Nyan Htoo Thet, as well as my peers and companions in Myanmar, for their encouragement and inspiration throughout this journey. A special note of appreciation goes to my peers in the cohort, who have become cherished friends. Their camaraderie and mutual encouragement were pivotal in my successful completion of this Master's degree and in fostering a profound cultural exchange and professional experience.

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.1.18) (Lourenço, 2021).

Deep Learning-based Documentation of Conflict Destroyed Buildings in Myanmar with VHR Satellite Images

Abstract

In the wake of the 2021 military coup in Myanmar, the nation has been subjected to severe human rights abuses, prominently featuring arson attacks by the military on villages exhibiting resistance. The challenge of documenting these attacks at the individual building level is substantial, necessitating an approach that surpasses the limitations of conventional methods in terms of precision, scalability, and speed to support accountability and transitional justice. This thesis introduces an automated workflow that leverages bitemporal high and very high-resolution satellite imagery, combining Mask R-CNN for instance-based building footprint extraction with spatial overlap analysis and a Siamese Network for change detection. This methodology not only identifies destroyed structures with high accuracy but also provides an interface for potentially linking documented damage to individual victims through auxiliary data such as cadastral databases in future research endeavors. By providing a method that integrates advanced deep learning techniques for change detection at the building level, this work contributes a potential tool for destruction documentation by different entities and actors, enhancing the efficiency, granularity and accuracy of their documentation work. The implications of this work improving the potential for accountability and aiding transitional justice processes by proving an efficient mean for documenting human rights violations in conflict-affected regions.

Sustainable Development Goals (SDGs):

Keywords: Conflict Destruction, Change Detection, Deep Learning, Remote Sensing, Transitional Justice

Index of the Text

Index of Figures

Index of Tables

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
AOI	Area of Interest
AP	Average Precision
AUC	Area under Curve
CBAM	Convolutional Block Attention Module
CD	Change Detection
CNN	Convolutional Neural Network
CVA	Change Vector Analysis
DL	Deep Learning
DNN	Deep Neural Network
DSM	Digital Surface Map
DT	Decision Trees
EO	Earth Observation
F-RCNN	Faster R-CNN
FCN	Fully Convolutional Network
FIRMS	Fire Information for Resource Management System
FN	False Negative
FP	False Positive
FPN	Feature Pyramids Network
FPR	False Positive Rate
GAN	Generative Adversarial Network
GIS	Geographic Information System
GLCM	Gray Level Co-occurrence Matrix

GSD Ground Sampling Distance
GT Ground Truth

HOG Histogram of Oriented Gradients

IoU Intersect over Union
ISODATA Iterative Self-Organizing Data Analysis Technique Algorithm

LR Learning Rate
LSTM Long Short-Term Memory Network

MAD Multivariate Alteration Detection
MCC Matthews Correlation Coefficient
MIMU Myanmar Information Management Unit
ML Machine Learning

NDVI Normalized Difference Vegetation Index
NIR Near-infrared

OBIA Object-based Image Analysis

P Precision
PBIA Pixel-based Image Analysis
PCA Principal Component Analysis
PCode Place Code
PL Pléiades
PLN Pléiades Neo
PR Precision-Recall

R Recall
R-CNN Region-based Convolutional Neural Network
ReLU Rectified Linear Unit
ResNet Residual Network
RF Random Forests
RGB Red, Green and Blue
RNN Recurrent Neural Network
ROC Receiver Operating Characteristic
ROI Region of Interest
RoIAlign Region of Interest Align
RPN Region Proposal Network

RS Remote Sensing

SAR Synthetic Aperture Radar
SIFT Scale-Invariant Feature Transform
SN Siamese Network
SSD Single Shot Detector
SVM Support Vector Machine

TanH Hyperbolic Tangent
TN True Negative
TP True Positive
TPR True Positive Rate

UAV Unmanned Aerial Vehicle
UINT8 8-bit Unsigned Integer

VHR Very High Resolution

YOLO You Only Look Once

1

Chapter 1

1.1 Introduction

The ongoing conflict in Myanmar, triggered by the military coup in 2021, has led to widespread destruction across the nation, predominantly through arson attacks on towns and villages. These acts, often carried out by military forces as retaliation or collective punishment (Chin Human Rights Organization, 2022) against local populations in regions of strong opposition, have resulted in the burning of hundreds of villages and affected tens of thousands of households(Human Rights Watch, 2023; UN High Commissioner for Human Rights, 2023; UN Independent Investigative Mechanism for Myanmar, 2023). This situation necessitates a comprehensive assessment and documentation of the destruction for both humanitarian reasons and transitional justice efforts. Particularly for justice efforts, it is crucial to document evidence and link it to individual victims affected by such destruction.

Given the extensive magnitude of destruction, the high frequency of occurrences, and their spatial dispersion and temporal prolongation across the country, precisely mapping and documenting these events poses a significant challenge. Moreover, these events continue to occur, underscoring the need for an innovative and automated solution. However, few existing tools are suited to address this unique situation, prompting the exploration and development of an automated workflow or system. This system aims to document destruction using established technology and open-source software components, with consideration for transparency, adoptability, accessibility, and with acceptable accuracy. Additionally, it is designed with the flexibility to adapt to diverse situations and contexts, such as post-disaster assessment.

Due to the limitations of in-situ data collection, the use of satellite imagery becomes indispensable. Furthermore, the objective to document destruction at the building level, to establish links back to individual victims, necessitates the employment of very high-resolution satellite imagery. This thesis explores and reports on an automated workflow that incorporates instance-based segmentation and change detection stages, employing deep learning components. This approach not only aims to meet the urgent

need for efficient documentation of conflict-induced destruction but also contributes to the broader field of remote sensing and machine learning in conflict and disaster management contexts. Through this work, an adaptable and feasible methodology is sought to be offered that enhances the accuracy and efficiency of documentation efforts, potentially informing accountability measures and supporting transitional justice initiatives.

1.2 Research Objectives

The central aim of this thesis is to develop an automated workflow for the documentation of building destruction in conflict-affected areas, with a focus on Myanmar following the 2021 military coup. This section outlines the specific objectives that guide the research towards achieving this aim:

1. To evaluate the **operational feasibility, efficiency, and accuracy** of an automated system that utilizes bitemporal very high-resolution satellite imagery and deep learning components for building-level documentation of destruction.

2. To assess the **enhancements** the proposed system offers over traditional manual documentation methods, particularly in terms of documentation speed, scale, and accuracy.

3. To explore the **applicability and adaptability** of the system across different geographic locations, ensuring the method's utility in diverse environmental and architectural contexts.

1.2.1 Approach to Achieving the Research Objectives

The research employs a methodological approach combining quantitative and qualitative analyses to meet the outlined objectives. Advanced deep learning techniques, including Mask R-CNN for instance-based segmentation, overlap analysis and Siamese Network (SN) for patch similarity comparison, form the technological foundation of the proposed workflow. The study also considers the potential for privision of an interface to auxiliary data, such as cadastral databases, to link documented damage to specific victims, thereby enriching the data's utility for accountability and transitional justice efforts.

1.3 Research Questions

Building upon the defined objectives, the research seeks to address the following fundamental questions, aiming to uncover insights into the proposed system's practicality and efficacy:

1. **How does the automated workflow perform in real-world scenarios concerning accuracy, efficiency, and feasibility?**
2. **In what ways does the automated system improve upon existing manual documentation efforts in conflict-affected areas?**
3. **Can the proposed automated workflow be effectively adapted to document building destruction across various geographic and contextual environments?**

1.4 Thesis Organization

The thesis is organized into six chapters, commencing with a comprehensive literature review in Chapter 2. This chapter explores various methodologies employed in documenting destruction from conflicts, natural disasters, and man-made calamities, alongside a detailed theoretical exposition of key concepts such as Mask R-CNN and Siamese Network, and foundational elements. Chapter 3 shifts focus to the study area, delineating the Area of Interest (AOI) and describing the conflict event data, very high-resolution satellite imagery, and administrative and village identification codes used. It further outlines the core methodology, including building extraction using Mask R-CNN, Change Detection (CD) through Overlap Analysis and Siamese Network Similarity Comparison, and auxiliary methods for tiling, polygon merging, and data preprocessing, complemented by a discussion on the evaluation metrics for model and workflow performance assessment.

Chapter 4 presents an in-depth analysis and the results of the individual components of the study, such as Building Extraction, Change Detection, model training, and the comprehensive evaluation of the workflow. Building on these findings, Chapter 5 delves into discussions, critically addressing the research questions, explaining the limitations of the study, and proposing directions for future research. The thesis culminates in Chapter 6 with a succinct conclusion that encapsulates the essence of the research findings and the capacity of the proposed methodology for the automated documentation of conflict-induced destruction.

2

Chapter 2

2.1 Literature Review

Conflict damage assessments and mapping via Remote Sensing (RS) and Earth Observation (EO) data are fundamentally categorized into manual and automated approaches. Witmer (2015) categorized these methodologies as encompassing a broad spectrum of destruction and damage types, including structural, environmental, population/humanitarian impacts, and land-use/land-cover changes resulting from conflicts. Within the structural damage category, the spatial and temporal granularity varies significantly, ranging from individual buildings or dwellings and village-level assessments (Klonus et al., 2012; Knoth and Pebesma, 2017) to city block evaluations (Al-Khudhairy, Caravaggi, and Giada, 2005), employing uni-temporal, bi-temporal, and multi-temporal analyses. This literature review places a particular emphasis on building-level assessments within bi-temporal contexts. It's noteworthy that the majority of structural damage assessment techniques originate from the natural hazards and post-disaster damage assessment domain (Brunner, Lemoine, and Bruzzone, 2010; Tomowski et al., 2010; Yusuf, Matsuoka, and Yamazaki, 2001), as well as change detection studies (J. Jensen and Im, 2007; M. Liu et al., 2022; Marin, Bovolo, and Bruzzone, 2015) irrespective of the causative factors. Consequently, although this thesis and literature review primarily concentrate on conflict-induced damage assessment, methodologies and insights from these related domains are considered applicable and have been examined for their potential transferability to conflict damage assessment tasks.

Manual approaches (Danti, Branting, and Penacho, 2017; Parks, 2009; Voigt et al., 2011) traditionally rely on the visual inspection and annotation of high and very high-resolution optical images, supplemented by various forms of supporting data. This auxiliary information encompasses ground-based investigations, open-source data, crowd-sourced insights, and other forms of photographic, annotative, contextual, and qualitative data. The spatial scope of these assessments can vary from localized areas to expansive city-wide projects. Manual inspections are inherently labor-intensive and time-consuming, constraints that are manageable for small-scale assessments but

become impractical for extensive areas or in contexts with frequent conflict occurrences. To address these limitations, the field has evolved towards crowd-sourced mapping methodologies (Barrington et al., 2011; Ghosh et al., 2011; Meier, 2013; Poiani et al., 2016; Weir, McQuillan, and Francis, 2019; Yuan and R. Liu, 2018). These strategies involve partitioning the area of interest into smaller, manageable segments distributed among a network of collaborators or volunteers. The feasibility and efficiency of this approach have significantly improved with the advancements in web technologies post-2000. Recent studies (Fechner, Wilhelm, and Kray, 2015; Hu, Janowicz, and Couclelis, 2017) have focused on refining these manual and collaborative mapping processes to enhance workflow efficiency while ensuring the quality of the outputs. This evolution has fostered the growth of global networks of communities and volunteers dedicated to mapping and assessment initiatives across various collaborative platforms(Team, 2024; *Ushahidi* 2024). Concurrently, the rapid development of artificial intelligence, web technologies, and the availability of open-source data and satellite imagery with higher spatial and temporal resolutions have fortified manual assessment methods. Moreover, there has been a marked shift towards developing assistive technologies and automated tools designed to expedite manual approaches. Research into semi-automated or fully automated methodologies for assessing man-made and natural disaster damage has progressed, with some innovations (*Hotosm/fAIr* 2024; Knoth, Slimani, et al., 2018; *UNOSAT S-1 FloodAI Monitoring Dashboard* 2024) being integrated into manual mapping workflows. This integration signifies a pivotal advancement in the field, blending traditional assessment techniques with cutting-edge technologies to enhance the effectiveness and efficiency of responses to conflict, disaster, and humanitarian crises.

Automatic assessment and documentation approaches exhibit significant diversity regarding data sources, analyzed information, analysis methods, and outputs. Common data sources in such approaches (Brunner, Lemoine, and Bruzzone, 2010) include medium, high, and Very High Resolution (VHR) satellite imagery, Unmanned Aerial Vehicle (UAV)-based optical and multispectral images, and Synthetic Aperture Radar (SAR) data. Additionally, some methodologies (Qin, 2014; Vetrivel et al., 2018; Zhou, Gong, and Guo, 2016) incorporate Digital Surface Map (DSM), three-dimensional (3D) representations, and point cloud data, alongside open-source auxiliary data such as building annotations from OpenStreetMap. Typically, these approaches either rely on a single data source or combine multiple sources, such as optical imagery and SAR data. Certain methodologies (Adriano et al., 2021; H. Du et al., 2024) employ data fusion techniques, integrating various data sets at different stages of the analysis.

Prior to the advent of Deep Learning (DL) in this field, analytical efforts (Gamanya, De Maeyer, and De Dapper, 2009; F. Huang et al., 2018; Kemper et al., 2011; Klonus et al., 2012; Konstantinidis et al., 2017; Leichtle et al., 2017; Prins, 2008; Xiao et al., 2016) concentrated on deriving meaningful information from original data—such as spectral pixel values—to provide a contextual foundation for analysis. Derived

information includes pixel attributes like color, intensity, and brightness, as well as indices such as the Normalized Difference Vegetation Index (NDVI) and albedo for pixel-based analysis. Principal Component Analysis (PCA) facilitates dimensionality reduction, further yielding edges, lines, and corner features through image processing, filtering, and morphological operations. Additionally, more contextual information is derived, including texture and geometrical properties, exemplified by the Gray Level Co-occurrence Matrix (GLCM), and metrics of shape and size. Moreover, visual-based features common in the computer vision domain, such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG), are also utilized in remote sensing analysis.

However, with the increasing popularity of deep learning within the domain, the focus of input information has shifted back to utilizing raw forms of data such as optical, thermal, multispectral, SAR, and DSM inputs, complemented with annotated data like classification samples and change maps. This shift is due to deep learning models' ability to autonomously extract and learn from both the detailed contextual and hierarchical information present in the original input data. Deep learning techniques, through their advanced architectures, can identify intricate patterns and relationships within the data, enabling a more nuanced and comprehensive analysis of satellite and UAV-based imagery for automatic assessment and documentation purposes.

Beyond uni-temporal approaches, which primarily focus on the direct detection or classification of damaged structures, Jiang et al. (2022) identified that most methodologies for damage documentation and change detection comprise two critical stages: feature extraction and change analysis. After the extraction of the aforementioned features, various change analysis techniques are applied to these extracted features and information to identify damaged or destroyed structures. In classical and conventional methods, Pixel-based Image Analysis (PBIA) features are subjected to straightforward techniques such as image differencing and ratioing. Change is discerned through thresholding or the application of rule-sets designed to pinpoint and highlight altered or damaged structures. Alternatively, statistical and probabilistic methods, including Fuzzy, Change Vector Analysis (CVA) and Multivariate Alteration Detection (MAD), offer sophisticated means of analyzing change. These change analysis techniques (Benz et al., 2004; Pagot and Pesaresi, 2008; Xiao et al., 2016) are equally applicable to Object-based Image Analysis (OBIA) or superpixel-based approaches, which place a greater emphasis on contextual information.

Progressing beyond this foundational stage, Machine Learning (ML)-based methods are increasingly adopted (Brunner, Lemoine, and Bruzzone, 2010; Im and J. R. Jensen, 2005; Klonus et al., 2012; Moya et al., 2019), showcasing a significant overlap between classical and contemporary approaches. These advanced techniques leverage both traditional feature analysis and modern computational algorithms to enhance the accuracy and efficiency of detecting and documenting changes and damages within structures. In this paradigm, features and information extracted using conventional

techniques are subsequently integrated into machine learning-based classification and change analysis phases. This integration employs both supervised methods, such as Support Vector Machine (SVM) and Random Forests (RF), and unsupervised techniques, including Decision Trees (DT) and Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA), to categorize and cluster pixels, superpixels, and objects. These classifications discern between intact structures and those compromised by damage, leveraging their shared characteristics for accurate labeling. An alternative strategy involves independently classifying each image into specific structural and land cover categories, followed by a comparative analysis across classifications to pinpoint changes. The spatial resolution of these methodologies extends from discrete components of individual structures and singular buildings to entire blocks, enhancing the granularity and detail of change detection efforts.

As deep learning methods have advanced and gained widespread adoption across various domains, the focus of research in conflict damage documentation has also shifted towards deep learning-based approaches, leading to the emergence of innovative solutions (Adriano et al., 2021; Daudt et al., 2018; Ghaffarian et al., 2019; Ma et al., 2020; Valentijn et al., 2020). Among these, Convolutional Neural Network (CNN) are pivotal, given their prevalent application in processing image data, a practice extensively employed in the field of remote sensing for analyzing satellite imagery. Owing to their architecture, CNNs excel at adaptively learning spatial hierarchies of features from input images, making them particularly effective for tasks such as semantic segmentation, e.g., U-Net (Ronneberger, Fischer, and Brox, 2015), and classification, applicable to both pixels and superpixels.

Moreover, CNNs are integrated with one-stage, for example You Only Look Once (YOLO)(Redmon et al., 2016) and Single Shot Detector (SSD)(W. Liu et al., 2016)), and two-stage detectors, for example Region-based Convolutional Neural Network (R-CNN) and Faster R-CNN (F-RCNN), for object detection tasks (Y. Li et al., 2019). These models are adept at identifying and categorizing objects within images, producing bounding boxes and class labels, which proves invaluable for detecting or enumerating specific entities like buildings within a study area. Enhanced versions of these architectures, such as Mask R-CNN, extend this functionality by segmenting target objects (Wen et al., 2019), thereby providing both masks and classes in addition to bounding boxes. These models can be configured and trained to directly identify damaged structures, often employing uni-temporal data analysis. Given that conflict and disaster events frequently involve multi-temporal datasets, Recurrent Neural Network (RNN) and their variants, such as Long Short-Term Memory Network (LSTM), have been adopted (Bai et al., 2022; H. Chen, Wu, B. Du, L. Zhang, and L. Wang, 2020; Jiang et al., 2022) for their proficient handling of sequential data. This makes them suitable for tasks focusing on temporal changes and developments. Similarly, Generative Adversarial Network (GAN) have made significant contributions (H. Chen, W. Li, and Shi, 2022; Saha, Bovolo, and Bruzzone, 2021) to the field. GANs excel at generating augmented images, addressing

the challenge of training data scarcity in destruction damage scenarios, and demonstrating remarkable capabilities in domain adaptation. This includes applications in super-resolution and spatial refinement of conflict event imagery, serving as assistive components. GANs can also be tailored to directly detect changes and damages by training the network to generate difference images between two temporal states.

Siamese Networks represent another critical deep learning strategy, functioning as both feature extractors and comparators. They can discern similarities and dissimilarities between two inputs, showcasing their relevance for change detection between pre-event and post-event images. The integration of these approaches (Bandara and Patel, 2022; H. Chen, Wu, B. Du, and L. Zhang, 2019; J. Chen et al., 2021; Ding et al., 2021; Y. Liu et al., 2021; M. Wang et al., 2020; H. Zheng et al., 2022; Z. Zheng et al., 2022), along with adaptations and modifications like Feature Pyramids Network (FPN) (for multi-scale analysis), Residual Network (ResNet), attention mechanisms (to emphasize relevant features), and Transformer models (for enhancing feature adaptability and facilitating data fusion), at various stages of an architecture, significantly boosts the effectiveness and precision of automated change detection and damage documentation models. This amalgamation of advanced techniques and architectures underscores the dynamic evolution of deep learning in transforming the capabilities of remote sensing for comprehensive and accurate assessments of conflict and disaster-induced damages.

Despite the abundance of recent research into change detection and destruction documentation, particularly within the natural disaster domain, few studies have been operationalized and integrated into practical workflows and platforms. Additionally, many studies focus on specific localities, data sources, and circumstances, which limits their generalizability and applicability. There is a growing demand for a universal, generalizable, and operational method that can be seamlessly integrated into existing processes. An exemplary initiative in this direction is the *Hotosm/fAIr* (2024) Artificial Intelligence (AI)assisted mapping tool developed by Humanitarian OpenStreetMap Team. This tool represents a significant advancement by incorporating AI-based automated tools within the framework of manual collaborative mapping approaches, showcasing the potential for enhancing the efficiency and effectiveness of mapping efforts through the integration of AI technologies.

In summary, the field of conflict damage assessment has transitioned from manual inspections and traditional methods to automated deep learning techniques using satellite and UAV imagery. Yet, the integration of these innovations into practical settings is still scarce. These methodologies often aim to develop new techniques, evaluated against standard datasets and benchmarks, or focus intensely on specific events or regions, constrained by particular data and conditions. Notably, ML and DL methods rely on samples of damaged buildings, a factor limited by data scarcity. This thesis introduces an alternative strategy, emphasizing the identification of intact buildings and conducting bi-temporal change analysis. This approach circumvents the challenge of insufficient specific data on buildings affected by conflict, a significant

barrier for adoption of ML and DL-based approaches due to the diverse nature of destruction and geographic variances, which limit the diversity and availability of training data.

2.2 Theoretical Background

This section explains the theoretical and conceptual underpinnings of various components that form the bedrock or critical aspects of the methodology and workflow proposed in Chapter 3. Central to the proposed methodology are the Mask R-CNN for building extraction and the Siamese Network for the similarity comparison of detected buildings in pre- and post-event scenarios for the purpose of change detection. These deep learning components draw on foundational concepts such as artificial neural networks, deep neural networks, and convolutional neural networks, further enhanced by variations like fully convolutional networks, region-based convolutional neural networks, region proposal networks, residual networks, and feature pyramid networks, among others. Noteworthy enhancements such as the region of interest (ROI) pooling and ROI align in Mask R-CNN, along with the convolutional block attention module in the Siamese Network, significantly boost the accuracy and efficacy of their respective networks. Additionally, this section delves into complementary components integral to the workflow, including the use of directed graphs for constructing a hierarchical structure of extracted building polygons for polygon merging, and R-tree spatial indexing to expedite and optimize the process of identifying overlapping polygons in a one-to-many setting for polygon merging and overlap analysis.

