# LARGE-SCALE BUILDING FOOTPRINT EXTRACTION AND DAMAGE LEVEL CLASSIFICATION: DEEP LEARNING APPROACHES WITH SATELLITE IMAGERY AND LIDAR DATA

**Chang Liu**

A Thesis in Fulfilment of the Requirements for the Degree of

Doctor of Philosophy


School of Civil and Environmental Engineering

Faculty of Engineering

The University of New South Wales, Sydney, Australia

September 2023

Thesis submission for the degree of Doctor of Philosophy

Thesis Title and Abstract | Declarations | Inclusion of Publications Statement | Corrected Thesis and Responses

ORIGINALITY STATEMENT

☑ I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

☑ I hereby grant the University of New South Wales or its agents a non-exclusive licence to arcl
thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public ver

☑ I certify

## Thesis Details

Thesis Title and Abstract | Originality Statement | Inclusion of Publications Statement | Thesis Submission

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **some of the work described in their thesis has been published and has been documented in the relevant Chapters with acknowledgement**.

A short statement on where this work appears in the thesis and how this work is acknowledged within chapter/s:

Chapter 4 of the thesis is partially adopted from " A Novel Attention-Based Deep Learning Method for Post-Disaster Building Damage Classification " in Expert Systems with Applications in which I was the first author. I designed and conducted the experiments under the supervision of the other authors and drafted the paper. I acknowledge all co-authors of the work, including Sepasgozar S, Zhang Q, and Ge L.
Chapter 5 of the thesis is partially adopted from "The Influence of Changing Features of Point Clouds on the Accuracy of Deep Learning-based Large-scale Outdoor Lidar Semantic Segmentation" which has been accepted for publication in 2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). I was the first author of the article. I designed and conducted the experiments, and wrote and revised the paper with the assistance of co-authors. I acknowledge all co-authors of the work, including Zhang Q, Shirowzhan S, Bai T, Sheng Z, Wu Y, Kuang J, Ge L.
Chapter 6 of the thesis is partially adopted from "Channel Attention and Normal-based Local Feature Aggregation Network (CNLNet): A Novel Deep Learning Method for Pre-Disaster Large-scale Outdoor Lidar Semantic Segmentation" which has been submitted to IEEE Transactions on Geoscience and Remote Sensing. I was the first author of the article. I designed and conducted the experiments, and wrote and revised the paper with the assistance of co-authors. I acknowledge all co-authors of the work, including Ge L, Xiang W, Du Z, and Zhang Q.
Chapter 7 of the thesis is partially adopted from "Rapid Large-scale Building Damage Level Classification after Earthquakes using Deep Learning with Lidar and Satellite Optical Data" which has been submitted
I designed and conducted the experiments, and wrote and revised the paper with the assistance of co-authors. I acknowledge all co-authors of the work

Acknowledgements of these works have been made in the footnotes of their corresponding chapters in the thesis.

### Candidate's Declaration

**I declare that I have complied with the Thesis Examination Procedure.**

# Acknowledgement

I would like to express my gratitude to several collaborating scholars. I also would like to thank the opportunity of studying at Massachusetts Institute of Technology (MIT) Senseable City Lab during my Ph.D. candidature. This experience has been an invaluable journey in my life. Furthermore, I would like to express my gratitude for the financial support provided by the SmartSat CRC, whose activities are funded by the Australian Government's CRC Program. I was also lucky to have received GRS DRTG and the school's conference funding for my international conference attendance. Moreover, thanks to all Civil and Environmental Engineering Research Students Association (CERSA) executive members who voluntarily organised several events together with me for HDR students from our school. I am also deeply grateful to many others for their support and contributions, although there is not enough space to list them all.

I would also be grateful to the Operational Satellite Applications Programme, Joint Research Centre, and World Bank for the 2010 Haiti building damage assessment shapefiles. I would like to thank OpenTopography for Lidar data and the Geospatial Information Authority of Japan (GSI) for the 2016 Kumamoto land cover shapefiles. I would express my gratitude to Bilibili content creators who taught me how to understand several renowned deep learning models and networks.

Finally, I sincerely thank my family for their unwavering support and encouragement throughout this study.

# Abstract

Severe earthquakes always lead to catastrophic building damage. Post-earthquake building damage level classification (BDLC) is an important task to rescue persons and make rapid earthquake responses for the reduction of severe injuries and casualties. To reduce data processing time for post-earthquake disaster response, pre-earthquake building data are always prepared, because pre-existing information about building locations and characteristics can reduce the time of post-earthquake localising buildings. Therefore, both pre-earthquake building information preparation and post-earthquake building damage information collection facilitate swift BDLC. Compared with conventional labour-intensive, time-consuming, and possibly dangerous in-situ observations, remote sensing technology provides a rapid and efficient approach to these pre- and post-event data collections because of its capability to acquire large-scale data remotely and rapidly. There are several remote sensing data types with their own advantages. For instance, two-dimensional (2D) optical satellite images provide large-scale information of the earth. Three-dimensional (3D) Light Detection and Ranging (Lidar) point clouds, as another type of remote sensing data, provide additional information on elevations of ground and non-ground points, including the heights of buildings.

Among the methods for processing 2D and 3D remote sensing data, deep learning semantic segmentation (DLSS) technology has a high potential in applications for BDLC on remote sensing data. However, the potential of these methods for BDLC has not been thoroughly studied in previous research. Indeed, there are four gaps in the literature in this domain. Firstly, few DLSS methods have been applied to 2D satellite images or 3D point clouds for

building damage classifications specifically related to earthquakes. Secondly, several well-known DLSS algorithms were proposed and tested only on small or indoor case studies in 2D and 3D applications. The large-scale outdoor scenarios have yet to be fully discussed or tested. Thirdly, most current post-earthquake BDLC studies lack detailed multi-level classification methods in the remote sensing field. Fourthly, for the training of the DLSS methods, there is a lack of labelled datasets for multi-level BDLC at large study extents in pre- and post-earthquake events.

This study solved these problems by applying these methods to both 2D and 3D remote sensing data on large-scale outdoor areas and by proposing novel DLSS approaches to classify building damage into four levels. To overcome the lack of training data, this study prepared and developed labelled datasets for the training of the proposed DLSS methods.

Ablation studies have been designed to test the performance of these proposed DLSS methods. The results in this study show the good performance of these methods at large-scale building footprint extraction and four-level BDLC with either satellite or Lidar data. Indeed, these novel methods have increased the accuracy of the chosen backbones in large-scale outdoor study areas. The channel attention mechanism helps to improve the accuracy of building information extraction in both 2D and 3D methods with higher Intersection over Union (IoU) values compared to the chosen backbones. Overall, this study overcomes the issues of the current methods of BDLC and will benefit society by providing a safe and speedy post-earthquake BDLC method that requires minimal fieldwork to support rescue teams for quick response. It will also help disaster management systems to store information efficiently for post-earthquake recovery planning.

# Publication List

## A. Journal articles

**Articles related to the thesis:**

**Liu C**, Ge L*, Xiang W, Du Z, and Zhang Q, **2023**. Channel Attention and Normal-based Local Feature Aggregation Network (CNLNet): A Deep Learning Method for Predisaster Large-scale Outdoor Lidar Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*. 62, pp. 1-12. DOI: 10.1109/TGRS.2023.3339475

**Liu C**, Sepasgozar S, Zhang Q, and Ge L*, **2022**. A Novel Attention-based Deep Learning Method for Post-Disaster Building Damage Classification. *Expert Systems with Applications.* 202, p.117268. DOI: 10.1016/j.eswa.2022.117268

**Other journal articles during the candidature:**

**Liu C**, Zhang Q*, Ge L, Sepasgozar S, Sheng Z, **2023**. Dielectric Fluctuation and Random Motion over Ground Model (DF-RMoG): An Unsupervised Three-stage Method of Forest Height Estimation Considering Dielectric Property Changes. *Remote Sensing.* 15(7), p.1877. DOI: 10.3390/rs15071877

Wu Y, **Liu C**, Zhang Q*, and Ge L*, **2022**. Bibliometric Analysis of Interferometric Synthetic Aperture Radar (InSAR) Application in Land Subsidence from 2000 to 2021, *Journal of Sensors*, 2022, p.1027673. DOI: 10.1155/2022/1027673

Zhang Q, Hensley S, Zhang R*, **Liu C**, and Ge L, **2022**. Improved Model-Based Forest Height Inversion Using Airborne L-Band Repeat-Pass Dual-Baseline Pol-InSAR Data. *Remote Sensing*, 14(20), p.5234. DOI: 10.3390/rs14205234

Zhang Q, Ge L, Hensley S, Metternicht G, **Liu C**, and Zhang R*, **2022**. PolGAN: A Deep-Learning-Based Unsupervised Forest Height Estimation Based on the Synergy of PolInSAR and LiDAR Data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 186, pp.123-139. DOI: 10.1016/j.isprsjprs.2022.02.008

Zhang Q, Ge L, Zhang R*, Met G, **Liu C**, and Du Z, **2021**. Towards a Deep-Learning-Based Framework of Sentinel-2 Imagery for Automated Active Fire Detection. *Remote Sensing,* 13(23), p.4790. DOI: 10.3390/rs13234790

## B. Selected conference papers

**Liu C**, Zhang Q, Shirowzhan S, Bai T, Sheng Z, Wu Y, Kuang J, Ge L*, **2023**. The Influence of Changing Features of Point Clouds on the Accuracy of Deep Learning-based Large-scale Outdoor Lidar Semantic Segmentation, In *2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Pasadena, U.S., pp. 4443-4446. DOI: 10.1109/IGARSS52108.2023.10281898

**Liu C**, Ge L*, and Sepasgozar S, **2021**. Post-Disaster Classification of Building Damage Using Transfer Learning, In *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* Brussels, Belgium, pp. 2194-2197. DOI: 10.1109/IGARSS47720.2021.9554795

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| Abbreviations | Full name |
|---|---|
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| AP | Attentive Pooling |
| BDLC | Building Damage Level Classification |
| BN | Batch Normalisation |
| CA | Channel Attention |
| CNLNet | Channel Attention and Normal-based Local Feature Aggregation Network |
| CNN | Convolutional Neural Network |
| DL | Deep Learning |
| DLSS | Deep Learning-based Semantic Segmentation |
| DSM | Digital Surface Models |
| ELM | Extreme Learning Machine |
| EMS | European Macroseismic Scale |

| | |
|---|---|
| FCN | Fully Convolutional Neural |
| GIS | Geographic Information System |
| GPU | Graphics Processing Unit |
| GSD | Ground Sample Distance |
| HRAI | High-resolution Aerial Imagery |
| HRNet | High-resolution Network |
| HRSI | High-resolution Satellite Imagery |
| ID | Identification |
| IoU | Intersection over Union |
| K3 | KOMPSAT-3 |
| KNN | K-Nearest Neighbours Algorithm |
| L1C | Level-1C |
| LFA | Local Feature Aggregation |
| Lidar | Light Detection and Ranging |
| LocSE | Local Spatial Encoding |
| LOLSS | Large-scale Outdoor Lidar Semantic Segmentation |
| LRN | Local Response Normalisation |
| mIoU | Mean Intersection over Union |

| | |
|---|---|
| MLP | Multi-Layer Perceptron |
| OA | Overall Accuracy |
| OSM | Open Street Map |
| pxl | Per pixel |
| RandLA-Net | Random Sampling and an Effective Local Feature Aggregator Network |
| ReLU | Rectified Linear Unit activation function |
| RGB | Red, Green, Blue |
| RQ | Research Question |
| RS | Remote Sensing |
| S2 | Sentinel-2 |
| SE | Squeeze-and-Excitation |
| SSA | Semantic Segmentation Accuracy |
| SVM | Support Vector Machine |

# Chapter 1

# Introduction

## 1.1 Background

Earthquakes have the potential to trigger catastrophic building damage and inflict casualties on numerous communities (Ji et al., 2018a). For instance, a $M_w$ 6.4 ($M_L$ 6.2) earthquake occurred in Croatia in December 2020 where several individuals were reported wounded and deceased, and roughly half of the city was left devastated (United States Geological Survey, 2020). Furthermore, as a severe earthquake that caused the most significant number of deaths in the last 15 years, the 2010 Haiti Earthquake resulted in an official death toll of about 230,000. Nearly half of all buildings collapsed or were severely damaged in the epicentral area in this Haiti earthquake, including more than 300,000 homes (Desroches et al., 2011). The damaged buildings were among the key factors contributing to the high number of casualties in that earthquake.

The seismic cycle is typically categorised into four stages: inter-seismic, pre-seismic, co-seismic and post-seismic. In the post-seismic stage, the primary objective is to rescue individuals and safeguard properties. In order to accomplish that, post-earthquake analysis becomes essential. Earthquake analysis generally contains three phases: disaster information analysis, post-earthquake emergency rescue decision-making, and post-earthquake recovery and reconstruction decision-making. The initial phase invariably involves the collection and analysis of disaster information. One of the foremost

responsibilities within this phase is the evaluation of post-earthquake building damage since it offers information for rescue, safety, and recovery. However, it is often hard for rescue teams to decide where to begin the rescue operation due to the dearth of prompt building damage information immediately following an earthquake. The lack of this information is caused by the difficulty of rapidly judging the levels of building damage due to the differences in structure and lack of rapid methods. Moreover, the search and rescue resources of the stricken area are usually insufficient in the first several hours. Consequently, the need arises for a fast, reliable, and efficient approach to building damage level classification (BDLC) aimed at rapidly identifying the most critical areas requiring rescue efforts and facilitating a prompt post-earthquake disaster response.

To reduce data processing time for post-earthquake disaster response, pre-earthquake building data preparation is advantageous. This is because pre-existing information about building locations and characteristics can help to assess the extent of building damage. Therefore, both the extraction of pre-earthquake building footprints and the evaluation of post-earthquake building damage are significant in facilitating swift emergency responses. The subsequent subsections delve into the contextual underpinnings of these two aspects.

### 1.1.1 Pre-earthquake building footprint extraction

As mentioned above, to avoid disastrous and chaotic aftermath, it is prudent to take pre-emptive measures before earthquakes, thereby streamlining data processing for post-earthquake emergency responses to rescue individuals and recover communities rapidly. By having accurate information about building locations before earthquakes, emergency

responders and relief organisations can quickly prioritise areas for search and rescue operations, allocate resources, and plan recovery efforts. As a result, it is necessary to conduct a meticulous building footprint extraction from the pre-earthquake source data. Remote sensing provides a rapid and efficient approach to pre-earthquake data collection, owing to its capability to rapidly acquire large-scale data.

After the initial pre-earthquake data collection, the need for a rapid and accurate method to extract buildings from remote sensing data becomes evident. Deep learning (DL) may be a suitable choice. With the rapid advancements in DL, it has been gradually applied in the remote sensing field for urban observation. Several papers applied DL to pre-event building extraction for the purpose of earthquake disaster analysis (Gupta et al., 2021, Zhang et al., 2022). DL-based pre-earthquake building footprint extraction using remote sensing involves identifying and mapping buildings in an area using remote sensing data, such as satellite imagery or Lidar. This process generates a baseline dataset that provides information about the spatial distribution, size, and shape of buildings before the earthquake.

### 1.1.2 Post-earthquake building damage level classification

Several elements influence the rescue decision making such as the damage degree of buildings, the number of trapped victims and the number of rescue labourers. The immediate damage estimation and classification after the occurrence of earthquakes will help emergency response plans to save human lives. In order to classify how serious the damage is, several scholars propose classification methods. Three common methods are

widely used for post-earthquake BDLC including 1) on-site investigation to calculate the loss, 2) evaluating the loss according to loss prediction models, and 3) evaluating loss based on remote sensing data.

The first investigation method is accurate, but it puts the lives of field investigators in danger of strong aftershocks and tsunamis. Further, the first method is time consuming, often weeks or even months, and very labour intensive. Detecting all buildings manually on the ground is time-consuming and ineffective, so it cannot be applied for rapid evaluation of earthquake damage. Because of the long processing time, rescue teams may miss the best time for rescuing the trapped and wounded. Meanwhile, some places cannot be reached by terrestrial vehicles after an earthquake. This increases the difficulty of investigation.

The second method most relies on the accuracy of prediction models, which might not be realistic. Due to the field restriction, the tools should be able to work remotely and automatically in terms of detection and classification.

The third remote sensing-based method usually represents the method using airborne or spaceborne remote sensing data. This method can collect large-scale data in the target area within a shorter time compared with the first in-situ method. This is because remote sensing devices can scan land from a great height, covering a much larger area than what can be obtained through in-situ observations. Moreover, this collection method is safer as it avoids potential on-site dangers. Another advantage is that the presence of multiple earth observation satellites allows for more frequent image information through them. In addition to these benefits, unlike the second method which generates estimated damage degrees, the

third method provides real damage degrees obtained from collected data. Therefore, among these three methods, the third one is considered superior due to its convenience, safety, and accessibility for frequent earth imaging coverage. This method has gained increasing interest on a global scale. With the fast development of artificial intelligence (AI), as a subset of AI, DL techniques have started to widely be utilised in various fields. DL teaches computers to process data automatically in a way that is inspired by the human brain (Guo et al., 2020). To accelerate the process of the remote sensing data, DL algorithms are being applied for BDLC (Su et al., 2020, Ji et al., 2018a, Yang et al., 2021, Wheeler and Karimi, 2020).

Based on the introduction of Sections 1.1.1 and 1.1.2, both pre-earthquake building footprint extraction and post-earthquake building damage evaluation are crucial for a successful earthquake disaster response. DL-based methods using remote sensing data are handy to do these two tasks. DL models or networks can be trained using both pre-earthquake information and post-earthquake building damage level data to establish relationships with them. Thus, these trained DL models or networks can then be used to predict or estimate the likely damage levels for buildings in a specific area following an earthquake. The following section discusses the current research problems of DL applications related to this study.

## 1.2 Research problems of deep learning-based semantic segmentation applications in earth observation

In recent years, with the fast development of DL, its related technologies have been widely utilised in different applications in the remote sensing field. Semantic segmentation is a key application, which labels every pixel in images or every point in Lidar point clouds. Lidar is a type of remote sensing technology that provides three-dimensional (3D) information of land covers. DL-based semantic segmentation (DLSS) using remote sensing data provides an accurate approach to both pre-earthquake building footprints extraction and post-earthquake BDLC, because it can outline the detailed building locations and shapes.

Although there is a trend of applying DLSS in the remote sensing field, it is still a challenging task (Gupta et al., 2019a). This is because there are several key technical problems and data limitations that need to be solved. Remote sensing data that can be used for the purpose of building-related research are always categorised into 2D and 3D applications according to data sources. One of the most common 2D data sources is satellite imagery, and Lidar is one of the widely used 3D sources in the remote sensing field. Based on the literature review, there are several problems of DL-based building related semantic segmentation research that need to be solved. Therefore, the following subsections discuss these problems from literature categorised by different data sources, including 2D satellite images and 3D Lidar point clouds.

### 1.2.1 Problems using 2D satellite images

One significant challenge associated with the employment of 2D satellite images is the lack of DL models that are designed for earthquake-related purposes. In detail, DL-based methods of building footprint extraction are widely discussed in the computer vision field. Nonetheless, most of these DL methods are not designed for earthquake-related purposes (Krupiński et al., 2019, Majd et al., 2019). For instance, most studies apply DL methods for building footprint extraction. However, few of them are designed for pre-earthquake building footprint extraction. Similarly, most existing DL methods may not be suitable to be applied in post-earthquake BDLC. Moreover, several DL-based 2D imagery processing methods, such as UNet (Ronneberger et al., 2015) and ResNet (He et al., 2016a), in the remote sensing field are not initially designed for data analysis of large-scale outdoor scenarios, because these models were initially proposed for classifying or segmenting indoor or small objects from images in the computer vision field which are not visible in satellite images.

Another issue is that most current studies for analysing post-earthquake building damage levels only classify damages into two levels, i.e., collapsed and intact. This is not enough for either a rescue or recovery plan because the information is too general. Few studies have discussed multi-level BDLC.

Another problem is the insufficiency of labelled image datasets of post-earthquake damaged buildings. Very limited satellite image datasets for damaged building levels are accessible publicly for applying DL methods. Popular 2D semantic segmentation datasets

in computer vision, such as ImageNet (Deng et al., 2009), do not always contain damaged building information. There are some available datasets containing information about collapsed or not, which is not enough for post-disaster rescue and management. Therefore, labelled images with more than two levels (damaged or not) are needed for DL study.

### 1.2.2 Problems using 3D Lidar point clouds

With the increasing popularity of Lidar applications in remote sensing, there is a trend to apply Lidar data for urban observation and building related research. The current problems of 3D Lidar data applications are very similar to the problems of 2D related research mentioned in Section 1.2.1. First, there is a lack of DLSS methods for focusing on earthquake-related building footprint extraction or damage level classification. Second, although some methods can be applied or transfer learnt for the purposes of this thesis, they were not designed for large-scale or city-scale. The third problem is that few DL-based studies discuss multi-level BDLC. The fourth problem is the lack of suitable labelled data for DL model training. Most open-source Lidar datasets are published for indoor observation or small-area outdoor semantic segmentation.

### 1.2.3 Summary of the problems of 2D and 3D data applications

Although a DL-based remote sensing method could be an approach for both pre- and post-earthquake data collection and analysis, there are still some challenges that exist.

1) Most existing 2D optical imagery based DLSS methods are not designed for large-scale post-earthquake BDLC.

2) Although some 3D Lidar-based DLSS methods are applied for pre-earthquake building extraction analysis, the locations of these studies do not consider the possibility of earthquakes.
3) Few 3D Lidar-based post-earthquake DLSS studies discuss its application in large-scale BDLC after earthquakes.
4) Labelled multi-level building damage 2D imagery or 3D Lidar point clouds datasets are limited in the public domain.

A detailed review of the state-of-the-art literature in related research fields for finding research gaps is presented in Chapter 2.

## 1.3 Research aim and objectives

To address the abovementioned issues, this research aims to propose novel DL models to classify building damage into four levels with large-scale in-house labelled datasets considering both pre- and post-earthquake periods.

This thesis has four objectives to achieve this goal with 2D satellite images and 3D Lidar point clouds.

- Objective 1: To propose a 2D BDLC method considering both pre- and post-earthquake periods using DL with large-scale optical satellite images.
- Objective 2: To offer a DL-based pre-earthquake building footprint extraction method with large-scale Lidar data tested in the case studies whose locations have the possibility of earthquakes.

- Objective 3: To provide a DL-based post-earthquake BDLC method with large-scale Lidar data.
- Objective 4: To build 2D satellite and 3D Lidar in-house labelled datasets of pre-earthquake building footprints and post-earthquake multi-level damaged building information.

## 1.4 Significance of the study

One of the main tasks of rescue teams after an earthquake is to save lives and safeguard properties. The best time for rescuing people is in the first 36 hours from the time an earthquake happens, so several countries require researchers to provide an emergency response plan every several hours. For instance, the Chinese government requires China's Earthquake Administration to provide updated responses every six hours after an earthquake. The main reason for death in large-scale earthquakes is the collapse of buildings: "Earthquakes don't kill people, buildings do (Ross, 2021)." If rescue teams know where the location of severe building collapses are and respond fast, more lives will be saved. To save more people, one of the urgent worldwide issues is to know how to assess detailed building damages after an earthquake quickly. Therefore, BDLC as a detailed building damage assessment is one of the critical tasks that requires to be done quickly to rescue more people and formulate a more effective earthquake emergency response plan.

Conventional in-situ detailed building damage level estimations have several issues, including being labour-intensive, time-consuming, expensive, manual, potentially dangerous and imprecise due to risks to the field investigators and estimators in

catastrophic situations. This study addresses this issue by offering automatic, precise, and rapid building damage estimation using remote sensing data. Remotely sensed data collection, for this purpose, significantly reduces the risks involved in the conventional in-situ manual data collection. For both pre-earthquake building information update and post-earthquake BDLC, the proposed novel methods in this study can be applied to large human settlement areas, while maintaining the acceptable Intersection over Union (IoU) of the results.

With the capability of detecting, analysing, and classifying city-scale building damages using multi-source data, this study provides a valuable tool for emergency response teams and disaster management authorities to efficiently prioritise resources and aid efforts in affected areas. In addition to saving lives, this study offers rapid and timely information for the recovery planning of earthquake-affected areas. The results from a detailed BDLC can also help different countries design and update seismic building standards and codes (see Section 2.3 for more details of related standards and codes).

## 1.5 Structure of the thesis

This thesis comprises eight chapters. The first chapter introduces background information of this study. The second chapter discusses the existing relevant literature in detail. The third chapter states the research methodology and the specific methods for each main part of the thesis are presented in their respective chapter. The fourth to seventh chapters are the main parts of the thesis. The last chapter is the conclusion of this study. The structure of each chapter is outlined in Figure 1-1 below. A brief description is stated as follows:

| Chapter | Purpose |
|---|---|
| **Chapter 1: Introduction** | Introduction to the background of remote sensing and its application in the related research field, problem statement and aim of the thesis. |
| **Chapter 2: Literature review** | Statement of current development, identification of gaps in related research. |
| **Chapter 3: Research design** | Research focus, methodology design, workflow, data, models and scenarios. |
| **Chapter 4: 2D-based building damage estimation** | Proposing a four-level building damage classification model with 2D optical images; Creating a manually labelled 2D dataset. |
| **Chapter 5: 3D-based test of influence of feature selection** | Testing influence of feature selections on the accuracy of urban semantic segmentation with Lidar; Creating a manually labelled 3D urban object dataset. |
| **Chapter 6: 3D-based building footprint extraction** | Proposing a 3D deep learning model for urban semantic segmentation with Lidar focusing on building footprints ; Creating a manually labelled 3D urban object dataset. |
| **Chapter 7: 3D-based building damage estimation** | Propose a 3D deep learning model for building damage level classification with Lidar; Creating a manually labelled four -level post -disaster building damage dataset. |
| **Chapter 8: Conclusion** | Discussion on findings, contributions, limitations and suggestions for future studies . |

Figure 1-1. The structure of the thesis and the purpose of each chapter

Chapter 1 is the introduction to this study. It begins by presenting the background of DL applications for both pre-earthquake building footprint extraction and post-earthquake BDLC. Then, research problems are stated and turned into the research aim and objectives of this study, followed by the significance of the study.

Chapter 2 reviews the literature on the adoption of state-of-the-art remote sensing and DL techniques in building footprint extraction and BDLC. Significant gaps in DL adoption in both pre- and post-earthquake applications are identified. The literature review leads to the design of this research.

Chapter 3 presents the research focus and the workflow of the whole study. This chapter also describes general reasons for the design and explains how the design aligns with the research objectives. The detailed methods of each main chapter are explained in their own chapters.

Chapter 4 analyses the performance of the proposed DL model for BDLC using 2D satellite images. The model contains two steps, including pre-event building footprint extraction and post-earthquake damage level analysis. This chapter presents a comparison of the proposed model with other DL models that no one has previously compared. Four types of comparison experiments have been conducted with their advantages and limits.

Chapter 5 is the preparatory work for Chapter 6. It examines the performance of feature selections on the accuracy of DL-based large-scale outdoor Lidar semantic segmentation. Results of the DL models with and without surface normal vectors are tested. The down-sampling scales and numbers of down-sampling layers are designed for four different feature selection options. In total, the chapter compares eight selections to evaluate their performance in relation to the accuracy of semantic segmentation. The building class is the focus.

Chapter 6 proposes a novel method of pre-event semantic segmentation for large-scale colourised Lidar. Some parameters and features are designed according to the experiments of Chapter 5. It performs acceptably for the extraction of the building class. Satellite images are fused with Lidar data to provide colourised point clouds.

Chapter 7 introduces a new DL-based approach for classifying building damage levels in the aftermath of earthquakes. The method combines surface normal information and attention-based DL techniques to effectively identify various levels of building damage from post-disaster colourised Lidar data. The proposed method showcases the results that no damage or total story failure is easier to be tested than other damage levels.

Chapter 8 draws conclusions and states the contributions of this study. The feasibility of applying DLSS to classify building damage into four levels after earthquakes is discussed. Finally, this chapter provides an overview of the limitations and offers suggestions for future research.

## 1.6 Conclusion

The gist of this chapter is to explain the backgrounds, aim and objectives, and the significance of this study. The aim and objectives of this study are proposed based on the current gaps and problems from the literature review. The design of this study is formulated based on the research aim and objectives. An overview of the structure of this thesis is drawn in Section 1.5.

In summary, utilising both pre-existing building data and post-event information can assess and classify the building damage caused by an earthquake. This study applies 2D and 3D data, remote sensing technology, and DL data analysis techniques to produce intelligence that helps the rescue team allocate the limited resources optimally. It also provides valuable insights for emergency response, recovery, and future urban planning.

The next chapter will review previous related literature. State-of-the-art DL methods applied in this field have been discussed. This review helps this study discover research problems and design the aim and objectives for this research.

# Chapter 2

# Remote sensing and deep learning applications related to building damage level classification

## 2.1 Chapter introduction

This chapter offers a review of the relevant literature. First, definitions of key terms in this study are stated (Section 2.2). BDLC codes and standards of different countries and regions are listed (Section 2.3). Then, the literature on 2D satellite imagery applications for predisaster building extraction and post-disaster BDLC is reviewed (Section 2.4). Then, the background of Lidar is introduced and literature discussing 3D Lidar techniques in predisaster building extraction and post-disaster BDLC is reviewed (Section 2.5). Related datasets are introduced in Section 2.6. Finally, a summary of reviews explaining the gaps in the literature that need to be bridged (Section 2.7).

## 2.2 Definition of related terms

### 2.2.1 Remote sensing related terms

Optical satellites use optical sensors to detect the reflection of solar radiation by ground features in the optical part of the spectrum, such as visible and infrared waves. Most optical ones are in the passive mode.

**Atmospheric calibration**

Atmospheric correction is a step for optical image processing to characterise the surface reflectance, which removes the scattering and absorption effects from the atmosphere (Liang and Wang, 2020).

**Active and passive remote sensors**

Active sensors generate and detect electromagnetic energy. Passive sensors do not generate energy and only generate externally detected electromagnetic radiation, such as the emittance by or reflectance from a target (Gerke and Kerle, 2011).

**Lidar**

Lidar, also called LiDAR or LADAR, is an acronym for "light detection and ranging" or "laser imaging, detection, and ranging" (Shan and Toth, 2018). Lidar measures distance by launching a laser and measuring the reflection. Those laser beams propagate in a straight line with good directivity and very narrow beams, so it is hard to find. The limitation of Lidar is that it is influenced by weather. If the beams encounter heavy rains, smoke, fog, or other bad weather conditions, the beams are hard to detect.

**Laser**

Laser is an acronym for light amplification by stimulated emission of radiation (Shan and Toth, 2018). Unlike other lights, laser emits monochromatic light with a wavelength between 0.1 μm and 3 mm (infrared, visible light, or ultraviolet).

**Surface normal**

A surface normal is a vector that is perpendicular to a given surface at a point on the surface.

### 2.2.2 Earthquake-related terms

**Natural disaster definition**

There is no official definition of "natural disaster". A natural disaster can be an event brought about by the natural processes of the Earth that leads to widespread environmental destruction and an increase in mortality and morbidity.

**Earthquake definition**

United States Geological Survey (2023) defines "An earthquake is what happens when two blocks of the earth suddenly slip past one another. The surface where they slip is called the fault or fault plane".

**Seismic magnitude scales**

Earthquake or seismic magnitude is the most well-known measure of earthquake strength for historical reasons. There are several types of seismic magnitude scales, but all are proposed according to the largest recorded amplitude. Various magnitude scales represent different methods of deriving magnitude from such information as is available. Common types of magnitude include local magnitude ($M_L$), surface wave magnitude ($M_s$), body wave magnitude ($M_b$) and Moment magnitude ($M_w$). The calculation methods for these types are not the same, resulting in different meanings for the same number of various types. For instance, $M_S$ 6.0 is not the same as $M_W$ 6.0. Different countries and regions adopt different magnitudes.

Charles Richter introduced a well-known magnitude scale in 1935, what is called the "Richter" scale or local magnitude $M_L$ (Richter, 1935). This was determined by measuring the largest amplitude of seismic waves recorded on a standard instrument, the Wood–Anderson seismograph. It is logarithmic, so each increase of one unit is a tenfold increase in the amplitude, which is about 31.6 times more earthquake's energy release. The moment magnitude scale (Mw) estimates the seismic moment released by an earthquake to measure the earthquake's magnitude, as proposed in 1979 (Hanks and Kanamori, 1979). It is also a logarithmic scale. Another general magnitude scale used for global seismology is the body-wave magnitude, $M_b$. The surface wave magnitude ($M_S$) is also a well-known scale using Rayleigh wave measurements on vertical instruments.

**Seismic intensity scales**

The seismic intensity provides another quantitative measure of the earthquake's energy release. Seismic intensity refers to the assessment of the effects and damage caused by ground shaking at specific locations (Shearer, 2019). It focuses on the resulting impact on structures, people, and the environment. The seismic intensity is always measured using scales. Seismic intensity scales focusing on effects caused by an earthquake are different than seismic magnitude scales, which measure the overall strength of that earthquake. In other words, two earthquakes with the same magnitude scale may have different intensity scales. Intensity scale categorisation methods vary across different regions worldwide.

### 2.2.3 Deep learning related terms

**Training, validation, and testing**

Deep learning network experiments mainly contain three phases: training, validation, and testing (Liu et al., 2022). These phases have separate datasets. “Training” means the training stage where the network learns features. “Validation” represents the process of evaluating the performance of a trained network during the training on another dataset. The purpose of validation is to assess how well the model generalises to new data and then adjust or optimise the trained network. Finally, a fully trained network with good validation results is generated for testing. “Testing” means the process where the performance of a fully trained model is evaluated on a testing set.

**Supervised and unsupervised learning**

The most significant difference between supervised and unsupervised learning in the AI field is the need for labelled data. Supervised learning relies on labelled input data. Unsupervised learning does not need the label of data, which can process unlabelled or raw data directly.

**Down-sampling**

Down-sampling decreases the size of chosen data, such as an image or a point cloud dataset, by a specific rule or algorithm.

**Semantic segmentation**

Semantic segmentation in this thesis represents a task that associates a label or category with every pixel or voxel in images or Lidar point clouds.

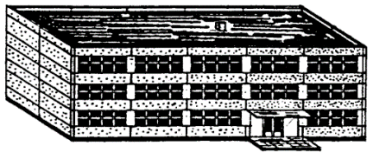
## 2.3 Building damage level classification codes and standards worldwide

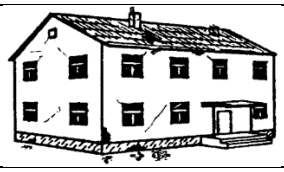
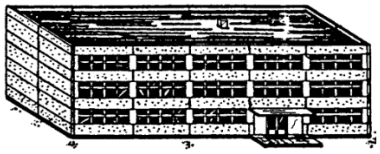
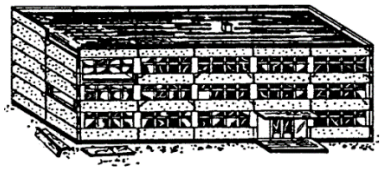
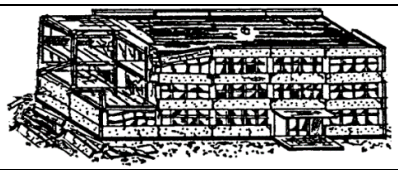
One of the critical effects of earthquakes is ground shaking. Buildings are damaged by the shaking itself or by the ground subsidence beneath them after earthquakes. The damages of buildings vary from each other, so a code or standard is necessary to classify them for disaster response and recovery. BDLC codes and standards vary across different countries and regions, and some are established by local authorities. Moreover, most of them are based on structural engineering and in-situ observations. These codes are usually implemented as guidelines for assessing and categorising the level of damage to buildings, which is helpful for future building designs, repairability, and construction. Below are some examples of building damage level codes in different countries.

**European countries**

European countries often adopt the European Macroseismic Scale (EMS-98) as the basis for assigning building damage levels (Grünthal, 1998). EMS-98, published in 1998, classified single building damage into five levels for both masonry buildings and reinforced concrete buildings, as shown in Table 2-1.

Table 2-1. Classification of damage in EMS-98

| Building Type | | Damage level |
|---|---|---|
| Masonry buildings | Buildings of reinforced concrete | |
|  |  | Level 1: Negligible to slight damage |

| | | |
|---|---|---|
|  |  | Level 2: Moderate damage |
|  |  | Level 3: Substantial to heavy damage |
|  |  | Level 4: Very heavy damage |
|  |  | Level 5: Destruction |

**China**

signs of damage, as shown in Table 2-3. This DB/T 75-2018 standard also designs a building damage levels classification standard for group buildings. Since group building-related topics are not the focus of this thesis, this study does not introduce relevant codes.