2.2.1 Artificial Neural Network

Artificial Neural Network (ANN) are sophisticated computational frameworks that mimic the neurological structure of the human brain to process complex data patterns. These networks comprise interconnected layers of nodes, akin to neurons, that transmit and transform input data through a series of weighted connections, emulating synaptic functionality(Uhrig, 1995). ANNs are distinguished by their ability to learn from data iteratively, adjusting the synaptic weights through algorithms like backpropagation. This learning process enables ANNs to perform a myriad of tasks, ranging from pattern recognition to predictive modeling, by generalizing from the input data without explicit programming for specific tasks. Their versatility and adaptability have made ANNs foundational in advancing the field of ML and AI.

2.2.2 Deep Neural Network

Deep Neural Network (DNN) extend the concept of Artificial Neural Networks by incorporating multiple hidden layers between the input and output layers, as contrasted in Figure 1, enabling the modeling of complex data with high levels of abstraction.

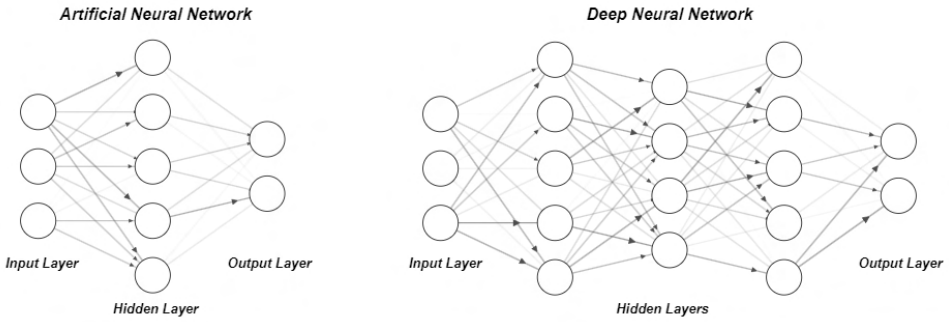

Figure 1: Architecture of artificial and deep neural networks with contrasting layers of hidden neuron.

DNNs are characterized by their deep architecture(Simon J.D. Prince, 2023), which allows them to learn hierarchical representations of data, making them particularly effective in fields such as image and speech recognition, natural language processing, and more. This depth facilitates the extraction of intricate patterns and features that simpler models might overlook, significantly enhancing predictive and analytical capabilities.

2.2.3 Activation Functions

Activation functions are mathematical equations that determine the output of a neural network node or neuron, given an input or set of inputs. They are fundamental to the ability of neural networks to model complex nonlinear relationships that exist between features in datasets. Activation functions introduce non-linearity into the network, enabling it to learn and perform more complex tasks than what would be possible with linear functions alone(Simon J.D. Prince, 2023). Activation natures of each functions are illustrated in Figure 2. Commonly used activation functions are Rectified Linear Unit (ReLU), Sigmoid, Hyperbolic Tangent (TanH), Leaky ReLU, Softmax and Swish. ReLU keeps positive values unchanged while setting negative values to zero, improving training efficiency and reducing the risk of vanishing gradients. ReLU is defined as in Equation 2.1. The sigmoid function, defined in Equation 2.2, outputs values between 0 and 1. It is particularly useful for binary classification tasks. Tanh extends the output range of sigmoid to between -1 and 1. This function, as defined in Equation 2.3, is commonly used in hidden layers. Leaky ReLU allows a small gradient when inactive to prevent neurons from dying as defined in Equation 2.4 where α is a small constant. Softmax is used in the output layer for multi-class classification, converting logits to probabilities as defined in Equation 2.5 for each input x_i in a vector of length N, making the output sum to 1. Swish is a self-gated function as defined in Equation 2.6 where β is a learnable parameter or fixed constant, and σ represents the sigmoid function.

$$f(x) = \max(0, x) \tag{2.1}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.3}$$

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{otherwise} \end{cases} \tag{2.4}$$

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{2.5}$$

$$f(x) = x \cdot \sigma(\beta x) \tag{2.6}$$

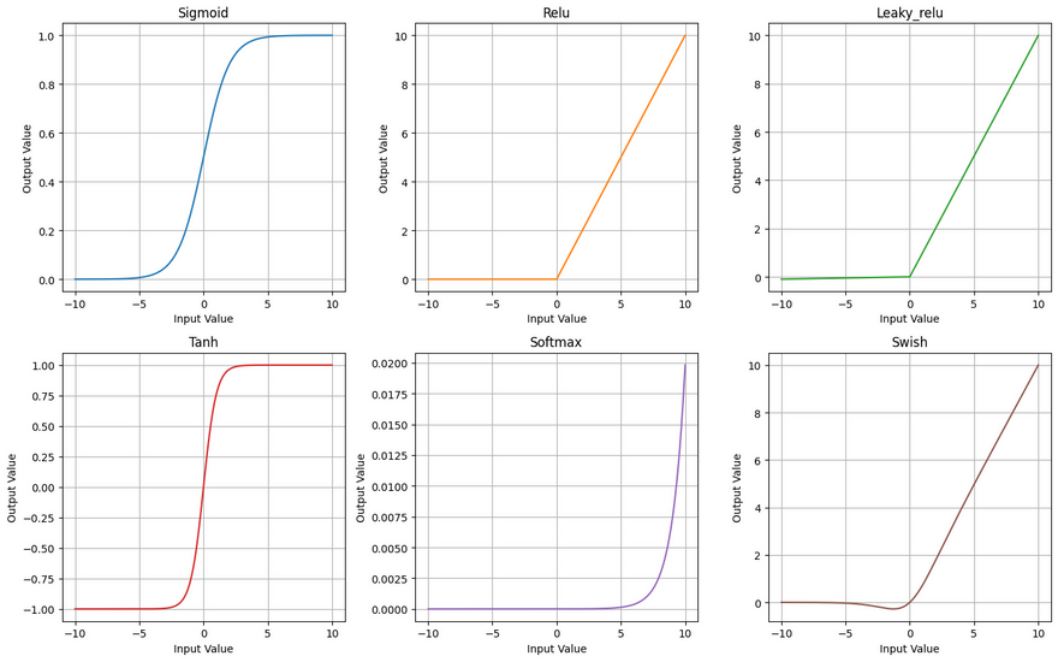

Figure 2: Activation or output characteristics of different activation functions. (Balkrishna Pandey, 2023)

2.2.4 Convolutional Neural Network

CNN are a class of deep neural networks, primarily used in processing data that has a grid-like topology, such as images. CNNs utilize a mathematical operation called convolution in at least one of their layers. They are particularly known for their ability to capture spatial and temporal dependencies in an image through the application of relevant filters, enabling them to efficiently recognize visual patterns directly from pixel images with minimal preprocessing. This capability makes CNNs extensively

applicable in image and video recognition, image classification, medical image analysis, and natural language processing tasks.

2.2.5 Fully Convolutional Networks

Fully Convolutional Network (FCN) represent a pivotal advancement in deep learning, particularly in the realm of semantic segmentation tasks. Originating from the adaptation of conventional convolutional neural networks (CNNs), FCNs have been innovatively restructured to enable spatially dense prediction tasks without the necessity for any fully connected layers. This modification permits an FCN to accept input images of arbitrary size and produce correspondingly sized output maps that categorize each pixel, making them exceptionally suitable for image segmentation. FCNs are generally composed of downsampling and upsampling paths.

Downsampling Path In the downsampling path, a series of convolutional and pooling layers extract and condense feature information from the input image, effectively reducing its spatial dimensions while increasing the depth of the feature maps. This process mirrors that of traditional CNNs, capturing the essential features required for understanding the content of the input image.

Upsampling Path The upsampling path reversely enlarges the feature maps to the original input image size using deconvolutional (transpose convolutional) layers. This procedure is crucial for mapping the learned dense feature representations back to the original pixel space, facilitating precise pixel-wise classification.

Skip Connections A significant innovation in FCNs is the integration of skip connections, which concatenate feature maps from the downsampling path with those from the upsampling path. These connections help to recover the spatial resolution lost during downsampling, ensuring that fine details necessary for accurate segmentation are preserved.

2.2.6 Region-based Convolutional Neural Network

Girshick et al. (2014) introduced R-CNN which is an advanced approach in object detection technology. They cleverly combine region proposal methods with the deep learning strength of CNNs. The process starts by identifying potential areas in an image where objects are likely found. Then, it uses CNNs to classify these regions into different object categories and fine-tunes their borders. This method significantly improves the precision of detecting and classifying objects in images, making R-CNNs a cornerstone for further advancements in image recognition tasks.

2.2.7 Faster R-CNN

F-RCNN (Ren et al., 2016) improves upon the earlier R-CNN models by integrating the region proposal step directly into the network, making it faster and more efficient. Unlike its predecessors that handle region proposals separately, Faster R-CNN uses a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus speeding up the process. This unified model significantly enhances object detection tasks by reducing the time taken to identify potential object locations while maintaining high accuracy in detecting and classifying objects.

2.2.8 Region Proposal Network

The Region Proposal Network (RPN) is introduced together with Faster R-CNN and designed for generating object proposals efficiently within the network. It scans the feature map obtained from the input image with a sliding window and outputs object bounds and objectness scores for each location. RPN enables the model to identify object locations and sizes at various scales and aspect ratios, significantly enhancing the efficiency and accuracy of the object detection process by integrating proposal generation and object detection into a single network.

2.2.9 Mask R-CNN

Mask R-CNN (He, Gkioxari, et al., 2018) is an extension of Faster R-CNN that adds a branch for predicting segmentation masks on each Region of Interest (ROI), in parallel with the existing branch for classification and bounding box regression. This framework allows for the precise localization of objects in an image while simultaneously generating a high-resolution object mask, making it highly effective for tasks requiring detailed object segmentation, such as instance segmentation. Mask R-CNN maintains a balance between speed and accuracy, making it a preferred choice for many computer vision applications.

2.2.10 Region of Interest Pooling

Region of Interest Pooling (ROI Pool) is a technique used in object detection tasks to extract a fixed-size feature vector from an arbitrary-sized ROI. It divides the ROI into a grid of cells and applies max pooling to each cell, ensuring the output feature map has a consistent size regardless of the ROI's dimensions. This process allows the subsequent layers of a neural network to operate on inputs of uniform size, facilitating the detection and classification of objects across varying scales and positions within the image.

2.2.11 Region of Interest Align

Region of Interest Align (RoIAlign) is a technique in Mask R-CNN that accurately extracts feature maps from ROIs, avoiding the misalignments caused by the quantization in ROI Pooling. Unlike ROI Pooling, which approximates the spatial locations of features, ROIAlign uses bilinear interpolation to compute the exact values of input features at four regularly sampled locations in each RoI bin, and then aggregates the results. This leads to more precise object detection and segmentation, particularly for tasks requiring fine-grained spatial understanding.

2.2.12 Feature Pyramid Network

The FPN is a backbone architecture introduced by Lin et al. (2017) and designed to improve the segmentation and detection performance by efficiently creating a pyramid of features at multiple scales. FPN enhances the network's ability to detect objects at various sizes by utilizing a top-down architecture with lateral connections, allowing for the integration of high-level semantic information from deep layers with high-resolution information from shallow layers. This structure significantly boosts the performance of Mask R-CNN in tasks requiring precise object localization and segmentation.

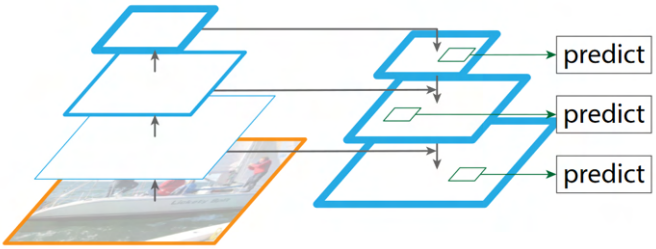

Figure 3: A feature pyramid with predictions made independently on feature maps at all spatial scale levels encompassing different levels of contextual and semantic features.

2.2.13 Residual Network

Residual Networks (ResNets), such as ResNet50 and ResNet101, are deep learning models introduced in He, X. Zhang, et al. (2015) and designed to address the vanishing gradient problem in very deep networks. They introduce "skip connections" allowing the network to learn residual functions with reference to the layer inputs. ResNet50 and ResNet101 denote variants with 50 and 101 layers, respectively. These architectures enable training of substantially deeper networks by facilitating the flow of gradients during the learning process, significantly improving performance on tasks like image classification and recognition.

2.2.14 Siamese Network

A Siamese Network is a neural network architecture that contains two or more identical subnetworks designed to generate feature representations of different inputs. The idea of Siamese Network is first introduced by Bromley et al. (1993). These subnetworks share the same parameters and weights. Siamese Networks are particularly useful in applications requiring a measure of similarity between inputs, such as face recognition or signature verification. They are trained on pairs of inputs to learn which features to compare and how to differentiate between them, effectively learning an embedding space where similar inputs are brought closer and dissimilar ones are pushed apart.

2.2.15 Contrastive Loss

Contrastive Loss is a function, defined in Equation 2.7, used in Siamese Networks to learn embeddings by bringing closer the representations of similar pairs while pushing apart those of dissimilar pairs. It penalizes embeddings based on their similarity, effectively training the model to distinguish between pairs of inputs. This loss function plays a crucial role in tasks like face verification and anomaly detection, where the model needs to understand and quantify the similarity between inputs.

$$L(x_i, x_j, y) = (1 - y)\frac{1}{2}(D)^2 + y\frac{1}{2}\max(0, m - D)^2 \tag{2.7}$$

where D is the distance between the embeddings of x_i and x_j, often computed using the Euclidean distance, m is the margin, a hyperparameter that defines how far apart the dissimilar points should be pushed. It acts as a threshold for dissimilar pairs. L is the contrastive loss for the pair (x_i, x_j).

2.2.16 Euclidean Distance

In Siamese Networks, Euclidean Distance is used to measure the similarity between embeddings. This distance metric calculates the root of the sum of the squared differences between the components of two vectors as described in Equation 2.8. By applying Euclidean Distance to embeddings generated by the network for a pair of inputs, the network can quantify how similar or dissimilar those inputs are. A smaller distance indicates high similarity, whereas a larger distance suggests dissimilarity. This approach is fundamental in tasks like verification, where the goal is to determine whether two inputs are the same or different.

$$D(p, q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \tag{2.8}$$

2.2.17 Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) introduced by Woo et al. (2018) is an attention mechanism for Convolutional Neural Networks (CNNs) that sequentially infers attention maps along two dimensions: channel and spatial. CBAM enhances the representational power of CNNs by focusing on informative features and suppressing less useful ones. It does this first by channel attention, determining the importance of each channel, followed by spatial attention, highlighting salient regions in the spatial domain as visualized in Figure 4. This module can be easily integrated into existing CNN architectures, significantly improving performance on various visual tasks with minimal computational overhead.

Channel attention in neural networks focuses on identifying the most informative feature channels within a given input. It dynamically weights the contribution of each channel based on its relevance to the task at hand. By doing so, channel attention mechanisms can enhance the model's ability to prioritize features that significantly impact the output, thereby improving the network's overall performance and efficiency in processing complex data.

Spatial attention in neural networks refers to the mechanism that allows the model to focus on specific regions of the input data, emphasizing areas that are more relevant for the task. This approach mimics human visual attention, where not all parts of a scene are processed equally, but focus is directed towards parts of interest. In the context of CNNs, spatial attention enhances feature representation by modulating the importance of different spatial locations within the input, leading to improved performance in tasks requiring detailed visual understanding.

2.2.18 Directed Graphs

Directed graphs, or digraphs, are fundamental structures in graph theory used to model entities and their directed relationships. A directed graph G is defined as an ordered pair $G = (V, E)$ where:

- V represents a set of vertices or nodes,
- $E \subseteq \{(u, v) | u, v \in V\}$ represents a set of directed edges, with each edge being an ordered pair (u, v) indicating direction from node u to node v.

The key feature distinguishing directed graphs from undirected graphs is the orientation of edges, which implies directionality in the relationships modeled by the graph. This directionality is crucial for applications involving asymmetric relationships, such as web link structures, citation networks, and various hierarchical or flow-oriented models. Directed graphs are widely used for modeling data structures, network flows, dependencies among tasks, and more.

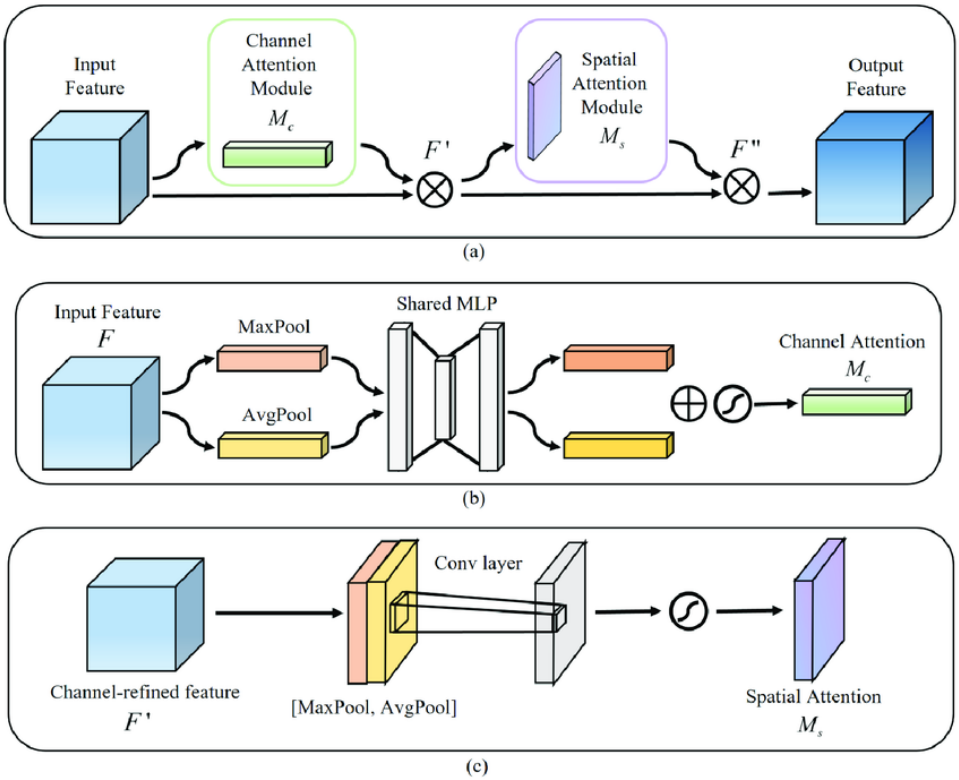

Figure 4: Architecture of (a) Convolutional Block Attention Module (CBAM) with (b) channel and (c) spatial attention modules. Channel attention puts emphasis on feature channels and spatial attention amplifies emphasis on spatial locations which correspond to objects of focus. (Woo et al., 2018)

2.2.19 R-Tree: Spatial Indexing Structure

An R-Tree, is a balanced tree data structure optimized for indexing spatial data, such as geographical coordinates and polygons(Guttman, 1984). It organizes spatial objects using hierarchical, nested minimum bounding rectangles (MBRs), enabling efficient search, insertion, and deletion operations. The structure of an R-Tree allows for variable node capacities, with each node containing entries that either point to child nodes (non-leaf nodes) or directly to spatial objects (leaf nodes). This hierarchical division of space into MBRs facilitates quick spatial queries by significantly reducing the search space through bounding rectangle overlap checks.

RTrees support various operations including spatial querying, where the tree is traversed to find objects within a specific area, and dynamic dataset modifications through object insertion and deletion, maintaining tree balance to optimize query performance. Due to their efficiency in handling spatial data, RTrees are widely applied in Geographic Information System (GIS), spatial databases, and for indexing multi-dimensional information in diverse applications requiring rapid spatial querying and data retrieval.

3

Chapter 3

3.1 Data and Methodology

The study area of this thesis is Myanmar, chosen for its ongoing conflict, extensive reporting and documentation of conflict events by various entities, the author's local and background knowledge, and the considerable availability of very high-resolution satellite images. The conflict and arson events, exacerbated nationwide since the military coup in 2021, as reported by Myanmar Witness(Myanmar Witness, 2024), are predominantly concentrated in the Sagaing, Magway, and Mandalay regions, along with the Chin, Karenni, and Kachin states. Among these potential study locations, Kachin State, Magway, and Sagaing regions were selected based on the reporting coverage and the availability of pre- and post-event satellite imagery.