Table 2-2. Damage characteristics of single buildings in DB/T 75-2018

| | |
|---|---|
| **No collapse** | The building structure is intact, and the main structure has not collapsed or partially damaged. |
| **Partial collapse** | Some parts of the building have collapsed, or the roof has been partially damaged, or the retaining wall is damaged, with 10% to 50% deformation. |
| **Collapse** | The entire building has completely collapsed, or the roof has completely collapsed, or more than 50% of the main structure has collapsed, twisted, deformed, or tilted. |

Table 2-3. Sub-levels of single buildings that have no collapse

| | |
|---|---|
| **No significant signs of damage** | The building structure is intact, and the main structure has not collapsed, but the roof has fallen tiles and collapsed, or the roof ridge has been partially damaged, or parapet walls have collapsed. |
| **Showing significant signs of damage** | The building structure is intact and has not collapsed with no obvious damage to the roof and retaining walls. |

Two damage assessment methods have been listed in DB/T 75-2018. The first method is visual interpretation, which needs manual drawing of building footprints. The second method is automatic or semi-automatic extraction, which requires advanced unsupervised or supervised methods for saving workload and time.

Besides the above damage assessment method, GB/T 24335-2009 (2009) code, published by the General Administration of Quality Supervision (2009), designs another code for categorising damage to buildings into four levels, including slight/no damage slight, moderate, heavy, and collapse.

**U.S.**

Damage Assessment Operations Manual (Federal Emergency Management Agency, 2016) lists the building damage level assessment metrics for manufactured and conventionally built homes in Appendix E in Pages 113 and 114. This document classifies damage to buildings into four levels, including affected, minor, major, and destroyed. “Affected”

means that there is no damage affecting habitability, and only cosmetic damages. Minor damage represents that the damage does not affect the structural integrity of the residence. A major damaged building has sustained significant structural damage. “Destroyed” means that this building is a total loss.

**Japan**

Akkar et al. (2021) introduce the damage levels from the rapid inspection method of structure engineering aspects published by the Japan Building Disaster Prevention Association, including “none”, “minor”, “slight”, “moderate”, “severe”, and “collapse”.

**Classification of building damage from scholars**

Besides governments, there are also some codes and classification categories proposed by scholars. For instance, Nakano et al. (2004) proposed a definition of damage of reinforced concrete columns and walls into five levels, from visible, narrow, to huge cracks. Schweier and Markus (2006) proposed a detailed ten-level catalogue of damaged buildings with airborne scanning techniques. Some of the levels have more detailed sub-levels to describe different damage statuses. The classification is based on spatial geometry, such as the change of height, outlines, and volume. Besides this classification method, scholars from Japan also proposed other classification methods.

It is important to note that building damage level codes also exist for other natural disasters, such as tsunamis. Early studies in Japan collected damage data from historical tsunamis and proposed threshold depths for collapse as damage criteria. For example, wooden houses and

reinforced concrete buildings may collapse if the tsunami inundation depth is 2 and 8 meters, respectively. Further studies on building damage due to tsunamis were conducted after the 2004 Indian Ocean tsunami. The data were analysed for all buildings, categorised by structural material, number of stories, and location along the coast to capture and explain potential variations in damage predictions. The survey covered over 250,000 structures (Suppasri et al., 2013). Unlike damage levels after an earthquake, these studies defined damage levels with a "washed away" category in addition to minor, major, or collapsed levels. Therefore, it can be found that different natural disasters have distinct building damage assessment standards. Because of that, building damage level codes for other natural disasters are not considered in this thesis.

**Conclusion of worldwide codes**

Most of these category codes and standards are designed based on their own regions or countries, so the classification results are often different from different investigation groups for the same event. Some researchers also classify damage degree by building groups. Areas with natural distribution, such as building blocks and natural villages, are generally selected as the seismic damage unit. The degree of damage in unit depends on the damage degree of most buildings.

## 2.4 2D image-based predisaster building extraction and post-disaster BDLC

Building extraction from RS data is still not automated, and most building damage mapping is only based on visual interpretation, which is time-consuming and labour-intensive (Gerke

and Kerle, 2011). The large amount of data after disasters increases the difficulty and cost of data interpretation. DL is a possible solution. Therefore, this section states some well-known DL models and their applications with four subsections from Section 2.4.1 to Section 2.4.4. Section 2.4.1 introduces the development of DLSS for 2D images in computer vision. Section 2.4.2 reviews current studies using DLSS methods for land cover classification with 2D optical images, whose classes include the building class. Section 2.4.3 reviews the studies that only focus on pre-earthquake building footprint extraction from land cover objects using DLSS methods with 2D data. Section 2.4.4 introduces the DLSS methods that are designed for post-earthquake BDLC.

### 2.4.1 Development of deep learning methods for 2D semantic segmentation

Optical image segmentation has developed in remote sensing for several years, as shown in Figure 2-1. Initially, various methods focused on individual pixels, applying computer vision techniques. Different domains use specific methods, such as vegetation index methods for greenery-related research and principal component analysis for classifying various features. Later, in the 1900s, machine learning methods like artificial neural network (ANN), support vector machine (SVM), and extreme learning machine (ELM) gained widespread adoption. Additionally, geographic information system (GIS) integrated methods are also widely spread, such as spectral mixing analysis, fuzzy cluster analysis, and other multi-data fusion analysis methods. Afterwards, since the 2000s, object-level analysis methods have found significant application. Subsequently, As AI rapidly advanced, several AI-related methods were proposed in the 2010s.

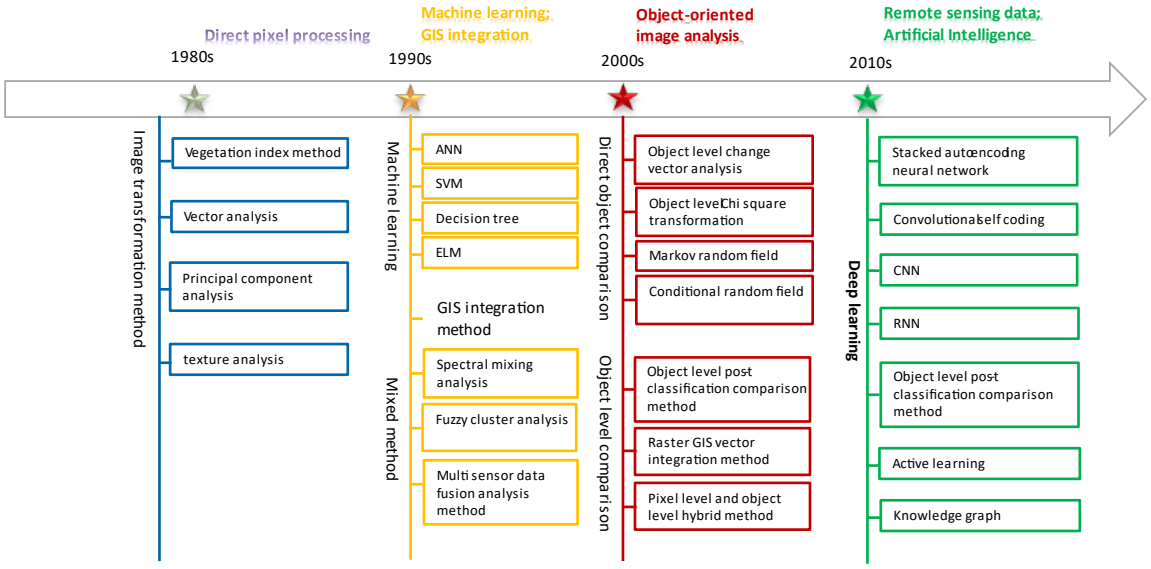



Figure 2-1. Development of optical image segmentation in remote sensing field

As a part developed from AI and machine learning, DL has been increasingly applied for image classification purposes recently, particularly when AlexNet was introduced in the literature (Krizhevsky et al., 2012). AlexNet improved image classification accuracy from 70%+ of conventional computer vision methods to 80%+, which is a breakthrough in terms of accuracy. The dominance of AlexNet in the classification contest was acknowledged by the ImageNet Large Scale Visual Recognition Challenge 2012 (LSVRC 2012) as a well-known competition in computer sciences, and thus, the application of DL for image classification in various contexts has been further increased.

One of its groundbreaking contributions is that it was the first to use the graphics processing unit (GPU) to accelerate training speed. Second, it applies the Rectified Linear Unit (ReLU) activation function instead of conventional activation functions to increase accuracy. Third, local response normalisation (LRN) was proposed to improve

generalisability. One of the most critical tasks of DL is to improve the generalisability (or generalisation ability). Generalisability means the ability of DL models to react to new data. If a model has a higher generalisability, it means the accuracy of this model will be higher for new data. Fourth, its first two fully connected layers use the "dropout" method to decrease the possibility of overfitting.

After that, another famous net, VGGNet, was proposed by the Visual Geometry Group of the University of Oxford (Simonyan and Zisserman, 2014). It was the winner of the LSVRC 2014 localisation task and the second place in the LSVRC 2014 classification task. Its highlight is that it applies two 3×3 kernels replacing one 5×5 kernel and three 3×3 kernels replacing one 7×7 kernel. They have the same receptive field. This can reduce required parameters during computing. The receptive field is the size of the region on the input layer corresponding to one feature (cell) on the output feature map.

The residual block was proposed in ResNet (He et al., 2016a). ResNet achieved the winner of image classification, localisation, and detection in the ILSVR Challenge 2015. It was also the first place of the object detection task and image segmentation task of the MS COCO Challenge 2015. ResNet-34 reduced top-1 error by 3.5% on ImageNet validation compared to its plain counterpart (He et al., 2016a). One of its significant contributions is the "residual block". Another one is that it applies batch normalisation (BN) to accelerate the training instead of the dropout.

The residual block was proposed for addressing two problems: 1) gradient vanishing problem and gradient exploding problem; and 2) degradation problem. Residual blocks are

skip-connection blocks to improve the accuracy of DL models and show good results based on the ResNet test. The output is the addition of an identity function and residual blocks.

In recent years, the use of attention mechanisms for image classification and semantic segmentation has developed quickly. Examples of the attention mechanism in the literature are SENet (Hu et al., 2018) and Vision Transformer (Dosovitskiy et al., 2020). The attention mechanism was initially applied in the natural language processing field. It has been increasingly used for other applications such as image processing, since several computer vision researchers realised its advantages (Dosovitskiy et al., 2020).

High-Resolution Network (HRNet) was proposed for human pose estimation initially (Sun et al., 2019a). Later, its authors applied HRNet for image classification, semantic segmentation, object detection, and facial landmark detection. Since it has been tested for both classification and segmentation, this research chooses this model. One great advantage of HRNet is that it maintains high-resolution representations in the network. Several conventional DL models decrease input sizes by losing information to decrease the amount of calculation, such as UNet (Ronneberger et al., 2015). Compared with those methods, HRNet can obtain features from the original high-resolution input images while keeping information. A higher-resolution image contains more features and information than a lower-resolution image.

### 2.4.2 DLSS methods for 2D land cover classification

Predisaster land cover object data collection and analysis have recently gotten more attention worldwide (Liu et al., 2022). With the rapid development of remote sensing

technologies, the resolution of no matter public or commercial satellite images has increased fast in recent years. These remote sensing images have opened several opportunities for new applications using DL techniques, such as land cover classification (or called land cover semantic segmentation).

Researchers have made tremendous efforts to develop accurate, fast, and automatic land cover classification methods. Among these, DL is considered one of the most promising and evolving approaches. Several DL methods have been applied to land cover classification. For instance, Zhang et al. (2019a) proposed a land cover classification method for the task of classifying land cover semantics, such as buildings and grassland, incorporating multi-layer perceptron (MLP) and convolutional neural network (CNN). Helber et al. (2019) presented a patch-based land cover classification approach using DL with its own labelled geo-referenced dataset, EuroSAT. The target classes vary in different studies and some of them do not include the building class. However, the building class is the primary class of this study. As a consequence, the review of land cover classification literature was narrowed down to the literature focused on building footprint extraction, which is introduced in the subsequent subsection 2.4.3.

### 2.4.3 DLSS methods for pre-earthquake building footprint extraction

In the remote sensing field, with the increasing accuracy of land cover classification, some recent studies have focused on applying DL methods only to extract buildings from various land cover objects and ignore other objects using optical satellite images because only building information has been the target in their studies. For example, Liu et al. (2019)

developed a fully convolutional neural (FCN) method for building extraction on high-resolution aerial imagery (HRAI). They conducted several experiments on two public datasets, the Inria Aerial Image Labelling Dataset and the WHU Aerial Building Dataset, to showcase the effectiveness of their proposed model in building footprint extraction. Four metrics (i.e., precision, recall, F1, and IoU) were employed for evaluation. Li et al. (2019) introduced a U-Net-based semantic segmentation method that explored the potential of integrating three public GIS map datasets (i.e., OpenStreetMap, Google Maps, and MapWorld) with WorldView-3 satellite datasets in four cities (Las Vegas, Paris, Shanghai, and Khartoum). The F1 score was used for performance evaluation. Wei et al. (2019) utilised an FCN method for building extraction using the WHU Aerial Building Dataset, assessing performance with IoU, recall, and precision metrics. Zhang et al. (2020) utilised GF-2 satellite images with a developed Mask R-CNN method. The average value of IOU served as an evaluation metric. Shao et al. (2020) introduced a novel network, the Building Residual Refine Network, using the Massachusetts Building dataset. The evaluation was based on IoU and F1. Wei et al. (2021) employed the $U^2$-net on the WHU building dataset, an international open-source dataset. Evaluation metrics included IoU, recall, precision, and F1 score.

In summary, although several methods proposed enhanced approaches for building extraction, they always used those four well-known accessible building datasets without other datasets. Therefore, it is necessary to develop more building datasets for research in this field. Those well-known datasets with some map databases are introduced in the

following paragraphs. Moreover, IoU and F1 are the first two most popular evaluation metrics for trained DL models and networks, according to the reviewed literature.

### 2.4.4 DLSS methods for post-earthquake BDLC

As mentioned in Section 2.3, classification levels of damaged buildings in the standards and codes vary from region to region. There is no global standard or code to unify all detailed damage levels. Therefore, there is no unified classification method for damage assessment. There are some typical building damage and impact assessment methods, such as self-reporting, fly-over, windshield surveys, door-to-door and site assessments, geospatial analysis and geographic information systems, and modelling. They are introduced on pages 69-73 of the Damage Assessment Operations Manual (Federal Emergency Management Agency, 2016). This thesis focuses on the classification methods that applied geospatial analysis and geographic information systems using remote sensing, as introduced in Chapter 1. Although other methods are feasible, they always require a large number of labours, which is hard to achieve, especially in countries where labour cost is high. Besides the academic field, various industry companies have proposed post-disaster building damage estimation methods using remote sensing. For instance, Cloudeo Company (2023) applied their own collected multispectral imagery for the January 2023 Turkish earthquake and generated a heatmap of damaged buildings, with a case study conducted in Adiyaman City, Turkey. Therefore, BDLC with remote sensing will be a trend in both academia and industry.

DL techniques with remote sensing data provide solutions to avoid in-situ time-consuming observations. With the fast improvement of performance on all types of optical remote sensing sensors, several studies are increasingly focused on their applications for BDLC using remote sensing images with DL-based methods. For instance, Xie et al. (2016) applied a crowdsourcing approach to recognise and classify collapsed buildings rapidly after an earthquake based on remote sensing images with the case study of an earthquake in Yushu, China. Ji et al. (2018a) identified building damage with four levels according to EMS-98 using CNN and SqueezeNet with 2D post-disaster optical satellite images. Building footprints were manually labelled using ArcGIS 10.4. Four metrics have been adopted, including producer accuracy, user accuracy, overall accuracy (OA), and Kappa. Valentijn et al. (2020) applied a CNN-based method to test its performance for detecting damage to buildings after a natural disaster with the open-source xBD dataset. All these studies show that the DL method can be applied for remote sensing-based building damage estimation after the earthquake.

While previous attempts have aimed to enhance the accuracy of BDLC, many have solely focused on applications. They overlooked the crucial aspect of developing DL algorithms or networks to suit the specific scenarios of damaged buildings following earthquakes.

## 2.5 3D Lidar-based predisaster building footprint extraction and post-disaster BDLC studies

### 2.5.1 Background of Lidar

Despite Lidar being proposed in the 1960s, its widespread adoption for topographic applications using laser profiling and scanning systems only took off in the mid-1990s (Shan and Toth, 2018). Unlike most passive optical sensors applied in the remote sensing field, Lidar is an active remote sensing technique. A Lidar sensor emits laser pulses toward a target and measures the distance from the sensor to this target (Shan and Toth, 2018).

Presently, various types of Lidar sensors are employed, such as terrestrial, airborne, and spaceborne laser scanners. Airborne Lidar, also referred to as airborne laser scanning, is a laser scanning system that uses a drone, aeroplane, or helicopter to collect laser data. The spaceborne Lidar system, or named satellite-based Lidar system, that is attached to a satellite to detect global surface 3D information. Terrestrial Lidar, also called topographic Lidar, collects 3D coordinates of targets, including numerous points on land. There has been an increasing spread of laser-related applications for the last 30 years. This can be seen in the incorporation of lasers into several surveying instruments, such as total stations. Lidar techniques have found application in various remote sensing domains.

All types of Lidar sensors have their own benefits, but airborne and spaceborne sensors may be more suitable for large-scale case studies, which can provide an efficient means to rapidly implement 3D information mapping. For example, Yuan et al. (2018) conducted

research on wheat height estimation using Lidar technology. In another study, Zhang et al. (2022) applied Lidar for forest height estimation. Additionally, Zheng et al. (2021) utilised Lidar in conjunction with machine and DL analyses to detect fruits and flowers in strawberry farming. Furthermore, researchers have successfully integrated Lidar and camera information to develop a real-time road scene 3D semantic map with large-scale and high precision (Li et al., 2020a). This innovative approach holds promising potential for enhancing road scene understanding and navigation in real-world scenarios. Considering the focus of this study, airborne and spaceborne Lidar are applied for collecting data.

Apart from the commercial or academic applications of Lidar for airborne laser scanning services, there have been notable contributions from U.S. government research agencies, particularly NASA, where intriguing Lidar systems have been designed and utilised primarily for scientific research pursuits. These endeavours have advanced the field, leading to a deeper understanding and improved applications of Lidar technology for a multitude of purposes.

As introduced in Section 2.5.1, besides satellite images, Lidar technology is also widely used in the remote sensing field with its own advantages, such as height information. This section has three subsections to introduce Lidar applications for building extraction in both pre- and post-earthquake situations using DLSS methods. Section 2.5.2 reviews the studies of DLSS methods for 3D Lidar point clouds in the computer vision field. Section 2.5.3 introduces current studies of pre-earthquake building footprint extraction using DLSS, and Section 2.5.4 states the existing studies for post-earthquake BDLC using DLSS.

### 2.5.2 Development of deep learning methods for 3D Lidar semantic segmentation

Lidar semantic segmentation plays a pivotal role in various applications, including 3D modelling, building maintenance, and urban planning. To achieve accurate segmentation, the process involves extracting relevant features and global geometric structures from the point cloud data (Guo et al., 2020). Therefore, point cloud semantic segmentation is a complex task that requires ample data and computational resources. DL could be an appropriate approach.

DL-based Lidar semantic segmentation methods are usually categorised into four types: projection-based, discretization-based, point-based, and hybrid methods (Guo et al., 2020). "Projection-based" means that these methods always project 3D data into 2D images. Discretization-based methods usually convert a 3D point cloud into a dense or sparse discrete representation such as lattices. Point-based methods directly work on each point in point clouds. Hybrid Methods learn and utilise multi-modal features from 3D scanning.

Point-based segmentation works on those unordered point clouds directly. Point-based methods in this field were first introduced in PointNet in 2017 (Qi et al., 2017a). After PointNet, PointNet++ was proposed to improve the structure of PointNet to share more features between each point (Qi et al., 2017b). Then, other point-based methods were proposed, such as PointSIFT and PointWeb (Jiang et al., 2018, Zhao et al., 2019a). Within these six years, researchers are gradually widening applications of point-based DL methods from small-scale indoor to large-scale outdoor thanks to the increasing number of online

free large-scale outdoor datasets. However, related research for large-scale outdoor data is in its early stages.

### 2.5.3 DLSS methods for building extraction with 3D Lidar point clouds

Building extraction using the DLSS method with Lidar data has been discussed in several studies. For instance, as early as 2006, Verma et al. (2006) proposed a building detection method based on roof topology analysis. After that, Dos Santos et al. (2019) optimised parameter α of the alpha-shape algorithm for building roof extraction from Lidar point clouds. Zhao et al. (2019b) proposed a filter for improving the accuracy of distinction between buildings and tree canopies based on digital surface models (DSM) from Lidar point clouds and aerial images. The test area is the Vaihingen area of Germany. According to the evaluation methods utilised by Rutzinger et al. (2009), Zhao et al. (2019b) applied completeness, correctness, and quality as metrics, comparing their proposed method with seven other methods. However, the results show that some low buildings or low parts of buildings cannot be detected as buildings. Huang et al. (2019) developed FCN networks by fusing HRAI and Lidar data for building extraction. The ground truth of building footprints was extracted from OSM. Wierzbicki et al. (2021) investigated the application of the modified U-Net for segmenting high-resolution aerial orthoimages and Lidar data to extract building outlines automatically.

It has been noticed that several studies are focusing on indoor 3D modelling of buildings. For instance, Chen (2018) applied airborne Lidar and Google Maps to provide information for increasing the accuracy of indoor 3D modelling from mobile and terrestrial point clouds.

However, indoor observation is not the focus of this study, so only research related to large-scale building extraction from land is considered.

In conclusion, while various DLSS methods have been introduced for building extraction from land cover objects using Lidar, detecting buildings with low heights remained challenging. Furthermore, the availability of building datasets derived from Lidar data is notably limited compared to those derived from 2D images. While several public 2D building datasets exist, there is a noticeable scarcity of Lidar datasets specifically designed for building extraction purposes.

### 2.5.4 DLSS methods for post-earthquake BDLC with 3D Lidar point clouds

There are several studies discussed about Lidar-based structural damage assessment in the remote sensing field. BDLC using remote sensing technologies can be mainly accomplished through three approaches, including employing multiple feature extraction methods, incorporating geometric and topological features of the buildings, or adopting DL techniques.

Although the first two can provide detailed damage information, most of them require terrestrial and mobile laser scanners with in-situ observations. Those kinds of time-consuming, unsafe, and labour-intensive methods may not be suitable for rapid response after natural hazards. For instance, Akhlaghi et al. (2021) presented a post-earthquake damage identification and performance assessment study of a single four-story building in Nepal from the structural engineering perspective, using ambient vibration and point cloud data. However, this is only suitable for a post-disaster investigation with no time limitation.

Rescue teams have no time to observe all damaged buildings one by one, so a speed automatic large-scale building damage level classification is necessary for rescue plans and strategies. With the fast development of AI, DL provided a possible solution to the above issue. Therefore, this study only focuses on the literature review of airborne and spaceborne Lidar applications.

Most related studies for BDLC using remote sensing technologies applied DLSS methods. For instance, Yang et al. (2019) proposed an inversion method to detect building heights using vertical information from the Geoscience Laser Altimetry System waveform and auxiliary horizontal information of QuickBird optical images. Ma et al. (2020) proposed an improved Inception-V3 method with CNN that combined aerial images and block vector data for evaluating the damage degree of groups of buildings. The case study was the 2010 Yushu Earthquake. Xiong et al. (2020) adopted a fine-tuned CNN-based VGGNet to study damage assessment of buildings after earthquakes using UAV-captured aerial images and GIS data containing building height information. The case study was the damaged Beichuan town after the $M_S$ 8.0 Wenchuan earthquake in 2008 with a 66 multi-story buildings investigation. Although existing studies discussed building damage levels, most of them lack the details. However, most earthquake disaster responses require detailed multi-level building damage information.

## 2.6 Datasets

### 2.6.1 2D building datasets and map databases

#### 2.6.1.1 2D building datasets

All the following datasets for building footprint detection have a very high spatial resolution from aerial or satellite images.

- **WHU Building Dataset (Ji et al., 2018b)**

  This dataset is an aerial and satellite imagery dataset. The aerial sub-dataset contains more than 220, 000 buildings with 0.075 m spatial resolution and covering an area of 450 $km^2$ in Christchurch, New Zealand. The satellite imagery sub-dataset consists of two parts, containing 204 and 17,388 images, respectively. The images in the first part are collected from cities in different countries with various resources such as QuickBird, Worldview series, and IKONOS. The other part consists of 6 neighbouring satellite images covering 550 $km^2$ in East Asia with 2.7 m ground resolution.

- **SpaceNet series building datasets (Van Etten et al., 2018)**

  SpaceNet series building datasets consist of Space 1 and Space 2 datasets. This series dataset was first provided in the DeepGlobe Satellite Challenge of the IEEE Conference on Computer Vision and Pattern Recognition 2018 (CVPR 2018). Some labelled files were not published in the challenge, and the prediction results could only be evaluated during the challenge. Thus, some studies only selected some parts of image scenes with labelled files as the dataset for their studies (Li et al., 2019).

SpaceNet 1 dataset contains 382,534 buildings, covering an area of 2,544 $km^2$ of WorldView-2 imagery with 0.5 m spatial resolution in Rio de Janeiro, Brazil.

SpaceNet 2 dataset includes 302,701 building footprints of Worldview-3 satellite imagery at 0.3 m spatial resolution across five cities. These cities are Rio de Janeiro, Las Vegas, Paris, Shanghai, and Khartoum.

- **INRIA Aerial Image Labelling Dataset (Liu et al., 2018)**

  This dataset comprises orthographic aerial images in ten cities with green, red, and blue (RGB) bands worldwide. Each tile contains 5000×5000 pixels at a spatial resolution of 0.3 m, covering about 2.25 $km^2$.

- **Massachusetts Buildings Dataset (Mnih, 2013)**

  This dataset consists of 151 aerial images of the Boston area, Massachusetts, U.S. Each image has 1500×1500 pixels with 1 m spatial resolution.

  https://www.kaggle.com/datasets/balraj98/massachusetts-buildings-dataset

#### 2.6.1.2 Map databases

- **OpenStreetMap (Steve Coast, 2004)**

  It is a free and open geographic database updated and maintained by volunteers since 2004.

- **Google Maps (Google, 2005)**

  Google Maps is a web mapping platform and consumer application offered by Google.

- **Map World (or "Tian Di Tu" or 天地图) (National Administration of Surveying, 2011)**

  It is a comprehensive Chinese geographic information service website built by the National Administration of Surveying, Mapping and Geoinformation of China.

### 2.6.2 Point cloud datasets for DLSS

There are some published open-source point cloud datasets for DLSS. This section introduces some well-known ones. While these popular datasets include building-related information, most of their point clouds are not acquired through airborne or spaceborne remote sensing devices, making them less aligned with the focus of this study. Nonetheless, these datasets are included here for potential future research within the domain of DLSS.

- **KITTI dataset (Geiger et al., 2012)**

  Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) dataset is a large-scale outdoor dataset applied in semantic segmentation (Geiger et al., 2012). It consists of traffic scenarios recorded with various sensor modalities, so it is mostly applied in the field of mobile robotics and autonomous driving. Despite its popularity, the dataset itself does not contain ground truth for DLSS. Several studies have manually labelled parts of the dataset to fit their necessities.

- **Semantic3D dataset (Hackel et al., 2017)**

  Semantic3D is one of the most popular point cloud datasets for DLSS (Hackel et al., 2017). It includes scanned outdoor scenes with over 4 billion labelled points with

various labelled classes collected by static terrestrial laser scanners. The building class is one of the labelled classes. This benchmark has greatly bridged the gap of the lack of large-scale labelled datasets.

- **Toronto-3D (Tan et al., 2020)**

  Toronto-3D is a large-scale urban outdoor point cloud dataset acquired in Toronto, Canada, for DLSS. This dataset covers approximately 78.3 million points with eight labelled classes, including unclassified, road, road marking, natural, buildings, utility lines, cars, and fences.

- **SensatUrban (Hu et al., 2021)**

  The SensatUrban dataset is an urban-scale photogrammetric point cloud dataset with nearly 3 billion labelled outdoor points. The dataset consists of large areas from two cities covering about 6 $km^2$ of the landscape. In the dataset, each 3D point is annotated as one of the 13 semantic classes, such as ground, vegetation, and building. The publisher of this dataset also proposed a light-wise DLSS method called RandLA-Net (Hu et al., 2020), which will be introduced in the following subsection.

## 2.7 Summary and the remaining gaps

A literature review is presented in this chapter to support the statement of the status of DL technologies and their adoptions for earthquake-related building analysis, including pre-earthquake building footprint extraction and post-earthquake BDLC.

Figure 2-2 summarises existing studies and illustrates the research gaps existing in them. It reflects how these gaps were found gradually. The following paragraphs state these gaps in detail.

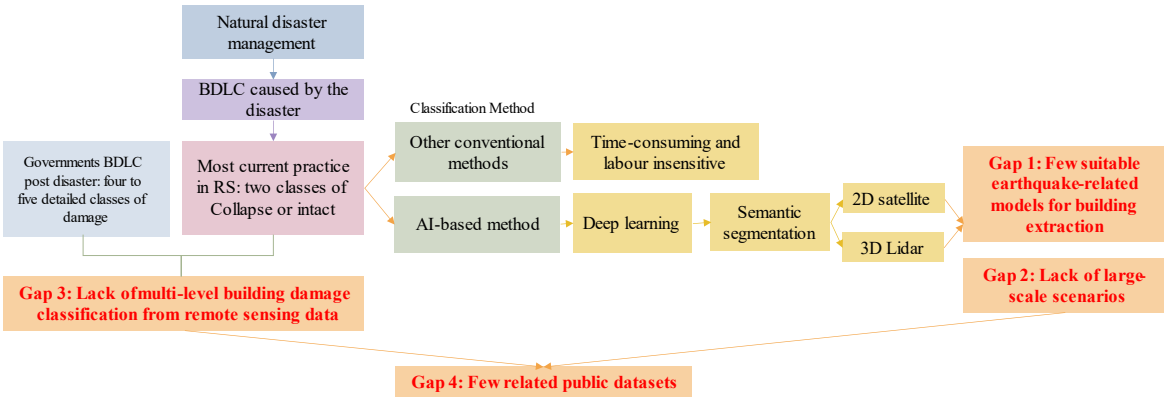



Figure 2-2. Research gaps in the reviewed literature

There are four issues that should be pointed out. Firstly, only some DLSS models are designed for earthquake-related buildings that can be applied for model training and are in short supply, no matter 2D or 3D data. Secondly, since BDLC based on DL just quickly developed in recent years, there is a lack of analyses of large-scale outdoor scenarios. Thirdly, although several detailed damage patterns and levels have been defined in the standards and codes worldwide, most cases in the real world only adopt two levels (collapsed/intact). Fourthly, there are few related large-scale datasets, including detailed labelled building damage levels. These four key gaps are illustrated as follows.

**Gap 1**: There is little evidence of applying DLSS models for building-related classifications considering data sources related to earthquakes, no matter whether 2D satellite images or 3D Lidar point clouds.

**Gap 2**: Several well-known DL algorithms were proposed based on small or indoor case studies in both 2D and 3D applications. The large-scale scenarios were few discussed or tested.

**Gap 3**: Most current post-earthquake BDLC studies lack detailed damage levels in the remote sensing field. This is because of the lack of rapid, detailed assessment methods. It is hard to provide all request parameters for those methods in a short time, and the parameters should be corrected several times.

**Gap 4:** There is a lack of large-scale earthquake-related pre-event building footprint information and post-event multi-level building damage data.

Considering the above gaps, efforts should be made to improve the detailed classification of post-earthquake building damages in the remote sensing field. It is necessary to develop current DL methods to provide a new approach to detecting building damages from remote sensing data semi-automatically or automatically. The following chapters take on this role.

# Chapter 3

# Research design

This chapter presents the methodological approach and research design steps adopted in this study to achieve research objectives. Section 3.1 introduces the research focus of this thesis. Section 3.3 has a description of the research methodology in the remote sensing field. Section 3.4 presents the justification for the research design, the workflow and the logic of the whole study, and the selection of particular methods. Finally, Section 3.5 summarises the research design and methodology applied in this study.

## 3.1 Research focus

Firstly, the focus of this study is earthquake-related damage assessment. Among natural disasters, major earthquakes can always lead to high casualties, so this study chooses earthquakes as the focus. As the damaged buildings after different natural disasters have various classification standards and damage statuses, studying building damage classifications for all natural disasters will be huge and long-term research beyond the scope of this thesis. Secondly, this study focuses on single-building damage assessment. Although some research articles analyse damage at the building cluster level, this study focuses only on the damage at each building level. The reason is to provide more accurate and higher resolution information for designing rescue and recovery plans after an earthquake.

Thirdly, the proposed DL-based semantic segmentation methods focus on large-scale outdoor scenarios. Considering this focus, datasets in this study are collected from either airborne or spaceborne remote sensing devices. It should be noted that 2D and 3D data have different formats, but this study only discusses 2D satellite optical images and 3D Lidar point clouds.

Fourthly, to analyse a large number of buildings, urban areas are the main study extents in this thesis. This is because urban areas usually include more buildings and more sophistications of building shapes and arrangements than regional or rural areas.

## 3.2 Research methodology of remote sensing

The remote sensing method is suggested as either scientific, technological or a combination of both (Bhatta, 2013). The scientific method relies on observations. It encompasses a range of techniques for exploring phenomena, acquiring novel insights, or correcting and integrating existing knowledge. For example, examining the spectral reflectance characteristics of greenery reveals a conclusion that greenery exhibits the highest reflectance within the near-infrared band of the optical electromagnetic spectrum. The technological method is more related to applications but not to specific products or processes. It is aimed at developing tools, models, or procedures, as well as testing equipment and procedures, all geared towards providing solutions to specific technical problems. As the example provided by Bhatta (2013), developing a model for predicting forthcoming urban expansion is technological research. Considering the objectives, a

combination of both is adopted in this study. Details of methods are introduced in the following section.

It should be mentioned that a widely applied classification of research is quantitative research and qualitative research. Since qualitative designed methods are always subjective, they are not suitable for this study. A quantitative design is adopted in this study to evaluate the accuracy and performance of the approaches proposed by this study.

## 3.3 Research design and methodology

This section offers an overview of the methods employed in this research. As discussed in Section 2.7, there are significant gaps in the research on evaluating and identifying appropriate applications of DL for BDLC. This thesis intends to fill those gaps by evaluating DL methods using different data sources. Semantic segmentation is applied for both extracting pre-earthquake building locations and classifying post-earthquake building damage levels.

Achieving pre- and post-earthquake building information is the target of the study, including pre-earthquake building extraction and post-earthquake BDLC. Chapters 4-7 are the main chapters of experiments to achieve this target with four objectives in this study.

The connections between them are shown in Table 3-1. Chapter 4 is designed for Objectives 1 and 4. Chapters 5 and 6 are designed to achieve Objectives 2 and 4. Chapter 7 focuses on achieving Objectives 3 and 4. Since all these chapters have their own labelled datasets, all of them are related to Objective 4.

As mentioned in Section 1.4, the scope of this study is designed using 2D and 3D remote sensing data. Therefore, this study separates chapters according to the data sources they applied. The detailed workflow of this study is introduced as follows, as shown in Figure 3-1. All of these chapters apply DLSS methods, but there are differences between them. As explained in Figure 3-1, Chapter 4 is designed to focus on 2D-related BDLC. It mainly consists of two sections: pre-earthquake building footprint extraction and post-earthquake BDLC. Chapter 5, Chapter 6, and Chapter 7 focus on 3D-related research. Chapter 5 and Chapter 6 discuss the pre-earthquake-related research in this study. Chapter 7 focuses on post-earthquake BDLC. Detailly, Chapter 5 discusses the influence of changing features on DL-based large-scale outdoor Lidar semantic segmentation. After that, Chapter 6 evaluates the performance of the proposed DL network from this study. The network was designed considering the results of Chapter 5. Based on Chapter 5 and Chapter 6, Chapter 7 proposed a DL-based BDLC method to detect buildings into four damage levels.