The pivotal information for this thesis encompasses conflict and arson attack event details, specifically the locations and timings of such incidents. This crucial data is derived from Myanmar Witness's reports, which aggregate and cross-reference information from diverse sources including news outlets(Democratic Voice of Burma, 2024; Khit Thit Media, 2024; Mizzima, 2024; Myanmar Now, 2024; The Irrawaddy, 2024), citizen journalism, social media platforms, and NASA's Fire Information for Resource Management System (FIRMS) fire data(NASA, 2024), alongside optical and thermal satellite images. Approximately a hundred reported events in the study areas were narrowed down to 20 areas of interest or villages, selected based on the availability of very high-resolution satellite images, the quality of images, and spatial coverage for each area as demonstrated in Figure 5. To validate the reported locations and timings of fire/arson events, PlanetScope (Planet Team, 2017) images were visually inspected for their high temporal resolution (daily coverage) and high spatial resolution (3m). Figure 6 illustrating the pre-, peri-, and post-event areas captured by PlanetScope imagery demonstrate this verification process. Subsequent to this verification, very high-resolution satellite images, 30cm and 50cm Ground Sampling Distance (GSD), from Pléiades (PL) and Pléiades Neo (PLN) satellites were acquired. Pléiades images,

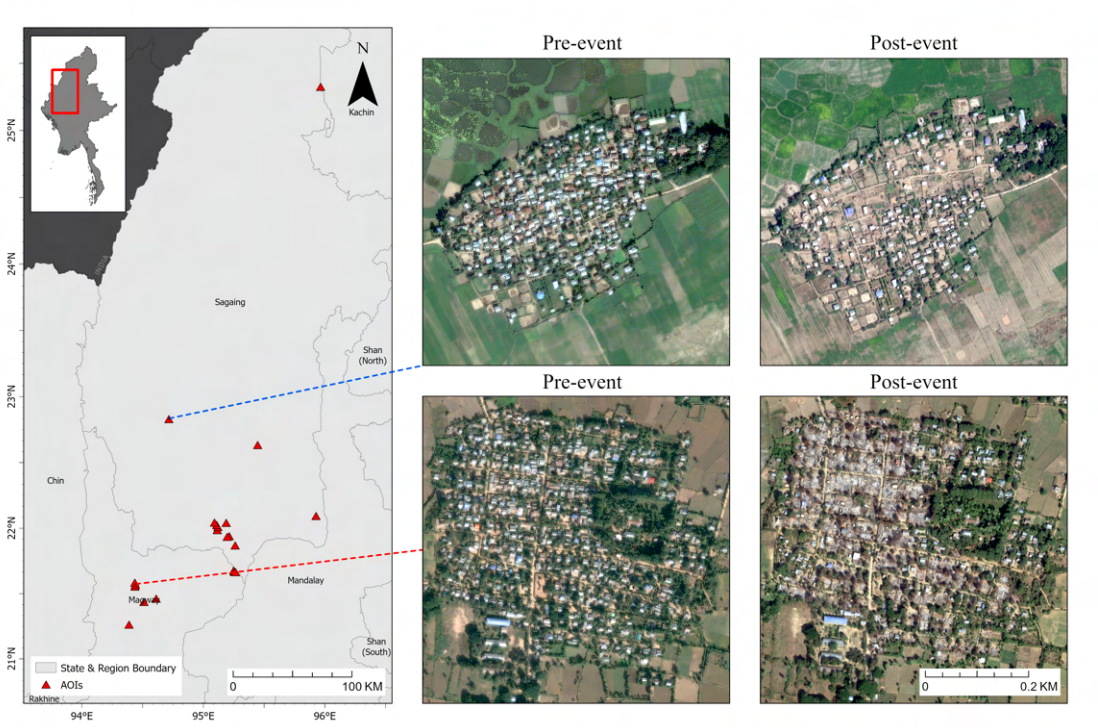

Figure 5: Areas of interest (AOIs) in Northwestern Myanmar (Sagaing, Magway and Kachin) that are involved in this thesis work. Callouts present visible conflict induced changes in residential areas.

Figure 6: (A) Pre-, (b) post- and (c) peri-events captured by PlanetScope imagery due to its high (daily) temporal resolution. This characteristic contributes to the conflict event time, location and extent verification process.

pansharpened reflectance products, and Pléiades NEO images, analytic level products, were prepared for analysis using ArcGIS Pro software for pansharpening and preprocessing. Although these products contain Red, Green and Blue (RGB) and Near-infrared (NIR) bands, only RGB data was extracted for the analysis. Furthermore, pre- and post-event images were registered and aligned using the Similarity transformation method in ArcGIS Pro software, rendering the data ready for subsequent analysis stages.

Auxiliary information includes systematic and official data regarding the areas of interest or villages. The Myanmar Information Management Unit (MIMU)(*Myanmar Information Management Unit* 2024) compiles and shares administrative geospatial information widely recognized and utilized in Myanmar. MIMU assigns each registered village or area of interest a unique identification code known as a Place Code (PCode)(*Place Codes (Pcodes) | MIMU* 2024), along with information on hierarchical administrative regions. This PCode serves both as a unique identifier for analysis areas and as a means to link analysis results with other spatial and administrative data. The study areas and data sources utilized are detailed in Table A1 and Table A2 in the Appendix A.

3.2 Methodology

3.2.1 Overview

The methodology outlined in this thesis, as demonstrated in Figure 7, presents a systematic approach for automating the documentation of buildings impacted by conflict, utilizing satellite imagery enhanced by moderately advanced deep learning techniques. The primary challenge addressed by this methodology is the scarcity of specific data on buildings damaged in conflicts, which significantly hampers the direct application of machine learning models for identifying such structures. This scarcity is further complicated by the diverse nature of destruction caused by both man-made and natural disasters, as well as the variance in geographic locations of these events, leading to a pronounced limitation in the availability and diversity of training data.

To counter these issues, the research strategy focuses on the identification of intact buildings—a domain where abundant data exists—thus facilitating the training of the Mask R-CNN model. Mask R-CNN, recognized for its robustness in segmenting and identifying distinct objects within a wide array of imaging data, is specifically adapted here for the recognition of undamaged buildings in both pre- and post-conflict satellite imagery. This approach not only enables detailed documentation of structural integrity but also lays the groundwork for potentially connecting this information to individuals affected by the conflict in subsequent analyses, thereby enhancing the data's utility for accountability and justice efforts.

The differential in temporal resolution, as seen in Figure 8, posed by the very high spatial resolution satellite images introduces an additional layer of complexity, making it challenging to distinguish between alterations caused by the conflict and those resulting from unrelated man-made or natural changes. This dilemma is exemplified in instances where temporary shelters are constructed on the sites of previously destroyed buildings, potentially leading to misclassification of these structures as original, undamaged buildings using basic spatial overlap analysis which is demonstrated in Figure 9.

To augment the accuracy of change detection, the methodology incorporates an

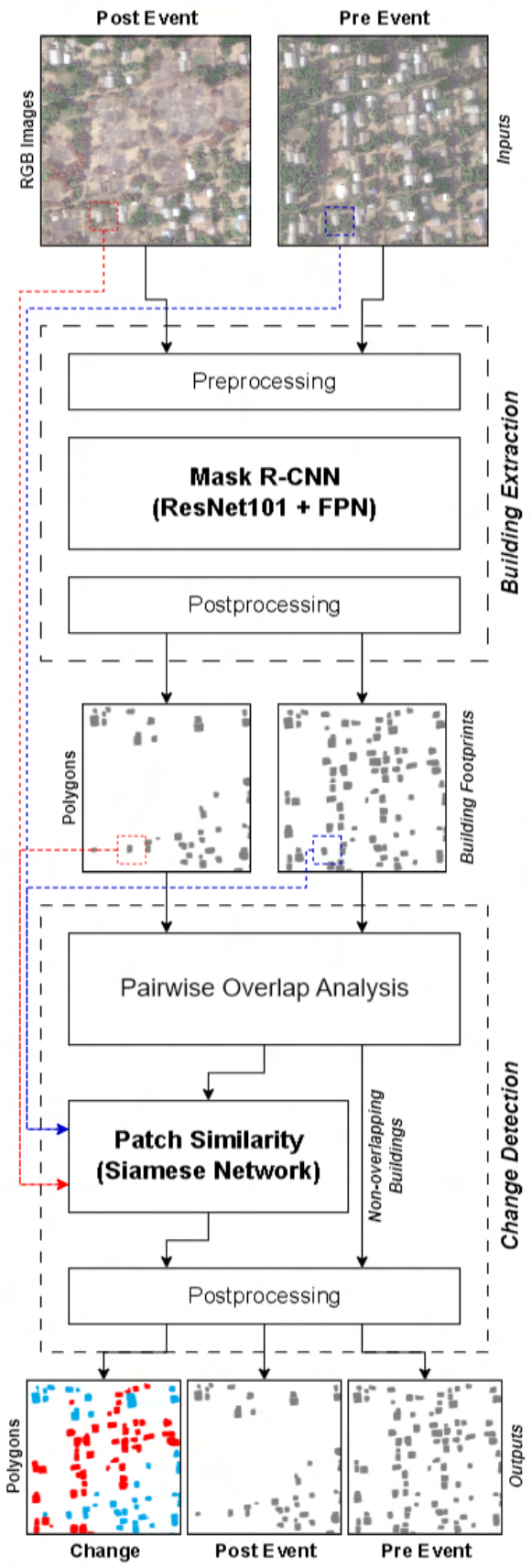

Figure 7: The proposed workflow composed of two major steps: building extraction and change detection. Mask R-CNN with ResNet101 and FPN as backbone is applied to extract buildings from bi-temporal RGB images. Overlap analysis of extracted building polygons determines presence or disappearance of buildings post-event while Siamese Network patch similarity component examine the change status of buildings.

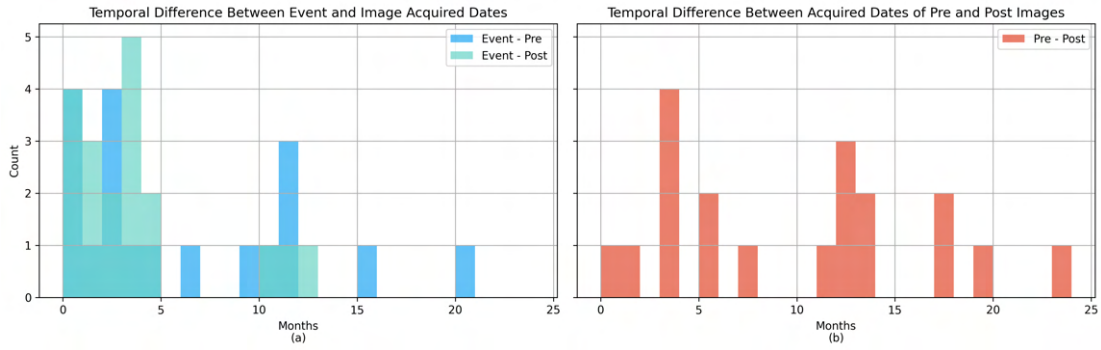

Figure 8: Difference between (a) event and pre-/post-event image acquired dates and (b) pre- and post-event image acquired dates.

Figure 9: Rebuilt structures and temporary shelters in destructed locations post-event. Temporal window between pre- and post-event images introduce building changes which renders sole reliance on overlap analysis unreliable.

additional phase of similarity comparison utilizing a Siamese Network, following the preliminary spatial overlap analysis. The Siamese Network, with its proven efficacy in various comparative analyses, including image similarity, object tracking, and face recognition, serves as an ideal tool for assessing the agreement between building instances before and after a conflict event. Through the pairwise comparison of image patches, the network assigns a similarity score that effectively quantifies the status of change. In essence, this integrated workflow combines the capability of Mask R-CNN for intact building detection with the nuanced change analysis provided by spatial overlap analysis and Siamese Network-based similarity comparison. This multifaceted approach adeptly addresses the challenges of limited data availability and the intricacies of temporal resolution, striving to produce a high-quality, reliable output. The forthcoming sections will further dissect each component of this workflow, providing a detailed exploration of the methodologies and technologies implemented to meticulously document buildings affected by conflicts.

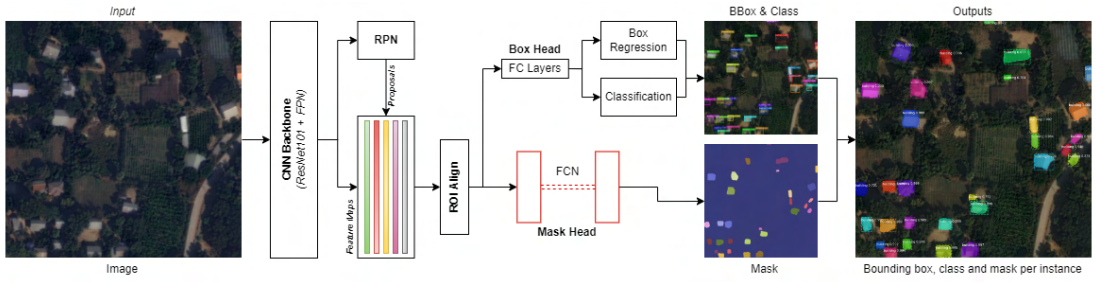

Figure 10: Architecture of Mask R-CNN with ResNet101 backbone with FPN applied in building extraction. Building bounding boxes and segmentation masks are extracted from input RGB raster images.

3.2.2 Building Extraction

The Building Extraction phase of the workflow is a critical step, utilizing Mask R-CNN to identify and outline building instances from satellite imagery. This process not only detects buildings but also generates detailed segmentation masks for each, which are subsequently transformed into precise building footprint polygons as demonstrated in Figure 10.

Before deploying Mask R-CNN, a vital data preprocessing step is undertaken to ensure the satellite images are compatible with the model's requirements. Given the varying dimensions of satellite imagery, each image is segmented into a grid of uniform tiles, typically sized at 512x512 pixels, with an overlap of approximately 20% between adjacent tiles. This tiling ensures the model receives input in a consistent format, addressing potential limitations of the computing platform and maintaining the structural integrity of the Mask R-CNN model. Additionally, to standardize data from different satellite sources, each tile undergoes min-max normalization to a 8-bit Unsigned Integer (UINT8) range, ensuring uniformity across the dataset.

Following preprocessing, these prepared tiles are fed into the Mask R-CNN detection stage, which outputs bounding boxes and segmentation masks for individual building instances within each tile. These outputs are localized to their respective tile coordinates. The post-processing stage addresses the model's segmentation of buildings into multiple entities based on visual or structural differences. This is achieved through a sophisticated polygon merging strategy, detailed in Figure 12. Utilizing an rtree index, this strategy identifies overlapping or intersecting polygons, calculating the intersection area between each pair. Subsequently, the ratio of the intersection area to the area of each polygon is calculated. If the ratio exceeds a certain threshold, typically 50%, the polygons are considered related, with the larger polygon deemed the parent in cases where both exceed the threshold. This relationship is codified in a directed graph, with edges representing the parent-child connections.

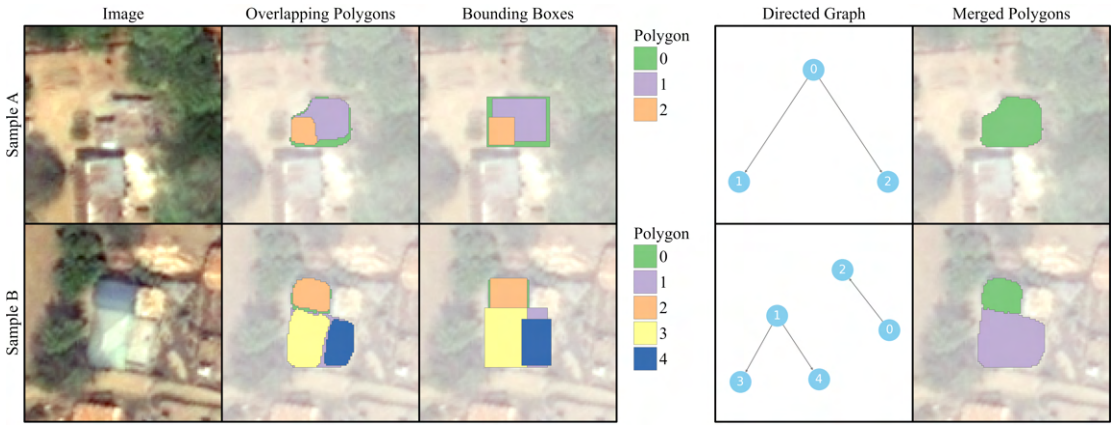

Figure 11: Polygon merging strategy based on the directed graph is applied to merge overlapping and redundant polygons from Mask R-CNN output into single building entities.

This iterative process results in a directed graph containing unconnected nodes (individual polygons) and clusters of interconnected polygons, which are then processed to identify root nodes. For each cluster, children or leaf nodes are recursively merged into their parent nodes, culminating in a single, merged polygon that accurately represents each building instance. This approach not only refines the segmentation masks produced by Mask R-CNN but also ensures that buildings segmented into multiple parts are accurately consolidated into coherent entities as demonstrated in Figure 11. The outcome of this building extraction process is a dataset of building polygons, prepared for subsequent overlap analysis.

3.2.3 Change Detection

3.2.3.1 Overlap Analysis

This phase employs a straightforward spatial overlap analysis to compare building instances, or polygons, delineated from pre- and post-event imagery as output by the Building Extraction stage. In this analysis, the bounding boxes of post-event building polygons are cataloged within an rtree index for efficient spatial querying. Subsequently, for each building polygon identified in the pre-event dataset, overlapping or intersecting polygons from the post-event dataset are retrieved from the rtree index. The Intersect over Union (IoU) score is calculated for each pre-event polygon against its overlapping post-event counterparts. Post-event polygons exhibiting an IoU score below a predetermined threshold, such as 0.5, are excluded from consideration. Among the remaining candidates, the post-event polygon with the highest IoU score is selected as the corresponding match for the pre-event building instance. Should a pre-event polygon lack any overlapping post-event polygons or fail to identify a suitable match, it is classified as destroyed. This is outlined in Algorithm Figure 13.

Algorithm 1: Polygon Merging Strategy

Input: A set of polygons P
Output: A set of merged polygons M
Initialize an empty directed graph G
foreach *polygon* p_{target} *in* P **do**
  foreach *polygon* p_{other} *in* $P \setminus \{p_{target}\}$ **do**
    Calculate area of overlap $O(p_{\text{target}}, p_{\text{other}})$
    if $O(p_{target}, p_{other}) > 0$ **then**
      Calculate ratio $r_{\text{target}} = \frac{O(p_{\text{target}}, p_{\text{other}})}{\text{Area}(p_{\text{target}})}$
      Calculate ratio $r_{\text{other}} = \frac{O(p_{\text{target}}, p_{\text{other}})}{\text{Area}(p_{\text{other}})}$
      if $r_{target} > 0.5$ ***and*** $r_{other} > 0.5$ **then**
        if $Area(p_{target}) > Area(p_{other})$ **then**
          Add edge $p_{\text{target}} \rightarrow p_{\text{other}}$ to G
        else
          Add edge $p_{\text{other}} \rightarrow p_{\text{target}}$ to G
        end
      else
        if $r_{target} > 0.5$ **then**
          Add edge $p_{\text{other}} \rightarrow p_{\text{target}}$ to G
        end
        else if $r_{other} > 0.5$ **then**
          Add edge $p_{\text{target}} \rightarrow p_{\text{other}}$ to G
        end
      end
    end
  end
end
Identify clusters in G and merge polygons within each cluster to form M
return M

Figure 12: Algorithm for merging redundant and overlapping polygons resulting from grid tiling of input image and imperfections of Mask R-CNN model. Individual parts of a coherent building detected as separate buildings are merged through overlap area ratioing and the analysis of hierarchical relationship between overlapping polygons.

This methodological step effectively segregates the dataset into two distinct categories: pre-event building polygons without corresponding post-event matches, indicative of destruction, and those with matching post-event polygons. The latter group is subsequently analyzed using the Siamese Network patch similarity comparison stage to further assess the integrity of the structures, while the former group proceeds directly to the post-processing stage of the change detection phase.

3.2.3.2 Patch Similarity Test

The CD phase incorporates a critical Siamese Network Patch Similarity Comparison to assess similarities between pre- and post-event building instances. This network consists of three main components: a feature extraction module that translates images into a vector space, a differencing mechanism for calculating vector distances, and a sigmoid neuron in the final head that determines similarity scores between 0 and 1, indicating dissimilarity and perfect similarity, respectively. The feature extraction utilizes dual identical branches, or "twin towers," each processing an input independently with shared weights. ResNet101 with a FPN is selected as the feature extractor

Algorithm 2: IoU Based Pre and Post Polygons Matching Strategy

Input: Two sets of polygons: Pre polygons P_{pre} and Post polygons P_{post}
Output: Matched pairs of Pre and Post polygons, and unmatched polygons from both sets
Initialize an empty list M for storing matched pairs
foreach *polygon* p_{pre} *in* P_{pre} **do**
  Initialize an empty list L for storing IoU values and corresponding Post polygons
  foreach *polygon* p_{post} *in* P_{post} **do**
    Calculate IoU between p_{pre} and p_{post}: IoU = CalculateIoU($p_{\text{pre}}, p_{\text{post}}$)
    if $IoU > 0.5$ **then**
      Add (IoU, p_{post}) to L
    end
  end
  if L *is not empty* **then**
    Find $p_{\text{post}}^{\text{max}}$ with the highest IoU in L
    Add ($p_{\text{pre}}, p_{\text{post}}^{\text{max}}$) to M
  end
end
foreach *polygon* p_{pre} *in* P_{pre} **do**
  if p_{pre} *is not in any pair in* M **then**
    Mark p_{pre} as unmatched in P_{pre}
  end
end
foreach *polygon* p_{post} *in* P_{post} **do**
  if p_{post} *is not in any pair in* M **then**
    Mark p_{post} as unmatched in P_{post}
  end
end
return *Matched pairs* M, *Unmatched in* P_{pre}, *Unmatched in* P_{post}

Figure 13: Algorithm to determine and match pre- and post-event building polygons based on IoU. Outputs from this algorithm is a set of paired pre- and post-event building polygons and sets of unmatched polygons for each temporal sets.

backbone as they were already implemented and trained in the Mask R-CNN model for identifying relevant building features. Pre-trained weights from this portion of Mask R-CNN model can be transferred to Siamese Network's feature extractor portion.

Attention mechanisms are usually integrated in tasks such as image similarity comparison, change detection and object tracking tasks (T. Chen et al., 2022; J. Huang et al., 2022; Sun et al., 2022). In this thesis work, CBAMs are employed to refine feature maps through Channel Attention and Spatial Attention, emphasizing areas pertinent to building comparison. Post-feature extraction, Euclidean distances between feature vectors are computed, and the sigmoid neuron outputs a similarity score. The whole architecture is outlined in Figure 14. For each pair of matching pre- and post-event building polygons identified in the overlap analysis, image patches are extracted based on a square bounding box defined by the union of each pair. These patches are resized to fit the network's input specifications, such as 224x224 pixels, and normalized to a UINT8 data range. The Siamese Network then processes these patches, with the sigmoid neuron activation providing a similarity score. A threshold, typically set at 0.5, classifies each pair as similar (intact) or dissimilar (changed or destroyed). This process results in a set of pre-event building polygons classified as "intact" or "changed/destroyed," merged with polygons lacking post-event matches.