Table 3-1. Mapping objectives to chapters

| Objective / Chapter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **4** | ✓ | | | ✓ |
| **5** | | ✓ | | ✓ |
| **6** | | ✓ | | ✓ |
| **7** | | | ✓ | ✓ |

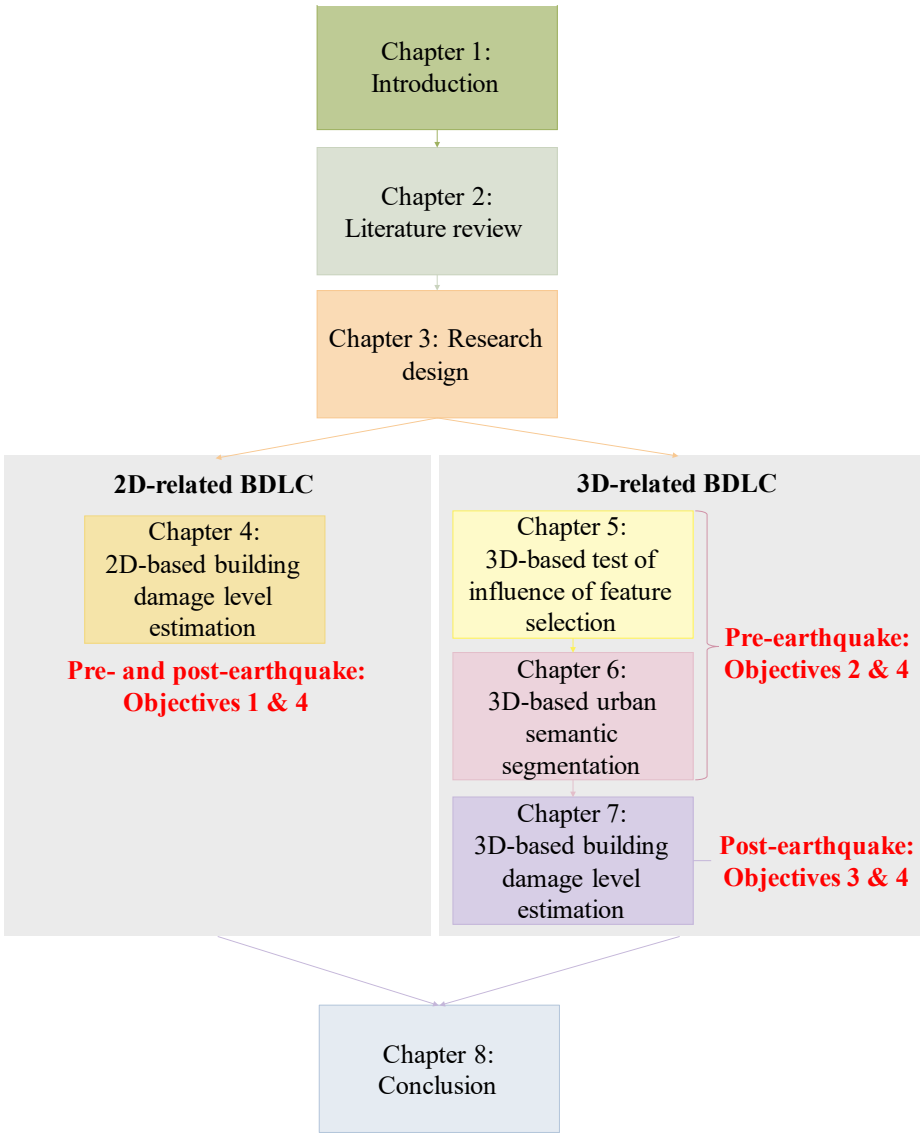



Figure 3-1. Research workflow of this study

To introduce the details of the research design for each objective, Table 3-2 is listed below. It lists the data, methods, and corresponding chapters of each objective.

Table 3-2. The research design for each objective

| Objective | Chapter | Description | Data | | Method |
|---|---|---|---|---|---|
| **1** | 4 | To propose a BDLC method considering both pre- and post-earthquake periods using DL with large-scale optical satellite images. | **2D optical data:**<br>xBD dataset; 2010 Haiti Earthquake dataset. | | Propose a DL method for four-level BDLC with 2D data. |
| **2** | 5,6 | To offer a DL-based pre-earthquake building footprint extraction method with large-scale Lidar data tested in the case studies whose locations have the possibility of earthquakes. | **2D optical data:**<br>KOMPSAT-3 & Sentinel-2 satellite images | **3D Lidar point clouds:**<br>2021 Kapiti Coast, New Zealand; 2022 Tasman, New Zealand; 2022 Nelson, New Zealand;<br>2016 Kumamoto (Pre-earthquake). | Propose a DL method for land cover object semantic segmentation with Lidar and satellite data. |
| **3** | 7 | To provide a DL-based post-earthquake BDLC method with large-scale Lidar data. | **2D optical data:**<br>KOMPSAT-3 satellite images | **3D Lidar point clouds:**<br>2016 Kumamoto (Post-earthquake). | Propose a DL method for four-level building damage classification with Lidar and satellite data. |
| **4** | 4,5,6,7 | To build 2D satellite and 3D Lidar in-house labelled datasets of pre-earthquake building footprints and post-earthquake building multi-level damage information. | • Create a 2D 2010 Haiti Earthquake dataset with manually labelled four-class building damage levels.<br>• Create 3D colourised Lidar building footprint datasets fusion with optical satellite RGB bands, including three locations in New Zealand (Kapiti Coast, Tasman, Nelson) and one in Japan (Kumamoto).<br>• Create a 3D colourised Lidar building damage level dataset fusion with optical satellite RGB bands of the 2016 Kumamoto Earthquake. | | |

The datasets utilised in this study include public and in-house labelled 2D optical images and Lidar point clouds. As shown in Table 3-2, besides the public xBD dataset, this study also collects optical images from KOMPSAT-3 (K3) and Sentinel-2 (S2) satellites. Only RGB bands are applied to all 2D images for colour information because other bands are not the focus of this study. Post-earthquake datasets are included in Chapters 4 and 7, and pre-earthquake datasets are applied in all main chapters from Chapter 4 to Chapter 7. This is because post-earthquake BDLC requires pre-earthquake information, such as pre-earthquake building footprints, but pre-earthquake building footprint extraction does not need post-earthquake information. This is also the reason that although this study focuses on post-earthquake multi-level BDLC, related pre-earthquake research also needs to be discussed.

The 2010 Haiti Earthquake and the 2016 Kumamoto Earthquake are two key case study areas because both have several destroyed buildings in various degrees. Due to the lack of training data, datasets from other types of natural disasters are also collected for training, such as xBD. It should be noted that these datasets are only mainly utilised for increasing the number of inputs in the training stage. This might be helpful to improve the accuracy of those DL models and networks.

Since DLSS techniques provide possible solutions for Objectives 1-3, all related chapters apply it for either pre-earthquake building extraction or post-disaster BDLC. Therefore, all these chapters should build many inputs for the training. The detailed designs of experiments of the main chapters are stated in the following corresponding chapters, respectively, including data collection and pre-processing methods, the architecture of the

designed DL model/network, deep learning model training and evaluation design, the design of evaluation metrics to test the trained model/network.

## 3.4 Chapter summary

This chapter provided the details of the research design of this study according to the objectives. The research workflow explains why this study designs Chapters 4-7. The order of these chapters is designed according to the objectives of this study and the data sources they apply, i.e., 2D or 3D datasets. In Chapters 4-7, each of them has one objective to achieve from Objectives 1 to 3. Moreover, they all have in-house labelled datasets to achieve Objective 4.

# Chapter 4

# Building damage evaluation from satellite images with an attention-base deep learning method[1]

## 4.1 Background and scope

As one of the most common approaches after disasters, on-site damage investigation can provide detailed information, but it is time-consuming and laborious, with a high risk of working in the field (Tanjung et al., 2020). If building damage levels can be obtained using remote sensing techniques with minimal delay, rescue teams and governments can make post-event decisions with the least on-site observation. Therefore, a quick post-event building damage classification method is critical to post-disaster management. Remote sensing can help resolve this issue by obtaining building data remotely (Ji et al., 2018a). Since pre-earthquake information is also essential for post-damage assessment, this chapter

[1] The content presented in this chapter is partially adopted from the following publication: "Liu C, Sepasgozar S, Zhang Q, and Ge L*, 2022. A Novel Attention-Based Deep Learning Method for Post-Disaster Building Damage Classification. *Expert Systems with Applications*. 202, p.117268. DOI: 10.1016/j.eswa.2022.117268". It has been acknowledged and detailed in the "Inclusion of Publications Statement" for this thesis.

presents a post-earthquake BDLC approach using 2D remote sensing images taking into account not only post-event information but also pre-event building data.

The fast development of DL in computer vision provides a pathway to offer quick classification (Yang et al., 2021, Su et al., 2020, Wheeler and Karimi, 2020). DL is widely applied in remote sensing and computer vision-related applications. It has the potential to overcome several limitations of conventional geoscience methods (Reichstein et al., 2019). Therefore, this chapter applied DL methods for segmenting building areas and categorising damage into four levels, no-, minor-, major-, and total damage.

As mentioned in Chapter 2, there is a limited availability of 2D open-source image datasets for labelled post-disaster building damage levels. Another challenge is the scarcity of DLSS models designed specifically for detecting damaged buildings. Moreover, most studies have primarily focused on two-level classification, distinguishing between collapsed or intact structures.

To address these issues, the current chapter aims to propose and assess a novel 2D DL for quickly classifying detailed post-event building damage levels. This approach utilises both publicly available and internally labelled datasets of damaged buildings. The model comprises two primary steps: pre-event building localisation and post-event damage classification. The initial step involves identifying building footprints using predisaster images, while the subsequent step categorises damage levels using post-event images based on the footprints determined in the first step.

This chapter employs two optical satellite image datasets: the publicly available xBD dataset and our internally labelled 2010 Haiti Earthquake dataset. Both datasets are annotated across four damage levels as mentioned above. The xBD dataset was published for 2019 Defense Innovation Unit Experimental (DIUx) xView2 Challenge of building damage classification (Gupta et al., 2019a). However, images after earthquakes are insufficient in the xBD dataset, though it contains several post-event images of different natural disasters such as tsunamis, bushfires, and tornados. Because of that, the second dataset, the 2010 Haiti Earthquake dataset, is added to the study. Building footprints were drawn and damage levels were labelled manually. Building damage levels in it are categorised based on the analysis in 2010 (UNITAR/UNOSAT/EC/JRC/WB, 2010). The number of its damage levels is the same as that in the xBD dataset. This study drew outlines of buildings and labelled damage levels manually. The detailed experiments are explained in the following paragraphs.

Considering the advantages of the aforementioned DL models as discussed in Section 2.4.1, this chapter employs three advanced strategies, including residual blocks, Squeeze-and-Excitation (SE) attention mechanism, and GPU training strategy (Krizhevsky et al., 2012, He et al., 2016a, Hu et al., 2018). ). The selection of HRNet is based on its ability to retain the high-resolution quality of input images, a crucial aspect for sensitive earthquake analyses. Given the two requisite analysis steps, namely, building footprint localisation and damage level classification, the present experimentation employs a dual HRNet to encompass both stages.

The literature presented in Chapter 2 shows that studies applying DL models in different contexts are rich. However, there are fewer resources and evidence evaluating different versions of the models using different functions or blocks applicable to earthquake data sources. A comprehensive set of comparative evaluations is imperative to fulfil the requirement of assessing the performance efficiency of these models. The following four aspects are crucial for assessing the accuracy of a model: the point at which a block is inserted into the architecture of a model (Hu et al., 2018), the model's accuracy with and without pre-trained weights (Liu et al., 2021), the flexibility to handle different input image sizes (Wang et al., 2020), and the efficiency of activation functions within the SE block.

Firstly, it is recommended to consider the comparative evaluation of different insertion points for a block within a model for the purpose of evaluation. For instance, Hu et al. (2018) compared the accuracy of different DL structures for the image segmentation task by adding a block in different parts of the same backbone. Therefore, in this chapter, the SE block is inserted at various positions within the basic unit of HRNet as the initial comparative experiment, aiming to analyse the optimal location for its incorporation.

Secondly, a crucial aspect that needs to be scrutinised for each model is its ability to function with or without pre-trained weights, aiming to evaluate the accuracy of outcomes within the context of earthquake building damages. For example, Koo et al. (2020) employed the ImageNet dataset, but they neither conducted any comparisons nor elucidated the rationale behind utilising pre-trained weights in their experimentation.

Thirdly, this chapter addresses the accuracy of the output concerning variations in the sizes of input images, mirroring the diversity found in real-life events. Limited published experiments have reported such comparisons due to the constraints of processing merged and substantial images on computers lacking high specifications and costly GPUs. This comparative analysis holds significance as it aims to ascertain whether smaller-sized images can yield outcomes of comparable accuracy to those generated from larger-sized images.

The fourth item refers to the evaluation of different activation functions that can affect the optimisation of a model. Due to the availability of different activation functions and the gap of sources reporting the effect of each function in different contexts, it is required to examine any selected functions for earthquake building detection purposes. All in all, this chapter will bridge this gap by conducting a set of four comparisons, which will be discussed in the following subsections.

## 4.2 Datasets of 2D building damage classification

The dataset of this study contains pre-event non-damaged and post-event damaged building images. Building damage is classified into four levels, including no, minor, major, and total damage. All images in this dataset are high-resolution optical satellite images with RGB bands. They are collected from multiple types of natural disasters, including earthquakes, volcanic eruptions, hurricanes, floods, tsunamis, and wildfires. This dataset contains 8,664 images in total, which are 4,332 image pairs. Each image pair has two images, namely, a pre-event image and a post-event image. This dataset is a mix of two datasets, the xBD

dataset, and the 2010 Haiti Earthquake dataset. 1,200 images (600 pairs) come from the Haiti earthquake, and 7,464 images (3,732 pairs) are chosen from the xBD dataset.

The image number for training is 6,498, including 3,249 image pairs, which are 75% of the whole dataset. Among these training images, 900 images (450 pairs) are collected from the Haiti earthquake, and 5,598 images (2,799 pairs) are chosen from xBD dataset. Among them, 10% of training images are chosen for validation randomly. Test images are 25% of the whole dataset, including 300 Haiti earthquake images (150 pairs) and 1,866 xBD images (933 pairs). The details of xBD dataset and the Haiti earthquake dataset are stated in the following sections.

### 4.2.1 XBD Data

Online free xBD dataset was published in 2019 for the xView2 Challenge (Gupta et al., 2019a). The dataset for training is important for DL methods. Because of the publication of the online free xBD satellite dataset for xView 2 Challenge in late 2019, the automatic processing of assessing post-event building damage attracts more attention (Gupta et al., 2019b). Hence, this chapter chose the xView2 dataset for training. This study chose 5,598 images from this dataset for training (2,799 pairs of pre- and post-event images) and 1,866 images for testing (933 pairs of pre- and post-event images). The chosen images in this study cover several natural disaster events, including volcanic eruptions, hurricanes, earthquakes, floods, tsunamis, and wildfires. These images were collected from different satellites, including GeoEye-1, WorldView-2, WorldView-3_VNIR, and QuickBird-2 (Su et al., 2020). The chosen images do not contain any images from the 2010 Haiti Earthquake.

The size of each image is 1024×1024 pixels below a 0.8 m ground sample distance (GSD) mark. The damage level number is decided by experts invited by the committee that held the xView2 Challenge, which contains four levels mentioned at the beginning of Section 4.2.

### 4.2.2 2010 Haiti Earthquake data

The 2010 Haiti damage building data were labelled in ArcMap manually at the University of New South Wales (UNSW). An $M_w$ 7.0 earthquake happened in Haiti on 12 January 2010, causing serious building damage (Ji et al., 2018a). All images were chosen from Port-au-Prince, which is one of the seriously damaged provinces in Haiti. The size of each image is set as 1024×1024 pixels to have the same size as xBD imagery. Images have some overlapping areas with each other. These optical satellite images were downloaded from the Maxar/DigitalGlobe Open Data Program (Maxar, 2010). The GSD is 0.8 m. The damage level of each building is according to the damage report of this earthquake provided by UNITAR/UNOSAT/EC/JRC/WB (2010). This report also categorised building damage into four levels, which is the same number as the damage level number of xBD data. The building location shapefile is provided by the Operational Satellite Applications Programme (UNOSAT), Joint Research Centre (JRC), and World Bank (WB). The shapefile contains the location point of each building. A building may contain more than one point if it has more than one roof with different heights. These building damage assessment points are projected to the optical images, as shown in Figure 4-1. The area in red is the chosen area for labelling. Footprints were drawn and building damage levels were labelled according to these provided damage levels and location points information.

Figure 4-2 shows an example of the labelled buildings. The colours representing no, minor, major, and total damaged levels are white, yellow, orange, and red, respectively. It should be noticed that some buildings contain more than one location point. If these points of a building were assessed at different levels, the labels would separate a building into parts with different damage levels according to the assessment.

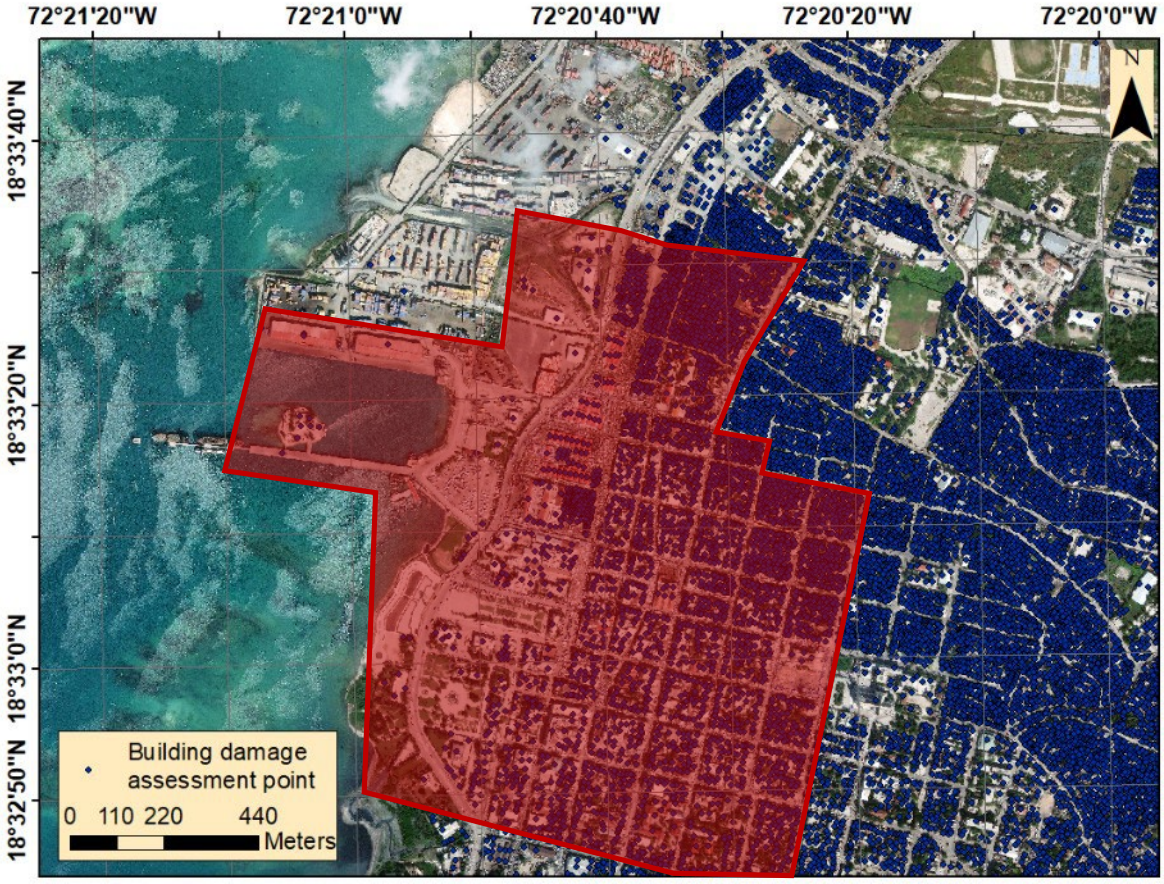



Figure 4-1. The selected Haiti Earthquake dataset area (red area)

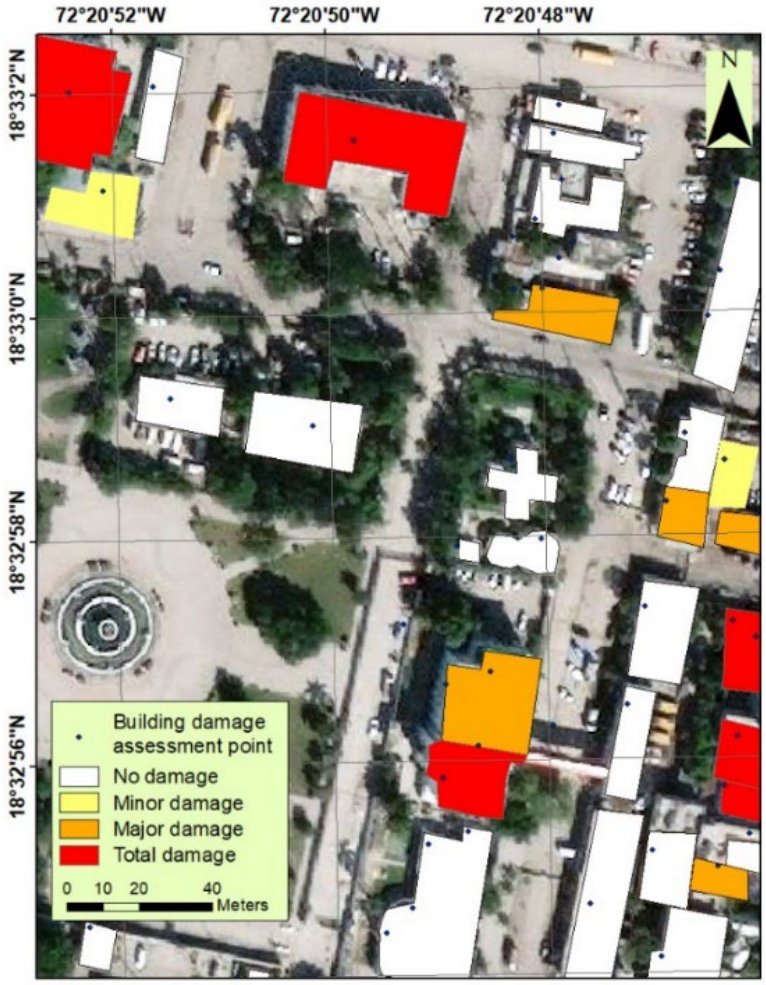



Figure 4-2. Example of labelled buildings with location points and labelled building damage levels

### 4.2.3 Data pre-processing for 2D building damage classification

The size of each image for training is cropped from 1024×1024 pixels to 256×256 pixels randomly. The choice of 256×256 is according to several reasons. First, several widely applied DL models are trained with the test of small input images smaller than 300 when these models are proposed initially. For instance, both SENet (Hu et al., 2018) and ResNet (He et al., 2016a) adopt 224×224 for training cropping. The original size of each image is

1024×1024, and 1024 can be evenly divisible by 256, not 224. Hence, this thesis selects 256 instead of 224. The 256×256 size is enough to cover the identified target range, and the use of smaller input is conducive to reducing the number of parameters, reducing the risk of overfitting, and increasing the operation speed.

Second, based on several attempts, the hardware of this study can afford the training with the input of 256×256 pixels for all five structure options. The detailed design will be introduced in Section 4.6. If the cropped input size is larger, the training time is very long, and the GPU memory is not enough. This is also why this study only uses the standard SE model other than SE-PRE, SE-POST, and SE-identity for the comparative experiment with different pixels. If 512×512 input size is applied in the other structure options, the GPU memory is not enough.

This chapter applied data augmentation. To reduce the possibility of overfitting, data augmentation, including random horizontal and vertical flipping, random 90-degree rotation, and random scaling (between 0.8 and 1.2), are adopted. Effective data augmentation can not only increase the number of images in the training set but also enrich the diversity of samples. On the one hand, it can avoid the overfitting phenomenon. On the other hand, it can improve the performance of DL models based on prior experience.

## 4.3 2D Building Damage Segmentation method

### 4.3.1 Workflow

This chapter proposes a SE-based dual-HRNet model for building damage classification. There are three main stages in this chapter, including data pre-processing, model training, and model testing. Four comparative experiments are implemented in the test stage. The workflow is shown in Figure 4-3.

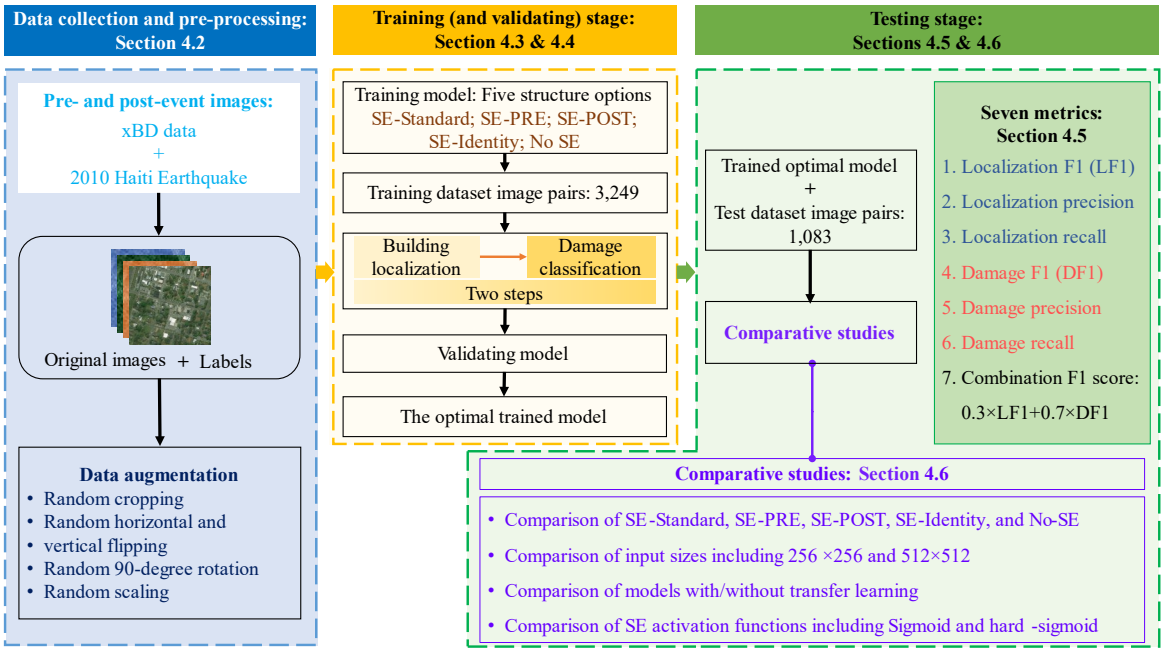



Figure 4-3. Workflow of the proposed method

### 4.3.2 Data pre-processing stage

The first stage is data pre-processing. The input data are pre- and post-event images from disasters. These images are the xBD dataset and the 2010 Haiti Earthquake dataset. Each image is labelled with all building footprint coordinates and the building damage level of

each building. All images are cropped, flipped, rotated, and scaled randomly for data augmentation. The detailed augmentation information of these two datasets is stated in Section 4.2.

### 4.3.3 Training stage

At the second stage, five structure options of the model are trained. They are SE-Standard, SE-PRE, SE-POST, SE-Identity, and No-SE. The training stage contains two steps, including localising buildings with pre-event images and classifying building damage into four levels with post-event images (representing no, minor, major, and total damage, respectively).

The dataset for training is important for DL methods. The number of image pairs at this stage is 3,249 for training and validating the model. Each pair contains a pre-event image and a post-event image. The structure of the model applies the backbone model twice to connect these two steps, which is the dual-HRNet model. The cross-validation method is applied to find the optimal parameter configuration. The information on the model is stated in Section 4.4. To be specific, key blocks in the proposed model are shown in Section 4.4.1, and the detailed structure of the model is stated in Section 4.4.2.

### 4.3.4 Testing stage

At the third stage, the optimal model of each option from the training stage is tested with the test dataset based on seven metrics, including a combination F1 score, three metrics for building localisation (localisation F1, localisation precision, localisation recall), and three metrics for damage classification (damage F1, damage precision, and damage recall). The

optimal model is chosen from the model with the highest combination F1 score. The number of test dataset image pairs is 1,083. The results show the damage level of each building. The results of these seven metrics are recorded for the following comparative studies. The details of metrics are stated in Section 4.5.

As shown in Figure 4-3, four types of comparison experiments are implemented in this study during the test stage. The first comparison is to find where the best is to place the SE channel attention (CA) block. Four models with different SE added places and one model without SE block are compared. Second, to judge the influence of cropped input size, a comparison of 256×256 and 512×512 is made. Third, models with and without transfer learning are trained to judge whether the pre-trained ImageNet image classification weights can improve the model performance or not. The last one is the comparison of two activation functions in the SE block, including Sigmoid and Hard-Sigmoid. Details of each experiment are stated in Section 4.6.

## 4.4 Proposed model in this chapter

This section introduces the proposed model. First, the structure of this proposed model is explained in Section 4.4.1. Second, the key blocks applied in the model are stated in Section 4.4.2.

### 4.4.1 Structure of the proposed model

The structure of the model proposed in this chapter is shown in Figure 4-4. A dual HRNet with added SE model is designed. The “dual HRNet” in this model means a parallel

structure with two HRNet. This chapter kept using “dual-HRNet” which is from fifth place in the xView 2 Challenge (Koo et al., 2020). This dual HRNet structure is used for the two steps in this model, including building localisation and damage classification, respectively, as shown in the blue and the orange rectangles of Figure 4-4. The first HRNet (shown in blue) gives the outputs of building locations. Only pre-event images are used in it. The second HRNet (shown in orange) accesses building damage levels. Building footprints in it are according to the locations from the results of the first HRNet model. Post-event images are exploited in the second HRNet with the building location results based on pre-event images. Hence, the results of this structure contain both building locations and damage levels.

These two HRNet are fused by adding the output channels together of one stage at the beginning of the next stage. The main function of the convolution layer is to extract features, which can provide deeper features through multi-layer convolution. Therefore, this study keeps all output layers of the two HRNet for saving information. SE CA block is added at each basic residual block in the dual HRNet. The detailed structures of the four adding SE options are given in Section 4.6.1.

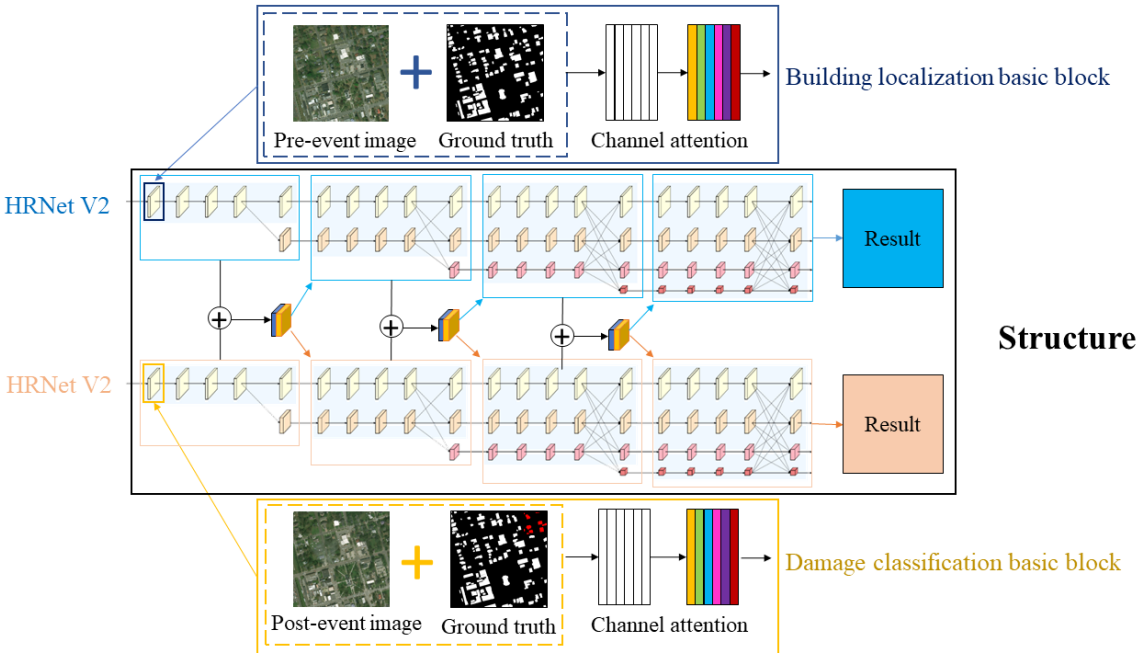



Figure 4-4. Structure of the proposed model

The version of HRNet in this study is HRNetV2W32. The structure of HRNetV2 has been explained in Section 4.4.2. "W32" means that the numbers of convolutional layers in the four stages are 32, 64, 128, and 256, respectively.

Pretrained ImageNet classification weights are added in this chapter. The training with pre-trained weights is called transfer learning. Pretrained weights are often be used in DL to save training time and have good results (Kolar et al., 2018). These weights are the most suitable ones for a completed task which is similar to the task of this chapter. For instance, pretrained ImageNet classification weights suit the image classification task with ImageNet dataset. If one task is similar to that, these weights can be added during the training, and the results may be better than training from scratch.

The output size of a convolutional layer is shown in Equation 4-1 (Dumoulin and Visin, 2016).

$$n_{out} = \frac{n_{in} + 2p - k}{s} + 1 \quad \text{4-1}$$

$n_{in}$: number of input features. $n_{out}$: number of output features. $k$: kernel size. $p$: padding size. $s$: stride size.

Stochastic Gradient Descent (SGD) optimizer is applied with the base learning rate of 0.05. It is a hyper parameter to adjust the weights of the model, and how to use it is shown in Equation 4-2. The momentum is 0.9, and the weight decay is 0.0001. The training epochs are 500, and the trained models are recorded every 50 epochs. Training 500 epochs is according to the number of training epochs from other papers which also use the xBD dataset. For instance, Koo et al. (2020) trained 250 epochs, and Wheeler and Karimi (2020) trained 100 epochs. Hence, 500 epochs are enough, and this chapter checks the validation results during the training to avoid overfitting. Recording the model every 50 epochs is for observing the training details. The model was trained on one Nvidia 2080 Ti GPU server with 60G memory in the Linux system. Considering the condition of the hardware, the batch size per GPU is 4 for both training and testing. The training loss is the sum of localisation loss and damage classification loss by calculating Lovasz-softmax loss (Berman et al., 2018).

$$new_{weight} = existing_{weight} - learning_{rate} \times gradient \quad \text{4-2}$$

### 4.4.2 Key blocks in the model

***Residual block***

There are two most applied structures of residual blocks in this chapter, as shown in Figure 4-5, whose input sizes are the same as the output sizes. Structure A is the basic residual block, which was designed for ResNet with 18 or 34 layers initially (He et al., 2016a). Structure B is called the bottleneck residual block. It was designed for the network with more layers, including ResNet with 50/101/152 layers (He et al., 2016a). Bottleneck residual blocks are variants of basic residual blocks. These bottleneck blocks utilise 1×1 convolutions to reduce the number of parameters and calculating times. The design of the bottleneck helps to increase the depth with fewer parameters of a DL model than basic residual blocks.

There are also some other structures of residual blocks applied in this chapter that require their output sizes to be different from the input sizes. Down-sampling is added in the identity part to change the number of channels in these structures. The down-sample parts contain convolutional and BN layers.

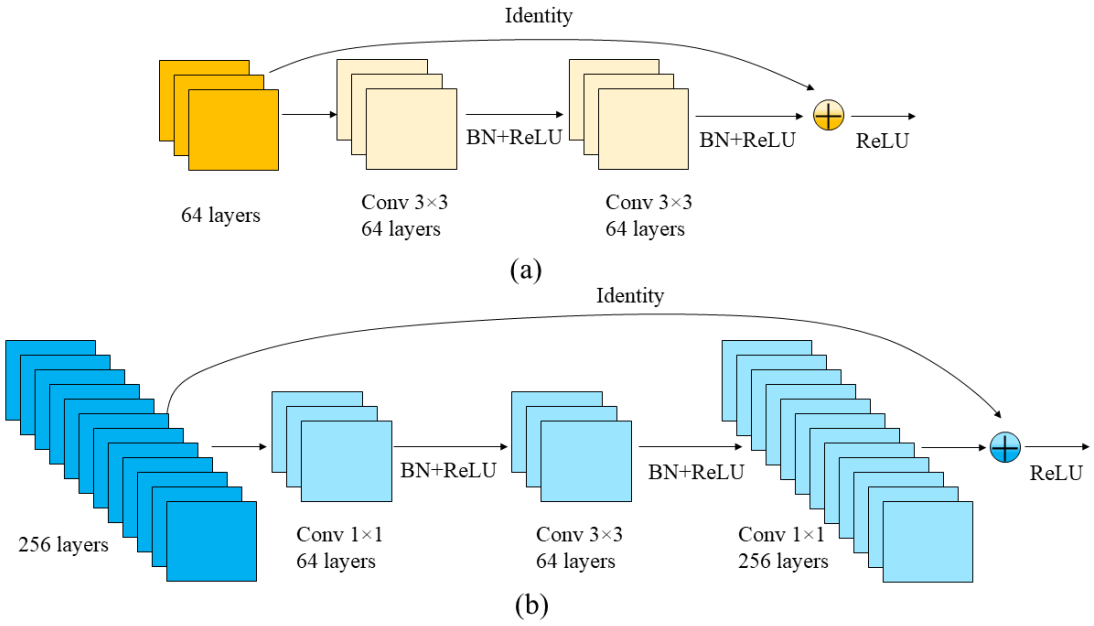



Figure 4-5. Residual block structures. (a) Structure A: basic residual block; (b) Structure B: bottleneck residual block

BN is widely applied before activation functions to speed up the training (Koo et al., 2020), so this chapter also applied BN. BN was proposed in 2015 (Ioffe and Szegedy, 2015). BN is a technique to normalise inputs to one layer in each batch. If it is added before activation functions, activation functions usually have better performance than those without BN. A model with BN does not need to set the bias in convolutional layers before BN, because bias is useless before BN.

The calculation details of BN are shown in Equation 4-3. The output of BN is $y$. $E[x]$ is the mean, and $Var[x]$ is the variance. $\varepsilon$ is the number to avoid the devisor being zero. $\gamma$ and $\beta$ are learnable parameter vectors of size C. $\gamma$ is 1 and $\beta$ is 0 by default. The input shape should be (N, C, H, W), which represents the batch size, channel number, height, and

weight, respectively. The performance of BN is better with larger batch sizes. With considering the hardware condition, this chapter chooses 4×4 as the batch size.

$$y = \frac{x - E[x]}{\sqrt{Var[x] + \varepsilon}} * \gamma + \beta \quad \text{4-3}$$

In this chapter, a defined BN layer in the Torch library (Paszke et al., 2021) is added between convolutional layers and the ReLU activation function. ReLU is a piecewise linear function. Its output is positive if the input is positive. Otherwise, the output will be zero.

***Channel attention block***

The attention mechanism in computer vision draws lessons from the visual attention mechanism in the human visual system. The attention block used in this chapter, the SE block, was proposed in the SENet DL model (Hu et al., 2018). SENet won first place in the ILSVRC 2017 Classification Challenge. It mainly studies the correlation between channels and selects the attention for channels. Although it slightly increases the amount of calculation in computers, the effect is better shown in SENet. Figure 4-6 shows the structure of the SE block applied in this chapter. First, a $H \times W \times C$ block is converted to a $1 \times 1 \times C$ block after global pooling. After that, with the fully connecting operation, ReLU activating, and another fully connecting operation, a $1 \times 1 \times C$ block with attention channels is achieved. Then, this chapter chooses both Sigmoid and Hard-Sigmoid activation functions for training to compare the performance of each other. The details of these two functions and the comparative study will be introduced in Section 4.6. After that, the block

is scaled up to the original size. The hyper parameter r is 16 in this chapter according to the number in SENet.

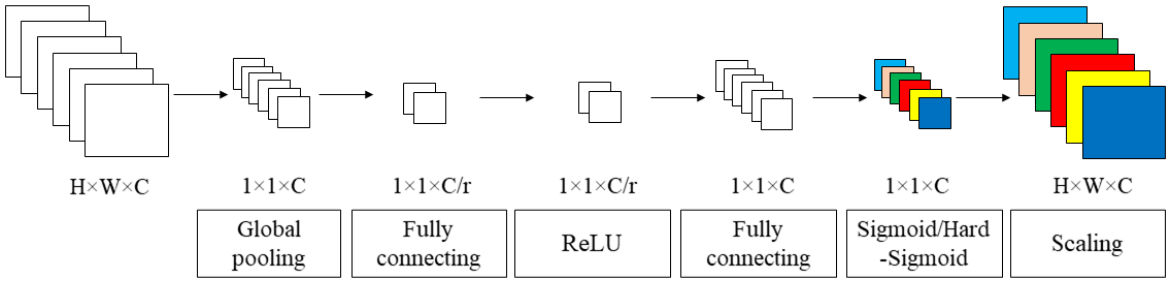



Figure 4-6. SE block in this chapter. *H*: Height; *W*: Width; *C*: Number of channels

***HRNet***

HRNet structure contains four stages. High-to-low resolution convolutions are connected in parallel in this structure. As shown in Figure 4-7, the light-yellow block, the original high-resolution input with all image information, is kept from the beginning to the end of the structure. Gradually scaled-down images are added at the first layer of each stage except the first stage. To be specific, the sizes of the light-orange, light-red, and dark-red blocks decrease gradually with reduced resolution.

This chapter adopts HRNetV2 (Sun et al., 2019b), which is the second version of HRNet. HRNetV2 adds all channels with different resolutions together at the end of the structure. HRNetV1, the first version, was designed for human pose estimation at the beginning, so it is not suitable for image classification and segmentation. In order to be more adaptive for image semantic segmentation, HRNetV2 was designed by concatenating the (upsampled) representations soon after HRNetV1. The structure of HRNetV2 is shown in Figure 4-7.

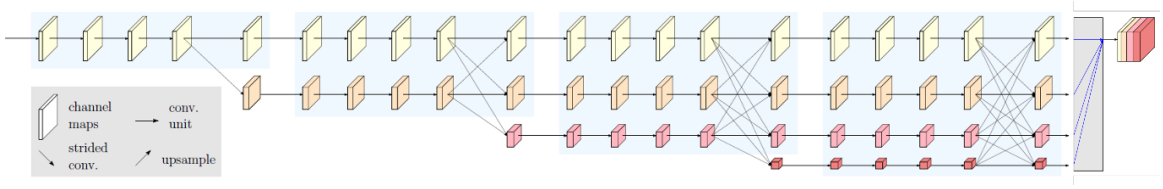



Figure 4-7. Structure of HRNetV2 (Sun et al., 2019b)

## 4.5 Performance metrics

This chapter attempts to test the model performance for each model using seven metrics, including a combination F1 score, three metrics for localisation, and three metrics for damage classification. The combination F1 score is the main metric to judge the performance of the model. This is because it contains both building localisation and damage classification results. The three metrics for building localisation are localisation F1 (LF1) score, localisation precision, and localisation recall. The three metrics for damage classification are damage F1 (DF1) score, damage precision, and damage recall. Since this study recorded a model every 50 epochs ending at the 500$^{th}$ epoch as mentioned in Section 4.4.1, each model has ten recorded test results with these seven metrics.

The equations of the first three metrics, including F1 score, precision, and recall for localisation, are presented from Equations 4-4 to 4-6. True positive (TP) represents a building pixel that is segmented correctly. False positive (FP) means a non-building pixel segmented as building. True negative (TN) is a non-building pixel that is correctly segmented as non-building, and false negative (FN) is a building pixel segmented as non-building wrongly.

$$F1 = \frac{2}{precision^{-1} + recall^{-1}} = \frac{2TP}{2TP + FP + FN} \quad 4\text{-}4$$

$$precision = \frac{TP}{TP + FP} \quad 4\text{-}5$$

$$recall = \frac{TP}{TP + FN} \quad 4\text{-}6$$

The next three metrics, F1 score, precision, and recall for damage classification, are computed as the harmonic mean of those scores of the four damage levels, as shown from Equations 4-7 to 4-9, respectively. In these three equations, $\varepsilon$ is $10^{-6}$ to avoid the denominator being 0 and *i* means the building damage level from one to four.

$$DF1 = \frac{4}{\sum_{i=1}^{4} \frac{1}{F1_i + \varepsilon}} \quad 4\text{-}7$$

$$Damage\ precision = \frac{4}{\sum_{i=1}^{4} \frac{1}{precision_i + \varepsilon}} \quad 4\text{-}8$$

$$Damage\ recall = \frac{4}{\sum_{i=1}^{4} \frac{1}{recall_i + \varepsilon}} \quad 4\text{-}9$$

The F1, precision and recall of each damage level from one to four use the same equations as the equations for building localisation step as shown from Equations 4-4 to 4-6. However, the meanings of TP, TN, FP, and FN for damage classification are different than their meanings for building localisation. TP means a pixel contained in this damage level is categorised by the model correctly. TN means one pixel that is not included in this damage

level is categorised correctly. FP means that a building pixel which is not contained in this damage level but wrongly categorised as this level, and FN represents one pixel in this level is wrongly categorised as another level by the model.

The last metric, the combination F1 score, is applied according to the xView2 Challenge (Diux, 2019). The combination F1 score calculates a weighted average of LF1 and DF1, as shown in Equation 4-10. The numbers are chosen as 0.3 and 0.7 because xView2 Challenge applied these numbers. The percentage of LF1 or DF1 is designed by experts from that xView2 Challenge. Parts of images applied in this chapter are collected from xBD dataset published for xView2 Challenge as mentioned in Section 4.2.1.

$$Combination\ F1\ score = 0.3 \times LF1 + 0.7 \times DF1 \quad 4\text{-}10$$

## 4.6 Comparative experimentations at the test stage

This section discusses a set of appropriate strategies to test the performance of models with different features in earthquake contexts.

### 4.6.1 Experimentation of SE block integrations

Four structure options with different SE added places in the model are compared with the original backbone model in this experiment. Since the final output size of the SE block is the same as its input size, it can be added anywhere in the model backbone without the need to change other layers. In this chapter, SE is added in the following four options in each residual block of HRNet shown from Figure 4-8 (a) to (d) for the comparative study. In other words, SE is added both in basic and bottleneck residual blocks in the codes. Figure

4-8 (e) is the initial block without SE block. The bottleneck residual block mentioned in Section 4.4.1 is chosen as the example to show the added places of the SE block shown in Figure 4-8. The first option shown in Figure 4-8 (a) is adding SE after residual blocks, and then the results are added with the identity function. The second option is adding SE before residual blocks shown in Figure 4-8 (b). The third one is adding SE after the addition of the identity function and residual blocks shown in Figure 4-8 (c). The last one is adding SE and residual blocks together, so the SE block takes the place of the identity part, as shown in Figure 4-8 (d). The input size of this comparative study is 256×256.

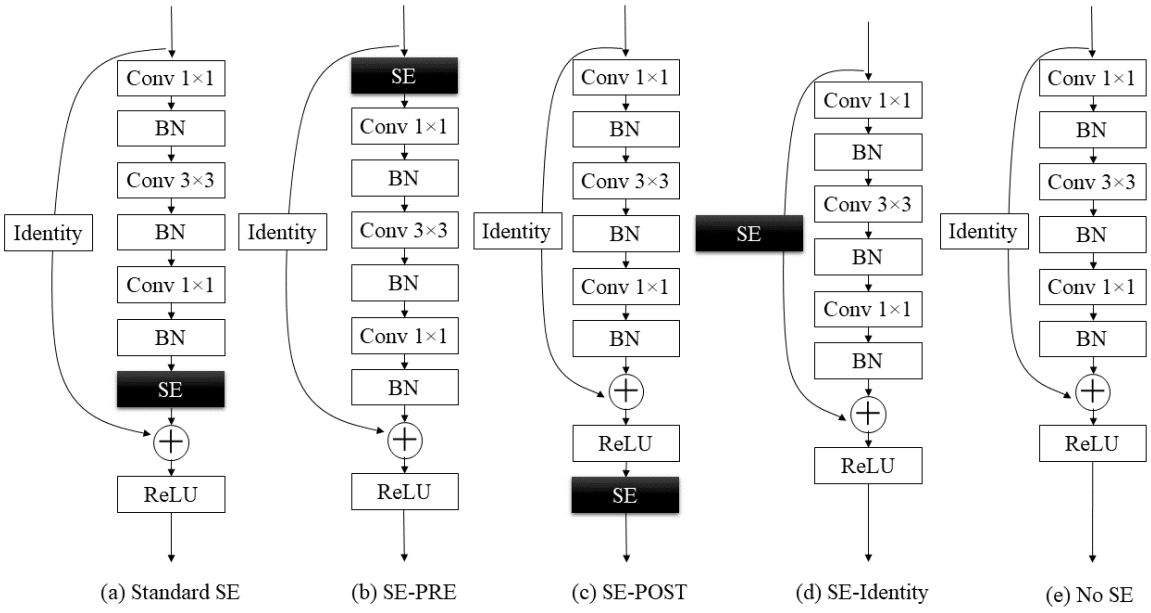



Figure 4-8. The blocks with different options of inserting SE attention

Although two papers (Li et al., 2020b, Li et al., 2020c) have also applied the SE attention mechanism, the targets of their papers are scene classification and human pose estimation, respectively. These applications are different from the applications in this chapter, and the

SE block added places are different. Another difference is that they added the SE block in only one place, while this chapter applies not only their methods but also others. Moreover, although no attention mechanism related codes can be found in the codes provided by (Li et al., 2020c) in GitHub (which is only HRNet codes), it added SE block parallel with residual block according to the figure in its paper. Its structure is shown in Figure 4-8 (d). SE-HRNet added SE block before the summation, as shown in Figure 4-8 (b). Hence, their model structures are different from the model of this chapter even though some parts are similar. This chapter attempts more possible combination modes than them.

### 4.6.2 Experimentation of other hyper parameters

Besides the comparison of SE added places, three types of comparative experiments are implemented in this chapter. The first one is the input size. Cropped input images with both 512×512 pixels and 256×256 pixels after data augmentation are trained for comparison of the model performance with different input sizes. As explained in Section 4.2.3, the choice of these input sizes is considering the hardware performance and previous experience. The "Standard SE" structure is chosen as the training model in this test experiment.

The second comparison is the results between training with and without transfer learning. To be specific, one model is training with HRNetV2W32_ImageNet_pretrained weights, and the other one without transfer learning is training from scratch. Although training with pre-trained weights always has a better result than without a pre-trained model theoretically, it is still necessary to check the results. Therefore, this comparative experiment is implemented in this chapter.

The third one is comparing Sigmoid and Hard-Sigmoid activation functions in the SE block. Sigmoid is shown in Equation 4-11. Here "$e$" is Euler's number.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4-11}$$

Hard-Sigmoid is the segment-wise linear approximation of Sigmoid, which is far less computationally expensive than Sigmoid both in software and specialised hardware implementations, as reported by Courbariaux et al. (2015). Equation 4-12 shows its formula.

$$\sigma(x) = clip\left(\frac{x+1}{2}, 0, 1\right) = \max\left(0, \min\left(1, \frac{x+1}{2}\right)\right) \tag{4-12}$$

## 4.7 Building damage results of the four comparative experiments

The results of the four comparative experiments (refer Section 4.6) are presented in this section. As mentioned in Section 4.4.2, all results are obtained based on the experiments with the test dataset. The performance of each model is compared based on its optimal model among 10 recorded models in this chapter.

The outputs of the test stage are presented as RGB images and evaluation scores. The RGB images contain the footprint of buildings and the damage level of each building. The evaluation scores are based on seven metrics as introduced in Section 4.5.

In each comparative experiment, the combination F1 score is the main metric to judge the performance of each model since it considers both building location and damage

classification results. This chapter not only computes the combination F1 score of each option but also visualises the scores as line charts of ten intermediate epochs from 50 to 500 that show the score of the training progress for each experiment. Moreover, F1 scores, precisions, and recalls for localisation and classification of the optimal model during training are also shown as line charts in this section. The details of all measures for each epoch are all recorded in supplementary documents for reliability checks and reference. Four comparisons are stated one by one as follows.

### 4.7.1 Comparison of SE block integrations

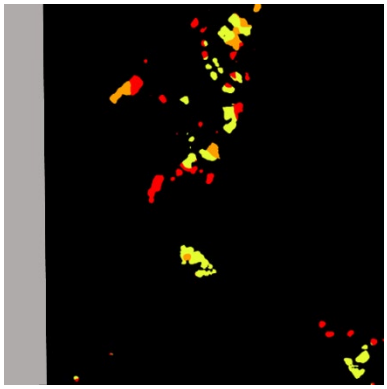
This section presents the outcome of the damage level classification using five structure options (refer to Section 4.6.1) as shown in the purple part of the workflow of Figure 4-3 in Section 4.3.1. In this section, the RGB image results are analysed first and shown in Figure 9. Then, the results of all the seven metrics are analysed, and the summary is presented in tables or as line charts. A sample of the pre-event image is shown in Figure 4-9 (a), which includes buildings and vast green vegetation. Figure 9 (b) shows what areas have been destroyed due to the disaster by applying the algorithm to the testing sample. Figure 4-9 (c) to (g) are visualised results of the five options (refer to Figure 4-8 in Section 4.6.1). To validate the performance of the results of each option with the ground truth, Figures 9 (c) to (g) are compared with Figure 4-9 (h). As mentioned in Section 4.2, white, yellow, orange, and red colours in these images represent no, minor, major, and total damage levels, respectively. Grey means no data.

Figure 4-9 (c) to (g) show that all the five options can segment buildings from a small 256×256 image, which is hard to be judged by human eyes. Standard SE, SE-PRE, and No-

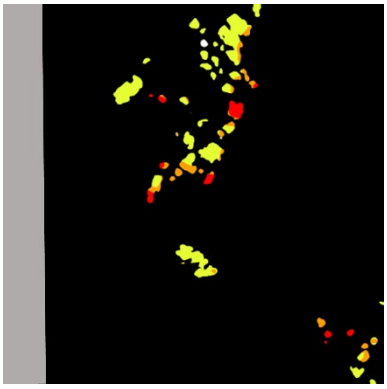
SE perform well for identifying both building locations and damage classifications among the five options. These three options can detect building footprints and classify damage into four levels. However, SE-POST and SE-Identity models are not desirable for both building localisation and building damage classification, as shown in Figure 4-9 (e) and (f). The models with these two options only detected rough locations of buildings, while the other three show more detailed location information. Besides, SE-POST did not detect any damages in this example. SE-Identity only detected two damage levels, including no damage and total damage. Hence, Standard SE, SE-PRE, and No-SE perform much better than SE-POST and SE-Identity.

(a) Pre-event image

(b) Post-event image

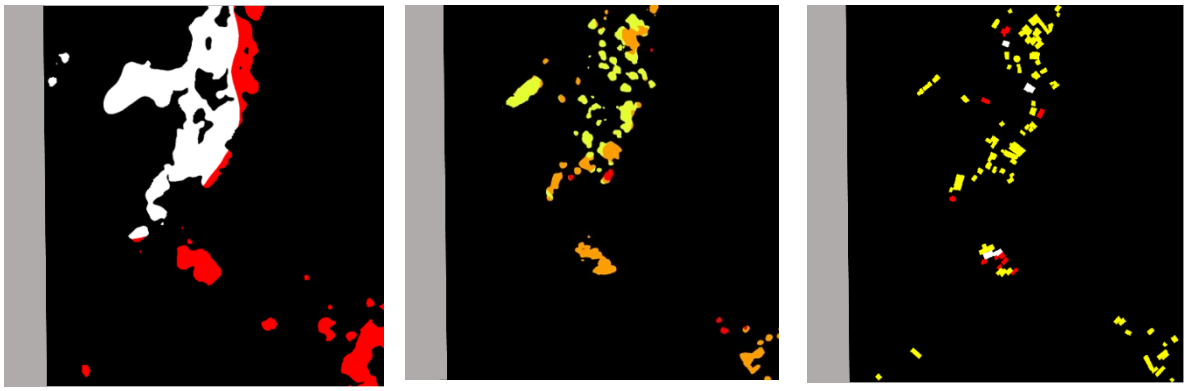


(c) Standard SE



(d) SE-PRE

(e) SE-POST



(f) SE-Identity (g) No-SE (h) Damage ground truth

Figure 4-9. Sample result: "hurricane-matthew_00000010" in xBD dataset

In addition to the above qualitative analyses of image results, the results of the chosen metrics give a quantitative analysis. Combination F1 scores are analysed first. Combination F1 scores of the five structure options are listed in Table 4-1. Since the model was recorded every 50 epochs, ten models were saved for each option. Only the optimal model among these ten in each structure option was used for comparison. Detailly, the model with the highest combination F1 score of each option during the test was chosen for the comparison with other options' models. The epoch number listed in Table 4-1 is the epoch of the optimal model during the training. Scores of all ten recorded models for each option are listed in Table 4-2. Combination F1 scores of the four models with SE are from 12.84% to 62.06%, as shown in Table 4-1. Two options, standard SE and SE-PRE, perform better than the No-SE option (the original residual block without SE). SE-PRE has the highest combination F1 score with 62.06%, which is 49.22% higher than the lowest SE-Identity with 12.84%. Hence, it could be stated that SE-PRE performs best of these five options. SE-PRE offers the best result, which is 5.41% higher than the result of the standard SE model. This may be because SE-PRE gives the CA before the convolution. The scores also

show that the performances of SE-POST and SE-Identity are not as good as expected because their F1 scores are lower than those of No-SE.

Table 4-1. Combination F1 scores of all five options

| Structure option | Standard SE | SE-PRE | SE-Post | SE-Identity | Original No-SE |
|---|---|---|---|---|---|
| **Combination F1 Score** | 56.65% | 62.06% | 16.49% | 12.84% | 53.93% |
| **Epoch** | 500 | 500 | 350 | 400 | 500 |

Table 4-2. Results of evaluating the robustness of models

| Epoch | Transfer Learning | Score | Localisation | | | Damage classification | | |
|---|---|---|---|---|---|---|---|---|
| | | | LF1 | Precision | Recall | DF1 | Precision | Recall |
| 50 | Standard SE | 13.73% | 45.78% | 31.78% | 81.80% | 0.00% | 0.00% | 0.00% |
| | SE-PRE | 12.13% | 40.42% | 26.94% | 80.93% | 0.00% | 0.00% | 0.00% |
| | SE-Post | 7.00E-07 | 0.00% | 0.00% | 0.00% | 1.00E-06 | 1.00E-06 | 1.00E-06 |
| | SE-Identity | 7.00E-07 | 0.00% | 0.00% | 0.00% | 1.00E-06 | 1.00E-06 | 1.00E-06 |
| | No-SE | 0.15% | 0.51% | 31.24% | 0.26% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| 100 | Standard SE | 18.01% | 60.05% | 47.82% | 80.66% | 0.00% | 0.00% | 0.00% |
| | SE-PRE | 22.77% | 75.91% | 76.81% | 75.03% | 0.00% | 0.00% | 0.00% |
| | SE-Post | 13.33% | 44.45% | 34.83% | 61.41% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 7.00E-07 | 0.00% | 0.00% | 0.00% | 1.00E-06 | 1.00E-06 | 1.00E-06 |
| | No-SE | 11.61% | 38.68% | 24.83% | 87.47% | 4.00E-06 | 4.00E-06 | 4.00E-06 |
| 150 | Standard SE | 21.52% | 71.72% | 74.41% | 69.21% | 0.00% | 0.00% | 0.00% |
| | SE-PRE | 24.36% | 75.01% | 78.17% | 72.08% | 2.66% | 19.89% | 1.42% |
| | SE-Post | 12.89% | 42.98% | 30.30% | 73.91% | 1.33E-06 | 1.33E-06 | 1.33E-06 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SE-Identity | 10.30% | 34.35% | 21.46% | 85.94% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | No-SE | 21.54% | 60.82% | 49.24% | 79.53% | 4.70% | 21.87% | 2.64% |
| 200 | Standard SE | 20.42% | 61.18% | 50.03% | 78.75% | 2.95% | 20.38% | 1.59% |
| | SE-PRE | 23.59% | 78.64% | 82.88% | 74.82% | 0.00% | 0.00% | 0.00% |
| | SE-Post | 14.11% | 47.03% | 35.33% | 70.31% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 7.00E-07 | 0.00% | 0.00% | 0.00% | 1.00E-06 | 1.00E-06 | 1.00E-06 |
| | No-SE | 19.61% | 60.98% | 76.70% | 50.61% | 1.88% | 20.75% | 0.99% |
| 250 | Standard SE | 31.87% | 73.03% | 81.99% | 65.83% | 14.23% | 29.95% | 9.33% |
| | SE-PRE | 37.82% | 76.41% | 84.78% | 69.53% | 21.28% | 50.95% | 13.45% |
| | SE-Post | 1.25E-06 | 1.48E-06 | 1.25% | 7.42E-07 | 1.15E-06 | 1.33E-06 | 1.10E-06 |
| | SE-Identity | 7.00E-07 | 0.00% | 0.00% | 0.00% | 1.00E-06 | 1.00E-06 | 1.00E-06 |
| | No-SE | 23.90% | 72.37% | 79.17% | 66.65% | 3.13% | 27.71% | 1.66% |
| 300 | Standard SE | 39.67% | 77.79% | 77.12% | 78.48% | 23.33% | 41.17% | 16.28% |
| | SE-PRE | 58.86% | 79.00% | 82.42% | 75.86% | 50.23% | 57.05% | 44.87% |
| | SE-Post | 15.72% | 52.41% | 39.89% | 76.38% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 10.28% | 34.26% | 27.62% | 45.12% | 2.00E-06 | 2.00E-06 | 2.00E-06 |
| | No-SE | 29.92% | 73.86% | 70.73% | 77.29% | 11.09% | 30.54% | 6.78% |
| 350 | Standard SE | 42.65% | 78.44% | 74.56% | 82.74% | 27.31% | 33.63% | 22.99% |
| | SE-PRE | 56.17% | 80.44% | 78.21% | 82.80% | 45.77% | 51.26% | 41.34% |
| | SE-Post | 16.49% | 54.98% | 46.25% | 67.76% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 11.95% | 39.83% | 32.47% | 51.49% | 2.00E-06 | 2.00E-06 | 2.00E-06 |
| | No-SE | 39.46% | 74.89% | 79.07% | 71.13% | 24.27% | 38.22% | 17.79% |
| 400 | Standard SE | 48.03% | 80.09% | 79.16% | 81.03% | 34.29% | 38.90% | 30.65% |
| | SE-PRE | 59.24% | 77.53% | 82.85% | 72.86% | 51.41% | 54.59% | 48.57% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SE-Post | 9.19% | 30.62% | 64.65% | 20.06% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 12.84% | 42.79% | 32.97% | 60.91% | 2.00E-06 | 2.00E-06 | 2.00E-06 |
| | No-SE | 47.93% | 78.03% | 79.92% | 76.23% | 35.03% | 41.04% | 30.55% |
| 450 | Standard SE | 52.86% | 80.44% | 76.40% | 84.94% | 41.04% | 41.00% | 41.08% |
| | SE-PRE | 61.92% | 81.63% | 81.53% | 81.73% | 53.47% | 57.17% | 50.22% |
| | SE-Post | 10.07% | 33.56% | 64.43% | 22.68% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 11.78% | 39.27% | 29.66% | 58.09% | 2.00E-06 | 2.00E-06 | 2.00E-06 |
| | No-SE | 51.04% | 79.51% | 80.10% | 78.94% | 38.83% | 44.32% | 34.55% |
| 500 | Standard SE | 56.65% | 80.79% | 77.32% | 84.57% | 46.31% | 43.41% | 49.63% |
| | SE-PRE | 62.06% | 82.07% | 82.39% | 81.77% | 53.48% | 56.42% | 50.84% |
| | SE-Post | 15.29% | 50.96% | 53.10% | 48.99% | 1.33E-06 | 1.33E-06 | 1.33E-06 |
| | SE-Identity | 12.46% | 41.54% | 34.89% | 51.32% | 2.00E-06 | 2.00E-06 | 2.00E-06 |
| | No-SE | 53.93% | 79.08% | 81.41% | 76.88% | 43.16% | 50.67% | 37.58% |

Figure 4-10 shows the combination F1 scores of the whole 500 epochs. It can be shown that standard SE and SE-PRE models have better results than No-SE results from the beginning to the end. Therefore, the addition of the SE block with these two options can help to improve the model performance. However, the scores of SE-POST and SE-Identity models are much lower than those of No-SE.

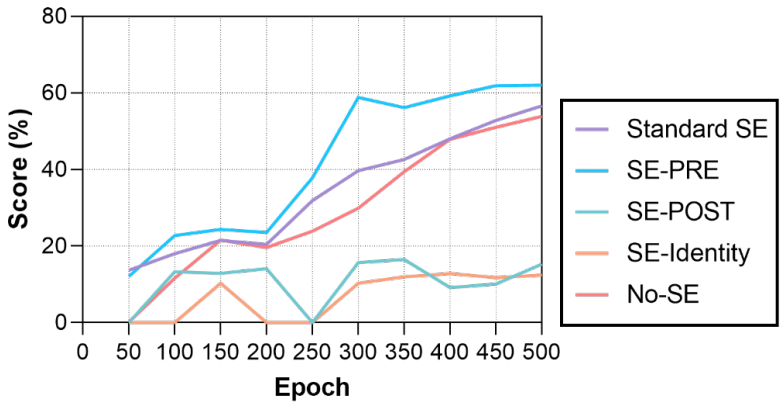
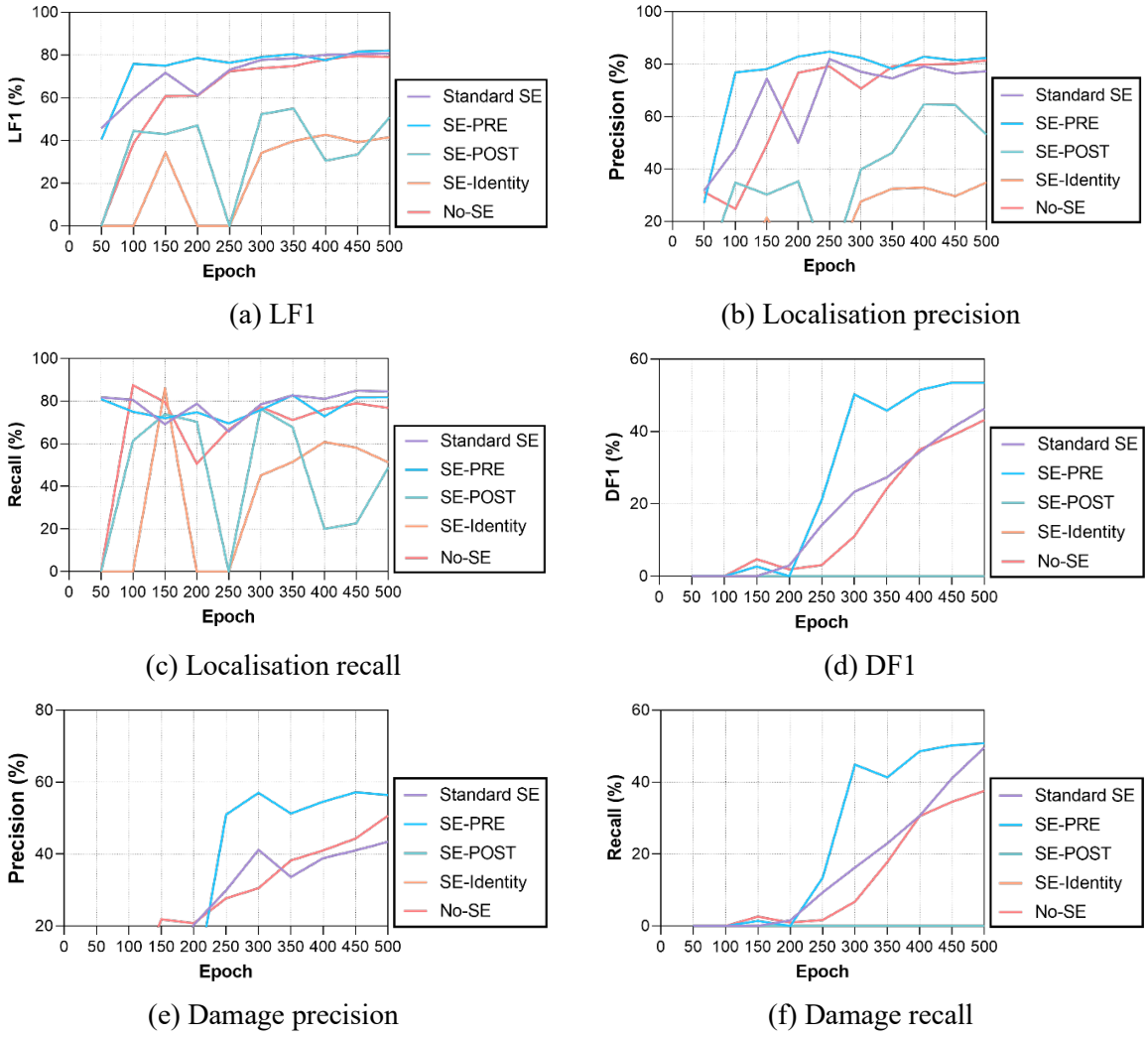



Figure 4-10. Combination F1 scores of all five options at every 50 epochs.

After the analysis of combination F1 scores, the results of other metrics are analysed. Figure 4-11 presents the results of all the other six metrics of these five options to compare their performances. Compared with the results of all these six metrics at the damage classification step, the results at the building localisation step have higher values no matter which model is used. Moreover, F1s, precisions, and recalls of SE-POST and SE-Identity models are nearly zero at the damage classification step. This reflects that these two models cannot be used for damage classification. Performances of these two models are also not good in the building localisation step. The performance of the SE-Identity model is the worst among all models according to these results shown in Figure 4-11 (a) to (f).

Figure 4-11 (a) shows that both standard SE and SE-PRE can achieve stable F1 scores from the 300th epoch. Their performances are better than No-SE for building localisation. The results of SE-POST and SE-Identity are much lower than No-SE. Hence, only standard SE and SE-PRE models can positively influence the model performance of building localisation. Localisation precision results are shown in Figure 4-11 (b). Similar to LF1, the values of localisation precision of SE-Identity are nearly zero from the first to around the

250th epoch. Localisation recalls of standard SE and SE-PRE do not increase much during the training, as shown in Figure 4-11 (c). Figure 4-11 (d) to (f) also show that standard SE and SE-PRE perform much better than SE-POST and SE-Identity.




(a) LF1

(b) Localisation precision

(c) Localisation recall

(d) DF1

(e) Damage precision

(f) Damage recall

Figure 4-11. F1s, precisions and recalls of all five options

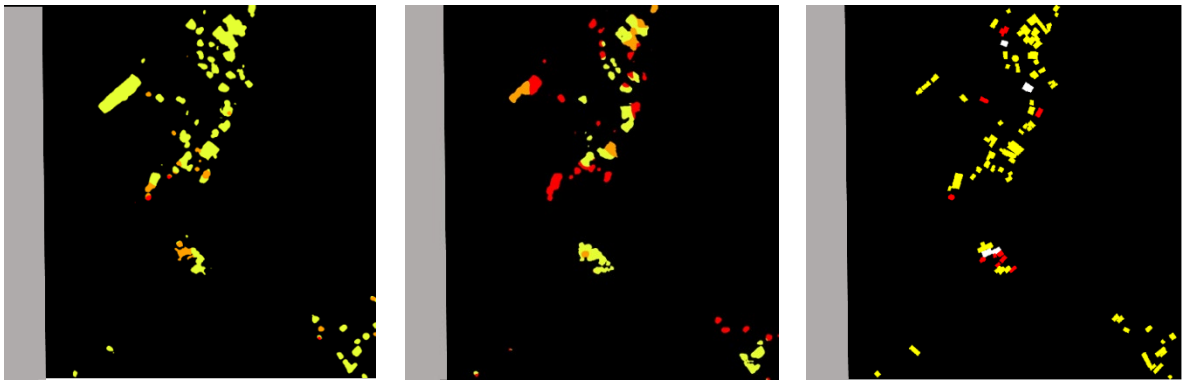
Possible reasons for undesirable results of SE-POST and SE-Identity are discussed in this paragraph. The reason that SE-POST has bad results probably is that the SE block does not have the benefit or even has a bad influence on the model if it is added after ReLU. The

reason for the SE-Identity model might be that the Identity block does not need CA. Therefore, the results show that the added attention block can help improve the accuracy of the model, but it depends on the place of the attention block.

### 4.7.2 Comparison of input size after data augmentation

In this section, the results of the standard SE model applied on two input samples of 512×512, and 256×256 sizes are presented. The method of this comparative experiment is stated in the first paragraph of Section 4.6.2. First, this section analyses the output visualisation image results. Second, the detailed results of seven metrics are analysed to have a comparison of model performances with different input sizes.

Figure 4-12 shows an example of visualisation results compared with the ground truth. The ground truth in Figure 4-12 (c) is the same as that in Figure 4-9 (h). The results show that the standard SE model can detect building footprints and building damage levels using both input samples with different sizes. While the RGB images are useful to show the level of damages in an efficient way, quantitative performance analyses are also carried out to compare the size effect on the performance of the model, and the quantitative analyses are discussed as follows.



(a) 512×512 (b) 256×256 (c) Ground truth

Figure 4-12. Image example with different input sizes.

Table 4-3 shows the combination F1 score results. Similar to Table 4-1, the epoch listed in Table 4-3 is the epoch of the optimal model during the training. The highest F1 score with 512×512 pixels is 70.17% at the 400$^{th}$ epoch. The highest F1 score with 256×256 pixels is 56.65% at the 500$^{th}$ epoch. F1 with 512×512 pixels is higher by 13% than that with 256×256 pixels. Hereafter, this chapter will use "512" and "256" to represent the two models with input sizes of 512×512 pixels and 256×256 pixels for brevity, respectively. The detailed results are shown in Table 4-4.

Table 4-3. Combination F1 score of each input size after data augmentation

| **Input size** | 512×512 | 256×256 |
|---|---|---|
| **Combination F1 Score** | 70.17% | 56.65% |
| **Epoch** | 400 | 500 |

Table 4-4. Standard SE block results with different input sizes

| **Epoch** | **Pixel** | **Score** | **Localisation** | | | **Damage classification** | | |
|---|---|---|---|---|---|---|---|---|
| | | | **LF1** | **Precision** | **Recall** | **DF1** | **Precision** | **Recall** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 50 | 512×512 | 37.77% | 49.94% | 34.70% | 89.07% | 32.56% | 37.53% | 28.75% |
| | 256×256 | 13.73% | 45.78% | 31.78% | 81.80% | 0.00% | 0.00% | 0.00% |
| 100 | 512×512 | 41.57% | 75.23% | 70.95% | 80.07% | 27.15% | 46.31% | 19.20% |
| | 256×256 | 18.01% | 60.05% | 47.82% | 80.66% | 0.00% | 0.00% | 0.00% |
| 150 | 512×512 | 45.67% | 80.15% | 76.65% | 84.00% | 30.90% | 39.00% | 25.58% |
| | 256×256 | 21.52% | 71.72% | 74.41% | 69.21% | 0.00% | 0.00% | 0.00% |
| 200 | 512×512 | 49.77% | 81.16% | 80.81% | 81.50% | 36.31% | 30.19% | 45.56% |
| | 256×256 | 20.42% | 61.18% | 50.03% | 78.75% | 2.95% | 20.38% | 1.59% |
| 250 | 512×512 | 64.73% | 82.77% | 83.73% | 81.84% | 56.99% | 61.21% | 53.32% |
| | 256×256 | 31.87% | 73.03% | 81.99% | 65.83% | 14.23% | 29.95% | 9.33% |
| 300 | 512×512 | 56.61% | 82.79% | 83.29% | 82.29% | 45.39% | 49.23% | 42.10% |
| | 256×256 | 39.67% | 77.79% | 77.12% | 78.48% | 23.33% | 41.17% | 16.28% |
| 350 | 512×512 | 68.28% | 84.08% | 81.76% | 86.53% | 61.51% | 62.06% | 60.98% |
| | 256×256 | 42.65% | 78.44% | 74.56% | 82.74% | 27.31% | 33.63% | 22.99% |
| 400 | 512×512 | 65.45% | 84.57% | 82.26% | 87.02% | 57.25% | 58.43% | 56.12% |
| | 256×256 | 48.03% | 80.09% | 79.16% | 81.03% | 34.29% | 38.90% | 30.65% |
| 450 | 512×512 | 70.17% | 84.90% | 82.89% | 87.01% | 63.86% | 65.27% | 62.50% |
| | 256×256 | 52.86% | 80.44% | 76.40% | 84.94% | 41.04% | 41.00% | 41.08% |
| 500 | 512×512 | 67.87% | 85.10% | 83.96% | 86.28% | 60.48% | 60.61% | 60.36% |
| | 256×256 | 56.65% | 80.79% | 77.32% | 84.57% | 46.31% | 43.41% | 49.63% |

Figure 4-13 shows the trend of combination F1s, DF1s, and LF1s. It is obvious that the "512" results are higher than the "256" results in all epochs with all three types of F1s. All LF1 lines are placed higher than DF lines, so it can be stated that this model shows better performance for building localisation than building damage classification disregarding the size of images.