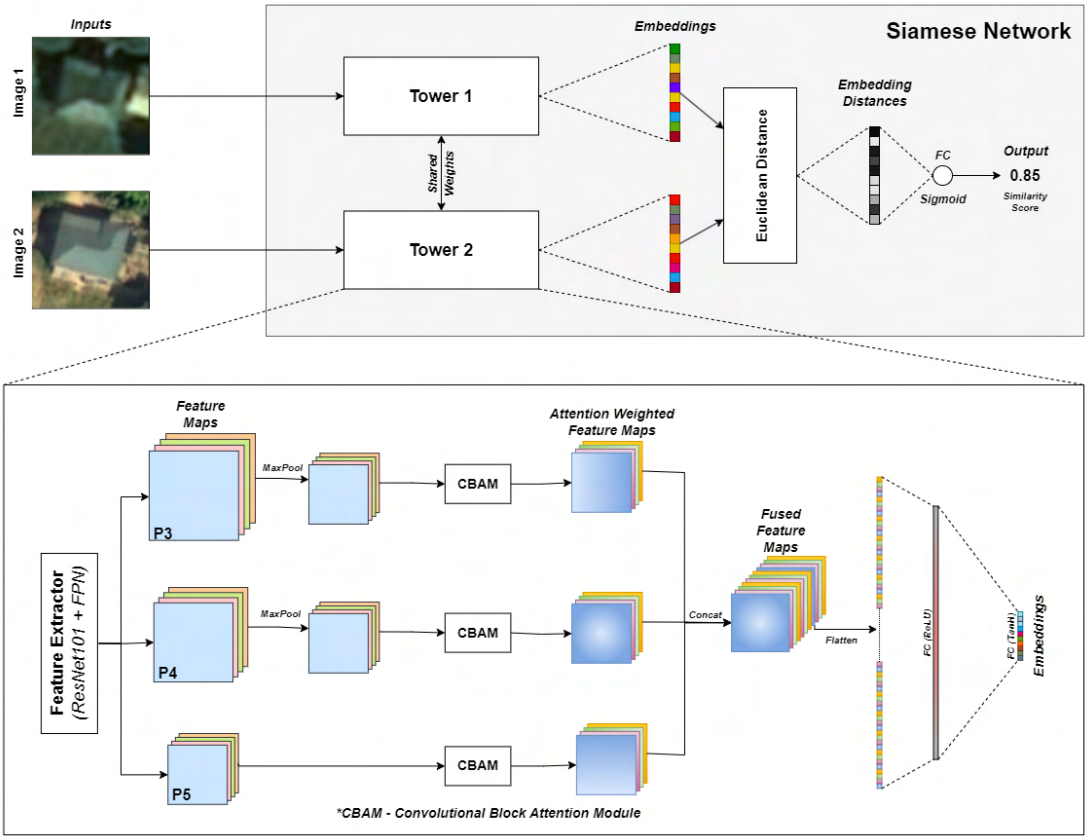

Figure 14: Architecture of Siamese Network similarity comparison component employing ResNet101 and FPN as feature extractor and integrating CBAMs to introduce attention mechanism. The similarity of two input image patches are determined by the score output of sigmoid neuron.

The final outputs include pre-event building polygons with change status annotations and identified post-event building polygons, transitioning from initial raster inputs to a detailed building delineations with structural change documentation.

3.2.4 Deep Learning Models

The proposed workflow is composed of two DL components: Mask R-CNN for building extraction and Siamese Network for image patch similarity comparison. The following sections describes data preparation and training processes for both DL components.

3.2.4.1 Data Preparation

For the training of Mask R-CNN model, pre- and post-event RGB raster images are aligned and georegistered in ArcGIS software using Spline Transformation. Both sets of images are resampled to the highest spatial resolution among input images, which is 30 cm per pixel. The Geo-SAM plugin(Zhao, Fan, and L. Liu, 2023) in QGIS facilitates the

swift and precise manual annotation of building footprints directly from these raster images, with each building delineated as a separate polygon. This process is applied to both pre- and post-event images across 20 areas of interest, ensuring comprehensive coverage. Figure 15 visualizes the raster images and annotated buildings. These pre- and post-event images, along with their corresponding building annotations, are segmented into 512x512 pixel tiles. Tiles containing a NODATA pixel count exceeding 50% of the entire raster tile are excluded from further processing. Subsequently, image tiles undergo min-max normalization to the UINT8 data range. These prepared tile sets and their annotations are then divided into three distinct datasets: training, validation, and testing, as outlined in Table 1. Given the architectural and environmental similarity of buildings across AOIs involved, the data was split by paying more attention to the equal distribution of different combinations of data sources, PL and PLN, of pre- and post-event images.

For the training of Siamese Network similarity comparison model, the input data constitutes anchor, positive and negative images. Anchor image is the image of an object which is deemed similar to the positive image while the negative image is the dissimilar image opposed to the anchor image as presented in Figure 16. In case of anchor and positive image (similar) pairs, image patches of the same building in the pre- and post-event images are extracted as pairs. In case of anchor and negative image (dissimilar) pairs, the anchor image of a building is paired up with a random image of another building as the negative image. The similar pairs are visually inspected and extracted from 12 AOIs. The bounding boxes (adjusted to square shapes) of building polygons are used as windows for extracting the image patches which later resized to 224x224 pixels dimension which is the input dimension of ResNet101 feature extractor and min-max normalized to UINT8 data range. Extracted image patches are further augmented by increamental rotation of 45 degrees. The resulting samples are further split into three datasets: train, validation and test as described in Table 2. For consistency of models' performance evaluation, respective samples belonging to train, validation and test datasets from Mask R-CNN are similarly put into same category of datasets for Siamese Network.

To assess the efficacy of this comprehensive workflow, GT data for six AOIs which are not involved in the training phases of both the Mask R-CNN and Siamese Network models were carefully prepared. This GT comprises pre- and post-event raster images and building annotations for both temporal states, along with the status of each building in the post-event scenario, delineating whether it has undergone destruction or change. The annotation process for establishing this GT was conducted manually, entailing a thorough visual comparison of images from both temporal states.

Figure 15: Input RGB raster tiles and building annotations used in the training process of Mask R-CNN model.

Figure 16: Image patches: (1-4) similar and (5-8) dissimilar pairs with corresponding labels for training Siamese Network similarity comparison model.

Dataset	Tiles	Total Area (Km^2)	Buildings	Split (%)
Train	695	16.40	14879	67.41
Validation	243	5.73	4860	22.02
Test	162	3.82	2334	10.57

Table 1: Train, validation and test datasets of image tiles and annotated buildings prepared for Mask R-CNN model training.

Dataset	Similar Pairs	Dissimilar Pairs	Split (%)
Train	35760	35760	63
Validation	11920	11920	21
Test	9120	9120	16

Table 2: Train, validation and test datasets of similar and dissimilar image patch pairs for Siamese Network model training.

3.2.4.2 Model Training

The Mask R-CNN model, as implemented by Matterport(Abdulla, 2017), incorporates a FPN and RoIAlign, and is employed to train the model. The ResNet101 architecture serves as the backbone for feature extraction. Given the limited volume of input data, a strategy of transfer learning from a pre-trained model is utilized to expedite the training process while enhancing model performance. The baseline pre-trained weights, released as part of the CrowdAI challenge, form the foundation for transfer learning. This model, having been trained on the SpaceNet challenge dataset (*Datasets* 2018)—comprising VHR (30-50cm) satellite imagery from the WorldView-2 and WorldView-3 satellites—bears a significant resemblance to the data utilized in this thesis, thereby providing a robust starting point. Data augmentation is applied through the incremental rotation of input image tiles by 30 degrees. The training procedure initiates with the loading of pre-trained weights from the CrowdAI challenge (Mohanty, 2018; Mohanty et al., 2020). Initial training phases freeze all the lower-level layers of the backbone, focusing solely on the head components—comprising RPN, object detection, and mask regression layers—with an initial Learning Rate (LR) of 0.001 over 30 epochs. Subsequent phases involve unfreezing layers above the fourth and third levels of the backbone, including the head components, for fine-tuning at a LR of 0.0001 for 30 and 20 epochs, respectively. The final stage entails unfreezing all layers for comprehensive fine-tuning at a LR of 0.000001 for 80 epochs, by which point the model typically achieves a plateau in training and validation loss values, approximately 0.75 and 0.7, respectively, as charted in Figure 17. Likewise, a model without transfer learning from a pre-trained model is also trained in order to contrast the performance between models with and without transfer learning applied. This contrast is obvious in learning curves of two models where the model with transfer learning have lower losses.

This culmination of losses—comprising the Mask R-CNN bounding box loss (mrcnn_bbox), class loss (mrcnn_class), mask loss (mrcnn_mask), and the RPN bounding box loss (rpn_bbox) and class loss (rpn_class)—indicates the model's performance, with the RPN bounding box loss being notably the most significant contributor. This is attributed to its role in proposing regions containing potential object candidates, which are then utilized by the Mask R-CNN for predicting bounding boxes and masks. The final deliverable of this training process is a TensorFlow-based Mask R-CNN model, proficient in identifying and segmenting buildings or houses from VHR satellite imagery as illustrated in Figure 18.

To ensure consistency across software dependencies, TensorFlow is utilized for training the Siamese network model. This network architecture generally comprises three main components: a feature extraction part that extracts features and embeddings into a vector space, a differencing part that calculates the distance between vectors, and a final head that produces a similarity or dissimilarity score for the two inputs.

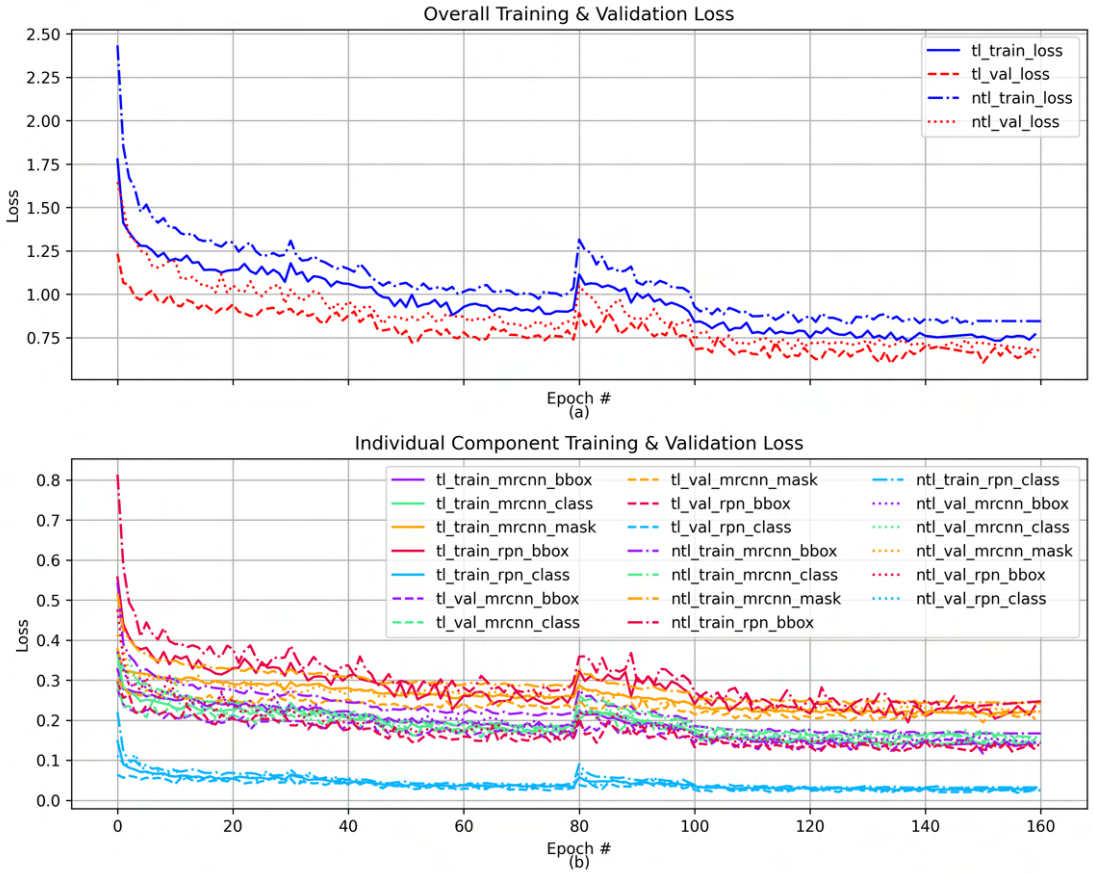

Figure 17: Learning curves of Mask R-CNN model training process: (a) overall total losses and (b) individual losses on train and validation datasets. Curves for models with and without transfer learning applied are denoted with TL and NTL prefixes respectively.

The feature extraction module is constructed as two identical branches, known as twin towers, which share the same weights and process each input separately. For this task, ResNet101 combined with FPN is selected as the feature extractor, leveraging its pre-existing implementation and training within the Mask R-CNN model to identify features associated with buildings and their background. The components of the ResNet101 backbone and FPN from the Mask R-CNN model, along with their pretrained weights, are integrated into the feature extraction portion of the Siamese network and set as non-trainable. The output feature maps from the FPN, which exist at varying spatial scales, are subsequently downsampled to a uniform dimension and processed through a CBAM to apply attention weights. The distance between the feature vectors extracted from each tower is calculated using Euclidean distance, which then feeds into a fully connected sigmoid neuron to yield a similarity score ranging between 0 and 1, where 0 indicates dissimilar image patches and 1 denotes perfect similarity.

The network is trained using a learning rate of 0.00001 and employs contrastive loss

Figure 18: Input images and building polygons extracted by trained Mask R-CNN model visualized against ground truth polygons.

across 45 epochs, at which point both the loss and accuracy tend to plateau around 0.15 and 0.85, respectively. The training trajectory in Figure 19 indicates a stable balance between training and validation performance up until the final epochs, where the model begins to show signs of overfitting to the training data.

3.2.5 Evaluation Metrics

In order to evaluate the feasibility, performance and accuracy of the proposed workflow as a whole or the individual stages of the workflow, a set of evaluation metrics are calculated as follow. It is imperative to recognize that the complexity of the integrated workflow precludes a single, unified metric that can encapsulize its overall performance and accuracy. The workflow is composed of two distinct deep learning stages: the extraction of buildings using Mask R-CNN and the subsequent similarity comparison via Siamese Network. Each stage has its potential for error; Mask R-CNN may yield both missed and false detections in a bi-temporal analysis, while Siamese Network's predictions are contingent upon the overlap analysis results, potentially leading to misclassifications regarding the change status of buildings. The compounded nature of these errors necessitates a multifaceted evaluation approach. The accuracy and

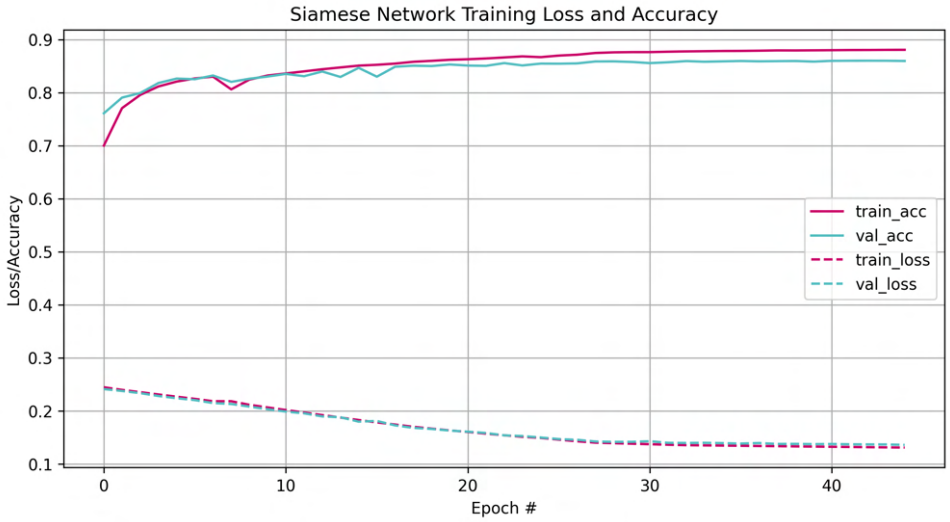

Figure 19: Learning curves, train and validation losses, of Siamese Network model training process over 45 epochs.

performance of the workflow are thus assessed through discrete analyses of each component, combined with a comparison of the workflow's results against the ground truth change status to form an overarching evaluation.

3.2.5.1 Intersection over Union

The IoU score is a relevant and widely utilized metric in object detection tasks. IoU calculates the ratio of the area of overlap between two polygons or bounding boxes to the area of their union, as defined in Equation 3.1. In this study, both the building extraction component and the overlap analysis of pre- and post-event buildings employ the IoU score for performance evaluation and analytical decision-making. For the evaluation of building extraction performance, the IoU between detected buildings ground truth buildings is calculated and compared against a threshold to assess agreement levels, identify missed detections, and false detections. In the context of overlap analysis, the presence of pre-event buildings in the post-event imagery is preliminarily determined by calculating the IoU between corresponding pre- and post-event building polygons.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{3.1}$$

3.2.5.2 True Positives and Negatives, False Positives and Negatives

These metrics form the basis for calculating additional scores and metrics, such as Precision (P), Recall (R), Average Precision (AP), F-beta, and Matthews Correlation Coefficient (MCC) scores. In the evaluation of the Mask R-CNN model, detected polygons with corresponding overlapping ground truth polygons whose IoU score exceeds a predefined threshold are classified as True Positive (TP) cases. Detected

polygons lacking a corresponding ground truth polygon are classified as False Positive (FP), while ground truth polygons without a corresponding detected polygon are considered False Negative (FN). Although the concept of a True Negative (TN) case is pertinent in scenarios involving the detection of multiple object classes, it is deemed irrelevant in the current context, where there is only one target object class considered as foreground and everything else as background.

For the Siamese Network patch similarity comparison component, the definitions for TP, TN, FP, and FN cases are clear and well-defined. Any input pairs labeled as similar and with a predicted similarity score above the threshold (e.g., 0.5) are classified as TP, whereas pairs labeled as similar but with a score below the threshold are considered FN. Similarly, for pairs labeled as dissimilar, those with a predicted score above the threshold are deemed FP, and those with a score below the threshold are TN.

In evaluating the overall workflow, the agreement, missed detections, and false detections between ground truth polygons and detected polygons are quantified. For all polygon pairs in agreement, the change status of GT polygons is compared against the change status of detected polygons. Situations where intact buildings in the ground truth correspond to detected polygons with an intact change status are categorized as TP, while instances where changed or destroyed buildings in the GT correspond to detected polygons with a changed or destroyed status are categorized as TN. Discrepancies between the GT polygon labels and detected polygons result in FN cases for intact GT polygons and FP cases for changed or destroyed GT polygons.

3.2.5.3 Precision, Recall (True Positive Rate) and False Positive Rate

In the context of object detection and binary classification, Precision, Recall or True Positive Rate (TPR), and False Positive Rate (FPR) are critical metrics used to evaluate the performance of a model. These metrics are essential for interpreting the effectiveness of a model in distinguishing between the classes (e.g., "object" vs. "no object" in object detection or "positive" vs. "negative" in binary classification).

Precision measures the accuracy of the positive predictions made by the model. It is defined as the number of true positive predictions divided by the total number of positive predictions (the sum of true positives and false positives) as described in Equation 3.2. In object detection, in case of this thesis work - building extraction, it quantifies how many of the detected objects were truly buildings.

Recall or TPR measures the model's ability to correctly identify all actual positives in the dataset. It is defined as the number of TP predictions divided by the actual number of positives in the dataset (the sum of TPs and FNs) as described in Equation 3.3. In scope of this proposed workflow, this metrics is relevant and important in evaluating how good the building extraction is in detecting all buildings existing in the input image. This is particularly important not to miss the presence of actual buildings in

both pre- and post-event images which affect the overlap analysis stage and the final output data quality.

The FPR measures the proportion of negative instances that are incorrectly classified as positive. It is defined as the number of FP predictions divided by the total number of actual negatives in the dataset (the sum of FP and TN) as described in Equation 3.4. Likewise to TPR, FPR is an important metric in the building extraction stage not to register false presence of buildings which contribute to the overlap analysis and final output quality.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.2}$$

$$\text{Recall} = \text{TPR} = \frac{TP}{TP + FN} \tag{3.3}$$

$$\text{FPR} = \frac{FP}{TN + FP} \tag{3.4}$$

3.2.5.4 Average Precision

The Average Precision, commonly used in object detection and instance segmentation tasks, is calculated by averaging the precision values at each recall level across the recall range (from 0 to 1) as described in Equation 3.5. This is done for each object class individually. Mean Average Precision (mAP), the average of the APs across all classes, is commonly used in cases of multiple classes which is irrelevant in current case where only building class is detected. AP provides a single figure that summarily reflects the performance of building extraction across all thresholds, making it a comprehensive metric for evaluating the model's detection and segmentation accuracy.

$$\text{AP} = \sum_{n} (R_n - R_{n-1}) P_n \tag{3.5}$$

3.2.5.5 F-beta

The F-beta score (Fβ) is a metric combining Precision and Recall to balance their importance, critical in object detection where precision and recall trade-offs are essential. The parameter β adjusts the emphasis on recall; values greater than 1 prioritize recall, highlighting the model's ability to detect all relevant entities, whereas values less than 1 prioritize precision, focusing on the model's accuracy in pinpointing true positives. In this study's context of building extraction, F1 and F2 scores are significant. The F1 score, aiming for a balance between precision and recall, is optimal for evaluating overall performance. The F2 score, giving more weight to recall, is essential for evaluating model's capability to capturing all buildings within a given input at a designated IoU threshold, ensuring comprehensive detection. Equation 3.6 details the calculation of (Fβ) score.