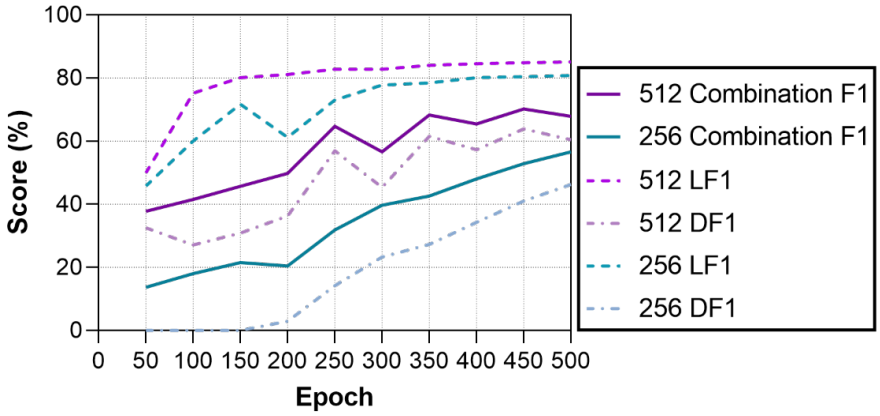



Figure 4-13. Combination F1 scores, LF1s, and DF1s with two different input sizes

Figure 4-14 shows the results of all the other metrics, including F1s, precisions, and recalls in both localisation and damage steps. Figure 4-14 (a) to (c) are localisation results, and (d) to (f) are results of building damage. All these six results show the "512" model has better performance than the "256" model, because all "512" results are higher than "256" results at every recorded epoch. It should be noted that all six "512" lines turn flat, or the fluctuation has decreased since the 350$^{th}$ epoch. The "256" lines do not have this phenomenon. The following paragraphs will analyse localisation results first and then results on the damage.

All "256" lines have more fluctuated than "512" lines in the localisation step as shown in Figure 4-14 (a). LF1s are analysed first. The highest "512" LF1 is 85.10% at the 500$^{th}$ epoch. LF1s from 350$^{th}$ to 500$^{th}$ are similar, which are 84.08%, 84.57%, 84.90% and 85.10%. The highest "256" LF1 is 80.79% at the 500$^{th}$ epoch. Similar to "512" LF1, the change in the results from 350$^{th}$ to 500$^{th}$ is less than that in previous epochs (78.44%, 80.09%, 80.44%, 80.79%). There is no decrease of LF1 for both "512" and "256" results.

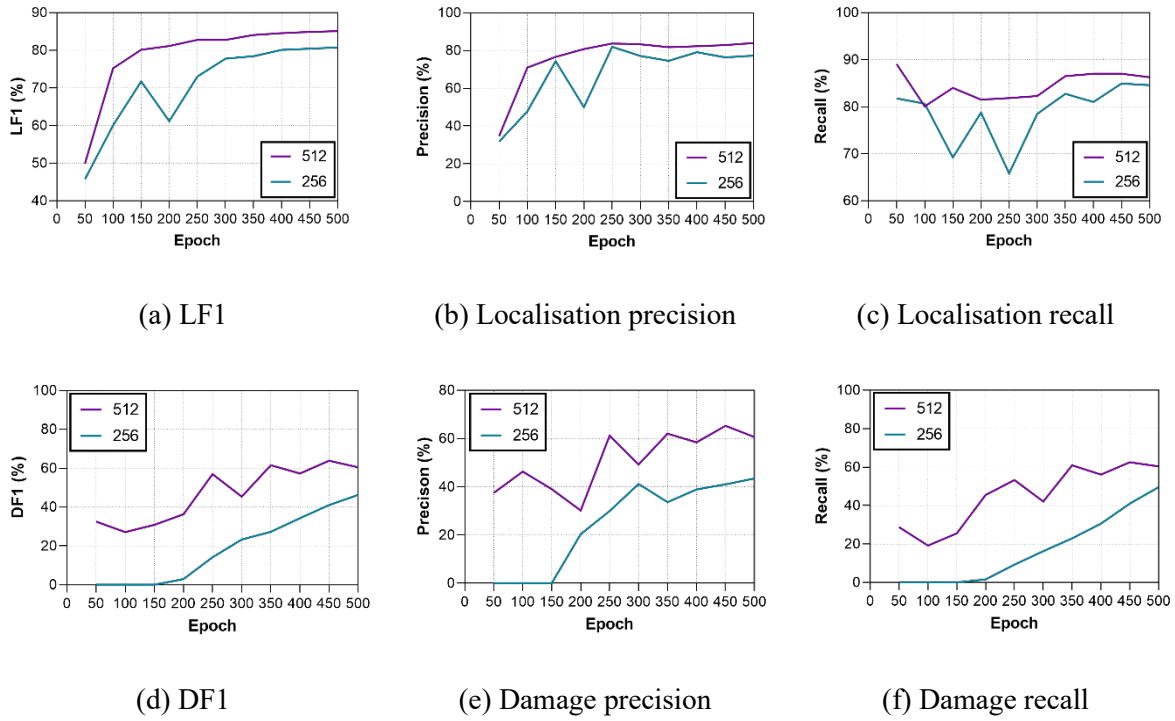
The second metric is localisation precision. As shown in Figure 4-14 (b), the line charts of both "512" and "256" localisation precision results have fluctuated. The highest "512" result is 83.96% at the 500$^{th}$ epoch, and the highest "256" result is 81.99% at the 250$^{th}$ epoch.

The result of the last metric for localisation is shown in Figure 4-14 (c). The two localisation recall lines fluctuated more than LF1 and localisation precision lines. The highest "512" recall is 87.02% at the 400$^{th}$ epoch, and the highest "256" recall is 84.94% at the 450$^{th}$ epoch.

In the second step, namely, the damage classification step, all the "256" lines of the three metrics are less fluctuated than the "512" lines. The trend of the DF1 line chart is similar to that of the combination F1 line chart in both "512" and "256" results. This is because the DF1 has a higher proportion than LF1 in Equation 4-10, which are 70% and 30%, respectively. The highest "512" DF1 is 63.86% at the 450$^{th}$ epoch, and the highest "256" DF1 is 46.31% at the 500$^{th}$ epoch. There is no decrease in the "256" results, while "512" results show the model performance is increasing in a fluctuation.

As for the damage precision results, "256" line increases more slowly than the "512" line, which is 0 from 0 to 150 epochs. The trend of "512" damage precision is similar to the trend of "512" DF1, especially from the 200$^{th}$ epoch. The highest "512" damage precision is 65.27% at the 450$^{th}$ epoch, and the highest "256" damage precision is 43.41% at the 500$^{th}$ epoch.

The trends of two damage recall lines in Figure 4-14 (f) are very similar to the trends of two DF1 lines in Figure 4-14 (d). The highest "512" recall is 62.5% at the 450$^{th}$ epoch, and the highest "256" recall is 49.63% at the 500$^{th}$ epoch.




(a) LF1 (b) Localisation precision (c) Localisation recall

(d) DF1 (e) Damage precision (f) Damage recall

Figure 4-14. F1s, precisions and recalls with different input sizes

The model of 512×512 pixels at the 450$^{th}$ epoch could be said to be the best model among all 512×512 pixels, because its value is the highest for five out of the seven results. The best localisation model of 512×512 pixels could be said between 450 to 500 epochs since their values are quite close to each other. The optimal damage classification model of 512×512 pixels is the model training with 450 epochs among all recorded models because its performances are the best with all damage classification metrics, including DF1, damage precision, and damage recall. As for the models of 256×256 pixels, the model at the 500$^{th}$

epoch is the best both in the localisation and damage classification since it performs best with all the metrics. Moreover, the results show that localisation results are higher than damage classification results with all metrics.

### 4.7.3 Comparison of transfer learning and non-transfer learning

This section compares the results of transfer learning and non-transfer learning models with 512×512 pixels. The method of this experiment is introduced in the second paragraph of Section 4.6.2. Combination F1 score results are discussed first, as shown in Table 4-5. The highest F1 score with transfer learning is 69.85% at the 400$^{th}$ epoch, and that with non-transfer learning is 70.17% at the 450$^{th}$ epoch. The detailed information is listed in Table 4-6. Figure 4-15 (a) shows the combination F1 scores of these models. The results reflect that the transfer learning model does not have obvious advantages over the non-transfer learning model. In Figure 4-15 (b), "Non" represents the trained model without transfer learning. "TransL" means that the model is trained with transfer learning. Figure 4-15 (b) shows that these two models have similar performance for building localisation. The model with transfer learning has better performance at the initial stage, but their performances turn similar gradually. That is, the highest score without transfer learning is even higher than that with transfer learning.

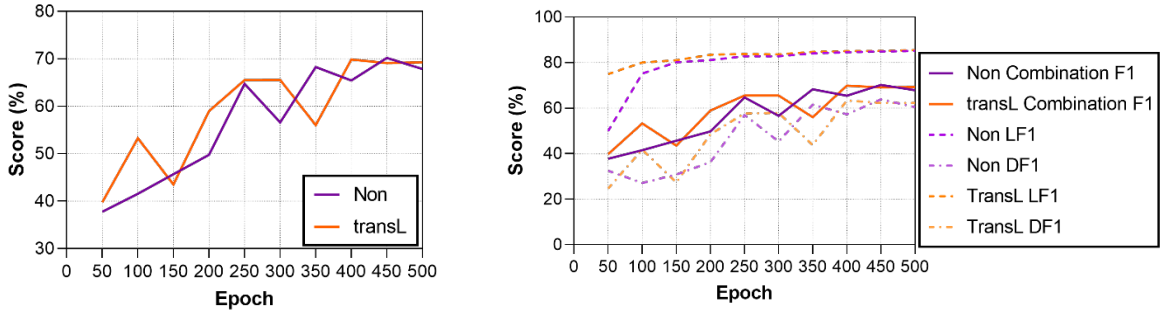
Table 4-5. Combination F1 scores with and without transfer learning

| Model | Transfer learning | Non-transfer learning |
|---|---|---|
| **Combination F1 Score** | 69.85% | 70.17% |

| **Epoch** | 400 | 450 |
|---|---|---|

Table 4-6. Standard SE block results with transfer and non-transfer learning

| **Epoch** | **Transfer learning** | **Score** | **Localisation** | | | **Damage classification** | | |
|---|---|---|---|---|---|---|---|---|
| | | | **LF1** | **Precision** | **Recall** | **DF1** | **Precision** | **Recall** |
| 50 | Yes | 39.78% | 74.97% | 85.59% | 66.70% | 24.70% | 34.59% | 19.21% |
| | No | 37.77% | 49.94% | 34.70% | 89.07% | 32.56% | 37.53% | 28.75% |
| 100 | Yes | 53.27% | 80.01% | 84.53% | 75.95% | 41.81% | 43.68% | 40.09% |
| | No | 41.57% | 75.23% | 70.95% | 80.07% | 27.15% | 46.31% | 19.20% |
| 150 | Yes | 43.50% | 81.06% | 84.78% | 77.65% | 27.40% | 43.57% | 19.98% |
| | No | 45.67% | 80.15% | 76.65% | 84.00% | 30.90% | 39.00% | 25.58% |
| 200 | Yes | 58.99% | 83.46% | 81.90% | 85.09% | 48.50% | 53.36% | 44.45% |
| | No | 49.77% | 81.16% | 80.81% | 81.50% | 36.31% | 30.19% | 45.56% |
| 250 | Yes | 65.52% | 83.81% | 84.51% | 83.12% | 57.68% | 60.89% | 54.79% |
| | No | 64.73% | 82.77% | 83.73% | 81.84% | 56.99% | 61.21% | 53.32% |
| 300 | Yes | 65.58% | 83.66% | 89.02% | 78.90% | 57.84% | 58.62% | 57.08% |
| | No | 56.61% | 82.79% | 83.29% | 82.29% | 45.39% | 49.23% | 42.10% |
| 350 | Yes | 56.04% | 84.73% | 86.62% | 82.93% | 43.75% | 46.23% | 41.53% |
| | No | 68.28% | 84.08% | 81.76% | 86.53% | 61.51% | 62.06% | 60.98% |
| 400 | Yes | 69.85% | 85.01% | 86.65% | 83.43% | 63.36% | 67.96% | 59.34% |
| | No | 65.45% | 84.57% | 82.26% | 87.02% | 57.25% | 58.43% | 56.12% |
| 450 | Yes | 69.11% | 85.10% | 86.71% | 83.54% | 62.25% | 64.49% | 60.16% |
| | No | 70.17% | 84.90% | 82.89% | 87.01% | 63.86% | 65.27% | 62.50% |
| 500 | Yes | 69.26% | 85.47% | 86.90% | 84.10% | 62.31% | 64.34% | 60.40% |
| | No | 67.87% | 85.10% | 83.96% | 86.28% | 60.48% | 60.61% | 60.36% |




(a) Combination F1 Score (b) Combination F1 scores, LF1s, and DF1s

Figure 4-15. F1 scores of transfer and non-transfer learning models

Figure 4-16 shows the detailed results at both localisation and classification steps. Figure 4-16 (a) to (c) display building localisation results. Figure 4-16 (a) shows the LF1s of these two methods. If the number of training epochs is less than 300, the benefit of the pre-trained model is obvious, but after 300 epochs, the LF1s difference is less than 1% in each recorded epoch. The highest LF1 with transfer learning is 85.47%, which is 0.37% higher than the highest LF1 without transfer learning (85.10%) as shown in Table 4-6. Figure 4-16 (b) and (c) show the localisation precisions and recalls, respectively. The trends of these results are similar to the trend of Figure 4-16 (a), whose difference is very large at the beginning, but the difference decreases quickly. All the results of these three metrics show that 350 training epochs may be enough no matter whether the model is with transfer learning or not. This is because the results improve slowly after 350 epochs.

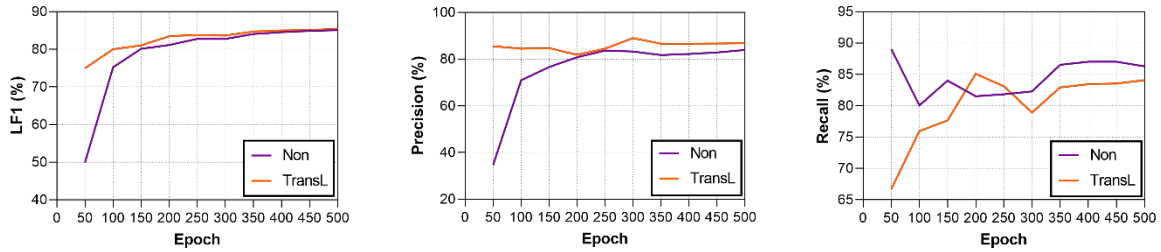
Figure 4-16 (d) to (f) show damage classification results. The trends of these results are different from the trends in the building localisation step. The initial values of the three metrics are not much different, and the results of the non-transfer learning model are even

better than those of the transfer learning model at the 50th epoch. The highest DF1, precision, and recall with transfer learning are 63.36% at the 400th epoch, 67.96% at the 400th epoch, and 60.40% at the 450th epoch. The highest DF1, precision, and recall without transfer learning are 63.86%, 65.27%, and 62.50%, respectively, all at the 450th epoch.

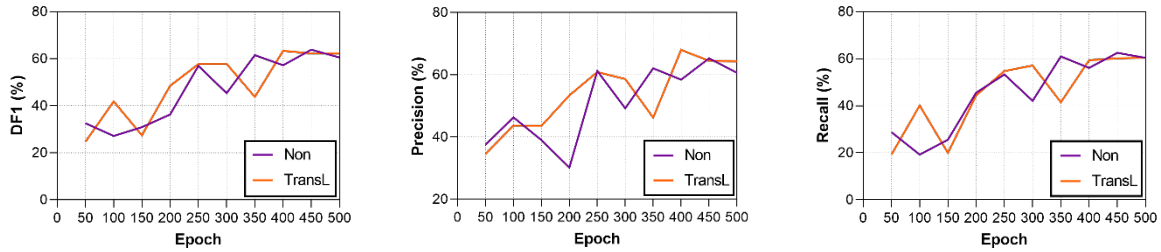
Based on the interpretation of all the results of these seven metrics, the transfer learning model with the pre-trained model is not much better than the non-transfer learning model. One benefit is that it can achieve higher precision than the non-transfer learning model at the beginning for only building localisation, which is faster, but the non-transfer learning model can also achieve this high precision after around 100 epochs. The transfer learning model does not show many advantages in damage classification.

The possible reason for that is listed as follows. First, the chosen pre-trained model is for image classification. That is why the pre-trained model has better performance of building localisation than damage classification. This pre-trained model may not be the best choice for damage classification. Second, the pre-trained model is trained with the ImageNet dataset. This dataset does not contain enough damaged buildings in the images. Hence, the pre-trained model does not show enough good performance as anticipated.

Although several papers adopt this pre-trained model based on experience (Koo et al., 2020), they do not compare its results with the non-transfer learning model. Before the testing, the hypothesis in this study was that its performance was better, but the test shows that this is not the case.




(a) Localisation F1 (LF1) (b) Localisation precision (c) Localisation recall




(d) Damage F1 (DF1) (e) Damage precision (f) Damage recall

Figure 4-16. F1s, precisions and recalls of transfer and non-transfer learning models

### 4.7.4 Comparison of Sigmoid and Hard-Sigmoid activation functions

As the fourth comparative experiment of the testing stage in the workflow, this section compares the performance of models with different activation functions placed in the SE block, as mentioned in the third paragraph of Section 4.6.2. The results of the seven metrics are analysed one by one.

The combination F1 score results are compared first with the optimal model of each situation. As shown in Table 4-7, the optimal model with the Sigmoid function is recorded at the 450$^{th}$ epoch with the highest F1 score of 70.17%. The F1 score of the optimal model with Hard-Sigmoid is 69.66%, which is at the 500$^{th}$ epoch. Table 4-8 shows the detailed results with different functions. Hence, the performances of models with Sigmoid and

Hard-Sigmoid functions are similar, as shown in Figure 4-17 (a). The performance of the Sigmoid model with Sigmoid is 0.51% higher than the Hard-Sigmoid model. Figure 4-17 (b) shows that building localisation results are much better than damage classification results no matter which activation function is used.

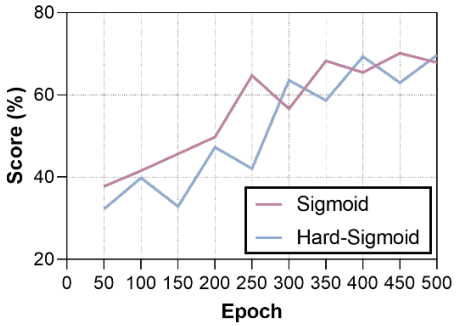
Table 4-7. Combination F1 scores with Sigmoid and Hard-Sigmoid activation functions

| **Activation function** | **Sigmoid** | **Hard-Sigmoid** |
|---|---|---|
| **F1 Score** | 70.17% | 69.66% |
| **Epoch** | 450 | 500 |

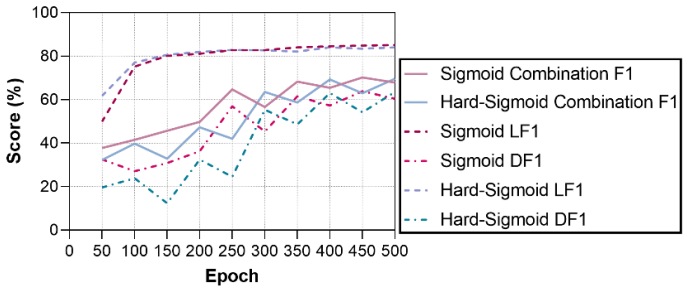
Table 4-8. Standard SE block results with different activation functions

| **Epoch** | **Activation function** | **Score** | **Localisation** | | | **Damage classification** | | |
|---|---|---|---|---|---|---|---|---|
| | | | **LF1** | **Precision** | **Recall** | **DF1** | **Precision** | **Recall** |
| 50 | Sigmoid | 37.77% | 49.94% | 34.70% | 89.07% | 32.56% | 37.53% | 28.75% |
| | Hard-sigmoid | 32.27% | 61.81% | 65.85% | 58.23% | 19.62% | 36.20% | 13.46% |
| 100 | Sigmoid | 41.57% | 75.23% | 70.95% | 80.07% | 27.15% | 46.31% | 19.20% |
| | Hard-sigmoid | 39.81% | 77.05% | 84.37% | 70.90% | 23.85% | 45.14% | 16.20% |
| 150 | Sigmoid | 45.67% | 80.15% | 76.65% | 84.00% | 30.90% | 39.00% | 25.58% |
| | Hard-sigmoid | 32.87% | 80.61% | 81.04% | 80.19% | 12.41% | 25.85% | 8.17% |
| 200 | Sigmoid | 49.77% | 81.16% | 80.81% | 81.50% | 36.31% | 30.19% | 45.56% |
| | Hard-sigmoid | 47.30% | 81.95% | 83.22% | 80.73% | 32.45% | 44.43% | 25.56% |
| 250 | Sigmoid | 64.73% | 82.77% | 83.73% | 81.84% | 56.99% | 61.21% | 53.32% |
| | Hard-sigmoid | 42.07% | 82.87% | 81.16% | 84.66% | 24.58% | 52.99% | 16.00% |
| 300 | Sigmoid | 56.61% | 82.79% | 83.29% | 82.29% | 45.39% | 49.23% | 42.10% |
| | Hard-sigmoid | 63.56% | 82.72% | 81.90% | 83.57% | 55.35% | 57.85% | 53.06% |
| 350 | Sigmoid | 68.28% | 84.08% | 81.76% | 86.53% | 61.51% | 62.06% | 60.98% |
| | Hard-sigmoid | 58.67% | 82.15% | 82.29% | 82.01% | 48.61% | 51.09% | 46.35% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 400 | Sigmoid | 65.45% | 84.57% | 82.26% | 87.02% | 57.25% | 58.43% | 56.12% |
| | Hard-sigmoid | 69.33% | 84.05% | 81.86% | 86.36% | 63.03% | 65.52% | 60.72% |
| 450 | Sigmoid | 70.17% | 84.90% | 82.89% | 87.01% | 63.86% | 65.27% | 62.50% |
| | Hard-sigmoid | 62.97% | 83.49% | 83.69% | 83.28% | 54.18% | 59.24% | 49.91% |
| 500 | Sigmoid | 67.87% | 85.10% | 83.96% | 86.28% | 60.48% | 60.61% | 60.36% |
| | Hard-sigmoid | 69.66% | 84.04% | 83.61% | 84.49% | 63.50% | 68.87% | 58.90% |




(a) Combination F1 Score




(b) Combination F1 scores, LF1s, and DF1s

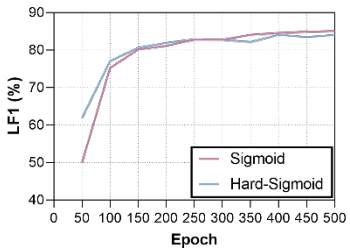
Figure 4-17. F1 scores of models with different SE block activation functions

Figure 4-18 shows all results at the localisation and classification steps. The highest LF1 with Sigmoid function is 85.10% at the 500$^{th}$ epoch, and the highest LF1 with Hard-Sigmoid function is 84.05% at the 400$^{th}$ epoch, as shown in Table 4-8. Figure 4-18 (a) shows that the Sigmoid model performs better than Hard-Sigmoid model from the 300$^{th}$ epoch. The results of another metric, localisation precision, are shown in Figure 4-18 (b). Its highest is 83.96% at the 500$^{th}$ epoch with Sigmoid and 83.69% at the 450$^{th}$ epoch with Hard-Sigmoid. Figure 4-18 (a) and (b) show that the trends of the two lines are similar. Despite the fact that the performance of the Hard-Sigmoid model is better at the beginning epochs, the performance of the Sigmoid model increases quicker and better than that of the

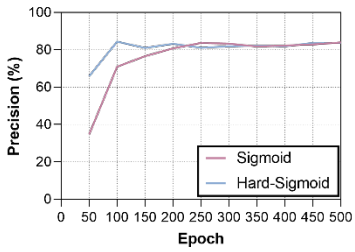
Hard-Sigmoid model at last. Figure 4-18 (d) shows the localisation recall results. The highest recalls are 87.02% with Sigmoid and 86.36% with Hard-Sigmoid, both at the 400$^{th}$ epoch. The recall of the Sigmoid model is 0.66% higher than that of the Hard-Sigmoid model.

Figure 4-18 (d) to (f) show the results at the damage classification step. The highest DF1 is 63.86% at the 450$^{th}$ epoch with Sigmoid, which is 0.36% higher than that with Hard-Sigmoid (63.50% at the 500$^{th}$ epoch). The highest damage precisions are 65.27% at the 450$^{th}$ epoch with Sigmoid and 68.87% at the 500$^{th}$ epoch with Hard-Sigmoid shown as Figure 4-18 (e). The highest damage recalls are 62.50% at the 450$^{th}$ epoch with Sigmoid and 60.72% at the 400$^{th}$ epoch with Hard-Sigmoid. Therefore, the damage classification results show that the performances of these two models are similar.

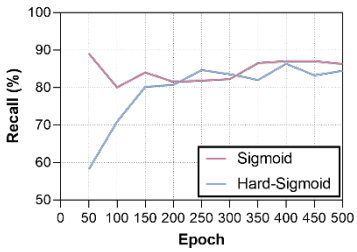
Based on the analysis, the Hard-Sigmoid model only has better results based on the damage recall metric, and the other five results show that the performance of the Sigmoid model is better. Therefore, it can be concluded that the model with Sigmoid performs slightly better than the model with Hard-Sigmoid.



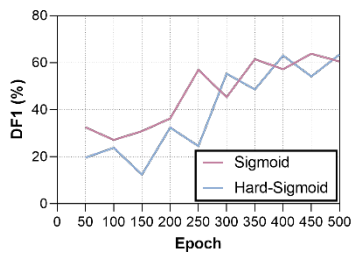

(a) Localisation F1 (LF1)



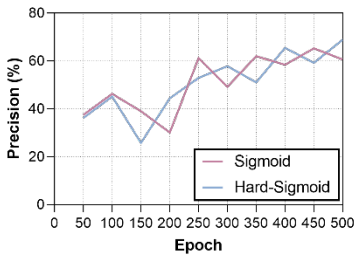

(b) Localisation precision



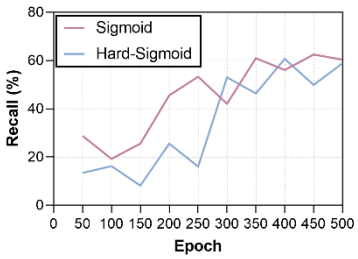

(c) Localisation recall




(d) Damage F1 (DF1)




(e) Damage precision




(f) Damage recall

Figure 4-18. F1s, precisions and recalls with different SE block activation functions

## 4.8 Discussion of 2D building damage classification results

Four types of experiments have been conducted for different comparison tasks. The first experiment tested the performance of five model options, including four modified models (SE, SE-PRE, SE-POST, SE-Identity) and one original model. In the second experiment, two models were tested for the performance comparison of different training input sizes. The next experiment was implemented for the comparison of models with and without transfer learning. The last experiment was implemented for comparison of different activation functions in the SE block. The results of these four experiments are discussed as follows.

First, the outcome of the experiment with five options (refer to Section 4.7.1) shows that the SE-PRE model has the best performance across all metrics except localisation recall among all these five models. One possible reason is that SE-PRE gives CA at the very beginning. Hence, the model could judge which feature needs more attention during the training and which feature is less important. The worst two models are SE-Post and SE-

Identity because they perform worst among all metrics, and their performances are much worse than the performances of the other two modified models and the original No-SE model. Therefore, although several papers state that a model would perform better after adding CA (Li et al., 2020b, Li et al., 2020c), the finding from this study is that different added places of SE in the original basic block have different effects on the model performances of building damage classification, sometimes improves the accuracy and sometimes decrease the accuracy.

In the second experiment, the results show that the standard SE model (Figure 4-8 (a)) with the input resolution of 512×512 by random cropping during the training can give better performance than that with 256×256. This might be because a 512×512 image contains more features than a 256×256 image. Hence, more information can be retained during the training with 512×512 size. Similarly, higher resolution images could have better results for models because they contain more features and information than lower resolution images. However, the larger size 512×512 spends more time and memory for training, so which size to be adopted depends on the time and hardware constraints of a task.

In the third comparison, the results of the models with and without transfer learning are compared. The model with transfer learning of pre-trained ImageNet weights does not perform better than model training from scratch. This may be because the pre-trained weights are not very suitable for this study since ImageNet does not contain damaged building labels. Moreover, the pre-trained weight is only used for image classification, whose task is easier for this study, including both image segmentation and classification.

Fourth, the difference in the performance of Sigmoid and Hard-Sigmoid is not obvious. Although the training time of the model with Hard-Sigmoid should be shorter than that with Sigmoid in theory, the time used in this experiment is similar.

The model performs better for building localisation than damage classification. The technical reason might be that the building localisation step is easier than damage classification. It only needs to segment two classes, buildings, and non-buildings, and the difference of features between these two classes are obvious such as outlines or colours. However, the feature difference of each damage level is small and complex between minor, major, and total damage. Moreover, the roof shapes of some totally damaged buildings did not change or break after disasters, especially after earthquakes. Hence, it increases the difficulty for models to assess the damage levels of buildings.

While the main target of this chapter is building damage classification, the observation shows that the SE-PRE also offers a high accuracy (82.07% of LF1 score) for building localisation or building footprint segmentation. Offering acceptable outcomes for both building damage classification and localisation will increase the practical implication of the SE-PRE model to guide disaster managers and practitioners for emergency actions by finding the most damaged buildings and their locations simultaneously.

This chapter contributes to the body of literature by addressing the challenging task of building damage classification utilising the novel approach of SE-PRE. Compared to the previous work, the present chapter shows an improvement in the classification tasks by classifying into not only collapse or not, but also four damage levels and providing a new

modified model. Several current papers just apply models to the building damage classification without modification of the structures of models. For example, Yang et al. (2021) categorised the damages by keeping the original structures of CNN.

## 4.9 Conclusion

The aim of this chapter was to provide a quick DL method for post-disaster multi-level BDLC with optical satellite images. This has been achieved by improving the performance of the DL method with SE added dual HRNet model and applying it to building damage classification with a total of 8,664 images from xBD and our newly created datasets. Four novel options of models with adding SE CA to different places of basic residual blocks in HRNet have been used to compare the original model without SE. These four are called standard SE, SE-PRE, SE-Post and SE-Identity. Four types of experiments applying seven metrics (refer to Section 4.5) were implemented on different model modifications to measure the effect on the model performance.

The results show that the DL model proposed in this chapter can classify building damage levels, which are hard or impossible to be achieved by human eyes. Four options of SE block added models are tested, and results show that SE-PRE has the best performance. A larger input size can have better results but use much more computing time. Transfer learning with a pre-trained ImageNet dataset does not have advantages because the dataset does not contain several damaged building images. The block with Sigmoid function has slightly better performance than that with Hard-Sigmoid. However, this chapter only

compared the proposed models with the backbone. The comparisons with relevant state-of-the-art models are suggested to be tested in further research.

This chapter created a building damage level dataset based on the official damage assessment document, but some limitations exist in it. One limitation is that the time duration between pre- and post-event images is large. Some post-event images were taken several months after the event happened. The imaging angle and the brightness are different between these two images. This limitation would not affect the comparison outcome of the experiment but cannot be avoided because this is very tough for remote sensing technology to take two images with the exact same air, sunlight, and imaging angle conditions on two different days. In the future, scholars can replicate the suggested method on new datasets with the development of remote sensing to avoid this limitation. Second, some pre- and post-event images contain a large area of clouds covering buildings, so the model will wrongly learn the white cloud area as buildings based on the ground truth map. Hence, future work would focus on reducing the limitations of data.

# Chapter 5

# Feature selection on the accuracy of deep learning-based large-scale Lidar semantic segmentation[2]

## 5.1 Background and scope

This chapter is the prepare work of Chapter 6 to focus on predisaster building segmentation. Lidar is a valuable technique for gathering and retaining 3D information. With the emergence of helicopters and drones capable of capturing 3D data through Lidar, the availability of extensive outdoor airborne Lidar data has rapidly expanded. Lidar can offer a broader range of land cover information compared to traditional 2D optical satellite image segmentation methods. Consequently, utilising urban land cover data from Lidar is recommended for applications such as urban planning, disaster data collection, and various earth observation purposes.

---

[2] The content presented in this chapter is partially adopted from the following work which has been accepted for publication: "Liu C, Zhang Q, Shirowzhan S, Bai T, Sheng Z, Wu Y, Kuang J, Ge L*, 2023. The Influence of Changing Features of Point Clouds on the Accuracy of Deep Learning-based Large-scale Outdoor Lidar Semantic Segmentation, In *2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Pasadena, U.S., pp. 4443-4446". It has been acknowledged and detailed in the "Inclusion of Publications Statement" for this thesis.

The studies in 3D semantic segmentation have experienced rapid development with the increasing accessibility of Lidar data in the last ten years. DL always provides a more efficient and accurate method than conventional computer vision methods. For instance, RandLa-Net is an effective and lightweight network for urban-scale Lidar data (Hu et al., 2020). However, 3D semantic segmentation is still a challenging task in the computer vision field, and the accuracy of these DL methods still needs improvement. Several feature setting selections of DL algorithms has not been tested to check they can increase semantic segmentation accuracy. Moreover, most DL methods were proposed without considering the case studies of the areas that potentially happen earthquakes. To resolve those problems, this chapter aims to test the influence of feature selections of DL networks on the accuracy of DL-based pre-earthquake large-scale Lidar semantic segmentation, which is Objective 2.

Two feature selections warrant exploration concerning their possible impact on the accuracy of DL networks, including surface normal information and the down-sampling layer configuration. Surface normal information is a crucial aspect widely utilised in the 3D visualisation field to enhance the fidelity of objects depicted in computers. A surface normal refers to a vector perpendicular to the surface of an object or a specific point on that surface. It furnishes insights into the orientation and direction of the surface at each point. Alongside surface normals, the design of random sampling layers emerges as an intriguing subject of exploration. Distinct down-sampling layer configurations can yield varying features assimilated by trained DL networks. Consequently, exploring their potential for augmenting accuracy holds significance.

In pursuit of the aim and objectives in this chapter, as mentioned in Chapter 3, a series of experiments have been conducted to assess the impact of incorporating surface normal information and altering the number of random sample layers on segmentation accuracy. It should be mentioned that this chapter serves as the preparatory groundwork for Chapter 6. Therefore, the design of the research methods in this chapter considers not only its own objectives but also those outlined in Chapter 6. The following sections gradually introduce the data collection (Section 5.2), designed methods (Section 5.3), results and discussions (Section 5.4), and conclusion (Section 5.5) of the work presented in this chapter.

## 5.2 Data collection and pre-processing

Lidar can be categorised into different types according to data collection methods, such as terrestrial, airborne, and spaceborne Lidar. This chapter chose airborne Lidar data because the focus of this whole study is to scan large-scale outdoor urban land cover objects, as mentioned in Section 1.4.

OpenTopography provides several freely available open-source pre- or post-disaster Lidar point cloud datasets. These datasets include several types of disasters, such as earthquakes, floods, bushfires, hurricanes, and cyclones (Asia Air Survey Co., 2018, Opentopography, 2022). Therefore, this chapter chose the datasets published by OpenTopography.

As this chapter serves as a precursor to the next chapter, and the subsequent chapter requires colour information, the inclusion of colour information in this chapter is crucial. However, many of the Lidar datasets provided by OpenTopography lack colour information. Consequently, colour information must be manually integrated into the data as part of the

data preparation process. This results in an in-house labelled dataset, which is Objective 4. The following paragraphs introduce how the colour information is added.