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}} \tag{3.6}$$

3.2.5.6 Areas under Curve: Receiver Operating Characteristics(ROC-AUC) and Precision-Recall (PR-AUC)

The ROC-Area under Curve (AUC) score is determined from the Receiver Operating Characteristic (ROC) curve, which plots TPR against FPR across varying threshold settings. This score is obtained by integrating the area under the ROC curve from 0 to 1, considering various binary classification thresholds. In the context of the Siamese Network similarity comparison stage, the ROC-AUC metric is particularly pertinent as it assesses the model's capability to correctly differentiate between positive and negative classes where both classes are equally important while both classes in training and evaluation datasets are well-balanced.

Furthermore, the Average Precision Score, associated with the PR-AUC (Precision-Recall Area Under the Curve), is calculated from the Precision-Recall (PR) curve. This curve delineates Precision against Recall at diverse threshold levels. The Average Precision (AP) encapsulates the PR curve as the weighted average of precision values at each threshold, employing the increment in recall from one threshold to the next as the weighting factor. Although the PR-AUC metric might not carry the same weight as the ROC-AUC in the Siamese Network's similarity comparison phase, it remains an essential measure for evaluating the model's proficiency in accurately identifying unchanged or intact buildings. This evaluation is crucial in real-world applications where a significant majority of detected buildings are likely to remain unchanged, with only a minority being altered or destroyed, emphasizing the model's precision and recall balance in identifying TPs and minimizing FPs.

3.2.5.7 Matthews Correlation Coefficient

The Matthews Correlation Coefficient (MCC) is a measure used in machine learning to evaluate the quality of binary (two-class) classifications. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The MCC is defined in Equation 3.7. The MCC is considered to be a reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset. This characteristic is relevant in evaluation of workflow's performance where the ability to distinguish both intact and changed/destroyed buildings is equally important.

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{3.7}$$

4

Chapter 4

4.1 Analysis and Results

In this chapter, the evaluation and analysis of the results for each deep learning component within the workflow, as well as the workflow as a whole as described in the previous chapter, will be presented. The initial focus will be on the Mask R-CNN model, followed by the Siamese Network similarity comparison, and concluding with the evaluation of the integrated workflow.

4.1.1 Mask R-CNN

The evaluation of the trained model hinges on conducting an IoU analysis between the Ground Truth (GT) and the polygons detected by the model. For every ground truth polygon, an IoU value is computed against any overlapping detected polygons. The list of potential matching candidates is then refined by excluding any polygons with an IoU score below the threshold of 0.5. Among the remaining candidates, the polygon boasting the highest IoU score is deemed the corresponding match for the GT polygon. This procedure culminates in a compilation of matched detections (TP), false detections (FP), and missed detections (FN). Subsequently, these metrics are utilized to calculate Precision and Recall scores, which are further employed to compute the AP and F-beta (F0.5, F1, and F2) scores. While Precision and Recall are pivotal, the imperative to identify and catalog all extant buildings accentuates the significance of Recall and the F2 score in the evaluation process.

To ascertain the model's efficacy across varying IoU thresholds, the aforementioned metrics are recalculated for IoU values spanning from 0.1 to 0.9, in increments of 0.05. The outcomes are illustrated in Figure 20, which plots the scores against IoU values, as shown below. The model without transfer learning is also included in the evaluation to contrast the performance difference between two models. It is apparent that the model with transfer learning perform noticeably better than the one without transfer learning. Additionally, Table 3 is provided to detail the specific metrics and scores achieved at the IoU threshold of 0.5.

Dataset	Tiles	Buildings Ground Truth	Buildings Detected	Precision	Recall	F0.5	F1	F2	AP
Train	595	14879	17160	0.81	0.94	0.84	0.87	0.91	0.94
Validation	206	4860	5446	0.81	0.91	0.83	0.86	0.89	0.89
Test	106	2334	2647	0.81	0.92	0.83	0.86	0.89	0.91

Table 3: Evaluation results of Mask R-CNN model for train, validation and test datasets. High scores in general display good performance of trained model.

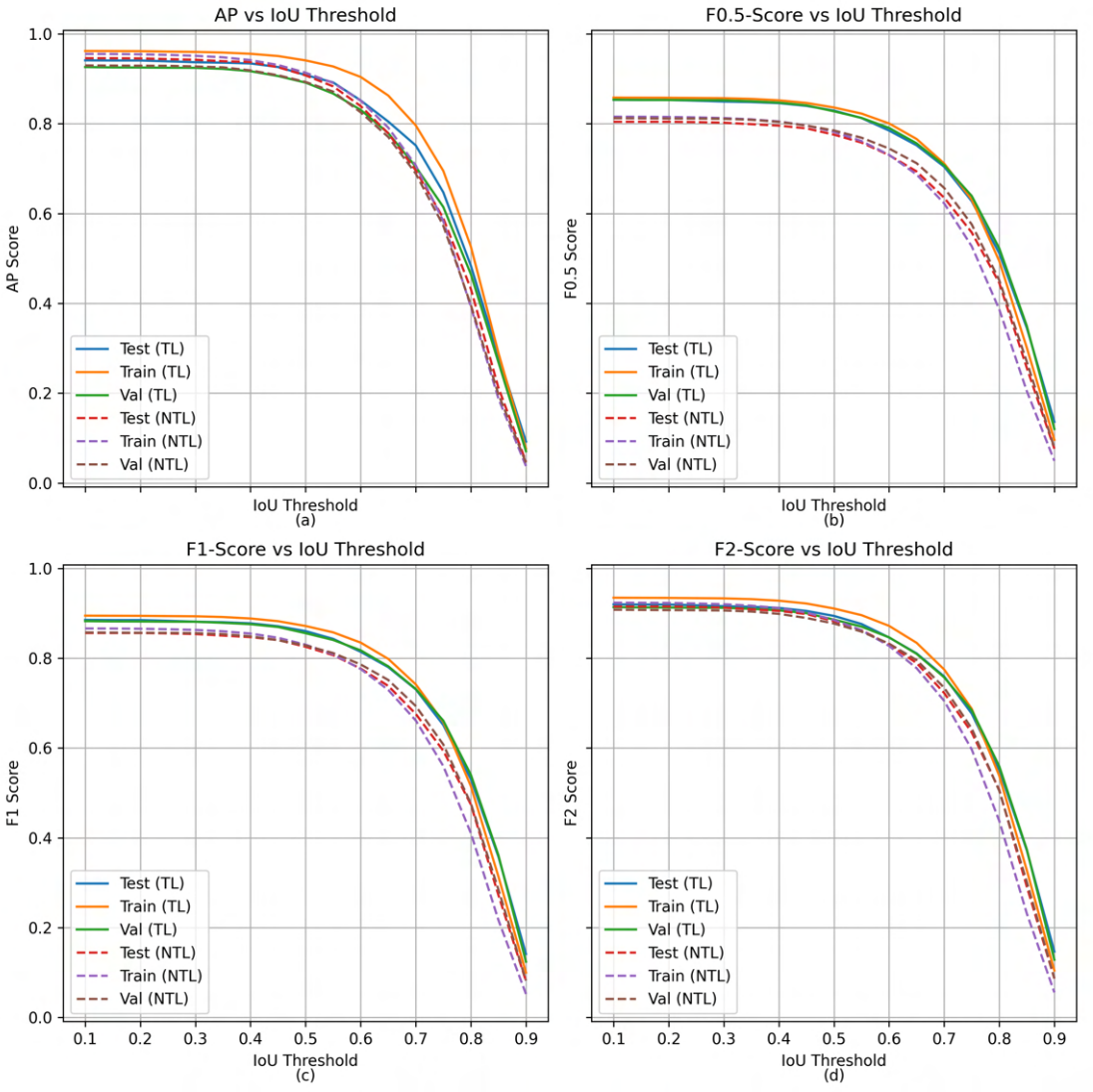

Figure 20: Mask R-CNN model evaluation metrics: (a) AP, (b) F0.5, (c) F1 and (d) F2 plotted over a 0.1-0.9 range of IoU thresholds. Scores above 0.8 is still observed till IoU threshold of 0.6. Metrics for models with and without transfer learning is denoted with prefixes TL and NTL respectively.

4.1.2 Siamese Network

The evaluation of the Siamese Network for similarity comparison involves directly contrasting the GT labels (similar or dissimilar) with the prediction results for image pairs as illustrated in Figure 21. An output score below a set threshold (e.g. 0.5) is categorized as dissimilar, while a score above this threshold is deemed similar. These output scores are compared against the GT labels, similar or dissimilar. This evaluation is conducted across all datasets—including training, validation, and testing—using a spectrum of threshold values from 0 to 1 to test the sensitivity across thresholds.

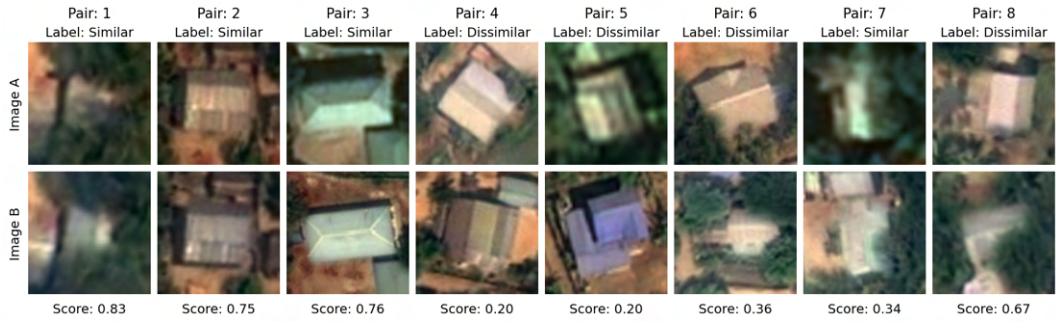

Figure 21: Examples of Siamese Network model's classification results. Pairs (1-3) and (4-6) present similar and dissimilar pairs with correct classification while pairs (7-8) present incorrect classification results.

Dataset	Similar (GT)	Dissimilar (GT)	Similar Classified	Dissimilar Classified	TPR	FPR	Precision	Threshold
Train	5957	5963	6309	5611	0.90	0.16	0.85	0.54
			6101	5819	0.88	0.15	0.86	0.55
			5875	6045	0.86	0.13	0.87	0.56
Validation	1990	1984	2084	1890	0.89	0.16	0.85	0.54
			2024	1950	0.88	0.14	0.86	0.55
			1968	2006	0.86	0.13	0.87	0.56
Test	1520	1520	1607	1433	0.87	0.19	0.82	0.54
			1546	1494	0.85	0.17	0.83	0.55
			1500	1540	0.84	0.15	0.85	0.56

Table 4: Evaluation results of trained Siamese Network model at three binary thresholds. High TPR scores present model's good performance in classifying intact buildings.

This approach generates metrics on TPs, TNs, FPs, and FNs, where positive cases represent similar pairs and negative cases denote dissimilar pairs. These metrics are further analyzed to calculate the TPR, FPR, Precision, and Recall. ROC and PR curves are plotted, with AUCs computed for all three datasets, as illustrated in Figure 22 and Table 4.

Moreover, the distribution of FPs and FNs across various threshold values is charted

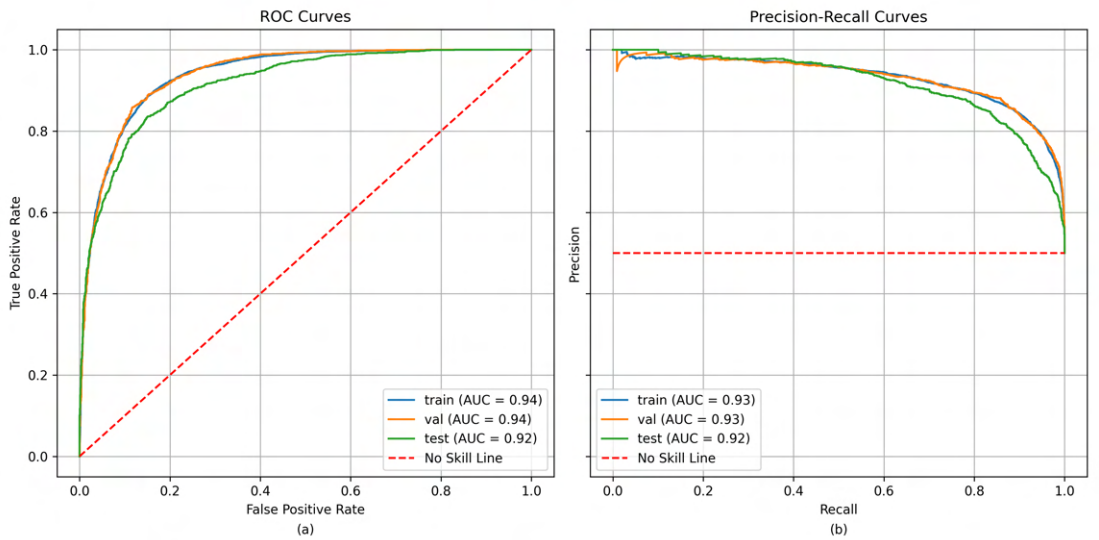

Figure 22: Siamese Network model evaluation curves (a) ROC and (b) Precision-Recall with respective AUCs plotted for train, validation and test datasets.

to identify the optimal threshold for achieving a balanced trade-off between FPs and FNs, as depicted in Figure 23 where a threshold between 0.5 and 0.55 seems optimal.

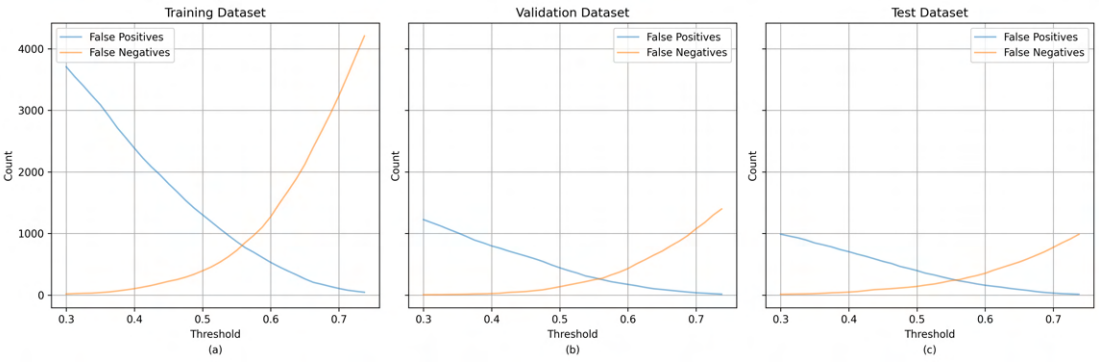

Figure 23: False positives and negatives charted over a threshold range for (a) train, (b) validation and (c) test datasets. A good balance between FP and FN can be observed around threshold value of 0.55.

4.1.3 Workflow

The integrated workflow, combining the processes of building extraction and change detection, is designed to process pre- and post-event images from specified AOIs. The outcome of this workflow includes two sets of polygons representing buildings for both pre- and post-events. Each polygon from the pre-event is appended with an attribute indicating whether the building has been destroyed, changed post-event, or remains intact.

Given that the workflow bifurcates into building extraction and CD, the evaluation is correspondingly dual-faceted. Initially, for the building extraction phase, the IoU metric is utilized to compare each ground truth building polygon with overlapping detected polygons. The selection of candidates is refined by excluding those with an IoU score below a predetermined threshold (0.55), and the highest-scoring candidate is identified as the corresponding prediction for each ground truth building. This phase's evaluation focuses on the accuracy of building detection, quantifying correct detections, missed detections, and FPs.

Subsequently, in the CD phase, only those delineated buildings that align with GT data are scrutinized. The comparison involves analyzing the status attributes—whether destroyed/changed or intact—of both extracted and GT polygons. Metrics such as TPs, TNs, FPs, and FNs are computed, which are then leveraged to derive the TPR, FPR, Precision, Recall, F1, and MCC scores. Among these metrics, MCC is emphasized as a crucial indicator of binary classification performance, highlighting the model's ability to correctly identify both TP and TN scenarios.

The evaluation findings are systematically presented in Table 5 and Table 6, complemented by illustrative examples comparing the workflow's detection outcomes with the ground truth, showcased in Figure 24. This structured approach enables a clear and concise assessment of the workflow's capability to accurately document building status changes in the aftermath of events. In summation, the workflow delineated herein exhibits a considerable degree of accuracy in identifying destroyed buildings, as substantiated by the results in Table 7. While there is a discernible presence of missed detections and erroneous classifications regarding the change status of buildings, the reduction in time and labor required to evaluate each AOI suggests that the workflow is viable for operational use. However, it is recommended that the workflow's outcomes undergo supplementary verification by a human expert to ensure precision and reliability in an applied setting.

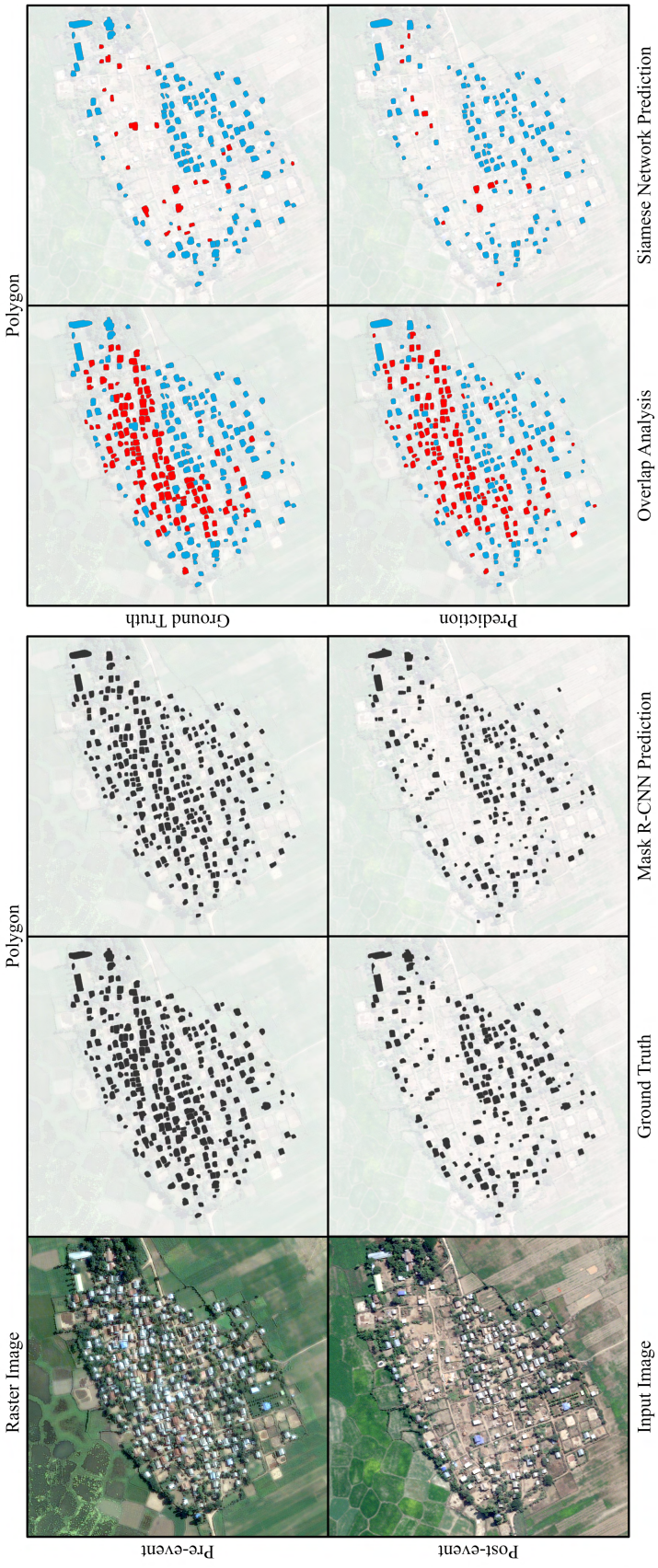

Figure 24: Visualization of extracted buildings and change status classified by the workflow with corresponding ground truths. Visualization of overlap analysis presents pre-event buildings with (blue) and without (red) overlapping post-event buildings. Changed or destroyed buildings and intact buildings according to Siamese Network similarity comparison are presented as red and blue polygons respectively.

AOI	Area (km^2)	Time Taken (Secs)	Buildings (GT)	Buildings (Detected)	Agreement (GT)	Missed Detections	False Detections
175981	1.05	92.81	297	298	272	25	26
188943	1.08	76.03	375	300	257	118	43
188837	0.63	45.24	380	321	288	92	33
171240	0.35	35.31	297	308	255	42	53
171241	0.77	68.75	511	527	450	61	77
171242	0.63	68.53	573	560	501	72	59

Table 5: Workflow's building extraction results over test dataset tabulated. A good agreement between ground truth and extraction is observed.

AOI	TP	TN	FP	FN	TPR	FPR	Precision	Recall	F1	MCC
175981	117	135	12	8	0.94	0.08	0.91	0.94	0.92	0.85
188943	162	69	9	17	0.91	0.12	0.95	0.91	0.93	0.77
188837	91	159	1	37	0.71	0.01	0.99	0.71	0.83	0.75
171240	130	80	4	41	0.76	0.05	0.97	0.76	0.85	0.67
171241	255	117	7	71	0.78	0.06	0.97	0.78	0.87	0.66
171242	310	138	7	46	0.87	0.05	0.98	0.87	0.92	0.77

Table 6: Siamese Network-based change classification results of the workflow over test dataset.

AOI	GT Destroyed	Correct	Incorrect	Missed Detections
175981	136	117	8	11
188943	257	162	17	78
188837	169	91	37	41
171240	201	130	41	30
171241	366	255	71	40
171242	402	310	46	46

Table 7: Summary of workflow's performance on detecting and classifying destroyed or changed buildings in test dataset.

5

Chapter 5

5.1 Discussion

In this chapter, the discussion revolves around the evaluation results and performance of key components, particularly the Mask R-CNN building extraction model and the Siamese Network patch similarity comparison model, in addition to the integrated workflow as a whole. The focus is on interpreting the evaluation outcomes, identifying the factors leading to misclassifications and deficiencies, and critically assessing each component along with the overall workflow. This is followed by addressing the research questions this study aimed to explore, detailing the limitations of the current methodology and the proposed workflow. The section concludes by outlining future directions, suggesting areas for potential extensions, further research, and enhancements to this study.

5.1.1 Building Extraction

The evaluation results for the Mask R-CNN model, as reflected in Figure 18 and Figure 20, focus on the model's capability to detect and delineate correct building instances. These results are critical as they form the foundational step for subsequent analyses in the assessment of building changes post-conflict. Table 3 indicates a consistent precision maintained around 0.8 with IoU threshold of 0.50 across training, validation, and testing datasets. This uniform precision suggests that the model is adept at accurately identifying intact buildings.

Recall is exceptionally high in the training set at 0.94, indicating a strong capability of the model to detect the majority of buildings truly present in bi-temporal images. This is crucial in the change detection. The F-beta scores, particularly the F2 score, demonstrate the model's sensitivity to capturing positive instances, which, in this context, are the intact buildings. The performance graphs reveal a commendable stability across varying IoU thresholds, suggesting that the model's detection capability is not overly sensitive to the exactness of the IoU cutoff till IoU 0.7. AP is high across all datasets, with the highest in the training set at 0.94, slightly tapering off to 0.91 in the

testing set, an indication that the model's ability to detect buildings translates well to unseen data. In terms of object detection evaluation metrics, both PASCAL VOC (IoU 0.5)and strict (IoU 0.75), as presented in (*COCO Dataset - Evaluation Metrics* 2024), AP and F-beta scores between IoU thresholds 0.5 and 0.75 presents a range between 0.65 and 0.9 which can be considered adequate.

In the specific context of conflict damage assessment, the Mask R-CNN model serves a pivotal role in detecting intact structures rather than the destroyed ones. The subsequent stages of change detection, which involve overlay analysis and Siamese network-based similarity comparison, rely on the accurate identification of buildings by the Mask R-CNN. The high recall and F2 scores are particularly significant, as they ensure that the model minimizes the chances of missing any intact buildings, which is a necessary condition for accurate change detection and assessment of destruction. The high level of precision and recall across all datasets underscores the model's robustness and its critical role in the workflow for conflict damage assessment.

5.1.2 Patch Similarity Comparison

The Siamese Network's performance for patch similarity comparison is critical for discerning the status of buildings, whether they remain intact or have been destroyed/changed post-event. The evaluation of the Siamese Network, as detailed in Table 4, indicates that the network is proficient at identifying both true positives and true negatives across training, validation, and test datasets. Notably, the TPR remains high across all datasets, signifying that the network reliably identifies similar image pairs—crucial for confirming buildings as intact which constitutes the majority of buildings in post-event after overlap analysis.

For the training set, the TPR is particularly robust at 0.90 with a threshold of 0.54. The FPR is managed well, staying around 0.15-0.16, which reflects the network's ability to discern dissimilar pairs effectively and avoid false alarms of building destruction. Precision and recall both exhibit high scores, hovering around 0.85-0.87 for precision and 0.86-0.90 for recall, again demonstrating the model's reliability.

In the validation dataset, a consistent pattern with TPRs of 0.89 at the 0.54 threshold is observed, indicating strong identification of true building pairs. The FPR remains controlled, suggesting that the model maintains a low rate of incorrectly identifying dissimilar pairs as similar. Precision and Recall metrics are comparable to the training set, ensuring confidence in the model's applicability to unseen data.

The test dataset, which is crucial for assessing the model's generalizability, shows a TPR of 0.87 at the 0.54 threshold and a slight decrease in Precision to 0.82, indicating a modest drop when faced with completely new data. However, the Precision and Recall values suggest that the network's performance is dependable, with only a minor decrease observed in the transition from training/validation to testing conditions.

The ROC and Precision-Recall curves provided in Figure 22 offer a visual representation of the model's performance across different thresholds. The areas under the curves (AUC) for both ROC and Precision-Recall are above 0.9 for the training and validation sets and slightly lower for the test set, reinforcing the robustness of the model. This also exhibits the model's good capability in differentiating both true positives and negatives. The curves exhibit a desirable steepness, especially in the ROC graph, which denotes a high true positive rate across all threshold levels.

In summary, the Siamese Network exhibits strong performance in identifying buildings that are either intact or have been destroyed/changed. The network's high true positive rate ensures that intact buildings are correctly identified as such, while the low false positive rate and high true negative rate confirm its capability to accurately recognize buildings that have undergone changes. These characteristics underscore the network's potential to be an effective component in the workflow for automatic documentation of conflict destruction in high-resolution satellite imagery.

In the evaluation of the Siamese Network for patch similarity comparison, a critical analysis of false predictions has revealed several underlying challenges. The misclassified pairs, presented in Figure 25, exhibit discordance between ground truth labels and the network's predicted similarity scores. Specifically, image pairs 1-4, annotated as 'Similar', have been assigned scores indicative of dissimilarity, while pairs 5-8, labeled as 'Dissimilar', have been erroneously scored as similar.

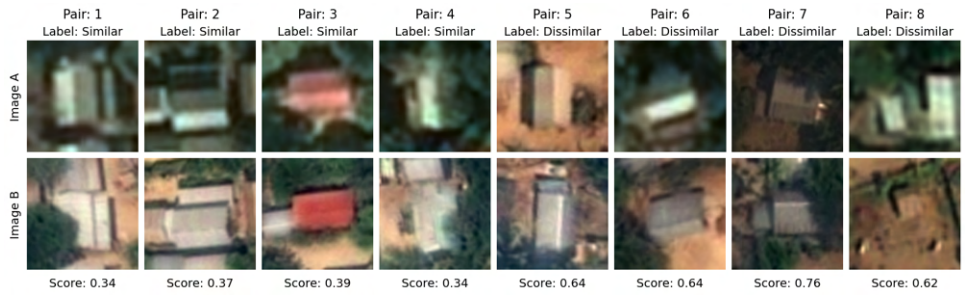

Figure 25: Examples of incorrect classifications by Siamese Network similarity comparison model.

Several contributory factors to these misclassifications could be identified as ambiguous visual features, occlusions and shadows, image quality and resolution, satellite look perspective and angle changes, and background noise. Shared architectural features, such as structural elements and color schemes, may lead to elevated similarity scores in inherently dissimilar building pairs (false positives), particularly in regions with homogenous building designs. Discrepancies in shadow casting or occlusions between temporally separated images can obscure critical building features, resulting in both false negatives and positives during the similarity assessment process. Variations in resolution, focus, and lighting conditions across the bi-temporal images may impair the model's ability to accurately extract and compare pertinent features, thereby

affecting the classification outcome. Alterations in perspective between image captures can modify the apparent structure of buildings, potentially distorting the feature vectors utilized for comparison and contributing to classification errors. Predominant background features may inadvertently influence the feature extraction, leading to a misjudgment in similarity, especially where background features overshadow those of the buildings themselves.

5.1.3 Workflow

In the comprehensive evaluation of an integrated workflow designed for the automatic documentation of conflict-induced destruction in buildings, a two-stage process was meticulously assessed using ground truth annotations for six areas of interest, with each region representing completely unseen data to the models. This rigorous approach is aimed at ensuring the robustness and applicability of the workflow in operational settings.

In the first stage of the evaluation, the Mask R-CNN model demonstrated a high degree of fidelity in detecting buildings from bi-temporal satellite images, as reflected in the substantial agreement with ground truth data presented in Table 5. Despite this, the process was not without its challenges, with certain instances of missed detections, ranging from 20 to 120 per AOI whereas total ground truth buildings range between 300 and 600, warranting further scrutiny. These missed detections are particularly critical, as they may represent instances of destruction that the workflow fails to capture, leading to potential underestimation of damage. Factors contributing to these omissions could include, but are not limited to, variations in building visibility due to occlusions, shadowing effects, the inherent diversity in structural forms, and limitations in image resolution or quality.

The second stage of the workflow focused on change detection through status comparison of building pairs, categorizing them as either intact or destroyed. Here, the Matthews Correlation Coefficient (MCC) was employed as a pivotal metric, providing a balanced measure of the model's performance in identifying true positive and true negative cases. The MCC scores, derived from the evaluation results, indicate a commendable level of accuracy in the binary classification of the buildings' post-event status. A high MCC value across the datasets as presented in Table 6 underscores the model's capability to discern and accurately classify buildings, taking into account the critical need to identify both intact structures and those that have been affected. Visual evidence provided in Figure 24 further illustrates the effectiveness of the workflow.

In conclusion, the evaluation results suggest a robust performance by the integrated workflow. However, they also highlight the necessity for continued enhancement, particularly in addressing the instances of missed detections in the initial extraction phase. Optimizing the model to mitigate these occurrences will be crucial for ensuring comprehensive damage documentation. Furthermore, maintaining high MCC values

in the change detection phase is essential for accurate post-conflict analysis.

5.2 Answering Research Questions

Research Question 1: How does the automated workflow perform in real-world scenarios concerning accuracy, efficiency, and feasibility? The proposed system's feasibility for documenting conflict-induced structural changes is evidenced by the accuracy and efficiency metrics presented. The processing platform, specified in Appendix C, adeptly processed and analyzed an area of 1 km^2 in 90 seconds as presented in Table 6. Such efficiency is noteworthy when contrasted with the time-consuming nature of manual documentation. Additionally, the proposed system is implemented based on open source software libraries and established technology which made it replicable and deployable. The responsibility inherent in documenting conflicts necessitates transparency in the tools, data, and workflows used. To this end, disseminating the implemented workflow as an open-source project, built upon open-source software components, alongside the publication of trained models, guarantees the transparency of the documentation tool. Despite these advantages, the presence of false detections and errors, albeit not significant, cannot be dismissed. These inaccuracies signify that while the system provides a robust tool for rapid assessments, it cannot be deemed infallible. The system's limitations suggest that it should serve as a supplementary tool in a more comprehensive documentation process.

Research Question 2: In what ways does the automated system improve upon existing manual documentation efforts in conflict-affected areas? When considering the enhancement over manual documentation efforts, the proposed system significantly reduces the resource and time expenditure. Specifically, the analysis per area of interest, encompassing hundreds of buildings, demonstrates a high level of agreement with ground truth data. However, the criticality of manual verification is highlighted in instances where the system's automated processes falter. The manual verification step remains essential to rectify and validate the system's output, ensuring the reliability of the final documentation. Thus, the system's role is best realized as an assistive technology that enhances, rather than replaces, human expertise.

Research Question 3: Can the proposed automated workflow be effectively adapted to document building destruction across various geographic and contextual environments? The architectural styles, construction materials, and environmental contexts of buildings vary significantly across different geographic locations and countries. This diversity often restricts the applicability or transferability of deep learning methodologies tailored to specific regions or contexts to other areas. However, the workflow introduced in this thesis is designed with a level of flexibility that allows for its deployment across various geographic settings, including but not limited to Myanmar,

by facilitating the exchange of data pertinent to any given location. Additionally, the incorporation of transfer learning within the deep learning components of the workflow mitigates the need for extensive training datasets. It also demonstrates the system's adaptability to diverse locations, leveraging a pre-trained model trained on data from varying geographic, architectural, and environmental backgrounds distinct from the study area. Crucially, the workflow's strategy for detecting destruction—centered on identifying intact buildings and their changes over time—avoids the necessity for samples of destroyed buildings, which may be scarce and vary across different regions and causes of destruction. The ability to fine-tune the model with moderate amounts of localized data points to the workflow's effective adaptability for geographic regions with different environmental and structural characteristics. Yet, the necessity for localized expertise becomes apparent, ensuring the system's outputs are contextually accurate and relevant. This need for human intervention is particularly pronounced given the system's considerable missed detections and misclassifications, reinforcing the premise that the workflow should be used judiciously as part of a larger toolkit.

5.3 Limitations

In the pursuit of refining the proposed system for the automatic documentation of conflict destruction, several limitations warrant consideration. The system's reliance on very high-resolution satellite imagery implies that any compromise in image quality could impact its performance while any shortage of image availability for a focus area could render the documentation impossible. Challenges such as obstructions from cloud cover, shadows, or dense foliage can obstruct the system's detection capabilities. Additionally, models' generalizability remains a concern; despite the advantages conferred by transfer learning, they may not perform uniformly across all environments. Architectural diversity and unique regional landscapes—particularly those not represented in the training set—pose potential hurdles to the system's applicability. Fine-tuning of models with localized data is a requirement.

Furthermore, the system, although accurate, is not immune to false positives and negatives. These inaccuracies are particularly critical in the realm of conflict damage assessment and documentation intended for justice efforts, where the stakes of misidentification are high. Another notable limitation is scalability; as the system is scaled up to process larger datasets or to span more extensive geographic regions, computational demands escalate, potentially posing logistical and resource-based challenges. The necessity of human oversight also emerges as a primary limitation. The current requirement for manual verification steps to uphold data integrity may impede the system's operational efficiency and scalability. Such a reliance underscores the importance of expert intervention in the validation of automated processes, which, while indispensable, could be seen as a bottleneck to achieving a fully automated workflow.

5.4 Future Works

Looking to the future, several avenues for advancement present themselves. The integration of additional data layers, such as multispectral imagery or temporal data, could significantly enhance the model's detection and classification accuracy. For example, integration of near-infrared band, could be beneficial to highlight and differentiate between man-made structures, vegetation and shadow. Additionally, since most of change detection errors are introduced from variations in color, pixel resolution, brightness and shadow conditions, components that are insensitive to such variations such as Transformers should be explored and introduced in the workflow. Furthermore, the low temporal resolution and high price of very high resolution satellite images may acts as a limitation in certain cases which can be navigated around by exploring a way to transfer models' capability to work on high resolution images with high temporal resolution. For example, GANs can be applied for both domain adaptation of models and refining high resolution images into very high resolution representations.

Customizable models that cater to specific regional characteristics and types of buildings, along with the development of real-time processing capabilities, could immensely benefit timely and responsive decision-making during or immediately after conflict events. Training the system on a more expansive and diverse range of datasets will also be instrumental in improving its generalizability. Lastly, future iterations of the system could aim to categorize not only the occurrence but also the severity of damage, offering a more detailed assessment of the destruction. As outlined in limitations, the need for human oversight is a bottleneck which cannot be completely excluded. An effort to accelerate this human-based review process can be made. Investigating into the development of a supplementary tool dedicated to a streamlined reviewing process is a potential future work. Moreover, in order for the proposed workflow to be accessible and applicable in operational scope, the proposed system will be packaged and redistributed either as a QGIS plugin or a standalone tool. However, current Python implementation of the workflow along with trained models can be found at (Nyan Lin, 2024).

In summation, while the proposed system marks a in leveraging technology for conflict damage documentation, it is clear that continued research and development to make the system operationally robust and ready are crucial. By addressing these limitations and exploring the outlined future work, the system can evolve to more effectively meet the intricate demands of documenting destruction in conflict zones, ultimately aiding humanitarian and justice efforts with greater precision and insight.

6

Chapter 6

6.1 Conclusions

The experimentation and evaluation results display that the development of an automated documentation system based on stable, well-established and publicly available technology is feasible. However, the system, while presenting its capability and efficiency in automating significant portions of the damage documentation workflow, must be approached with caution. Its current form, although deemed satisfactory, is not without errors, and reliance on this system alone could lead to unreliable documentation. Such unreliability is untenable in the critical and sensitive context of conflict destruction assessment. Therefore, the system is recommended as an assistive tool—a means to expedite the documentation process and to handle a substantial workload with satisfactory accuracy. The final step, however, must invariably involve human scrutiny, wherein experts can apply their nuanced understanding and critical evaluation to ensure the integrity and credibility of the documentation. The value of the proposed system lies in its support to these experts, not in its capacity to replace the essential human element of the assessment process. The proposed system, in any regards, can be concluded as a feasible, efficient and useful system which can be replicated and employed for various international causes. The workflow examined in this thesis presents an innovative approach that sidesteps the necessity for training samples specific to destruction. By harnessing transfer learning, it ensures adaptability across diverse geographic and environmental contexts, aiming for operational applicability. This method stands as an alternative to post-classification strategies and others reviewed in the literature, offering a practical solution to the challenges identified.

Biblographic References

Abdulla, W. (2017). *Mask R-CNN for Object Detection and Instance Segmentation on Keras and TensorFlow*. Github. URL: https://github.com/matterport/Mask_RCNN (cit. on pp. 31, 69, 72).

Adriano, B. et al. (2021). "Learning from Multimodal and Multitemporal Earth Observation Data for Building Damage Mapping". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 175, pp. 132–143. ISSN: 09242716 (ISSN). DOI: 10.1016/j.isprsjprs.2021.02.016. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102615227&doi=10.1016%2fj.isprsjprs.2021.02.016&partnerID=40&md5=db46dd13f40a0e3c19bdf60015267b29 (cit. on pp. 5, 7).

Bai, B. et al. (2022). "Edge-Guided Recurrent Convolutional Neural Network for Multitemporal Remote Sensing Image Building Change Detection". In: *IEEE Transactions on Geoscience and Remote Sensing* 60. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2021.3106697. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124390104&doi=10.1109%2fTGRS.2021.3106697&partnerID=40&md5=25aff386413408dd13f639d28ee00668 (cit. on p. 7).

Balkrishna Pandey (2023-06-12). *How Do Neural Networks Make Decisions? A Look at Activation Functions*. Goglides Dev. URL: https://www.goglides.dev/bkpandey/how-do-neural-networks-make-decisions-a-look-at-activation-functions-141e (visited on 2024-02-22) (cit. on p. 11).

Bandara, W. and V. Patel (2022). "A Transformer-Based Siamese Network for Change Detection". In: *Dig Int Geosci Remote Sens Symp (IGARSS)*. 2022 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2022. In collab. with The Institute of Electrical and Electronics Engineers Geoscience and Remote Sensing Society (GRSS). Vol. 2022-July. Institute of Electrical and Electronics Engineers Inc., pp. 207–210. ISBN: 9781665427920 (ISBN). DOI: 10.1109/IGARSS46834.2022.9883686. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140408393&doi=10.1109%2fIGARSS46834.2022.9883686&partnerID=40&md5=3e25894ebe1aa600b5c273425a8c3b8f (visited on 2022-07-17) (cit. on p. 8).

Barrington, L. et al. (2011). "Crowdsourcing Earthquake Damage Assessment Using Remote Sensing Imagery". In: *Annals of Geophysics* 54.6, pp. 680–687. ISSN: 15935213 (ISSN). DOI: 10.4401/ag-5324. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84856005946&doi=10.4401%2fag-5324&partnerID=40&md5=3e8aa5ef45f8f910a9c984afdd51c558 (cit. on p. 5).

Benz, U. C. et al. (2004-01-01). "Multi-Resolution, Object-Oriented Fuzzy Analysis of Remote Sensing Data for GIS-ready Information". In: *ISPRS Journal of Photogrammetry and Remote Sensing*. Integration of Geodata and Imagery for Automated Refinement and Update of Spatial Databases 58.3, pp. 239–258. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2003.10.002. URL: https://www.sciencedirect.com/science/article/pii/S0924271603000601 (visited on 2024-02-09) (cit. on p. 6).

Bossche, J. V. den et al. (2023-06-06). *Geopandas/Geopandas: V0.13.2*. Version v0.13.2. Zenodo. DOI: 10.5281/zenodo.8009629. URL: https://zenodo.org/records/8009629 (visited on 2024-02-22) (cit. on p. 72).

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools* (cit. on p. 72).

Bromley, J. et al. (1993-11-29). "Signature Verification Using a "Siamese" Time Delay Neural Network". In: *Proceedings of the 6th International Conference on Neural Information Processing Systems*. NIPS'93. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 737–744 (cit. on p. 15).

Brunner, D., G. Lemoine, and L. Bruzzone (2010). "Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery". In: *IEEE Transactions on Geoscience and Remote Sensing* 48.5, pp. 2403–2420. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2009.2038274. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77951295197&doi=10.1109%2fTGRS.2009.2038274&partnerID=40&md5=5f275244bc0edd151a2bd266c2b4bd7e (cit. on pp. 4–6).

Chen, H., W. Li, and Z. Shi (2022). "Adversarial Instance Augmentation for Building Change Detection in Remote Sensing Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 60. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2021.3066802. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103290011&doi=10.1109%2fTGRS.2021.3066802&partnerID=40&md5=7d57a49754c3bf2f20ce3dde637f8187 (cit. on p. 7).

Chen, H., C. Wu, B. Du, and L. Zhang (2019). "Deep Siamese Multi-scale Convolutional Network for Change Detection in Multi-Temporal VHR Images". In: *Int. Workshop Anal. Multitemporal Remote Sens. Images, MultiTemp*. 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2019. Institute of Electrical and Electronics Engineers Inc. ISBN: 9781728146157 (ISBN). DOI: 10.1109/Multi-Temp.2019.8866947. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074288310&doi=10.1109%2fMulti-Temp.2019

.8866947&partnerID=40&md5=93ec6d0e45a463b93b39a5f548676b07 (visited on 2019-08-05) (cit. on p. 8).

Chen, H., C. Wu, B. Du, L. Zhang, and L. Wang (2020). "Change Detection in Multisource VHR Images via Deep Siamese Convolutional Multiple-Layers Recurrent Neural Network". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.4, pp. 2848–2864. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2019.2956756. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85082938757&doi=10.1109%2fTGRS.2019.2956756&partnerID=40&md5=4da9b02d8cb0e4f0c1c351b6e9b9453a (cit. on p. 7).

Chen, J. et al. (2021). "DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp. 1194–1206. ISSN: 19391404 (ISSN). DOI: 10.1109/JSTARS.2020.3037893. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097951973&doi=10.1109%2fJSTARS.2020.3037893&partnerID=40&md5=6f6a79c09407042e84a0eeeb25a877ee (cit. on p. 8).