The original Lidar point cloud dataset was collected from Kapiti Coast, New Zealand by OpenTopography, which includes labelled land cover objects. The labels in the Lidar data utilised in this chapter included the ground, low vegetation, medium vegetation, high vegetation, and buildings. This location was chosen because New Zealand happened a very serious earthquake in Christchurch in 2011. Moreover, this location contains several possible natural disasters besides earthquakes (Kapiti Coast District Council, 2022). The collection dates were from 13/03/2021 to 15/03/2021. The original dataset covers the entire coast with an area of 292.63 $km^2$. 26 parts were clipped from the dataset because they contain both buildings and vegetation, as shown in Figure 5-1 (Opentopography, 2022). The training, validation, and testing stages consist of 21, 3, and 2 point clouds, respectively.

The colour information was introduced to the dataset using S2 images. Given that the original Lidar data lacked colour information, an S2 image with a 10m cell size, captured on 24/07/2021, was fused with the Lidar data through the utilisation of FME software (Safe Software, 2022). The incorporation of colours was deemed necessary as it often leads to improved accuracy in predictions when employed in DL networks, aligning with the preparatory work for Chapter 6. The choice of the S2 data acquisition date was made to closely match the date of the Lidar data collection. Moreover, a comprehensive examination of the S2 image was carried out to ensure minimal presence of clouds or smoke on that specific date. Since the coordinate systems of S2 imagery and Lidar data are different, the coordinate system of S2 imagery, UTM84-60S, was reproject to that of Lidar

point clouds. The detailed information on Lidar and satellite data are listed in Table 5-1. After the data pre-processing step, a colourised Lidar point cloud dataset with five labelled classes was created.

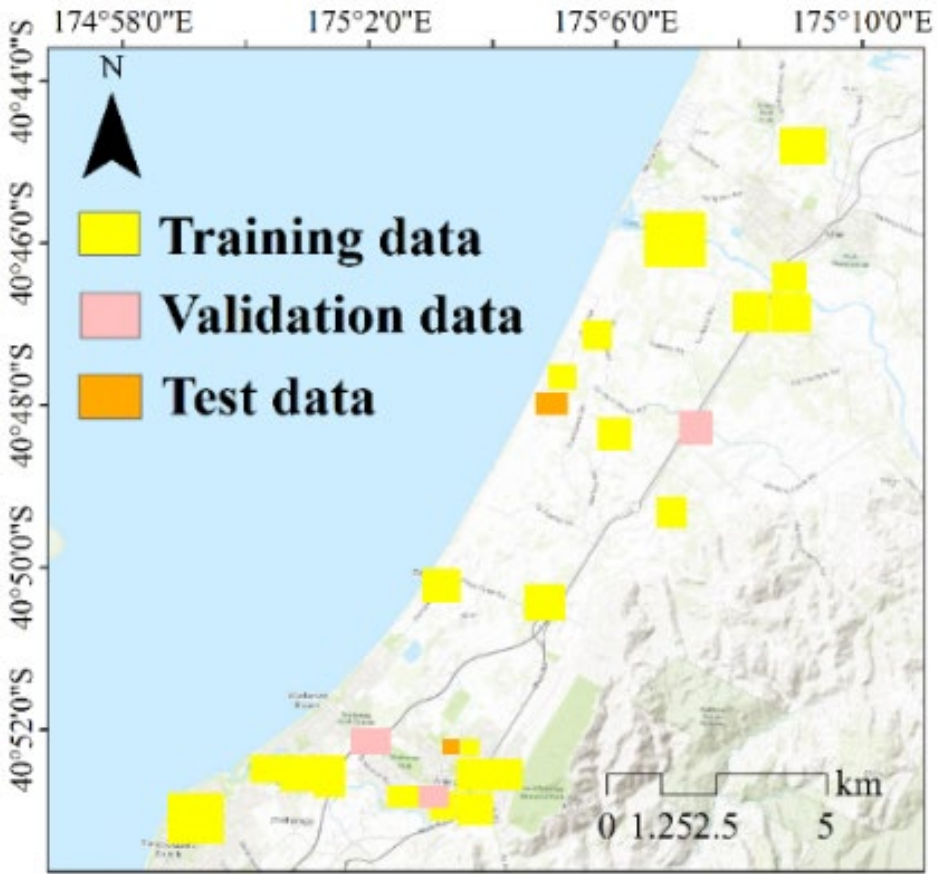



Figure 5-1. Kapiti Coast Lidar data location

Table 5-1. Detailed information of data applied in Chapter 5

| Data | 3D Lidar point clouds | 2D satellite imagery |
|---|---|---|
| **Collection date** | 13/03/2021 to 15/03/2021 | 24/07/2021 |

| | | |
|---|---|---|
| **Publisher** | OpenTopography | European Space Agency: Sentinel 2 |
| **Labelled class** | Ground, low vegetation, medium vegetation, high vegetation, buildings | None |
| **Coordinate system** | Horizontal: NZTM2000 NZGD2000 Meters [EPSG: 2193] Vertical: NZVD2016 [EPSG: 7839] | UTM84-60S |
| **Colour information** | None | Colours from red, green, and blue bands |

## 5.3 Method for the comparative experiment

This chapter presents the design of a comparative experiment involving eight feature selections to assess the impact of random down-sampling layers and surface normal information on pre-earthquake large-scale outdoor Lidar semantic segmentation, particularly focused on building footprint extraction. The choice of RandLA-Net as the backbone stems from its status as one of the pioneering point-based Lidar semantic segmentation networks developed for handling extensive datasets. Therefore, the comparative experiment involves the manipulation of two variables.

The first variable pertains to configuring random down-sampling layers within the network. Taking inspiration from a devised experiment whose backbone is also RandLA-Net (Huang, 2022), the first four selections were designed, including [4442], [4444], [44442], and

[444422]. Number 2 or 4 represents the random down-sampling ratio at each layer. In other words, ½ or ¼ points are saved at each layer after down-sampling points from its previous layer. The number of digits in each selection is the number of layers. For instance, [4442] and [44442] indicate that there are four and five layers, respectively. The layer configurations of [4444] and [44442] were chosen based on the original RandLA-Net backbone paper, which utilised these structures in their experimental setups. According to these two configurations, the other two, [4442] and [444422], were designed in this study.

The second variable is a Boolean value signifies whether to include surface normal information or not. The normals were calculated automatically by in-built algorithms in Python during the data processing stage. The detailed steps of calculating surface normals will be presented in Section 6.4.1. Considering the abovementioned random sampling layer variable, therefore, eight feature selections were designed in total, including [4442], [4444], [44442], [444422], [4442] + normals, [4444] + normals, [44442] + normals, and [444422] + normal.

Mean IoU (mIoU) was chosen as the main metric for validating and testing those networks in this chapter. This is because it is a widely applied metric in the DLSS field. IoU is always used to evaluate DL networks by estimating how well a predicted segmentation matches the ground truth. Moreover, the original DL backbone also applied IoU for the evaluation in its published open-source codes (Hu et al., 2020). The training epoch was 100, and the epoch with the highest mIoU during the validation stage was chosen and the trained network at that epoch was saved for the test. The meaning of mIoU is the average of the IoUs of all tested classes. equations are as follows (Huang, 2022):

$$mIoU = \frac{\sum_{k=1}^{i} IOU_k}{i} \quad 5\text{-}1$$

$$IoU = \frac{TP}{TP + FP + FN} \quad 5\text{-}2$$

where $k$ is the k$^{th}$ class and $i$ is the total number of labelled classes.

The IoU equation means the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. A higher IoU reflects a better predicted segmentation. TP, FN, and FP are true positive, false negative, and false positive, respectively. The definitions of TP, FN, and FP remain consistent with the explanations provided in Section 4.5.

## 5.4 Results and discussion

The experiments were implemented with one Nvidia RTX 2080Ti GPU card. As mentioned in Section 5.3, two of these 26 point clouds were chosen for the test. They were the 3$^{rd}$ and the 26$^{th}$ colourised point clouds, as shown in Figure 5-2. The points in them were 3,738,726 and 7,816,274, respectively.

(a) The 3$^{rd}$ point cloud (b) The 26$^{th}$ point cloud

Figure 5-2. Colourised Lidar data for testing the designed networks

The visualisation results of those two tested point clouds are shown in Figure 5-3 and Figure 5-4. Different colours represent different labelled classes. Only buildings can be visually recognized from these results directly. Additionally, visual interpretation alone does not provide a comprehensive analysis of the differences between each result. Therefore, conducting a quantitative analysis based on IoU is essential.

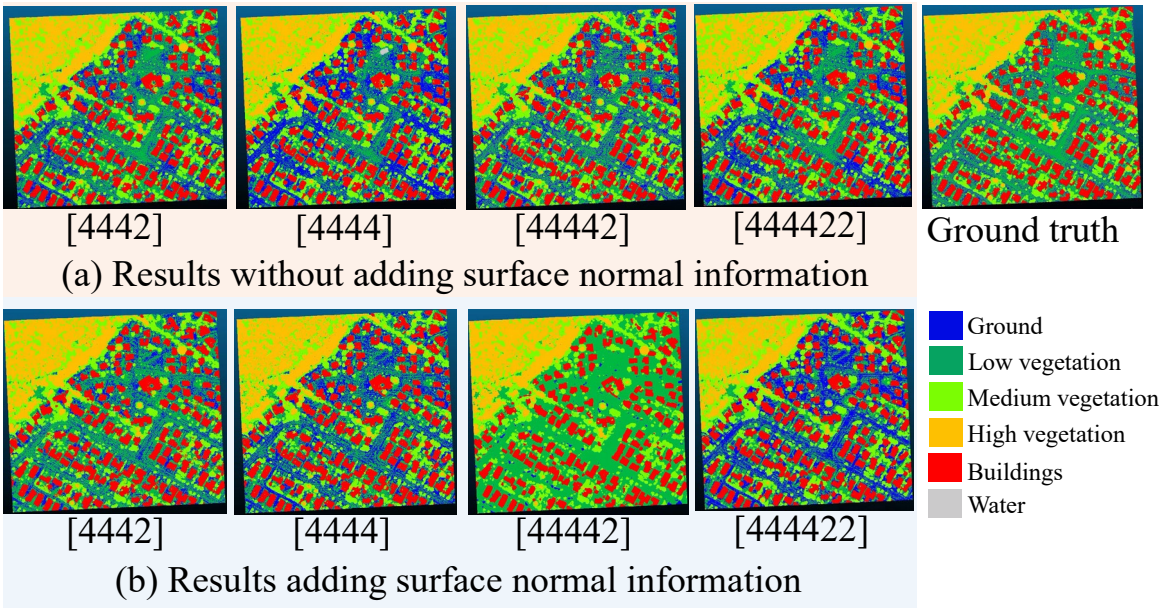



Figure 5-3. Results of the 3rd point cloud

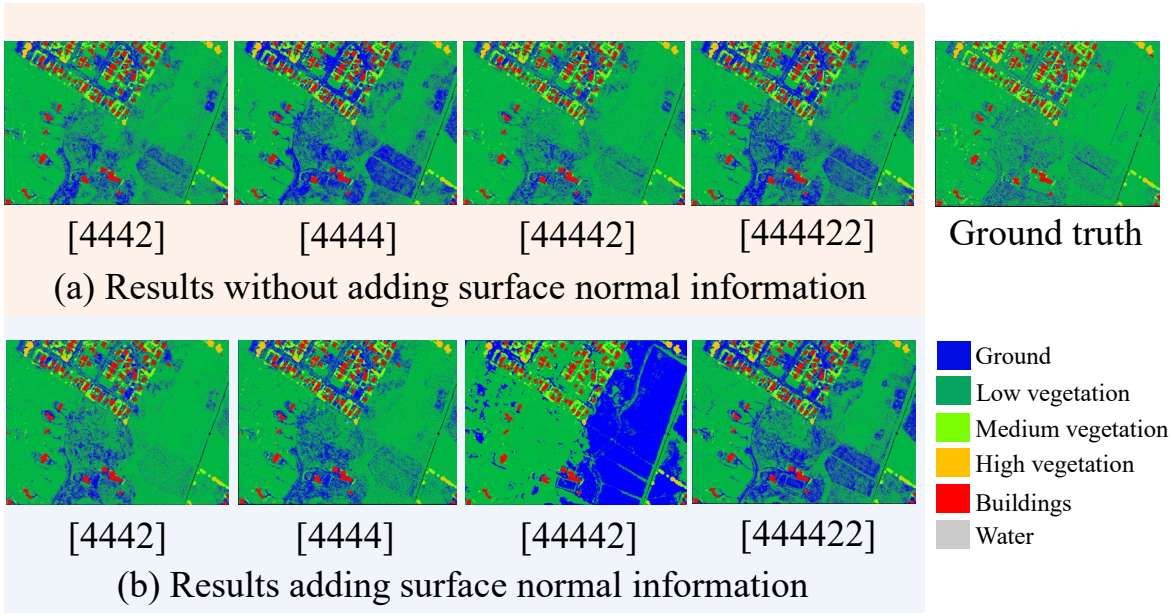



Figure 5-4. Results of the 26th point cloud

The mIoU of the two tested point clouds are listed in Table 5-2. '+ normals' means that the surface normal information has been calculated and added in the data pre-processing stage in that network. The mIoUs in the test results are the average IoU of five classes, including ground, low vegetation, medium vegetation, high vegetation, and buildings.

As shown in Table 5-2, the networks with '[4444]' and '[4444] + normals' structures have the highest mIoUs for the 3rd and 26th point clouds, respectively. The results also reflect that '[4444] + normals' network always has a high mIoU for each tested point cloud. It is an interesting finding that the networks adding surface normal information always have a higher mIoUs than the networks without it. The possible reason is that the network learned more features from adding surface normal information.

Table 5-2. Mean IoU of all tested point clouds

| Selection | Designed layer structure | mIoU | |
|---|---|---|---|
| | | **3rd point cloud** | **26th point cloud** |
| 1 | [4442] | 68.96% | 67.80% |
| 2 | [4444] | **72.37%** | 69.80% |
| 3 | [44442] | 62.52% | 63.66% |
| 4 | [444422] | 67.94% | 65.63% |
| 5 | [4442] + normals | 70.93% | 70.91% |
| 6 | [4444] + normals | 72.35% | **71.02%** |
| 7 | [44442] + normals | 70.26% | 68.78% |
| 8 | [444422] + normals | 68.57% | 67.22% |

The highest mIoU of both point clouds in Table 5-2 are both higher than 70%. All IoUs range between 62.52% and 72.37%. In order to explain the reason for this range, IoUs of all classes of the two tested point clouds are listed in Table 5-3 and Table 5-4. IoUs of three types of vegetation have been combined to be shown as a mIoU of them in these two tables. IoU of the building class is the highest and IoU of the ground class is the lowest no matter in which network. IoU of vegetation ranges from 66.98% to 76.88%, which is near the mIoU of all classes. It can be concluded that those designed networks are most suitable for building segmentation.

Table 5-3. IoU of all classes of 3rd point cloud

| Selection | Designed layer structure | Ground | Vegetation | Buildings |
|---|---|---|---|---|
| 1 | [4442] | 32.88% | 73.21% | 92.29% |

| | | | | |
|---|---|---|---|---|
| 2 | [4444] | 42.63% | 74.56% | 95.55% |
| 3 | [44442] | 1.37% | 74.80% | 86.84% |
| 4 | [444422] | 39.18% | 68.17% | 95.99% |
| 5 | [4442] + normals | 35.11% | 74.59% | 95.76% |
| 6 | [4444] + normals | 39.47% | 75.75% | 95.01% |
| 7 | [44442] + normals | 34.50% | 74.09% | 94.55% |
| 8 | [444422] +normals | 41.67% | 68.61% | 95.32% |

Table 5-4. IoU of all classes of 26$^{th}$ point cloud

| **Selection** | **Designed layer structure** | **Ground** | **Vegetation** | **Buildings** |
|---|---|---|---|---|
| 1 | [4442] | 27.33% | 73.38% | 91.55% |
| 2 | [4444] | 31.69% | 76.95% | 92.56% |
| 3 | [44442] | 27.95% | 74.76% | 91.68% |
| 4 | [444422] | 35.64% | 66.98% | 91.58% |
| 5 | [4442] + normals | 30.97% | 76.88% | 92.97% |
| 6 | [4444] + normals | 37.20% | 73.03% | 92.73% |
| 7 | [44442] + normals | 26.14% | 68.04% | 88.06% |
| 8 | [444422] +normals | 35.69% | 69.36% | 92.36% |

The detailed results of the three vegetation classes are listed in Table 5-5 for the 3$^{rd}$ point cloud and Table 5-6 for the 26$^{th}$ point cloud. Similar to the results shown in Table 5-2, the selections including the '[4444]' layer configuration always performed better results than other selections. Detailly, '[4444] + normals' performed best among all designed selections for the test of the 3$^{rd}$ point cloud. '[4444]' performed outstanding results than the rest of the selections for the 26$^{th}$ point cloud.

Table 5-5. Detailed results of the three vegetation classes in the 3rd point cloud

| **Selection** | | **[4442]** | **[4444]** | **[44442]** | **[444422]** | **[4442] + normals** | **[4444] + normals** | **[44442] + normals** | **[444422] +normals** |
|---|---|---|---|---|---|---|---|---|---|
| **Low vegetation** | **TP** | 774799 | 672045 | 1002111 | 738720 | 783470 | 686330 | 763643 | 622577 |
| | **FN** | 253665 | 356419 | 26353 | 289744 | 244994 | 342134 | 264821 | 405887 |
| | **FP** | 426896 | 270654 | 733806 | 335096 | 389997 | 314421 | 377692 | 273415 |
| | **IoU** | 53.24% | 51.73% | 56.86% | 54.18% | 55.23% | 51.11% | 54.31% | 47.82% |
| **Medium vegetation** | **TP** | 778993 | 841360 | 791805 | 853989 | 839397 | 815517 | 858775 | 830808 |
| | **FN** | 112433 | 50066 | 99621 | 37437 | 52029 | 75909 | 32651 | 60618 |
| | **FP** | 71173 | 86736 | 95798 | 180762 | 101418 | 49905 | 124594 | 140742 |
| | **IoU** | 80.93% | 86.01% | 80.21% | 79.65% | 84.54% | 86.63% | 84.52% | 80.49% |
| **High vegetation** | **TP** | 495281 | 487907 | 511340 | 390791 | 476795 | 515173 | 468958 | 440389 |
| | **FN** | 56592 | 63966 | 40533 | 161082 | 75078 | 36700 | 82915 | 111484 |
| | **FP** | 27564 | 15838 | 33689 | 913 | 15734 | 23693 | 10262 | 16184 |
| | **IoU** | 85.48% | 85.94% | 87.32% | 70.69% | 84.00% | 89.51% | 83.42% | 77.53% |
| **Mean IoU** | | 73.21% | 74.56% | 74.80% | 68.17% | 74.59% | **75.75%** | 74.09% | 68.61% |

Table 5-6. Detailed results of the three vegetation classes in the 26th point cloud

| Selection | | [4442] | [4444] | [44442] | [444422] | [4442] + normals | [4444] + normals | [44442] + normals | [444422] +normals |
|---|---|---|---|---|---|---|---|---|---|
| **Low vegetation** | **TP** | 4232093 | 4051103 | 4242372 | 3814868 | 4153723 | 3794905 | 3100359 | 3932447 |
| | **FN** | 657599 | 838589 | 647320 | 1074824 | 735969 | 1094787 | 1789333 | 957245 |
| | **FP** | 1539597 | 1330859 | 1501529 | 1138790 | 1385659 | 1067600 | 1270126 | 1175222 |
| | **IoU** | 65.83% | 65.12% | 66.38% | 63.28% | 66.19% | 63.70% | 50.33% | 64.84% |
| **Medium vegetation** | **TP** | 232557 | 258324 | 264187 | 249026 | 262487 | 260775 | 250942 | 254893 |
| | **FN** | 74469 | 48702 | 42839 | 58000 | 44539 | 46251 | 56084 | 52133 |
| | **FP** | 13471 | 12690 | 24525 | 32264 | 17647 | 23696 | 21787 | 28416 |
| | **IoU** | 72.56% | 80.80% | 79.68% | 73.40% | 80.85% | 78.85% | 76.32% | 75.99% |
| **High vegetation** | **TP** | 52990 | 55668 | 50743 | 41562 | 54126 | 49660 | 52013 | 43486 |
| | **FN** | 11681 | 9003 | 13928 | 23109 | 10545 | 15011 | 12658 | 21185 |
| | **FP** | 145 | 882 | 202 | 5 | 80 | 215 | 2480 | 1 |
| | **IoU** | 81.75% | 84.92% | 78.22% | 64.26% | 83.59% | 76.53% | 77.46% | 67.24% |
| **Mean IoU** | | 73.38% | **76.95%** | 74.76% | 66.98% | 76.88% | 73.03% | 68.04% | 69.36% |

## 5.5 Conclusion

To improve LOLSS using DL networks, this chapter aimed to analyse the possible benefits of adding surface normal information and changing layer structures of the random down-sampling stage. Eight feature selections were designed for the experiment and a manually labelled dataset was created. The results show that the network can always have acceptable predicted segmentation with a mIoU value over 70% that adds surface normal information with four random down-sampling layers whose sampling ratios are 4, 4, 4, and 4 of those layers. This structure is always higher by at least 1% than other combination selections. Moreover, all the designed networks are most suitable for building segmentation among all labelled classes. This may be attributed to the fact that the features of the building class are more prominent and readily learned by DL networks compared to other classes, namely ground and vegetation. The findings in Table 5-2 also show that the second structure is the best for ground segmentation and the sixth is the best for vegetation segmentation. This chapter is beneficial for building extraction from Lidar data in applications such as urban planning or predisaster 3D land cover information storage.

There are some suggestions for future studies. Firstly, additional datasets are suggested to be incorporated. The point clouds chosen from different places can be used to test the generalisability of the proposed DL networks in future. Secondly, other features can be added or revised in the structure of a DL network to exploit their potential benefits of improving segmentation accuracy. Since this chapter is the preparatory work for Chapter 6,

[4444] was chosen as the down-sampling layer configuration for the next Chapter according to the findings of this chapter.

# Chapter 6

# Large-scale predisaster colourised Lidar semantic segmentation[3]

## 6.1 Background and scope

The frequency of destructive natural disasters is on the rise due to the increasing occurrence of extreme weather events attributed to climate change. Natural disaster management has gained significant global attention recently (Liu et al., 2022). To avoid the disastrous and chaotic aftermath, pre-emptive measures are valuable before the impact of a disaster. Predisaster information storage allows post-disaster decision-makers to strategize rescue routes and determine suitable locations for temporary housing, thereby enabling swift disaster response.

[3] The content presented in this chapter is partially adopted from the following published paper: "Liu C, Ge L*, Xiang W, Du Z, and Zhang Q, 2023. Channel Attention and Normal-based Local Feature Aggregation Network (CNLNet): A Deep Learning Method for Predisaster Large-scale Outdoor Lidar Semantic Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*. 62, pp. 1-12. DOI: 10.1109/TGRS.2023.3339475." It has been acknowledged and detailed in the "Inclusion of Publications Statement" for this thesis.

As a component of predisaster information storage, the retention of predisaster urban land cover visualisation data is invaluable for disaster analysis reconnaissance (He et al., 2016b). These data should be stored and periodically updated to expedite post-disaster analysis and management processes. However, conventional in-situ data collection methods have several issues, including being labour-intensive, time-consuming, costly, and potentially dangerous. Remote sensing technology offers a swift and efficient alternative for urban land cover visualisation data collection due to its capacity to acquire extensive data on a large scale with relative ease.

Lidar has recently gained significant attention in remote sensing because of its 3D information and higher vertical accuracy with better penetration than conventional photogrammetry. Compared with conventional in-situ urban data collection methods, Lidar usually spends less time, which helps operators save time and labour costs (Zhang et al., 2022). Due to the rapid development of DL, there has been a burgeoning interest in its application to remote sensing-based Lidar semantic segmentation in recent years (Guo et al., 2020). Therefore, a DL-based Lidar semantic segmentation could solve rapid predisaster land cover visualisation data collection and storage.

Unlike 2D imagery, Lidar point cloud data belong to non-Euclidean geometry data. Therefore, semantic segmentation methods for 3D data cannot simply be decreased to 2D segmentation. Reducing the dimensionality from 3D to 2D inevitably results in the loss of information. To design DL methods suitable for 3D semantic segmentation while retaining the inherent 3D data, the development of point-based networks began in 2017.

In 2017, PointNet directly took points as its input, which was the first point-based network. It learns features with a shared MLP (Qi et al., 2017a). Nevertheless, the local structures and the mutual interactions between features cannot be extracted by a shared MLP in PointNet (Qi et al., 2017a). To learn richer local geometry in point clouds and capture a broader context for each point, several methods have been introduced to develop PointNet, such as neighbouring feature pooling. In particular, PointNet++ was proposed soon after the generation of PointNet to categorise points hierarchically and progressively learn from larger local regions. It achieved better results than PointNet according to the conducted experiments (Qi et al., 2017b). Following PointNet++, Jiang et al. (Jiang et al., 2018) introduced a PointSIFT module to stack and encode the point information from eight spatial orientations using a three-stage ordered convolution process.

Given the rapid advancements in point-based DL methods for 3D semantic segmentation, certain scholars have commenced discourse on the topic of large-scale outdoor Lidar semantic segmentation (LOLSS). For instance, RandLA-Net was proposed for LOLSS as a lightweight network for saving processing time (Hu et al., 2020). It applies random point down-sampling to attain a high level of efficiency in memory and computation. A local feature aggregation unit was further proposed to capture and retain geometric features. However, there is still a lack of enough studies for large-scale scenarios in the computer vision field. Most advanced networks are still only designed for small or indoor scenes.

Moreover, there is still a lack of full development of semantic segmentation methods for the predisaster land cover information storage purpose. In detail, several possible methods have not been fully discussed for disaster-related research, and DL networks have not been

extensively trained to account for the potential occurrence of natural disasters in the selected datasets' locations. Therefore, there is a lack of efficient and accurate Lidar semantic segmentation methods that can classify predisaster large-scale land cover classification.

In order to solve these problems, this chapter aims to provide a DL LOLSS network by creating a dataset tailored to the targeted task to store the 3D information of predisaster large-scale outdoor land cover objects.

## 6.2 Data and study extents

This study chose four own labelled places and one public dataset to test the proposed network. The own labelled places include Kapiti Coast, Tasman, Nelson, New Zealand, and Kumamoto, Japan. These four places were chosen because they are both tectonically active urban areas near the sea. They are difficult sites for in-situ observations and contain several potential natural hazards (Kapiti Coast District Council, 2022). Moreover, three of them have already caused serious natural disasters. Continuous heavy rain caused severe landslips and flooding in Tasman and Nelson in August 2022 (Nelson Government, 2023), and a severe earthquake occurred on 16/04/2016 in Kumamoto, Japan (Yamada et al., 2017). This study chose the data whose collection dates were near the floods and before the earthquake.

The original labelled classes from these own labelled datasets are listed in Table 6-1. All unlabelled point clouds were ignored during experiments. The original Lidar point clouds of these places do not include colour information, so this study needs to add corresponding

colours to Lidar data in the pre-processing step. The colour information of RGB bands from optical images is a viable choice for finishing this task. Therefore, this study collected optical images from the same places of the Lidar datasets to fuse 3D Lidar and 2D images. The detailed pre-processing steps for each place are introduced in Section 6.3.

Lidar data and optical images with RGB bands of these four places are shown in Table 6-2. S2 images of all places were collected. KOMPSAT-3 (K3) images for the 2016 Kumamoto pre-earthquake Lidar data were also collected to test the influence of image resolution on the performance of the proposed network (refer to Section 6.4.5). Since the Lidar data of the datasets were collected by different organizations, their parameters are different. Considering this, the Kapiti Coast, Tasman, and Nelson datasets were applied for both DL training (and validation) and testing stages, while the Kumamoto dataset was only utilized in the testing stage to test the generation capability of the networks trained with the other datasets.

Semantic3D is a large-scale open-source dataset. It was chosen to compare the accuracy of the proposed method and other well-known DL networks for Lidar semantic segmentation.

Section 6.2.1 introduces detailed information on the three datasets from New Zealand. Section 6.2.2 introduces the Kumamoto dataset collected before the 2016 Kumamoto Earthquake. Section 6.2.3 introduces the Semantic3D that is applied in this study.

Table 6-1. Original labelled classes of the data

| Lidar dataset | Number of labelled classes | Labelled classes |
|---|---|---|

| | | |
|---|---|---|
| Kapiti Coast Tasman, Nelson | 5 | Ground, low vegetation, medium vegetation, high vegetation, buildings |
| Kumamoto (pre-earthquake) | 4 | Ground, low vegetation, medium vegetation, high vegetation |

Table 6-2. Data information

| Location | Potential disaster | Point cloud | | | Optical image | | |
|---|---|---|---|---|---|---|---|
| | | Data | Point density | Acquired date | Satellite | SR, m/pxl | Acquired date |
| Kapiti Coast | Earthquakes, storms, floods, landslides | Lidar | 27.95 pts/m$^2$ | 13/03 to 15/03/2021 | S2 | 10 | 24/07/2021 |
| Tasman | Floods | Lidar | 15.05 pts/m$^2$ | 23/082022 to 06/09/2022 | S2 | 10 | 15/09/2022 |
| Nelson | | | | | | | |
| Kumamoto pre-earthquake | Earthquakes, storms, floods, landslides | Lidar | 2.94 pts/m$^2$ | 15/04/2016 | K3 | 0.5 | 15/04/2016 |
| | | | | | S2 | 10 | 03/03/2016 |

### 6.2.1 Kapiti Coast, Tasman, Nelson in New Zealand

49 selected patches of point cloud data from New Zealand were selected in this study, as shown in Figure 6-1. 26, 7, and 16 are from Kapiti Coast, Tasman, and Nelson, respectively. The number of Lidar data was chosen because of considering the number of Semantic 3D data (Hackel et al., 2017) applied in RandLA-Net (Hu et al., 2020) since this study is developed from RandLA-Net. The training, validation, and testing data are shown in indicolite green, olivine yellow, and sugilite sky colours in Figure 6-1. The dataset contains five classes labelled by experts from the data provider, as shown in Table 6-1, including ground, low vegetation, medium vegetation, high vegetation, and buildings

(Opentopography, 2022, Opentopography, 2023). All these labels are kept in this study as the information of all these classes is necessary for recovery plans.

This study chose S2 images for colour fusion because it is free and easy to access. After checking all S2 data with the date near the dates of Lidar collection, the dates of S2 images were chosen, as shown in Table 6-2. The images of other dates either contain several clouds or are in the dark.

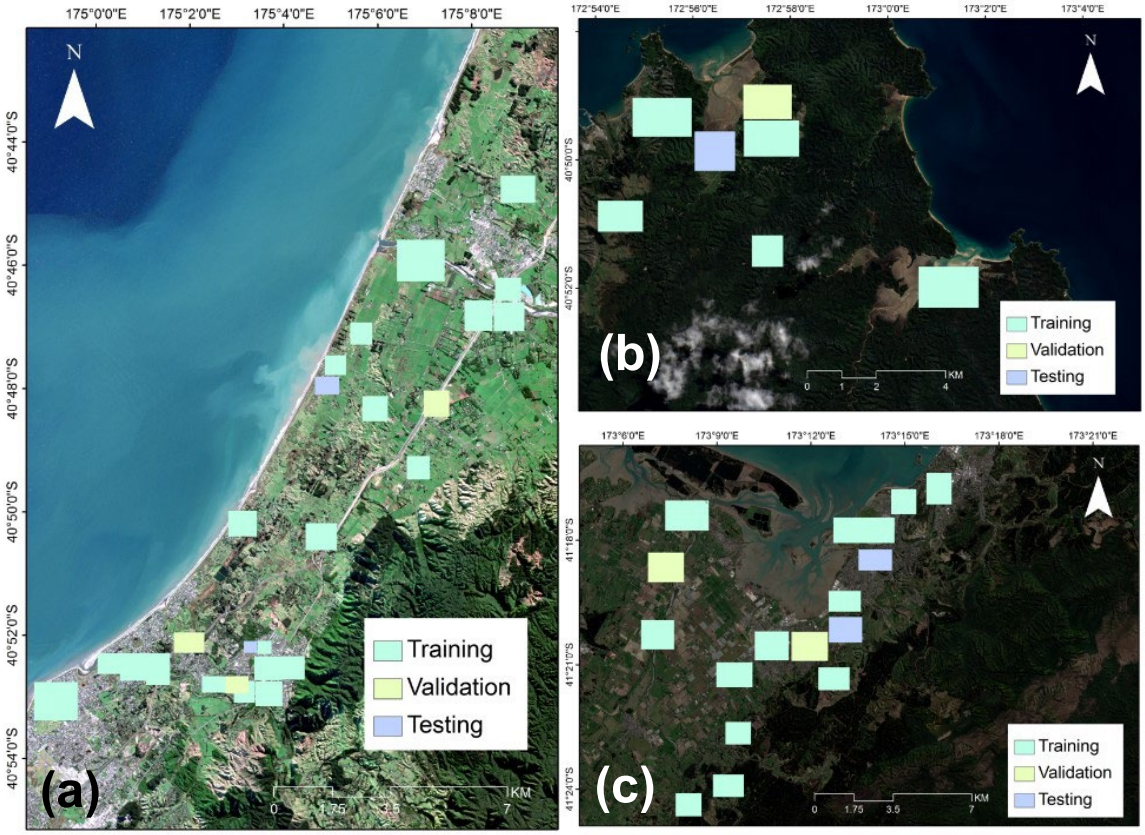
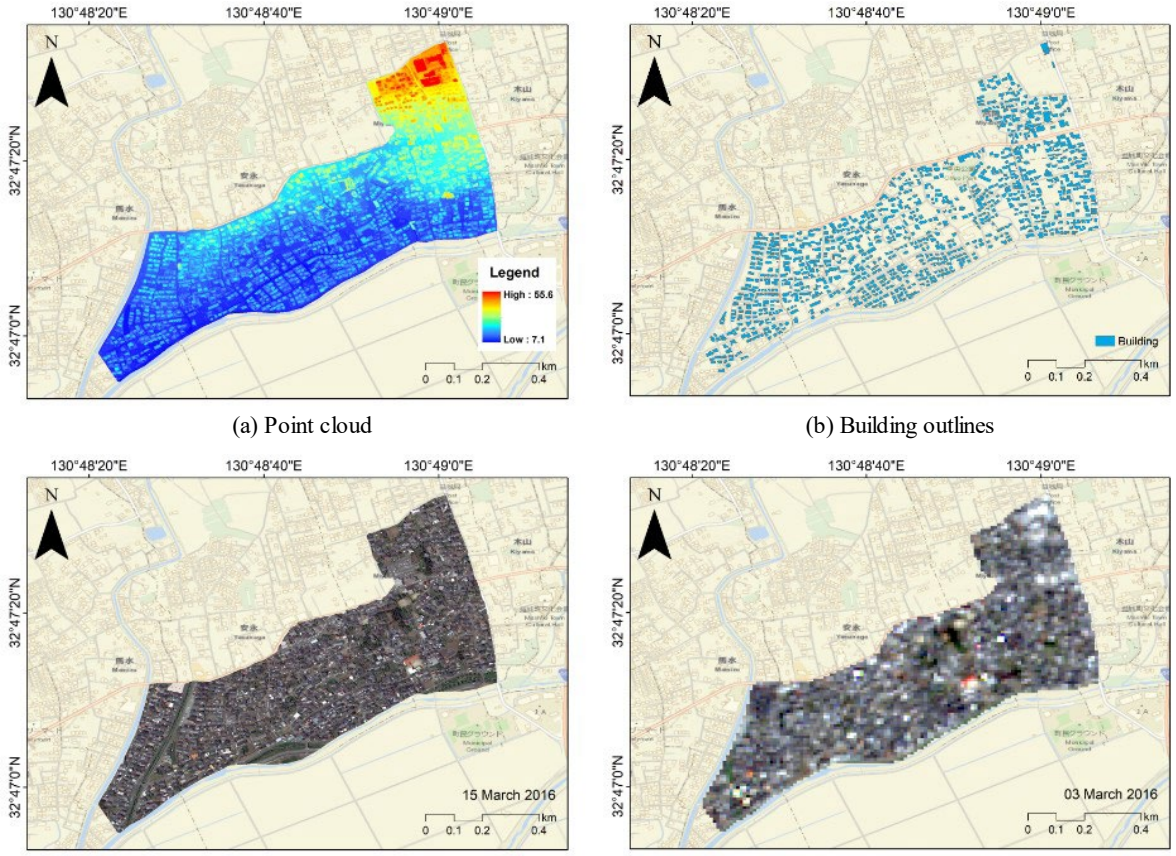



Figure 6-1. Locations of datasets in New Zealand; (a) Kapiti Coast; (b) Tasman; (c) Nelson

### 6.2.2 Kumamoto pre-earthquake dataset

A mainshock of the 7.0 $M_W$ Kumamoto earthquake struck on 16/04/2016. Four types of pre-earthquake data in Kumamoto were utilized in this study, as shown in Figure 6-2,

including a Lidar point cloud (Figure 6-2 (a)), a building outline shapefile (Figure 6-2 (b)), an optical image from K3 satellite (Figure 6-2 (c)), and an optical image from S2 satellite (Figure 6-2 (d)). The colour from blue to red shown in Figure 6-2 (a) represents the increase in elevation.