Chen, T. et al. (2022). "A Siamese Network Based U-Net for Change Detection in High Resolution Remote Sensing Images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, pp. 2357–2369. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2022.3157648. URL: https://ieeexplore.ieee.org/document/9733219 (visited on 2024-02-22) (cit. on p. 26).

Chin Human Rights Organization (2022-03-01). *Collective Punishment: Implementation of "Four Cuts"*. International Work Group for Indigenous Affairs, p. 30. URL: https://iwgia.org/en/resources/publications/4615-iwgia-chro-four-cuts.html?highlight=WyJjb2xsZWN0aXZlIiwiY29sbGVjdCIsInB1bmlzaG1lbnQiLCJwdW5pc2giLCJjb2xsZWN0aXZlIHB1bmlzaG1lbnQiXQ== (visited on 2024-02-21) (cit. on p. 1).

COCO Dataset - Evaluation Metrics (2024). Common Objects in Context. URL: https://cocodataset.org/#detection-eval (visited on 2024-02-11) (cit. on p. 47).

Danti, M., S. Branting, and S. Penacho (2017). "The American Schools of Oriental Research Cultural Heritage Initiatives: Monitoring Cultural Heritage in Syria and Northern Iraq by Geospatial Imagery". In: *Geosciences (Switzerland)* 7.4. ISSN: 20763263 (ISSN). DOI: 10.3390/geosciences7040095. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85030700104&doi=10.3390%2fgeosciences7040095&partnerID=40&md5=a4248fe911906d68ad8691d8dfdfc439 (cit. on p. 4).

Datasets (2018-10-01). The SpaceNet Catalog. URL: https://spacenet.ai/datasets/ (visited on 2024-02-11) (cit. on p. 31).

Daudt, R. et al. (2018). "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks". In: *Dig Int Geosci Remote Sens Symp*

(IGARSS). 38th Annual IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018. In collab. with Geoscience and Remote Sensing Society (GRSS); The Institute of Electrical and Electronics Engineers (IEEE). Vol. 2018-July. Institute of Electrical and Electronics Engineers Inc., pp. 2115–2118. ISBN: 9781538671504 (ISBN). DOI: 10.1109/IGARSS.2018.8518015. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062908111&doi=10.1109%2fIGARSS.2018.8518015&partnerID=40&md5=f0e3b33b8b349add3c0af9a0dc0fcdcf (visited on 2018-07-22) (cit. on p. 7).

Democratic Voice of Burma (2024). *DVB | Democratic Voice of Burma*. URL: https://burmese.dvb.no/ (visited on 2024-02-11) (cit. on p. 18).

Developers, T. (2021-05-13). *TensorFlow*. Version v2.5.0. Zenodo. DOI: 10.5281/zenodo.4758419. URL: https://zenodo.org/records/4758419 (visited on 2024-02-22) (cit. on p. 72).

Ding, Q. et al. (2021). "DSA-Net: A Novel Deeply Supervised Attention-Guided Network for Building Change Detection in High-Resolution Remote Sensing Images". In: *International Journal of Applied Earth Observation and Geoinformation* 105. ISSN: 15698432 (ISSN). DOI: 10.1016/j.jag.2021.102591. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85121622559&doi=10.1016%2fj.jag.2021.102591&partnerID=40&md5=83751c05f24b29b750fa02e0f55a19af (cit. on p. 8).

Du, H. et al. (2024). "A Single-Building Damage Detection Model Based on Multi-Feature Fusion: A Case Study in Yangbi". In: *iScience* 27.1. ISSN: 25890042 (ISSN). DOI: 10.1016/j.isci.2023.108586. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179609834&doi=10.1016%2fj.isci.2023.108586&partnerID=40&md5=2c9f985072009acaa6550f9205fdd73a (cit. on p. 5).

ESRI (2022). *ArcGIS Pro*. Version 3.1.3. Redlands, CA: Environmental Systems Research Institute, Inc. (cit. on p. 72).

Fechner, T., D. Wilhelm, and C. Kray (2015). "Ethermap - Real-time Collaborative Map Editing". In: *Conf Hum Fact Comput Syst Proc*. 33rd Annual CHI Conference on Human Factors in Computing Systems, CHI 2015. In collab. with ACM SIGCHI. Vol. 2015-April. Association for Computing Machinery, pp. 3583–3592. ISBN: 9781450331456 (ISBN). DOI: 10.1145/2702123.2702536. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84951058120&doi=10.1145%2f2702123.2702536&partnerID=40&md5=d4b36cca5bcd16d2a085fb1bca0980bb (visited on 2015-04-18) (cit. on p. 5).

Gamanya, R., P. De Maeyer, and M. De Dapper (2009). "Object-Oriented Change Detection for the City of Harare, Zimbabwe". In: *Expert Systems with Applications* 36.1, pp. 571–588. ISSN: 09574174 (ISSN). DOI: 10.1016/j.eswa.2007.09.067. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-53849142529&doi=10.1016%2fj.eswa.2007.09.067&partnerID=40&md5=955414755c87fcb08e328ea825636a43 (cit. on p. 5).

Ghaffarian, S. et al. (2019). "Post-Disaster Building Database Updating Using Automated Deep Learning: An Integration of Pre-Disaster OpenStreetMap and Multi-Temporal Satellite Data". In: *Remote Sensing* 11.20. ISSN: 20724292 (ISSN). DOI: 10.3390/rs11202427. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074215826&doi=10.3390%2frs11202427&partnerID=40&md5=c31262911bca16e75d7019c9c3bcd91a (cit. on p. 7).

Ghosh, S. et al. (2011). "Crowdsourcing for Rapid Damage Assessment: The Global Earth Observation Catastrophe Assessment Network (GEO-CAN)". In: *Earthquake Spectra* 27 (SUPPL. 1), S179–S198. ISSN: 87552930 (ISSN). DOI: 10.1193/1.3636416 (cit. on p. 5).

Gillies, S. et al. (2023-10). *Shapely*. Version 2.0.2. DOI: 10.5281/zenodo.5597138. URL: https://github.com/shapely/shapely (cit. on p. 72).

Girshick, R. et al. (2014-10-22). *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*. DOI: 10.48550/arXiv.1311.2524. arXiv: 1311.2524 [cs]. URL: http://arxiv.org/abs/1311.2524 (visited on 2024-02-03). preprint (cit. on p. 12).

Guttman, A. (1984-06-01). "R-Trees: A Dynamic Index Structure for Spatial Searching". In: *ACM SIGMOD Record* 14.2, pp. 47–57. ISSN: 0163-5808. DOI: 10.1145/971697.602266. URL: https://dl.acm.org/doi/10.1145/971697.602266 (visited on 2024-02-22) (cit. on p. 17).

Hagberg, A. A., D. A. Schult, and P. J. Swart (2008). "Exploring Network Structure, Dynamics, and Function Using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Ed. by G. Varoquaux, T. Vaught, and J. Millman. Pasadena, CA USA, pp. 11–15 (cit. on p. 72).

Harris, C. R. et al. (2020-09). "Array Programming with NumPy". In: *Nature* 585.7825 (7825), pp. 357–362. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: https://www.nature.com/articles/s41586-020-2649-2 (visited on 2024-02-22) (cit. on p. 72).

He, K., G. Gkioxari, et al. (2018-01-24). *Mask R-CNN*. DOI: 10.48550/arXiv.1703.06870. arXiv: 1703.06870 [cs]. URL: http://arxiv.org/abs/1703.06870 (visited on 2024-02-03). preprint (cit. on p. 13).

He, K., X. Zhang, et al. (2015-12-10). *Deep Residual Learning for Image Recognition*. DOI: 10.48550/arXiv.1512.03385. arXiv: 1512.03385 [cs]. URL: http://arxiv.org/abs/1512.03385 (visited on 2024-02-03). preprint (cit. on p. 14).

Hotosm/fAIr (2024-01-27). Humanitarian OpenStreetMap Team. URL: https://github.com/hotosm/fAIr (visited on 2024-02-09) (cit. on pp. 5, 8).

Hu, Y., K. Janowicz, and H. Couclelis (2017). "Prioritizing Disaster Mapping Tasks for Online Volunteers Based on Information Value Theory". In: *Geographical Analysis* 49.2, pp. 175–198. ISSN: 1538-4632. DOI: 10.1111/gean.12117. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gean.12117 (visited on 2024-02-07) (cit. on p. 5).

Huang, F. et al. (2018). "Object-Oriented Change Detection and Damage Assessment Using High-Resolution Remote Sensing Images, Tangjiao Landslide, Three Gorges Reservoir, China". In: *Environmental Earth Sciences* 77.5. ISSN: 18666280 (ISSN). DOI: 10.1007/s12665-018-7334-5. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042635192&doi=10.1007%2fs12665-018-7334-5&partnerID=40&md5=5875323607f38b1fbaf9ea49744b67d4 (cit. on p. 5).

Huang, J. et al. (2022). "Multiple Attention Siamese Network for High-Resolution Image Change Detection". In: *IEEE Transactions on Geoscience and Remote Sensing* 60, pp. 1–16. ISSN: 1558-0644. DOI: 10.1109/TGRS.2021.3127580. URL: https://ieeexplore.ieee.org/document/9612208 (visited on 2024-02-22) (cit. on p. 26).

Human Rights Watch (2023-01-12). "Myanmar: Events of 2022". In: *World Report 2023*. URL: https://www.hrw.org/world-report/2023/country-chapters/myanmar (cit. on p. 1).

Im, J. and J. R. Jensen (2005-11-30). "A Change Detection Model Based on Neighborhood Correlation Image Analysis and Decision Tree Classification". In: *Remote Sensing of Environment* 99.3, pp. 326–340. ISSN: 0034-4257. DOI: 10.1016/j.rse.2005.09.008. URL: https://www.sciencedirect.com/science/article/pii/S0034425705002919 (visited on 2024-02-09) (cit. on p. 6).

Jensen, J. and J. Im (2007). "Remote Sensing Change Detection in Urban Environments". In: *Geo-Spat. Technol. in Urban Environ. (Second Edition): Plcy., Pract., and Pixels*. Springer Berlin Heidelberg, pp. 7–31. ISBN: 3540222634 (ISBN). DOI: 10.1007/978-3-540-69417-5_2. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-77954407984&doi=10.1007%2f978-3-540-69417-5_2&partnerID=40&md5=34288a97f5c05cd1b20eeb6d173859a4 (cit. on p. 4).

Jiang, H. et al. (2022). "A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images". In: *Remote Sensing* 14.7. ISSN: 20724292 (ISSN). DOI: 10.3390/rs14071552. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85127538755&doi=10.3390%2frs14071552&partnerID=40&md5=218cfa87261e51ee8ee7d5c78aaac03b (cit. on pp. 6, 7).

Kemper, T. et al. (2011-03). "Enumeration of Dwellings in Darfur Camps From GeoEye-1 Satellite Images Using Mathematical Morphology". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.1, pp. 8–15. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2010.2053700. URL: https://ieeexplore.ieee.org/document/5546897 (visited on 2024-02-07) (cit. on p. 5).

Khit Thit Media (2024-02-11). *Yangon Khit Thit Media*. Khit Thit Media. URL: https://yktnews.com/ (visited on 2024-02-11) (cit. on p. 18).

Al-Khudhairy, D., I. Caravaggi, and S. Giada (2005). "Structural Damage Assessments from Ikonos Data Using Change Detection, Object-Oriented Segmentation, and Classification Techniques". In: *Photogrammetric Engineering and Remote Sensing* 71.7, pp. 825–837. ISSN: 00991112 (ISSN). DOI: 10.14358/PERS.71.7.825. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-31344454018&doi=10.143

58%2fPERS.71.7.825&partnerID=40&md5=d6e335a354f3256094cb0a10423c5107 (cit. on p. 4).

Klonus, S. et al. (2012). "Combined Edge Segment Texture Analysis for the Detection of Damaged Buildings in Crisis Areas". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.4, pp. 1118–1128. ISSN: 19391404 (ISSN). DOI: 10.1109/JSTARS.2012.2205559 (cit. on pp. 4–6).

Knoth, C. and E. Pebesma (2017-01-02). "Detecting Dwelling Destruction in Darfur through Object-Based Change Analysis of Very High-Resolution Imagery". In: *International Journal of Remote Sensing* 38.1, pp. 273–295. ISSN: 0143-1161. DOI: 10.1080/01431161.2016.1266105. URL: https://doi.org/10.1080/01431161.2016.1266105 (visited on 2024-02-07) (cit. on p. 4).

Knoth, C., S. Slimani, et al. (2018-08-01). "Combining Automatic and Manual Image Analysis in a Web-Mapping Application for Collaborative Conflict Damage Assessment". In: *Applied Geography* 97, pp. 25–34. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2018.05.016. URL: https://www.sciencedirect.com/science/article/pii/S014362281730783X (visited on 2024-02-07) (cit. on p. 5).

Konstantinidis, D. et al. (2017). "Building Detection Using Enhanced HOG-LBP Features and Region Refinement Processes". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.3, pp. 888–905. ISSN: 19391404 (ISSN). DOI: 10.1109/JSTARS.2016.2602439. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85027435893&doi=10.1109%2fJSTARS.2016.2602439&partnerID=40&md5=7098bf5671d4c68cad41851bc9f98dd0 (cit. on p. 5).

Leichtle, T. et al. (2017-02-01). "Unsupervised Change Detection in VHR Remote Sensing Imagery – an Object-Based Clustering Approach in a Dynamic Urban Environment". In: *International Journal of Applied Earth Observation and Geoinformation* 54, pp. 15–27. ISSN: 1569-8432. DOI: 10.1016/j.jag.2016.08.010. URL: https://www.sciencedirect.com/science/article/pii/S0303243416301490 (visited on 2024-02-09) (cit. on p. 5).

Li, Y. et al. (2019). "Building Damage Detection from Post-Event Aerial Imagery Using Single Shot Multibox Detector". In: *Applied Sciences (Switzerland)* 9.6. ISSN: 20763417 (ISSN). DOI: 10.3390/app9061128 (cit. on p. 7).

Lin, T.-Y. et al. (2017-04-19). *Feature Pyramid Networks for Object Detection*. DOI: 10.48550/arXiv.1612.03144. arXiv: 1612.03144 [cs]. URL: http://arxiv.org/abs/1612.03144 (visited on 2024-02-03). preprint (cit. on p. 14).

Liu, M. et al. (2022). "Super-Resolution-Based Change Detection Network with Stacked Attention Module for Images with Different Resolutions". In: *IEEE Transactions on Geoscience and Remote Sensing* 60. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2021.3091758. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85112178949&doi=10.1109%2fTGRS.2021.3091758&partnerID=40&md5=48703e9bc5766adbf461bd0008427191 (cit. on p. 4).

Liu, W. et al. (2016). "SSD: Single Shot MultiBox Detector". In: vol. 9905, pp. 21–37. DOI: 10.1007/978-3-319-46448-0_2. arXiv: 1512.02325 [cs]. URL: http://arxiv.org/abs/1512.02325 (visited on 2024-02-11) (cit. on p. 7).

Liu, Y. et al. (2021). "Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model". In: *IEEE Geoscience and Remote Sensing Letters* 18.5, pp. 811–815. ISSN: 1545598X (ISSN). DOI: 10.1109/LGRS.2020.2988032. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099104166&doi=10.1109%2fLGRS.2020.2988032&partnerID=40&md5=a395f223b6bc602aacabc7f09c1c321d (cit. on p. 8).

Lourenço, J. M. (2021). *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. URL: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. v).

Ma, H. et al. (2020). "Improved CNN Classification Method for Groups of Buildings Damaged by Earthquake, Based on High Resolution Remote Sensing Images". In: *Remote Sensing* 12.2. ISSN: 20724292 (ISSN). DOI: 10.3390/rs12020260. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081093363&doi=10.3390%2frs12020260&partnerID=40&md5=c90163a7fac0c132c184ac8a7b30c1d3 (cit. on p. 7).

Marin, C., F. Bovolo, and L. Bruzzone (2015). "Building Change Detection in Multitemporal Very High Resolution SAR Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.5, pp. 2664–2682. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2014.2363548. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84921032920&doi=10.1109%2fTGRS.2014.2363548&partnerID=40&md5=c0955de52a503a245174178ffa00f4b0 (cit. on p. 4).

Meier, P. (2013). "Human Computation for Disaster Response". In: *Handb. of Human Comput.* Springer New York, pp. 95–104. ISBN: 9781461488064 (ISBN). DOI: 10.1007/978-1-4614-8806-4_11. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85028856775&doi=10.1007%2f978-1-4614-8806-4_11&partnerID=40&md5=129ffef6aeba2bd9aa82d71082225468 (cit. on p. 5).

Mizzima (2024). *Mizzima*. URL: https://eng.mizzima.com/ (visited on 2024-02-11) (cit. on p. 18).

Mohanty, S. P. (2018). *CrowdAI Mapping Challenge 2018 : Baseline with Mask RCNN*. Version bac1cf19adbc9d078122c6933da6f808c4ee590d. URL: https://github.com/crowdAI/crowdai-mapping-challenge-mask-rcnn (visited on 2024-02-11) (cit. on p. 31).

Mohanty, S. P. et al. (2020). "Deep Learning for Understanding Satellite Imagery: An Experimental Survey". In: *Frontiers in Artificial Intelligence* 3. ISSN: 2624-8212. URL: https://www.frontiersin.org/articles/10.3389/frai.2020.534696 (visited on 2024-02-11) (cit. on p. 31).

Moya, L. et al. (2019-03-01). "3D Gray Level Co-Occurrence Matrix and Its Application to Identifying Collapsed Buildings". In: *ISPRS Journal of Photogrammetry and Remote*

Sensing 149, pp. 14–28. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2019.01.008. URL: https://www.sciencedirect.com/science/article/pii/S0924271619300085 (visited on 2024-02-09) (cit. on p. 6).

Myanmar Information Management Unit (2024). MIMU. URL: https://themimu.info/ (visited on 2024-02-11) (cit. on p. 20).

Myanmar Now (2024). *Myanmar Now*. Myanmar Now. URL: https://myanmar-now.org/en/ (visited on 2024-02-11) (cit. on p. 18).

Myanmar Witness (2024). *Fire Map*. Myanmar Witness. URL: https://www.myanmarwitness.org/fire-map (visited on 2024-02-11) (cit. on p. 18).

NASA (2024). *NASA-FIRMS*. URL: https://firms.modaps.eosdis.nasa.gov/map/ (visited on 2024-02-11) (cit. on p. 18).

Nyan Lin, N. N. (2024-02-26). *Automated Conflict Building Destruction Documentation Toolkit*. Version pre-release. Zenodo. DOI: 10.5281/ZENODO.10689997. URL: https://zenodo.org/doi/10.5281/zenodo.10689997 (visited on 2024-02-26) (cit. on pp. 52, 71).

Pagot, E. and M. Pesaresi (2008-06). "Systematic Study of the Urban Postconflict Change Classification Performance Using Spectral and Structural Features in a Support Vector Machine". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 1.2, pp. 120–128. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2008.2001154. URL: https://ieeexplore.ieee.org/document/4624555 (visited on 2024-02-07) (cit. on p. 6).

Parks, L. (2009). "Digging into Google Earth: An Analysis of "Crisis in Darfur"". In: *Geoforum* 40.4, pp. 535–545. ISSN: 00167185 (ISSN). DOI: 10.1016/j.geoforum.2009.04.004. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-67649945827&doi=10.1016%2fj.geoforum.2009.04.004&partnerID=40&md5=d3472f61deb21726a5fcc0c6bf87b334 (cit. on p. 4).

Pedregosa, F. et al. (2011). "Scikit-Learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830 (cit. on p. 72).

Place Codes (Pcodes) | MIMU (2024). URL: https://themimu.info/place-codes (visited on 2024-02-11) (cit. on p. 20).

Planet Team (2017). *Planet Application Program Interface: In Space for Life on Earth*. San Francisco, CA. URL: https://api.planet.com (cit. on p. 18).

Poiani, T. H. et al. (2016-01). "Potential of Collaborative Mapping for Disaster Relief: A Case Study of OpenStreetMap in the Nepal Earthquake 2015". In: *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2016 49th Hawaii International Conference on System Sciences (HICSS), pp. 188–197. DOI: 10.1109/HICSS.2016.31. URL: https://ieeexplore.ieee.org/document/7427206 (visited on 2024-02-07) (cit. on p. 5).

Prins, E. (2008-02-01). "Use of Low Cost Landsat ETM+ to Spot Burnt Villages in Darfur, Sudan". In: *International Journal of Remote Sensing* 29.4, pp. 1207–1214. ISSN:

0143-1161. DOI: 10.1080/01431160701730110. URL: https://doi.org/10.1080/01431160701730110 (visited on 2024-02-09) (cit. on p. 5).

QGIS Development Team (2009). *QGIS Geographic Information System*. manual. Open Source Geospatial Foundation. URL: http://qgis.osgeo.org (cit. on p. 72).

Qin, R. (2014). "Change Detection on LOD 2 Building Models with Very High Resolution Spaceborne Stereo Imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 96, pp. 179–192. ISSN: 09242716 (ISSN). DOI: 10.1016/j.isprsjprs.2014.07.007 (cit. on p. 5).

Rasterio Github Organization (2024). *Rasterio: Access to Geospatial Raster Data — Rasterio Documentation*. Rasterio Github Organization. URL: https://rasterio.readthedocs.io/en/stable/index.html (visited on 2024-02-22) (cit. on p. 72).

Redmon, J. et al. (2016-05-09). *You Only Look Once: Unified, Real-Time Object Detection*. DOI: 10.48550/arXiv.1506.02640. arXiv: 1506.02640 [cs]. URL: http://arxiv.org/abs/1506.02640 (visited on 2024-02-11). preprint (cit. on p. 7).

Ren, S. et al. (2016-01-06). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. DOI: 10.48550/arXiv.1506.01497. arXiv: 1506.01497 [cs]. URL: http://arxiv.org/abs/1506.01497 (visited on 2024-02-03). preprint (cit. on p. 13).