(a) Point cloud

(b) Building outlines

(c) K3 image shown in RGB bands

(d) S2 image shown in RGB bands

Figure 6-2. Data applied for the fusion of optical images and Lidar point cloud data in the Kumamoto pre-earthquake dataset

The original labelled classes in the Lidar point clouds were ground, low vegetation, medium vegetation, and high vegetation. The building class is an integral part of predisaster information collection, but the original Lidar dataset did not have this class, so the building footprint information was added to the Kumamoto dataset during the pre-processing, which

will be introduced in Section 6.3.2. Optical images from two satellites with different resolutions to test if the image spatial resolution will influence the accuracy of the proposed network. K3 is 0.5m per pixel (/pxl), and S2 is 10m/pxl.

Detailly, Figure 6-2 (c) is a K3 Ortho-ready Correction L1O image in PSG 32652 projected coordinate system. Its resolution is 0.5m/pxl. The L1O mode removes errors caused by the satellite's posture or position and matches the geographic coordinate system. Figure 6-2 (d) is an S2 image in UTM84-52N projected coordinate system. Only the S2 Level-1C (L1C) image is chosen because of its acquisition date. This study, therefore, converted the L1C product to its corresponding L2A product with SNAP software. This conversion included a scene classification and an atmospheric correction applied to L1C orthoimage products.

### 6.2.3 Semantic3D dataset

Semantic3D dataset is one of the most popular open-source point cloud datasets for DLSS. Eight labelled classes from this dataset were chosen in this study, including natural terrain, high vegetation, low vegetation, buildings, hard scape, scanning artifacts, and cars. Four point clouds were selected for the network test according to the design of the RandLA-Net backbone.

## 6.3 3D data pre-processing: Data fusion of Lidar data with satellite RGB data

The main task of this step was to incorporate colour information into Lidar data. Although the Lidar coordinate systems varied among different datasets, this study disregarded these

differences during training. However, it is essential to ensure that the coordinate systems of satellite images and Lidar data within the same dataset are consistent. Therefore, all other data, regardless of whether they were in projected or geographic coordinate systems, were transformed to match the coordinate system of the Lidar data.

### 6.3.1 Pre-processing of the three datasets in New Zealand

Some pre-processing steps were applied before training the DL network, as shown in Figure 6-3. Since the original Lidar point data do not have colours, this study fused 2D optical images and 3D point clouds to obtain the colour point cloud using Feature Manipulation Engine (Safe Software, 2022). First, the point clouds were loaded. Second, the colour optical data were reprojected from UTM84-60S to the same coordinate system as the Lidar. Thus, the colour was added to the top points in each point cloud according to the coordinate system. The optical image was collected from S2, and only RGB bands were applied in this study. Then, the point cloud and the RGB image were fused to obtain the colourised point cloud.

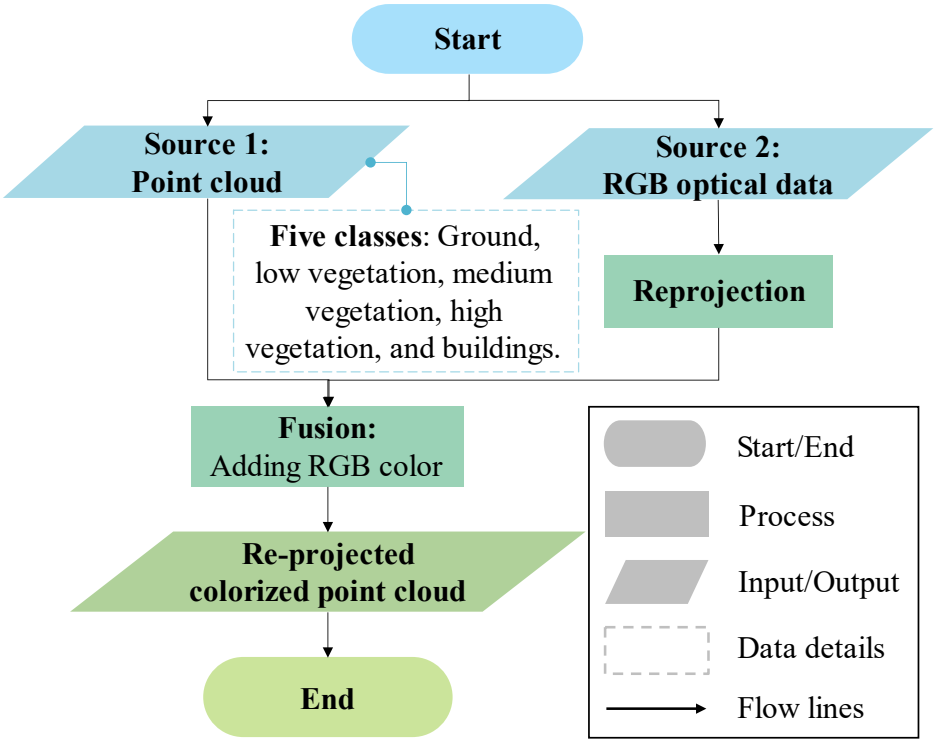



Figure 6-3. Workflow of Kapiti Coast data fusion with adding RGB information

### 6.3.2 Pre-processing of Kumamoto pre-earthquake data

The original non-colour Lidar dataset contains only four classes without the building class. Since information about the building class and colours is essential for this study, both colour bands and building outlines were fused into Lidar point clouds in Feature Manipulation Engine, as illustrated in Figure 6-4.

The first fusion is adding building outlines in Lidar. The originally downloaded coordinate system for building outline polygons is the LL-WGS 1984_0 geographic coordinate system. It was reprojected to JGD2K-02 projected coordinate system (the coordinate system of the Lidar data), and then point clouds in building outlines were classified as 'building'.

The second fusion is adding colour. K3 and S2 images were reprojected to the JGD2K-02 coordinate system to match the system of Lidar. Then, the reprojected RGB bands from clipped optical images were fused with Lidar. The fusion results were the reprojected colourised Lidar point cloud dataset with the five classes.

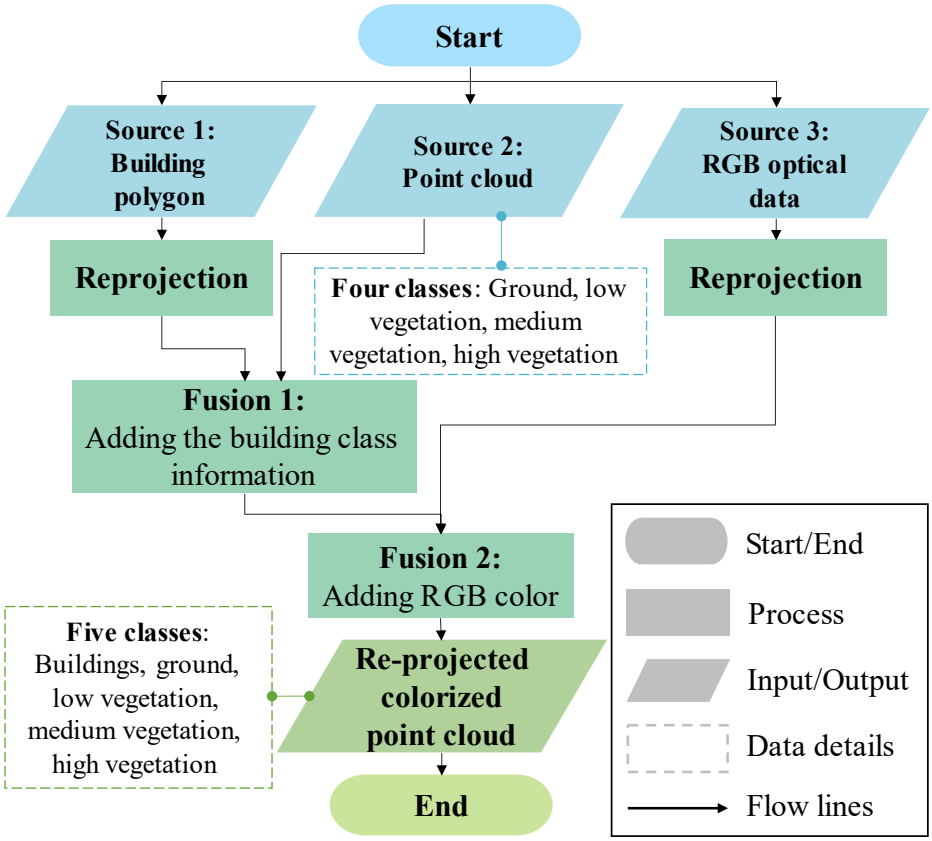



Figure 6-4. Workflow of Kumamoto pre-earthquake data fusion with adding the building class and RGB information

## 6.4 Methodology of the channel attention and normal-based local feature aggregation network (CNLNet)

This study proposed a DL method called CNLNet for LOLSS. The main improvements include adding normal information and the CA mechanism in the backbone.

CNLNet adds these two possible helpful approaches in the backbone to increase the accuracy of LOLSS. Surface normal information is important in several point cloud applications. The attention mechanism is widely confirmed effective in 2D or small-scale 3D DL networks. However, they are not always applied in large-scale predisaster scenarios. Therefore, this study added these two to enhance the backbone and developed a module in it.

### 6.4.1 Surface normal information addition and data preparation

This section introduces how to calculate the surface normal during data preparation. The collected, revised coloured Lidar data (refer to Section 6.3) required further processing for data preparation before the training stage.

Surface normal information is one of the essential properties of a geometric surface, and it finds applications in various research areas. For instance, in computer graphics, light rendering depends on normal information to generate shadings and other visual effects to look more realistic. Therefore, this chapter evaluates the impact of surface normal information on enhancing the accuracy of semantically segmenting large-scale point clouds. The process of adding normal information involves four main steps: data format

transformation, voxel grid down-sampling, computation of normal estimation, and data storage, as depicted in Figure 6-5. Voxel grid down-sampling is necessary because using the original massive point cloud data as inputs in the computer is impractical.

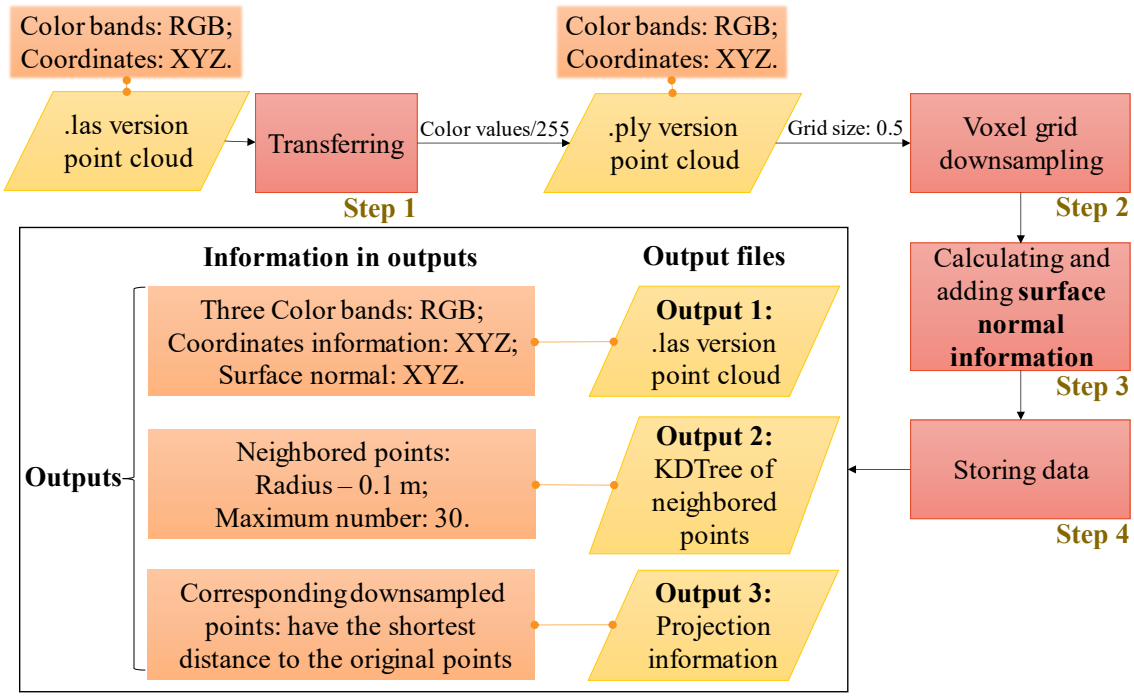



Figure 6-5. Workflow of data preparation with adding surface information

In the first step, in order to have the same data format for all point clouds, the clouds with the '.las' format were transferred to the '.ply' format. This is because the proposed network was designed for processing point clouds in the '.ply' version. The values of each colour band in the '.las' format were divided by 255 before transferring to the '.ply' clouds. To calculate the normal on a point, the local surface must be estimated to represent itself and its neighbours. Hence, the coordinate values of each point were necessary. Since colour information was also needed in this study, both the coordinate values and RGB colour information were stored for the next step.

In the second step, voxel grid down-sampling was applied to all points. The volume of the originally collected point cloud data is exceptionally large in most situations. Thus, the volume is always reduced by down-sampling without affecting the characteristics of a point cloud. This operation can help to save processing time and avoid out-of-memory during training networks. The grid size was 0.5 m in this study.

Thirdly, surface normal information was calculated. To add surface normal information, this study applied Open3D, an open-source Python library, to generate normals. This is because Open3D has already encapsulated the function. The built-in function 'estimate_normals' finds $K$ -nearest neighbour points within a radius and calculates the principal axis of the adjacent points using covariance analysis (Open3D). The function chooses a point and its $K$ -nearest neighbours (i.e., $1 + K$ points in total) to estimate a plane using the least square method and then makes a vertical line of the plane through that point, which is its normal vector. Specifically, the problem of estimating the surface normal of a point is simplified as an analysis of eigenvectors and eigenvalues of the covariance matrix calculated from the nearest neighbour of the point. In this study, the search radius was 0.1m, and the maximum nearest neighbour was 30 using KDTree search for neighbourhood search, which are default numbers in Open3D. Choosing default numbers because these parameters are not the focus of this study.

The normal orientation problem of surface normal calculation should be noted. Two normal candidates with opposite directions are produced from the covariance analysis algorithm. Without knowing the global structure of geometry, both can be correct, which could cause problems. Therefore, Open3D tried to orient the normal to align with the original normal if

it existed. Otherwise, Open3D made a random guess. Then, normal values were added to point cloud data.

In the fourth step, three types of outputs were stored, as shown in the black rectangle of Figure 6-5. Detailly, after producing the point cloud data from the third step, KDTree files and projection files are also generated and stored for each point cloud. Each KDTree file was named 'XX_KDTree.pkl'. KDTree files have the information of the nearest *N* points around each down-sampled point. Projection files have stored the number of the down-sampled points with the shortest distance from each original point. The original points are the points before the second step—downsampling. These numbers were stored in files named 'XX_proj.pkl'. Projection files are necessary because point clouds need to be restored to the original size after semantic segmentation for the down-sampled ones in the proposed network. The restoration needs these numbers for nearest neighbour interpolation.

Following the above steps, the final outputs include colourised point cloud data in the '.ply' format, KDTree files, and projection files. The information in point clouds contains the RGB bands, three values of coordinate systems, and three values of the corresponding normals.

### 6.4.2 The architecture of the proposed network

The architecture of the proposed CNLNet is shown in Figure 6-6. It is a conventional encoder and decoder architecture with skip connections. The inputs contain three types of files, including point clouds, KDTrees, and projected numbers of point clouds. The architecture has four encoding and decoding layers. As shown in those four encoding layers,

only a quarter of the point features are retained with the increased feature dimension after each layer for down-sampling. Random point sampling is applied for high efficiency of memory and computation, as its computational complexity is only O(1). After that, the point features are up-sampled gradually through a nearest-neighbour interpolation in the four decoding layers (Hu et al., 2020). The final output is obtained through shared fully connected layers. The final output is the predicted class of each point. To be noticed, this study adds contents in red rectangles, including normal information and CA in the local feature aggregation (LFA) module. The details of the backbone including LFA and CA are introduced as follows.

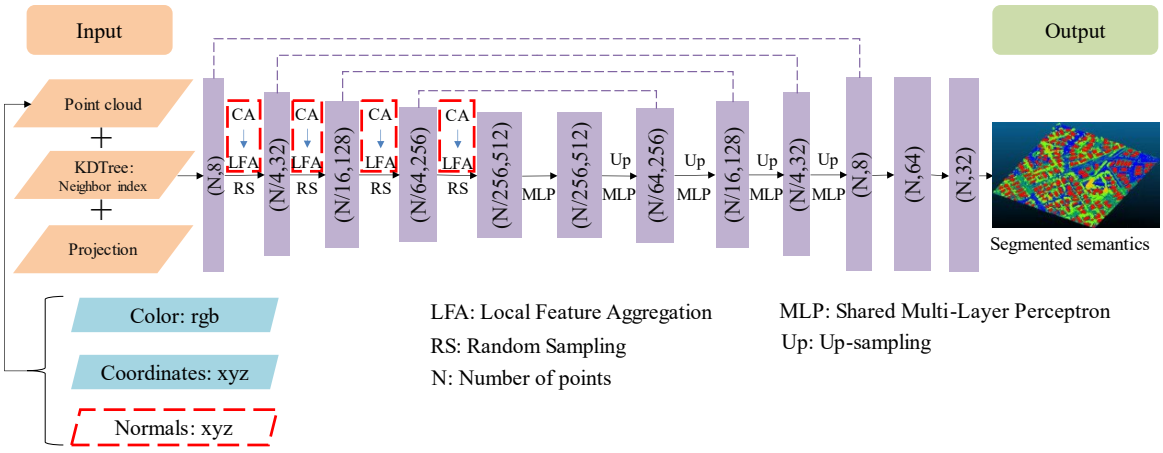



Figure 6-6. The architecture of the proposed with the amalgamation of the surface normal information and CA

Datasets in New Zealand were applied for training, validation, and testing with the number of point clouds 38, 6, and 5, respectively. Kumamoto data were only applied for the tests, because this area is too small to separate it into three parts for training, validation, and testing.

### 6.4.3 RandLA-Net backbone

The backbone of the proposed network is RandLA-Net, that is, 'random sampling and an effective local feature aggregator network' (Hu et al., 2020). Although several networks showed promising results for small point cloud semantic segmentation, most cannot directly scale up to large scenarios. This is because of their high memory and computational costs. The benefit of RandLA-Net is that it was designed for large-scale point cloud semantic segmentation with less memory and computation, which is suitable for predisaster tasks. Therefore, this study chose it as the backbone.

RandLA-Net is a lightweight point-wise MLP network. Point-based DL methods for semantic segmentation can be roughly divided into pointwise MLP, point convolution, RNN-based, and graph-based methods (Guo et al., 2020). MLP is a supplement of a feed-forward neural network, including the input layer, the output layer, and the hidden layers. A sample structure is shown in Figure 6-7 to explain the relationships between these three layers.

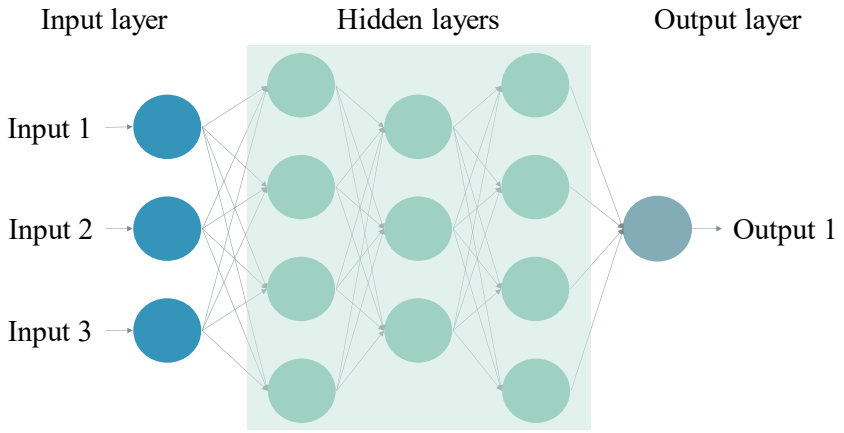



Figure 6-7. Sample structure of MLP

RandLA-Net designed a local feature aggregation (LFA) module with shared MLP preserving local geometric structures and other useful local features, as shown in Figure 6-8. The LFA module has two key units: Local spatial encoding (LocSE) and attentive pooling (AP). Their details are shown in Figure 6-9. The LocSE unit is applied for local geometric structures, and the attentive pooling unit is applied for saving those useful local features. In the LocSE unit, the *K* -nearest neighbours algorithm (KNN) is utilized to find neighbour points based on the point-wise Euclidean distances. *K* represents the number of neighbour points. *K* is 16 in this study. After finding the neighbour points, MLP is applied to encode the relative point positions between every centre point and its neighbouring points. Hence, the local geometric structures are encoded for every centre point to augment neighbouring point features by LocSE. After that, the attentive pooling unit is applied to aggregate the neighbouring point features. This unit applied shared MLP followed by SoftMax function to learn a unique attention score for every feature. Then, the features are weighted and summed. RandLA-Net stacks multiple LocSE and AP units with a skip connection as a dilated residual block. In order to avoid overfitting during the training stage and keep computation efficiency, only two sets of LocSE and AP are stacked (Hu et al., 2020).

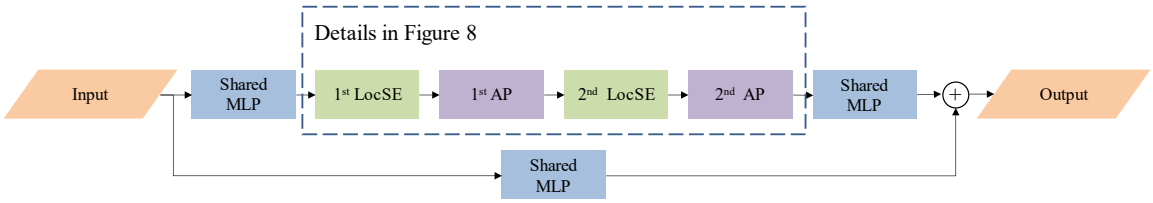



Figure 6-8. The architecture of the LFA module

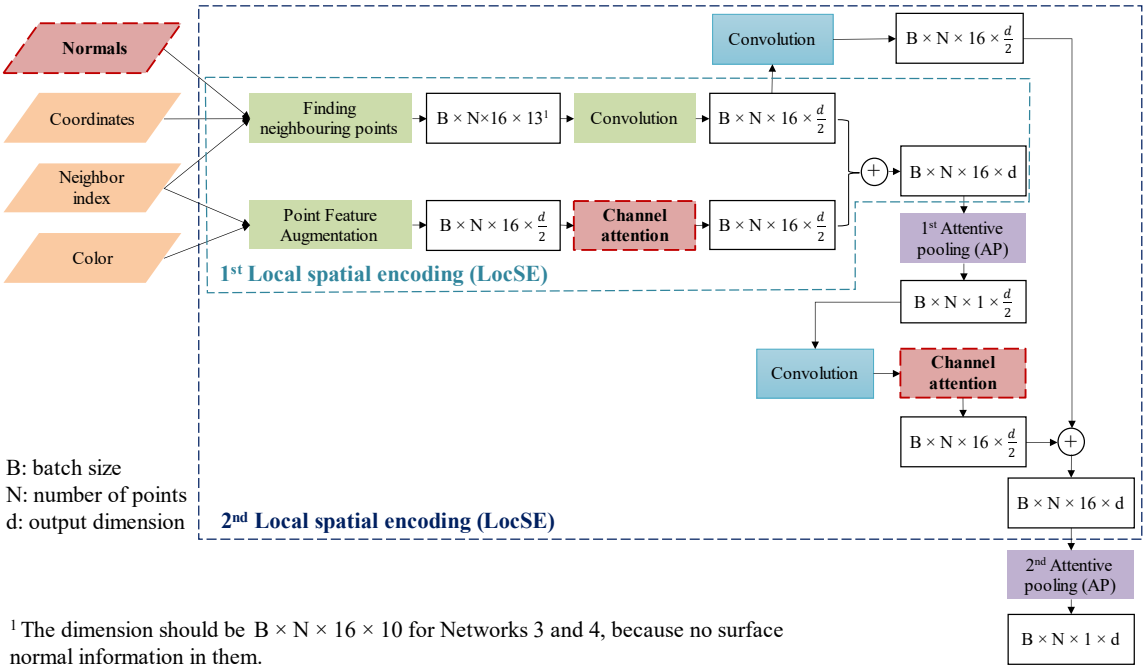



Figure 6-9. Structure details of the CA added to the modified LocSE and AP

### 6.4.4 Channel attention in CNLNet

CA was added to the proposed network to examine its effect on the final output. Adapting the previous work of a multi-branch network (Geng et al., 2021), CA applied this study that deduces the channel number to 1 and then recovers back to the original number. With this operation, the relevance between each channel and key information in channels can be more obvious and easier for the computer to learn. The benefit of this attention mechanism is that it can usually help to achieve significant improvement in accuracy in terms of 2D semantic segmentation (Hu et al., 2018). Moreover, since the accuracy needs to increase and LFA does not contain CA, this study applied CA in 3D semantic segmentation to test its effects.

The novel proposed network consists of LFA and CA with surface normals. CA is added into the LocSE unit of the LFA module as shown in red rectangles of Figure 6-9. The details are explained in Figure 6-10. *K* and *d* represent the number of neighbour points, and the feature dimension, respectively. First, the matrix is transposed from *(K, d)* to *(d, K)*. Then, the transposed is multiplied by the original matrix. The dimension of the multiplication result is squeezed to 1 by max pooling and restored to *d* by copying and subtracting an activation function. The multiplication of the original matrix and the restored one is operated after that. Last, the attentive result is the sum.

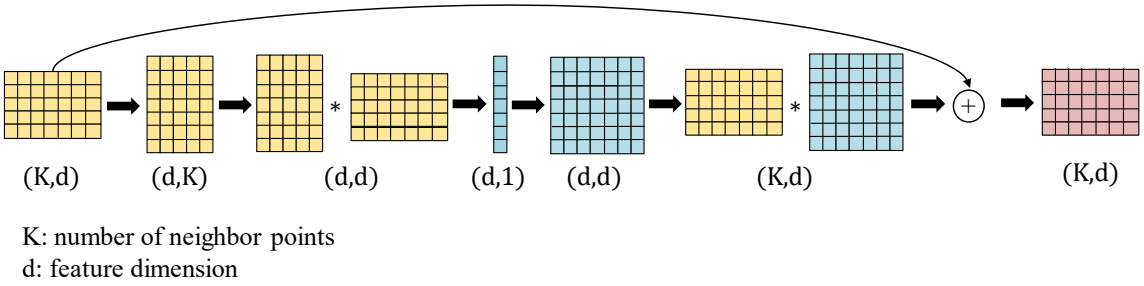



Figure 6-10. Details of CA

### 6.4.5 Ablation studies on four in-house labelled datasets

The design of an ablation study with five evaluation metrics for detecting the impact of information and CA on segmentation is introduced in this section.

Four networks were tested in the ablation study, as shown in Table 6-3. They were designed to demonstrate the benefit of adding surface normal information or CA in the backbone. The backbone added both normal information and the CA block is Network 1. The backbone with only normal information or CA was designed as Networks 2 and 3. The

original RandLA-Net backbone network was tested at last, which is Network 4. Each point in data is represented by its coordinates, normal, and colour information in Networks 1 and 2. It is represented by the coordinates and colours in Networks 3 and 4.

Five evaluation metrics were chosen. These five metrics were calculated for testing the segmentation performance of each network, including TP, FN, FP, IoU, and semantic segmentation accuracy (SSA). The summation of TP, TN, FN, and FP is the whole number of points in one point cloud. TP represents a point whose tested label is the same as its true label. TN in each class indicates the points that both its tested and true labels do not belong to that class. In the results of a class, FN refers to the point that its tested label does not belong to this class, but its true label does. On the other hand, FP in results of a particular class means the tested result is in this class but its true label does not.

The IoU shown in Equation 5-2 is a mathematical way to choose the best network by checking the degree of similarity of the output produced by the proposed networks with the ground truth. A higher IoU value, a better performance of the chosen network. After observing initial results, this study predominately discussed IoU rather than the other four metrics (i.e., TP, FN, TP, SSA). It is noted that the IoU applied in this chapter is the same as the IoU in Chapter 5. One reason is that IoU contains TP, FN, and FP. Discussing IoU would be more helpful for data analysis than only analysing a single TP, FN, or FP. Another reason is that several relevant papers utilised IoU as the metric.

SSA shown in Equation 6-1 was not considered as the main metric mainly because a high SSA cannot reflect a good result in this study. The number of each point cloud is huge. If

TP, FN, and FP were all very low in one class, TN would be very high in this class. This situation thus cannot demonstrate that the results are ideal, although SSA was nearly 1. Therefore, SSA can only be considered as a reference metric.

Mean IoU (mIoU) was also calculated for multi-class-based semantic segmentation. The mIoU represents the average between the IoU of all segmented classes over all the images of each tested point cloud. All networks were trained ten times, and the network that had the highest mIoU value for validation data was chosen to be applied to the test data. It shows the correctly segmented area over all the areas that the network segmented.

Table 6-3. Designed networks of ablation study

| **Network 1 (CNLNet)** | **Network 2** | **Network 3** | **Network 4** |
|---|---|---|---|
| Backbone + Normals + CA | Backbone + Normals | Backnoe + CA | Backbone |

$$Sematic\ Segmentation\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad 6\text{-}1$$

Besides ablation studies of the proposed network, this study tested the influence of 2D image resolution on segmentation results. The Kumamoto pre-earthquake dataset was applied for this test. This was because this dataset has two optical images with different resolutions.

### 6.4.6 Comparison on public dataset Semantic3D

To detect the performance of the proposed network, these four networks were trained and tested on the public dataset Semantic3D (Hackel et al., 2017). Only coordinates and RGB information with eight labelled classes from the dataset were used to train and test different methods. Some well-known and state-of-the-art networks were also tested for comparison, including PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), and ShellNet (Zhang et al., 2019b). The tested point clouds were chosen according to the selected test datasets provided by (Hu et al., 2020), which include four point clouds.

## 6.5 Results

This section presents the semantic segmentation results of the ablation studies, which contain results of the four networks using five metrics with the test data.

As mentioned above, this study set four datasets as test data. Five classes were tested, including buildings, ground, low vegetation, medium vegetation, and high vegetation. Visualisation results and quantitative results are stated in this section. Five point clouds were tested. The first two tested point clouds are from Kapiti Coast. The third is from the Tasman dataset, and the last two are in the Nelson dataset. These two datasets' visualisation results and quantitative results are stated in this section.

### 6.5.1 Hardware and environment

In this study, one Nvidia RTX 2080Ti GPU card, CUDA 11.3, Python 3.6, and TensorFlow 1.15 were applied. Since TensorFlow versions 1 and 2 have huge differences, it would be more convenient to use TensorFlow version 1 to fit its version in the original RandLA-Net

backbone. Usually, CUDA 11.0 or higher only support TensorFlow 2.X version, according to its official document (Tensorflow, 2022). It means TensorFlow 1.X version usually can only run with CPU but not GPU if CUDA is 11.0 version or higher in a computer. Hence, this chapter added some commands during the environment configuration step to enable CUDA 11.3 to the run TensorFlow 1.15 GPU version. “Batch size during training” is 2, and “Number of steps per epoch” is 1,000. It took 8 hours to run in the GPU version with 100 epochs.

### 6.5.2 Results of 2021 New Zealand datasets

Five patches of point clouds were chosen as the test data from the three New Zealand datasets. Their visualisation results are shown in Figure 6-11. Red, blue, dark green, bright green, and orange represent buildings, ground, low, medium, and high vegetation, respectively. Based on the visual observation, compared with the ground truths of the point clouds, most buildings were recognized correctly, but most medium and high vegetation points were mistakenly recognized as ground and low vegetation points.

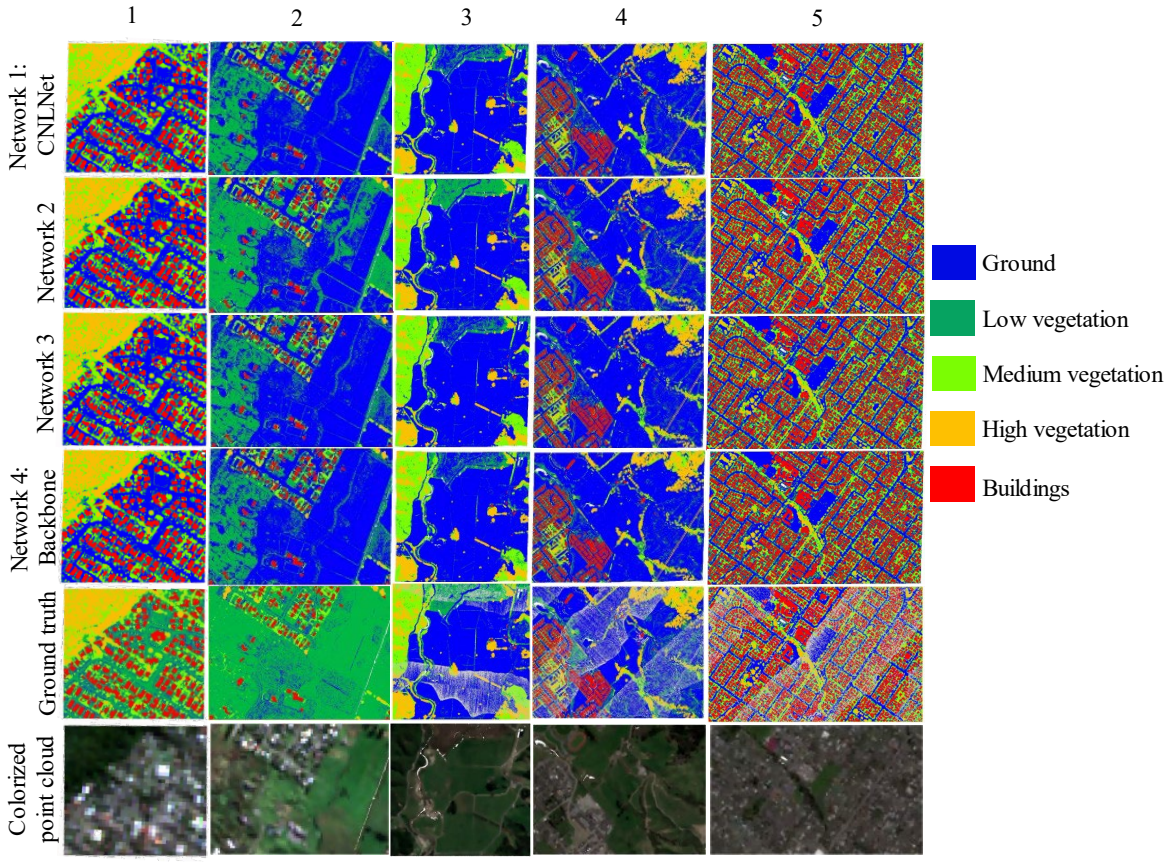



Figure 6-11. Visualisation results of ablation study for Kapiti Coast, Tasman, and Nelson datasets

Results for each class with the five metrics are shown in Table 6-4. The first class is the bare-ground class. The proposed Network 1 performed best for the ground segmentation according to IoU results. Networks 2 and 3 are nearly the same as the backbone.

The following three classes are the three vegetation classes, including low, medium, and high. Medium vegetation segmentation results performed best among these three classes in all networks according to their IoUs. In all low vegetation results, Network 4 performed best among the four networks with the highest IoUs. The highest IoUs of medium and high

vegetation results were also the results of Network 1. IoUs of low vegetation are always the lowest among the three vegetation classes.

The fifth class is the building class. Network 1 always performs best among all networks, but Networks 2 and 3 do not show any significant advantage over the backbone of Network 4. IoUs of the tested point clouds from the Kapiti Coast dataset in CNLNet are very high, which are 0.95 and 0.92, respectively. They are the top two highest among all tested point clouds with the top two highest TP values. A probable reason for this is that the number of point clouds belonging to the building class in the Kapiti Coast dataset accounts for a significant proportion of the total number of building points.