Ronneberger, O., P. Fischer, and T. Brox (2015-05-18). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. DOI: 10.48550/arXiv.1505.04597. arXiv: 1505.04597 [cs]. URL: http://arxiv.org/abs/1505.04597 (visited on 2024-02-11). preprint (cit. on p. 7).

Saha, S., F. Bovolo, and L. Bruzzone (2021). "Building Change Detection in VHR SAR Images via Unsupervised Deep Transcoding". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.3, pp. 1917–1929. ISSN: 01962892 (ISSN). DOI: 10.1109/TGRS.2020.3000296 (cit. on p. 7).

Simon J.D. Prince (2023). *Understanding Deep Learning*. MIT Press. ISBN: 978-0-262-04864-4. URL: http://udlbook.com (cit. on p. 10).

Sun, X. et al. (2022-04-01). "Two-Stage Aware Attentional Siamese Network for Visual Tracking". In: *Pattern Recognition* 124, p. 108502. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2021.108502. URL: https://www.sciencedirect.com/science/article/pii/S0031320321006786 (visited on 2024-02-22) (cit. on p. 26).

Team, H. O. (2024). *HOT Tasking Manager*. HOT Tasking Manager. URL: https://tasks.hotosm.org (visited on 2024-02-11) (cit. on p. 5).

team, T. pandas development (2020-02). *Pandas-Dev/Pandas: Pandas*. Version latest. Zenodo. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134 (cit. on p. 72).

The Irrawaddy (2024). *The Irrawaddy - Covering Burma and Southeast Asia*. The Irrawaddy. URL: https://www.irrawaddy.com/ (visited on 2024-02-11) (cit. on p. 18).

Toblerity (2024). *Toblerity/Rtree: Rtree: Spatial Index for Python GIS*. URL: https://github.com/Toblerity/rtree (visited on 2024-02-22) (cit. on p. 72).

Tomowski, D. et al. (2010-07-01). "Change Visualization through a Texture-Based Analysis Approach for Disaster Applications". In: vol. 38, CDROM (cit. on p. 4).

Uhrig, R. (1995). "Introduction to Artificial Neural Networks". In: *Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics*. IECON '95 - 21st Annual Conference on IEEE Industrial Electronics. Vol. 1. Orlando, FL, USA: IEEE, pp. 33–37. ISBN: 978-0-7803-3026-9. DOI: 10.1109/IECON.1995.483329. URL: http://ieeexplore.ieee.org/document/483329/ (visited on 2024-02-22) (cit. on p. 9).

UN High Commissioner for Human Rights (2023-03-02). "Situation of Human Rights in Myanmar since 1 February 2022 :: Report of the United Nations High Commissioner for Human Rights". In: URL: https://digitallibrary.un.org/record/4019889 (cit. on p. 1).

UN Independent Investigative Mechanism for Myanmar (2023-06-30). *Report of the Independent Investigative Mechanism for Myanmar*. A/HRC/54/19. UN, URL: https://digitallibrary.un.org/record/4017802 (visited on 2024-02-21) (cit. on p. 1).

UNOSAT S-1 FloodAI Monitoring Dashboard (2024). United Nations Satellite Centre UNOSAT | UNITAR. URL: https://unosat-rm.cern.ch/FloodAI/apps/BGD/ (visited on 2024-02-09) (cit. on p. 5).

Ushahidi (2024). Ushahidi. URL: https://www.ushahidi.com// (visited on 2024-02-11) (cit. on p. 5).

Valentijn, T. et al. (2020). "Multi-Hazard and Spatial Transferability of a CNN for Automated Building Damage Assessment". In: *Remote Sensing* 12.17, pp. 1–29. ISSN: 20724292 (ISSN). DOI: 10.3390/rs12172839. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092714373&doi=10.3390%2frs12172839&partnerID=40&md5=bf76055b5930c1daecd1338a9c481b45 (cit. on p. 7).

Van Rossum, G. and J. de Boer (1991-12-01). "Interactively Testing Remote Servers Using the Python Programming Language". In: *CWI Quarterly* 4.4, pp. 283–304. ISSN: 0922-5366. URL: https://ir.cwi.nl/pub/18204 (visited on 2024-02-22) (cit. on p. 72).

Vetrivel, A. et al. (2018). "Disaster Damage Detection through Synergistic Use of Deep Learning and 3D Point Cloud Features Derived from Very High Resolution Oblique Aerial Images, and Multiple-Kernel-Learning". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 140, pp. 45–59. ISSN: 09242716 (ISSN). DOI: 10.1016/j.isprsjprs.2017.03.001. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85014668630&doi=10.1016%2fj.isprsjprs.2017.03.001&partnerID=40&md5=c55c6594390a87f3387aa8eaf4592348 (cit. on p. 5).

Voigt, S. et al. (2011). "Rapid Damage Assessment and Situation Mapping: Learning from the 2010 Haiti Earthquake". In: *Photogrammetric Engineering and Remote Sensing* 77.9, pp. 923–931. ISSN: 00991112 (ISSN). DOI: 10.14358/PERS.77.9.923. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80052766644

&doi=10.14358%2fPERS.77.9.923&partnerID=40&md5=ec591ea6ccaf0e31907a0b8ae48fb53a (cit. on p. 4).

Wang, M. et al. (2020). "A Deep Siamese Network with Hybrid Convolutional Feature Extraction Module for Change Detection Based on Multi-Sensor Remote Sensing Images". In: *Remote Sensing* 12.2. ISSN: 20724292 (ISSN). DOI: 10.3390/rs12020205. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081083982&doi=10.3390%2frs12020205&partnerID=40&md5=bed7fd40d52d0cbb6773b77cac33b65c (cit. on p. 8).

Weir, D., D. McQuillan, and R. Francis (2019). "Civilian Science: The Potential of Participatory Environmental Monitoring in Areas Affected by Armed Conflicts". In: *Environmental Monitoring and Assessment* 191.10. ISSN: 01676369 (ISSN). DOI: 10.1007/s10661-019-7773-9. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071896724&doi=10.1007%2fs10661-019-7773-9&partnerID=40&md5=d3d1bc848d72f78a47e1e813bd320bfd (cit. on p. 5).

Wen, Q. et al. (2019). "Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network". In: *Sensors (Switzerland)* 19.2. ISSN: 14248220 (ISSN). DOI: 10.3390/s19020333. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85060140240&doi=10.3390%2fs19020333&partnerID=40&md5=068e26660ee156339435d19516a226bf (cit. on p. 7).

Witmer, F. D. W. (2015-05-03). "Remote Sensing of Violent Conflict: Eyes from Above". In: *International Journal of Remote Sensing* 36.9, pp. 2326–2352. ISSN: 0143-1161. DOI: 10.1080/01431161.2015.1035412. URL: https://doi.org/10.1080/01431161.2015.1035412 (visited on 2024-02-07) (cit. on p. 4).

Woo, S. et al. (2018-07-18). *CBAM: Convolutional Block Attention Module*. Version 2. DOI: 10.48550/arXiv.1807.06521. arXiv: 1807.06521 [cs]. URL: http://arxiv.org/abs/1807.06521 (visited on 2024-02-03). preprint (cit. on pp. 16, 17).

Xiao, P. et al. (2016-09-01). "Change Detection of Built-up Land: A Framework of Combining Pixel-Based Detection and Object-Based Recognition". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 119, pp. 402–414. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2016.07.003. URL: https://www.sciencedirect.com/science/article/pii/S092427161630168X (visited on 2024-02-09) (cit. on pp. 5, 6).

Yuan, F. and R. Liu (2018). "Feasibility Study of Using Crowdsourcing to Identify Critical Affected Areas for Rapid Damage Assessment: Hurricane Matthew Case Study". In: *International Journal of Disaster Risk Reduction* 28, pp. 758–767. ISSN: 22124209 (ISSN). DOI: 10.1016/j.ijdrr.2018.02.003. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042073366&doi=10.1016%2fj.ijdrr.2018.02.003&partnerID=40&md5=c0bf55b9dc7f954eecf35bf4115a17ae (cit. on p. 5).

Yusuf, Y., M. Matsuoka, and F. Yamazaki (2001). "Damage Assessment after 2001 Gujarat Earthquake Using Landsat-7 Satellite Images". In: *Journal of the Indian Society of Remote Sensing* 29.1-2, pp. 17–22. ISSN: 0255660X (ISSN). DOI: 10.1007/BF02989909 (cit. on p. 4).

Zhao, Z., C. Fan, and L. Liu (2023-07-27). *Geo SAM: A QGIS Plugin Using Segment Anything Model (SAM) to Accelerate Geospatial Image Segmentation*. Version 1.1.0. Zenodo. DOI: 10.5281/zenodo.8191039. URL: https://zenodo.org/records/8191039 (visited on 2024-02-11) (cit. on pp. 27, 72).

Zheng, H. et al. (2022). "HFA-Net: High Frequency Attention Siamese Network for Building Change Detection in VHR Remote Sensing Images". In: *Pattern Recognition* 129. ISSN: 00313203 (ISSN). DOI: 10.1016/j.patcog.2022.108717. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129262536&doi=10.1016%2fj.patcog.2022.108717&partnerID=40&md5=c7c5f7105879f0d922501468597cb778 (cit. on p. 8).

Zheng, Z. et al. (2022). "ChangeMask: Deep Multi-Task Encoder-Transformer-Decoder Architecture for Semantic Change Detection". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 183, pp. 228–239. ISSN: 09242716 (ISSN). DOI: 10.1016/j.isprsjprs.2021.10.015. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85119995073&doi=10.1016%2fj.isprsjprs.2021.10.015&partnerID=40&md5=9d1f9da4e7d31039b79ec09782e18cdc (cit. on p. 8).

Zhou, Z., J. Gong, and M. Guo (2016). "Image-Based 3D Reconstruction for Posthurricane Residential Building Damage Assessment". In: *Journal of Computing in Civil Engineering* 30.2. ISSN: 08873801 (ISSN). DOI: 10.1061/(ASCE)CP.1943-5487.0000480. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959017655&doi=10.1061%2f%28ASCE%29CP.1943-5487.0000480&partnerID=40&md5=0cafcaa333dc3a9bcbba291a25547af9 (cit. on p. 5).

A

Study Area and Data

Table A1 and Table A2 detail the unique place codes, village name, coordinates, event dates, image acquired dates and satellite sources of image for areas of interest involved in this thesis work.

PCODE	Village Name	Township	Region State	Latitude (N)	Longitude (E)
166804	Se Zin	Hpakant	Kachin	25.32962	95.96509
171081	Pa Du	Sagaing	Sagaing	22.09759	95.92734
171241	Na Nwin Kaing	Myaung	Sagaing	21.67091	95.25502
171240	Shwe Ta Chaung	Myaung	Sagaing	21.68133	95.2493
171242	Su Lay Kone	Myaung	Sagaing	21.6688	95.26508
172578	Na Myar	Tabayin	Sagaing	22.63742	95.44688
173038	Hta Naung Taw (South)	Monywa	Sagaing	22.04531	95.18803
173494	Ma Hti Thar	Chaung-U	Sagaing	21.87292	95.26181
173512	Inn Ma	Chaung-U	Sagaing	21.94339	95.21317
173516	Ta Pei Inn	Chaung-U	Sagaing	21.93832	95.19591
173908	Sar Htone	Salingyi	Sagaing	21.99043	95.11454
173913	Htan Zin	Salingyi	Sagaing	22.01357	95.11932
173971	Hpaung Ka Tar (North)	Salingyi	Sagaing	22.02902	95.10939
173978	Se Te - Zee Taw	Salingyi	Sagaing	22.04727	95.09101
175981	Bin	Mingin	Sagaing	22.83414	94.71625
188837	Kinma	Pauk	Magway	21.26701	94.38808
188848	Ohn Taw	Pauk	Magway	21.44122	94.51079
188870	Gway Gon	Pauk	Magway	21.5585	94.43727
188876	Sa Thein	Pauk	Magway	21.57209	94.44043
188877	Nyaung Yin	Pauk	Magway	21.58858	94.43562
188943	Lel Beit (Taung Su)	Pauk	Magway	21.46618	94.61146

Table A1: Areas of interest or villages involved in the dataset and the study with their respective information and coordinates.

PCODE	Image Source (Pre-event)	Image Source (Post-event)	Event Date	Pre-event Image Date	Pre-event Image Date
166804	PL	PL	08/10/22	04/18/21	11/24/22
171081	PL	PL	05/22/23	02/23/23	10/21/23
171241	PLN	PLN	11/07/22	11/26/21	11/12/22
171240	PLN	PLN	11/07/22	11/26/21	11/12/22
171242	PLN	PLN	11/07/22	11/26/21	11/12/22
172578	PL	PLN	12/06/22	10/09/22	12/08/22
173038	PLN	PLN	03/15/23	01/11/23	04/27/23
173494	PL	PLN	10/15/22	01/09/22	01/29/23
173512	PL	PLN	04/23/22	11/25/21	04/27/23
173516	PL	PLN	05/28/22	11/25/21	04/27/23
173908	PL	PLN	12/01/22	11/04/22	04/27/23
173913	PL	PL	11/30/22	11/04/22	02/22/23
173971	PL	PLN	01/10/23	11/04/22	04/27/23
173978	PLN	PLN	01/13/23	01/11/23	04/27/23
175981	PL	PLN	01/31/22	11/30/21	01/24/23
188837	PL	PL	06/15/21	02/24/21	06/21/21
188848	PLN	PLN	03/09/23	03/04/23	03/30/23
188870	PL	PLN	11/28/22	12/26/21	02/06/23
188876	PL	PLN	11/23/22	12/26/21	01/11/23
188877	PL	PLN	11/28/22	12/26/21	01/11/23
188943	PL	PLN	12/30/22	04/05/21	03/30/23

Table A2: VHR satellite imagery sources, reported event and image acquired dates per AOI.

B

DL Model Training Configurations

Table B1 details the configurations associated with training processes of Mask R-CNN and Siamese Network models. Hyper-parameters, configurations and architecture of Mask R-CNN model training are not described here since the default implementation by (Abdulla, 2017) is used. For Siamese Network model, a margin value of 1 is used for the contrastive loss function whose architecture is described in Figure B1 and Figure B2.

Model	Loss Function	Optimizer	Epochs	LR	Layers
Mask R-CNN	Cross Entropy (Binary, Sparse Categorical), Smooth L1	SDG	30	1E-03	Heads (RPN, Object Detection, Mask Regression)
			60	1E-04	ResNet101 Layer 4 and above
			80	1E-04	ResNet101 Layer 3 and above
			160	1E-05	All
Siamese Network	Contrastive	RMSProp	45	1E-05	All except ResNet101 and FPN

Table B1: Training configurations for Mask R-CNN and Siamese Network models.

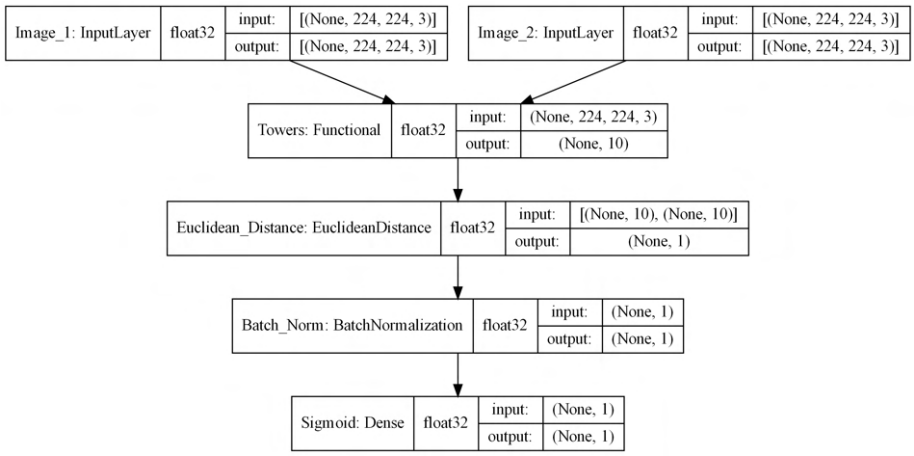

Figure B1: Overview architecture of Siamese Network model. Inputs to two feature extraction towers are two image patches while one Sigmoid activation neuron outputs the similarity score between two inputs.

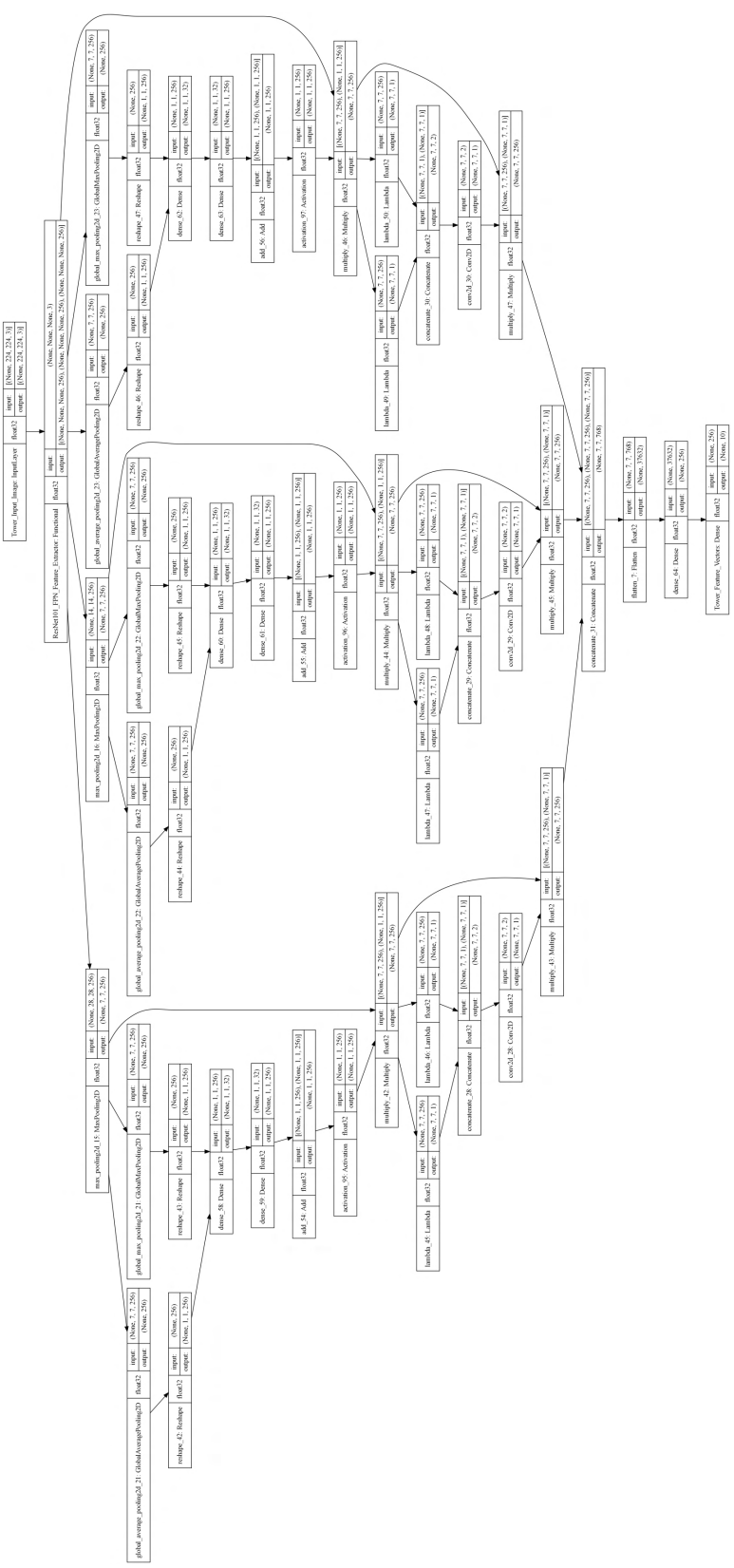

Figure B2: Detailed architecture of the feature extraction tower of Siamese Network model with CBAMs integrated.

C

Technical Specifications and Experimental Setup

The implementation of the workflow as well as complementary components proposed in this thesis work is based on open source software components for the transparency, reproducibility, and accessibility. Table C1 details the list of these components involved. Implemented workflow and supplementary tools for data processing, model training and evaluation as Python scripts as well as Mask R-CNN and Siamese Network models and weights as outputs from this thesis work can be can be found in this repository (Nyan Lin, 2024).

All processes were run in a Microsoft Windows 11 Pro machine, equipped with Intel Core(TM) i7-10750H @ 2.60GHz CPU with 48GB RAM and Nvidia Quadro T1000 Ti GPU with 4GB graphic RAM and 896 CUDA cores.

Software	Source	Version	Purpose
ArcGIS Pro	(ESRI, 2022)	3.1.3	Raster processing and georeferencing
QGIS	(QGIS Development Team, 2009)	3.32.0	Data annotation and vector features manipulation
Geo-SAM	(Zhao, Fan, and L. Liu, 2023)	1.1.1	Data annotation
Python	(van Rossum and de Boer, 1991)	3.8.18	Implementation language
TensorFlow	(Developers, 2021)	2.5.0	Deep learning model training
Scikit-Learn	(Pedregosa et al., 2011)	1.3.2	Model evaluation and metrics
Mask R-CNN	(Abdulla, 2017)	2.1	Building extraction model training
OpenCV	(Bradski, 2000)	4.8.1	Image processing and manipulation
Raserio	(Rasterio Github Organization, 2024)	1.3.9	Raster image processing and manipulation
Shapely	(Gillies et al., 2023)	2.0.2	Polygon feature processing and manipulation
Numpy	(Harris et al., 2020)	1.19.5	Numerical and mathematical operations on arrays
Pandas	(team, 2020)	1.3.4	Data manipulation and analysis
Geopandas	(Bossche et al., 2023)	0.13.2	Geospatial data processing and analysis
Networkx	(Hagberg, Schult, and Swart, 2008)	3.1	Graph construction and network analysis
R-tree	(Toblerity, 2024)	1.1.0	Spatial indexing for optimization

Table C1: Software components and versions utilized in this thesis work.

Masters Program in **Geospatial Technologies**