Table 6-4. Results of New Zealand datasets

| Class | Metric | Networks | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Network 1 CNLNet: Backbone + Normals + CA** | | | | | **Network 2: Backbone + Normals** | | | | | **Network 3: Backbone + CA** | | | | | **Network 4: Backbone** | | | | |
| | | '3_rgb' | '26_rgb' | Nelson_test1_1_rgb' | Nelson_test2_1_rgb' | Nelson_test2_2_rgb' | '3_rgb' | '26_rgb' | Nelson_test1_1_rgb' | Nelson_test2_1_rgb' | Nelson_test2_2_rgb' | '3_rgb' | '26_rgb' | Nelson_test1_1_rgb' | Nelson_test2_1_rgb' | Nelson_test2_2_rgb' | '3_rgb' | '26_rgb' | Nelson_test1_1_rgb' | Nelson_test2_1_rgb' | Nelson_test2_2_rgb' |
| **Number of total points** | | 3,738,726 | 7,816,274 | 9,783,954 | 14,060,465 | 12122253 | 3,738,726 | 7,816,274 | 9,783,954 | 14,060,465 | 12122253 | 3,738,726 | 7,816,274 | 9,783,954 | 14,060,465 | 12,122,253 | 3,738,726 | 7,816,274 | 9,783,954 | 14,060,465 | 12,122,253 |
| **Ground** | **TP** | 631,672 | 1,957,185 | 4,576,483 | 7,034,383 | 4,372,163 | 675,525 | 2,229,488 | 4,693,481 | 7,161,717 | 4,427,799 | 674,589 | 2,233,321 | 4,711,576 | 7,134,927 | 4,421,202 | 672,691 | 2,204,359 | 4,652,074 | 7,105,862 | 4,415,843 |
| | **FN** | 45,291 | 309,047 | 162,364 | 161,393 | 61,466 | 1,438 | 36,744 | 45,366 | 34,059 | 5,830 | 2,374 | 32,911 | 27,271 | 60,849 | 12,427 | 4,272 | 61,873 | 86,773 | 89,914 | 17,786 |
| | **FP** | 626,543 | 2,509,943 | 1,398,597 | 3,341,092 | 1,923,477 | 767,898 | 3,687,628 | 1,703,521 | 3,735,955 | 2,243,770 | 780,273 | 3,715,702 | 1,773,298 | 3,727,740 | 2,383,826 | 747,679 | 3,532,053 | 1,589,890 | 3,551,675 | 2,130,646 |
| | **IoU** | 0.48 | 0.41 | 0.75 | 0.67 | 0.69 | 0.47 | 0.37 | 0.73 | 0.66 | 0.66 | 0.46 | 0.37 | 0.72 | 0.65 | 0.65 | 0.47 | 0.38 | 0.74 | 0.66 | 0.67 |
| | **SSA** | 0.82 | 0.64 | 0.84 | 0.75 | 0.84 | 0.79 | 0.52 | 0.82 | 0.73 | 0.81 | 0.79 | 0.52 | 0.82 | 0.73 | 0.80 | 0.80 | 0.54 | 0.83 | 0.74 | 0.82 |
| **Low Vegetation** | **TP** | 385,217 | 2,384,564 | 584,640 | 620,595 | 721,492 | 267,591 | 1,231,741 | 416,570 | 427,263 | 563,431 | 264,609 | 1,196,848 | 403,186 | 444,091 | 553,922 | 270,415 | 1,372,237 | 494,279 | 534,275 | 637,690 |
| | **FN** | 643,247 | 2,505,128 | 206,667 | 456,679 | 282,411 | 760,873 | 3,657,951 | 374,737 | 650,011 | 440,472 | 763,855 | 3,692,844 | 388,121 | 633,183 | 449,981 | 758,049 | 3,517,455 | 297,028 | 542,999 | 366,213 |
| | **FP** | 82,336 | 364,362 | 517,203 | 615,262 | 792,197 | 109,964 | 96,265 | 278,401 | 366,119 | 802,561 | 52,625 | 141,177 | 335,287 | 366,933 | 665,901 | 93,338 | 144,401 | 370,356 | 484,053 | 868,455 |
| | **IoU** | 0.35 | 0.45 | 0.45 | 0.37 | 0.40 | 0.24 | 0.25 | 0.39 | 0.30 | 0.31 | 0.24 | 0.24 | 0.36 | 0.31 | 0.33 | 0.24 | 0.27 | 0.43 | 0.34 | 0.34 |
| | **SSA** | 0.81 | 0.63 | 0.93 | 0.92 | 0.91 | 0.77 | 0.52 | 0.93 | 0.93 | 0.90 | 0.78 | 0.51 | 0.93 | 0.93 | 0.91 | 0.77 | 0.53 | 0.93 | 0.93 | 0.90 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Medium Vegetation** | **TP** | 840,963 | 253,245 | 1,329,536 | 591,701 | 844,735 | 779,816 | 250,821 | 1,294,081 | 566,743 | 774,112 | 834,458 | 206,580 | 1,239,430 | 571,320 | 868,270 | 828,162 | 242,181 | 1,308,591 | 582,199 | 821,932 |
| | **FN** | 50,463 | 53,781 | 44,845 | 49,618 | 154,567 | 111,610 | 56,205 | 80,300 | 74,576 | 225,190 | 56,968 | 100,446 | 134,951 | 69,999 | 131,032 | 63,264 | 64,845 | 65,790 | 59,120 | 177,370 |
| | **FP** | 123,782 | 34,774 | 566,422 | 297,865 | 443,536 | 104,033 | 54,422 | 622,960 | 331,912 | 413,835 | 187,389 | 28,781 | 559,004 | 333,582 | 606,830 | 280,514 | 54,351 | 713,787 | 413,099 | 506,586 |
| | **IoU** | 0.83 | 0.74 | 0.69 | 0.63 | 0.59 | 0.78 | 0.69 | 0.65 | 0.58 | 0.55 | 0.77 | 0.62 | 0.64 | 0.59 | 0.54 | 0.71 | 0.67 | 0.63 | 0.55 | 0.55 |
| | **SSA** | 0.95 | 0.99 | 0.94 | 0.98 | 0.95 | 0.94 | 0.99 | 0.93 | 0.97 | 0.95 | 0.93 | 0.98 | 0.93 | 0.97 | 0.94 | 0.91 | 0.98 | 0.92 | 0.97 | 0.94 |
| **High Vegetation** | **TP** | 453,361 | 37,705 | 622,549 | 664,871 | 155,707 | 466,121 | 40,165 | 587,491 | 620,639 | 136,285 | 439,268 | 38,216 | 587,213 | 644,548 | 144,710 | 338,142 | 31,828 | 508,139 | 572,013 | 113,831 |
| | **FN** | 98,512 | 26,966 | 129,679 | 47,567 | 38,740 | 85,752 | 24,506 | 164,737 | 91,799 | 58,162 | 112,605 | 26,455 | 165,015 | 67,890 | 49,737 | 213,731 | 32,843 | 244,089 | 140,425 | 80,616 |
| | **FP** | 3,168 | 341 | 177,680 | 175,772 | 57,273 | 9,270 | 141 | 179,266 | 148,026 | 47,360 | 11,515 | 14 | 165,497 | 174,593 | 55,023 | 2,436 | 0 | 137,929 | 129,700 | 36,963 |
| | **IoU** | 0.82 | 0.58 | 0.67 | 0.75 | 0.62 | 0.83 | 0.62 | 0.63 | 0.72 | 0.56 | 0.78 | 0.59 | 0.64 | 0.73 | 0.58 | 0.61 | 0.49 | 0.57 | 0.68 | 0.49 |
| | **SSA** | 0.97 | 1.00 | 0.97 | 0.98 | 0.99 | 0.97 | 1.00 | 0.96 | 0.98 | 0.99 | 0.97 | 1.00 | 0.97 | 0.98 | 0.99 | 0.94 | 1.00 | 0.96 | 0.98 | 0.99 |
| **Buildings** | **TP** | 573,653 | 264,277 | 8,525 | 578,421 | 1,946,281 | 547,310 | 222,536 | 7,690 | 574,128 | 1,913,226 | 488,448 | 250,183 | 8,318 | 546,717 | 1,755,103 | 498,582 | 231,180 | 8,047 | 565,947 | 1,855,662 |
| | **FN** | 13,998 | 13,216 | 373 | 13,550 | 41,776 | 40,341 | 54,957 | 1,208 | 17,843 | 74,831 | 99,203 | 27,310 | 580 | 45,254 | 232,954 | 89,069 | 46,313 | 851 | 26,024 | 132,395 |
| | **FP** | 18,031 | 9,878 | 2,319 | 140,500 | 865,385 | 11,198 | 3,067 | 493 | 127,817 | 799,862 | 5,552 | 5,452 | 1,145 | 115,923 | 667,412 | 6,767 | 3,684 | 862 | 121,509 | 734,645 |
| | **IoU** | 0.95 | 0.92 | 0.76 | 0.79 | 0.68 | 0.91 | 0.79 | 0.82 | 0.80 | 0.69 | 0.82 | 0.88 | 0.83 | 0.77 | 0.66 | 0.84 | 0.82 | 0.82 | 0.79 | 0.68 |
| | **SSA** | 0.99 | 1.00 | 1.00 | 0.99 | 0.93 | 0.99 | 0.99 | 1.00 | 0.99 | 0.93 | 0.97 | 1.00 | 1.00 | 0.99 | 0.93 | 0.97 | 0.99 | 1.00 | 0.99 | 0.93 |
| **Mean IoU** | | 0.68 | 0.62 | 0.66 | 0.64 | 0.60 | 0.65 | 0.55 | 0.64 | 0.61 | 0.55 | 0.62 | 0.54 | 0.64 | 0.61 | 0.55 | 0.57 | 0.53 | 0.64 | 0.61 | 0.55 |

After comparing the performance of each network for every class, the result differences between the five classes should also be mentioned. Compared with the other classes, the building class always has the highest IoU among the five classes in the results of each network, which most are higher than 0.90 in some test results. It is convinced that the RandLA-Net backbone is suitable for building detection. Segmentation of ground and low vegetation performed worst in results according to IoUs. The likely reason is data imbalance. The numbers of Lidar points in these classes are lower than those of others.

### 6.5.3 Results of 2016 Kumamoto pre-earthquake data

The 2016 Kumamoto pre-earthquake point cloud dataset with both high and low resolutions of optical satellite images was tested. The number of total points in this point cloud is 1,438,042. As mentioned in Table 6-2, the resolution of the high-resolution image is 0.5m/pxl, and that of the low-resolution image is 10m/pxl. Figure 6-12 shows their visualisation results. Five classes were segmented. Red, blue, light green, bright green, and orange represent buildings, ground, low vegetation, medium vegetation, and high vegetation, respectively. It can be easily found that most high vegetation points were mistakenly segmented as other classes, such as low and medium vegetation.

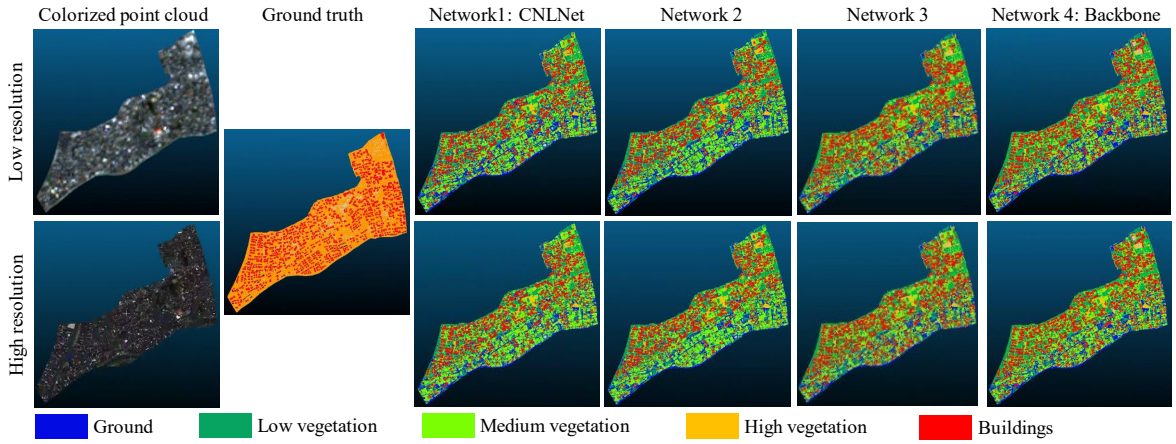



Figure 6-12. Visualisation results of ablation study for 2016 Kumamoto pre-earthquake dataset

Quantitative results of the 2016 Kumamoto pre-earthquake data are shown in Table 6-5. The results of the five classes are listed in it.

Table 6-5. Results of 2016 Kumamoto pre-earthquake dataset

| Class | Metric | Networks | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Network 1 CNLNet: Backbone + Normals + CA | | Network 2: Backbone + Normals | | Network 3: Backbone + CA | | Network 4: Backbone | |
| RGB image resolution (m/pxl) | | 0.5 | 10 | 0.5 | 10 | 0.5 | 10 | 0.5 | 10 |
| Ground | TP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | FN | 740 | 740 | 740 | 740 | 740 | 740 | 740 | 740 |
| | FP | 526,767 | 526,499 | 496,807 | 496,530 | 446,119 | 445,199 | 440,101 | 441,066 |
| | IoU | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | SSA | 0.63 | 0.63 | 0.65 | 0.65 | 0.69 | 0.69 | 0.69 | 0.69 |
| Low Vegetation | TP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **FN** | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| | **FP** | 319,168 | 319,951 | 356,297 | 356,347 | 377,855 | 378,374 | 383,788 | 383,946 |
| | **IoU** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | **SSA** | 0.78 | 0.78 | 0.75 | 0.75 | 0.74 | 0.74 | 0.73 | 0.73 |
| **Medium Vegetation** | **TP** | 3,924 | 3,830 | 3,732 | 3,702 | 3,935 | 3,962 | 4,133 | 4,083 |
| | **FN** | 51,581 | 51,675 | 51,773 | 51,803 | 51,570 | 51,543 | 51,372 | 51,422 |
| | **FP** | 269,520 | 268,099 | 260,032 | 259,495 | 237,357 | 237,781 | 248,369 | 247,419 |
| | **IoU** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | **SSA** | 0.78 | 0.78 | 0.78 | 0.78 | 0.80 | 0.80 | 0.79 | 0.79 |
| **High Vegetation** | **TP** | 10,836 | 10,686 | 13,183 | 12,673 | 11,603 | 11,725 | 8,588 | 8,513 |
| | **FN** | 988,345 | 988,495 | 985,998 | 986,508 | 987,578 | 987,456 | 990,593 | 990,668 |
| | **FP** | 5,317 | 5,167 | 6,262 | 6,288 | 5,385 | 5,445 | 5,107 | 4,733 |
| | **IoU** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | **SSA** | 0.31 | 0.31 | 0.31 | 0.31 | 0.31 | 0.31 | 0.31 | 0.31 |
| **Building** | **TP** | 194,893 | 195,114 | 185,324 | 185,713 | 221,889 | 221,844 | 214,358 | 214,034 |
| | **FN** | 134,076 | 133,855 | 143,645 | 143,256 | 107,080 | 107,125 | 114,611 | 114,935 |
| | **FP** | 106,641 | 107,774 | 114,300 | 115,356 | 131,326 | 131,207 | 127,371 | 128,054 |
| | **IoU** | 0.45 | 0.45 | 0.42 | 0.42 | 0.48 | 0.48 | 0.47 | 0.47 |
| | **SSA** | 0.83 | 0.83 | 0.82 | 0.82 | 0.83 | 0.83 | 0.83 | 0.83 |

The first test class is the ground class. All networks performed not so well for this class no matter with high or low resolution. The number of FP points is too high, no matter which

network. The second class is the low vegetation class. Similar to segmentation results for the ground class, IoUs were nearly zero for all networks. Thousands of points were detected as low vegetation wrongly. In other words, the FP values of these two classes are high. Moreover, nearly no TP points have been detected, as shown in the results of the ground and the low vegetation classes. The first probable reason is that the information difference of segmentation labels between the training data and these test data is large, which are from different datasets. The second possible reason is that the points from those two classes are too few to be detected in these test data.

The next two classes are medium and high vegetation. Although their IoUs were also nearly zero, the number of their TP points was much higher than those of ground and low vegetation. Besides IoUs, SSA results for medium vegetation were higher than those for high vegetation, while the numbers of FP points in medium vegetation results were higher than those in high vegetation for all these three networks.

The last class is the building class. The highest IoUs were the building class results among all five detected classes in all networks. This might be because the number of points labelled as buildings is high in the training dataset. According to IoUs, Network 3 performed the best, which shows its generalisability for building segmentation is the best of these networks.

Among all segmented classes, the generalisabilities of all tested networks in the ablation study are not ideal except for the building class. There are some possible reasons. Although the two datasets both have these five labelled classes and colours, the labelled information of the classes in the 2016 Kumamoto pre-earthquake dataset is much different from those of

the New Zealand datasets. As mentioned in Section 6.4, only New Zealand datasets are applied for training due to the small area of 2016 Kumamoto pre-earthquake data.

### 6.5.4 Results of Semantic3D

The results are listed in Table 6-6. Network 3 has the highest mIoU of the tested eight classes. Networks 1-4 all achieve acceptable results compared with the other networks.

However, Network 1 performed slightly worse than Networks 2 and 3 after adding both normal information and CA in the backbone, though it performed best in some class results. The probable reason is that Network 1 was overfitted. Overfitting might exist if the network is too complicated. In total, the results demonstrate that both surface normal information and channel attention have helped with large-scale outdoor point cloud semantic segmentation based on the RandLA-Net backbone. Each of them can improve mIoU by 1% to 2% than the backbone. This might be due to its overly complicated structure. The network's performance with adding both CA and surface information (Network 1) is not as good as the network with only adding one.

Table 6-6. Results of different methods on Semantic3D

| **Network** | **mIoU** |
|---|---|
| Network 1: Backbone + Normals + CA | 0.67 |
| Network 2: Backbone + Normals | 0.68 |
| Network 3: Backbone + CA | **0.70** |
| Network 4: Backbone | 0.67 |
| ShellNet (Zhang et al., 2019b) | 0.63 |
| PointNet++ (Qi et al., 2017b) | 0.42 |
| PointNet (Qi et al., 2017a) | 0.41 |

## 6.6 Discussion of 3D semantic segmentation

This study designed ablation studies to demonstrate the benefits of adding surface normal information and channel attention mechanism in LOLSS for predisaster information classification and storage.

IoUs of the building class were always the highest in the results in all own labelled datasets among all tested networks. This reflects that these networks are all suitable for segmenting buildings. Besides that, in the test of the Kumamoto dataset, the building segmentation IoUs were significantly higher than the results of other classes. The training and validation steps did not contain Kumamoto data. Hence, the generalisability of the trained network for building segmentation is the highest. Moreover, the results for the Kumamoto point clouds with different resolutions of optical satellite images were very similar. It can be concluded that the optical satellite image resolutions may have little influence on the performance of the proposed model.

In addition to the analysis of the IoU of each class, the overall IoU of all classes should be discussed. As mentioned in Section 6.4.5, the mIoU of each network was calculated to analyse its performance considering the results of all classes. The mIoU results for all classes in the five tested point clouds are shown in Table 6-7. It should be noted that mIoUs of Kumamoto data are not discussed because IoU values of the other four classes are nearly zero except the IoU of the building class due to the poor generalisabilities of these four classes. Table 6-7 shows that mIoU values of Network 1 are always the highest in these networks for all tested point clouds. Moreover, the mIoUs of Network 2 are higher than

those of Network 3, so it shows that adding normal information might be more helpful than CA to semantic segmentation.

Table 6-7. Mean IoU of the New Zealand test datasets

| No. | mIoU | | | |
|---|---|---|---|---|
| | **Network 1: Backbone + Normals + CA** | **Network 2: Backbone + Normals** | **Network 3: Backbone + CA** | **Network 4: Backbone** |
| 1 | **0.68** | 0.65 | 0.62 | 0.57 |
| 2 | **0.62** | 0.55 | 0.54 | 0.53 |
| 3 | **0.66** | 0.64 | 0.64 | 0.64 |
| 4 | **0.64** | 0.61 | 0.61 | 0.61 |

Other metrics also demonstrated that the designed network is suitable for segmenting buildings from the background. The TPs of the building class in Table 6-4 and Table 6-5 are very high. The SSA of the building class in Table 6-4 is nearly 1, and its SSA in Table 6-5 is the highest among SSAs of all classes. The results for Semnatic3D also demonstrated that the designed network is suitable for predisaster land cover object segmentation from the background.

Based on the abovementioned discussion, it can be concluded that surface normal information and channel attention can improve segmentation accuracy. The proposed CNLNet can improve mIoU by 1% to 11% compared to the backbone in different scenarios. Besides that, in contrast to the RandLA-Net backbone and other well-known networks, each of these two types of feature information (Networks 2 and 3) can help to improve the

accuracy of semantic segmentation. The network with only adding surface information (Network 2) is more effective between these two types of networks.

## 6.7 Conclusion

In this study, a network named CNLNet was proposed to enhance the precision of DL-based LOLSS for predisaster land cover information segmentation and preservation. Surface normals and CA were added to this network. A labelled large-scale land cover Lidar dataset was first created in this study considering potential natural disaster occurrences in selected datasets' places, including Kapiti Coast, Tasman, and Nelson in New Zealand and Kumamoto in Japan. Optical satellite images were integrated as inputs. Compared with the state-of-the-art RandLA-Net backbone and other renowned networks, the findings demonstrate the benefits of surface normal information and CA applied to LOLSS. Normal information can provide more feature information, and CA can emphasize key information in channels, so they can improve the accuracy of segmentation results. Furthermore, the proposed network exhibits the strongest generalisability for the building class. Interestingly, the network that incorporated either surface normals or CA alone slightly outperformed the one incorporating both during the test on the open-source Semantic3D dataset. The likely reason is that overfitting might occur if a network is too complex. With the potential to save labour and mitigate in-situ risks, the practical implication of this method lies in its applicability for urban land cover segmentation from 3D Lidar point clouds, particularly for building segmentation. The outcomes can be utilized for predisaster urban visualisation data information storage and update, thereby expediting post-disaster emergency response efforts. Further research is suggested to find an approach

to improve the segmentation accuracy of separating classes other than buildings, such as low, medium, and high vegetation. The setting of the direction of surface normals could also be discussed in future studies.

# Chapter 7

# Large-scale post-earthquake building damage level classification using colourised Lidar data[4]

## 7.1 Chapter introduction

Extreme natural disasters can have devastating consequences, resulting in significant loss of life and widespread destruction. For instance, as a severe earthquake that caused the most significant number of deaths in the last 15 years, the 2010 Haiti Earthquake resulted in an official death toll of about 230,000. Nearly half of all structures collapsed or were severely damaged in the epicentral area in this Haiti earthquake, including more than 300,000 homes (Desroches et al., 2011).

In the post-earthquake stage, the primary objective is to rescue individuals and safeguard properties. To save lives, rapid post-earthquake emergency response is necessary. The initial step of post-earthquake emergency response invariably involves the collection and analysis of disaster information. However, it is often hard for rescue teams to decide where

[4] The content presented in this chapter is partially adopted from the following submitted paper: "Rapid Large-scale Building Damage Level Classification after Earthquakes using Deep Learning with Lidar and Satellite Optical Data." It has been acknowledged and detailed in the "Inclusion of Publications Statement" for this thesis.

to begin the rescue operation first due to the dearth of prompt building damage information immediately following an earthquake. The lack of this information is caused by the difficulty of rapidly judging the levels of building damage due to the differences in structure and lack of rapid methods. Moreover, the search and rescue resources of the stricken areas are usually not sufficient in the first several hours. Most BDLC methods request in-situ observations, which are time-consuming, labour-intensive, and sometimes dangerous. Those detailed in-situ classifications are not suitable for rapid disaster rescue planning. Consequently, the need arises for a fast, reliable, and efficient approach to classify building damage levels, aimed at rapidly identifying the most critical areas requiring rescue efforts and facilitating a prompt post-earthquake disaster response.

To address these limitations, remote sensing was applied recently to assess building damage levels with the integration of multifarious advanced techniques for a rapid response. In the beginning, most remote sensing studies utilized elevation difference and texture difference to classify damage levels based on change detection. Recently, DL methods have been gradually applied to detect post-disaster information. Moreover, with the fast development of Lidar, Lidar data are widely applied in the disaster response field for providing building height information. Therefore, DL methods using Lidar data could be a rapid approach to quick BDLC. However, the limited availability of publicly available post-earthquake Lidar datasets is still an issue for training DL networks. This chapter aims to reveal the potential and limitations of Lidar-based DL approaches for BDLC by developing a DL network with an in-house labelled dataset. A disaster response system is also

proposed according to the information of detailed building damage levels obtained from the proposed method.

## 7.2 Assessing building damage with deep learning and remote sensing

In recent years, remote sensing has increasingly played a crucial role in BDLC, particularly in post-disaster scenarios such as earthquakes. DL-based automatic methods for post-earthquake BDLC can always be categorised into three types, including image-based techniques, Lidar-based techniques, and data fusion methods. Initially, owing to the high spatial resolution of available optical images, studies were interested in image-based techniques for post-event damage estimation. For instance, Kalantar et al. (2020) assessed the adaption of CNN for building damage detection based on pre- and post-earthquake orthophoto images. They categorised damage into four levels, including background, no damage, minor damage, and debris. Zhan et al. (2022) proposed a modified Mask R-CNN model to estimate building damage levels from high-resolution post-disaster aerial images. The case study is Mashiki Town, Kumamoto Prefecture, after the 2016 Kumamoto Earthquake. Wang et al. (2023) proposed a CNN-based seismic building damage level classification method. The experiment data were HRAIs collected in Beichuan town, Sichuan, after the 2008 Wenchuan Earthquake. Damaged buildings were categorised into three levels including destroyed, severely damaged, and others. These applications show that advancements in DL showed promising results in classifying building damage levels using optical images. However, most image-based methods utilised only one resolution input, and their information may not find slight building damages (Cotrufo et al., 2018), so multi-source data methods are suggested to be introduced to improve the accuracy.

With the fast development of AI, recent studies proposed various AI methods and architectures to assess building damages using Lidar point cloud data. For instance, Khodaverdi et al. (2019) developed building damage detection using Lidar for height information and high-resolution satellite images (HRSI) by comparing the difference between pre- and post-earthquake with the supervised KNN. An area in Port-au-Prince after the 2010 7.0 $M_W$ Haiti Earthquake was tested. Three levels were categorised, including surely damaged, probably damaged, and undamaged. Eslamizade et al. (2021) proposed an SVM-based method to generate the building damage map using both pre- and post-earthquake HRSI and Lidar data. The case study is also the 2010 Haiti Earthquake. Damaged buildings were categorised into four levels, including low damage, moderate damage, heavy damage, and destructed. While these 3D point cloud-based methods have made significant advancements, they mainly focused on machine learning methods. As far as we know from a thorough examination of recent literature, there appears to be a scarcity of studies focusing on applying DL for post-earthquake BDLC using Lidar point clouds (Xiu et al., 2020). Moreover, related recent studies still always chose the same earthquake that happened more than ten years ago as the case study, which is the 2010 Haiti Earthquake.

Considering the above knowledge gaps, this study aims to reveal the performance of the DL network applied in BDLC with Lidar data. A DL method is developed, and in-house labelled datasets are created for a case study where the earthquake happened later, other than the 2010 Haiti Earthquake, which is an earthquake that happened in Kumamoto, Japan,

in 2016. The building damage is categorised into four levels, from no damage to story failure.

## 7.3 Data and pre-processing

The mainshock of a 7.0 $M_W$ earthquake happened in Kumamoto, Japan, on 16/04/2016 (Asia Air Survey Co., 2018). This earthquake was chosen as the study area in this study. There were five types of sources applied in this study to build one colourised point cloud dataset, including pre-earthquake building footprint vector shapefiles, pre- and post-earthquake HRSI, post-earthquake Lidar point clouds, and post-earthquake building damage level geo-location files. The data collection dates are listed in Table 7-1. The available data closest to the date before the earthquake occurrence was chosen as the predisaster data for the building footprint shapefile and HRSI. After the pre-processing, building damage areas were extracted from both post-earthquake data and the pre-earthquake building footprint vector map.

Table 7-1. Data collection date of each source

| Source | Data collection date |
|---|---|
| Pre-earthquake building footprint shapefile | 01/04/2016 |
| Pre-earthquake HRSI | 15/04/2016 |
| Post-earthquake building damage level | 08/09/2016 (No detailed collection date, only published date) |
| Post-earthquake point clouds | 23/04/2016 |
| Post-earthquake HRSI | 20/04/2016 |

The workflow of adding these different types of information into the post-earthquake point cloud data is shown in Figure 7-1. Source 1 includes published pre-earthquake building footprint shapefiles from the Geospatial Information Authority of Japan (GSI). The shapes of the building footprints that they provided are slightly different. Because of that, firstly, this research labelled the pre-earthquake non-damage-level building footprints of the selected area according to Source 1 with reference to the building outlines in Source 2 - pre-earthquake HRSI. The image of Source 2 is shown in Figure 7-2 (a), which was collected from KOMPSAT-3 (K3) on 15/04/2016.

Secondly, the coordinate system of the building footprints was reprojected to the same system as Source 2 – post-earthquake non-damage-level point clouds. Thus, the building footprint information was added to the point clouds as a labelled class.

Thirdly, post-earthquake building damage level information was added to those non-damage-level building footprints according to Source 3 – building damage level information shown in Table 7-2 provided by Yamada et al. (2017). Yamada et al. (2017) provided detailed wooden building damage levels from in-situ observations and aerial photo analysis. They classified the building damage into four levels, including D0 (no/minor damage), D1-D3 (partially collapsed), D4 (totally collapsed), and D5 (story failure). The total number of buildings is 1,041. Descriptions of damage levels and the number of buildings on each level are listed in Table 7-2. The labelled building footprints with damage levels were achieved after this step, as shown in Figure 7-3.

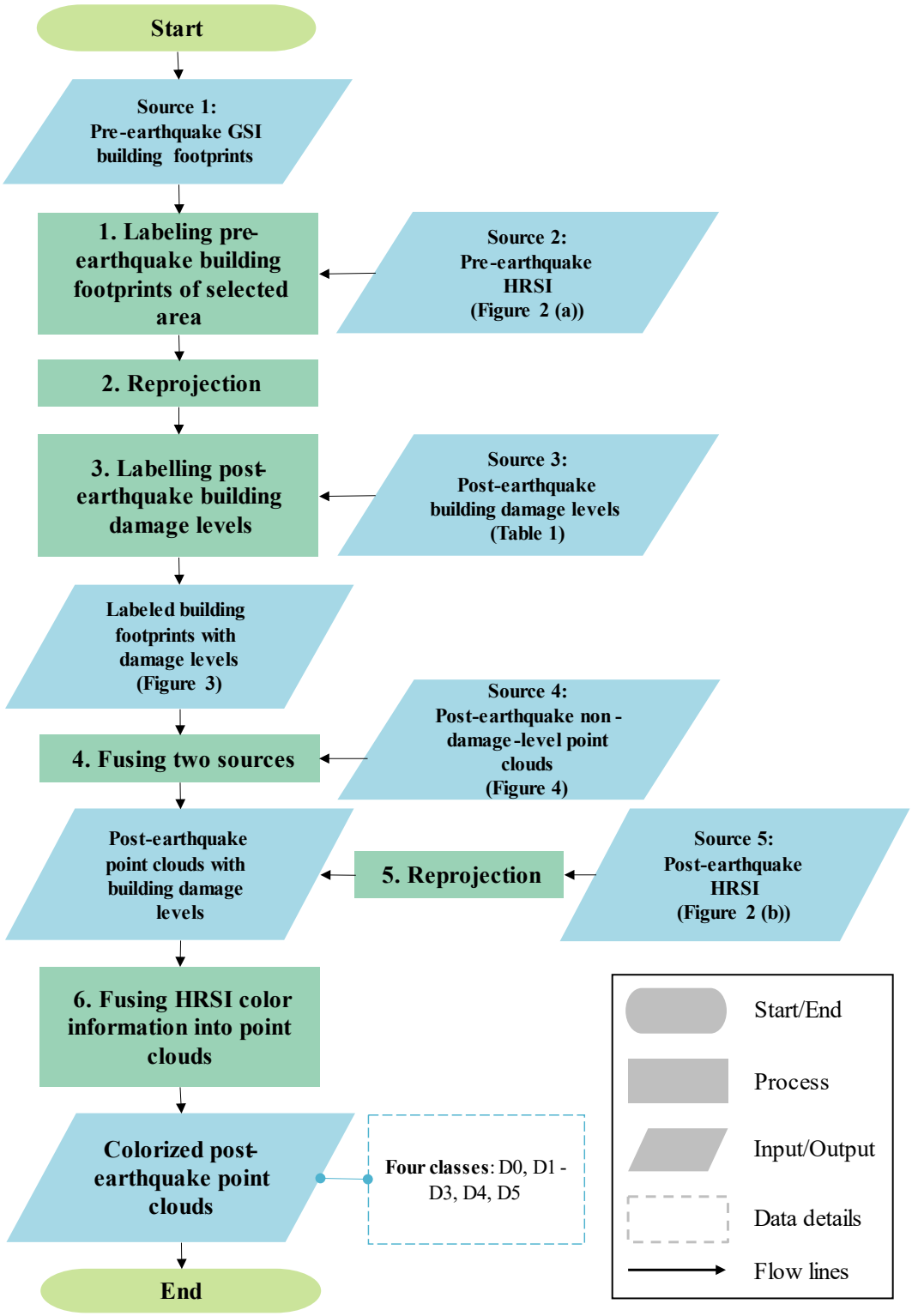
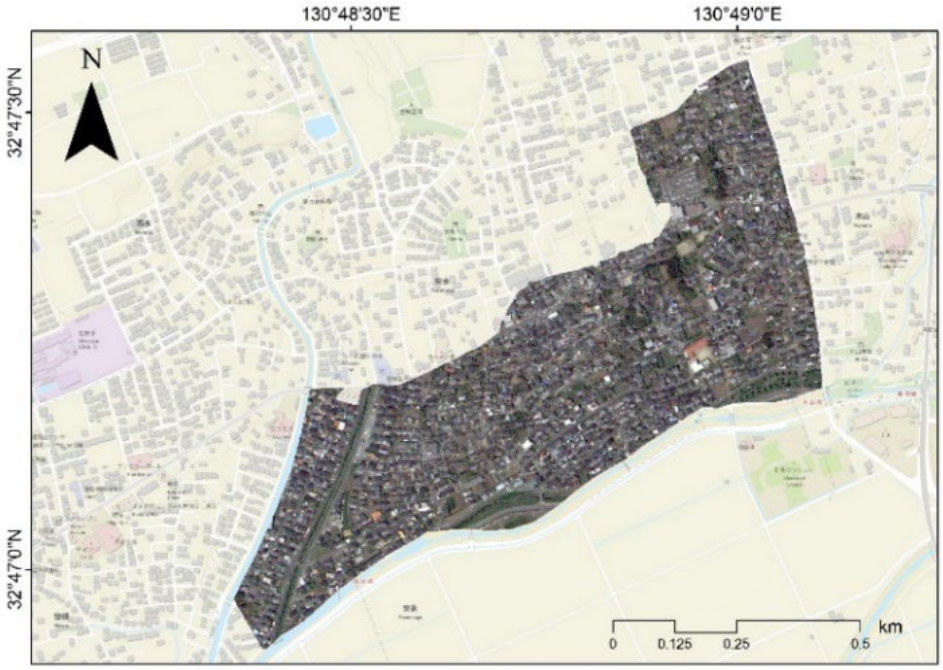
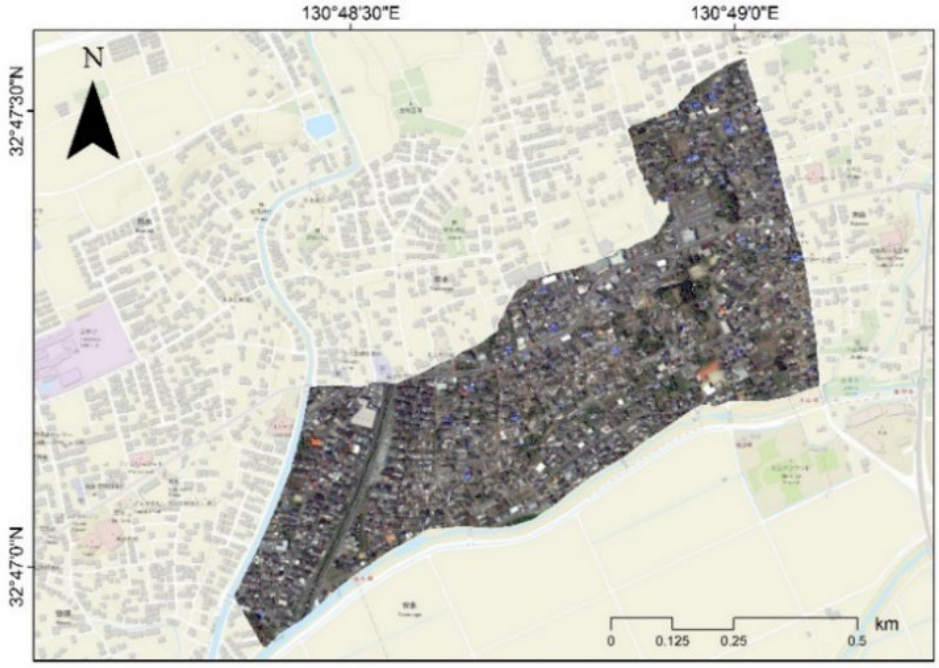



Figure 7-1. Workflow of data fusion for pre-processing




(a) Pre-earthquake optical image




(b) Post-earthquake optical image

Figure 7-2. Pre- and post-earthquake K3 images

Table 7-2. Number of buildings in each damage level

| Damage level | D0 | D1-D3 | D4 | D5 |
|---|---|---|---|---|
| **Description** | No/minor damage | Partially collapsed | Totally collapsed | Story failure |
| **Number of buildings** | 371 | 231 | 158 | 281 |

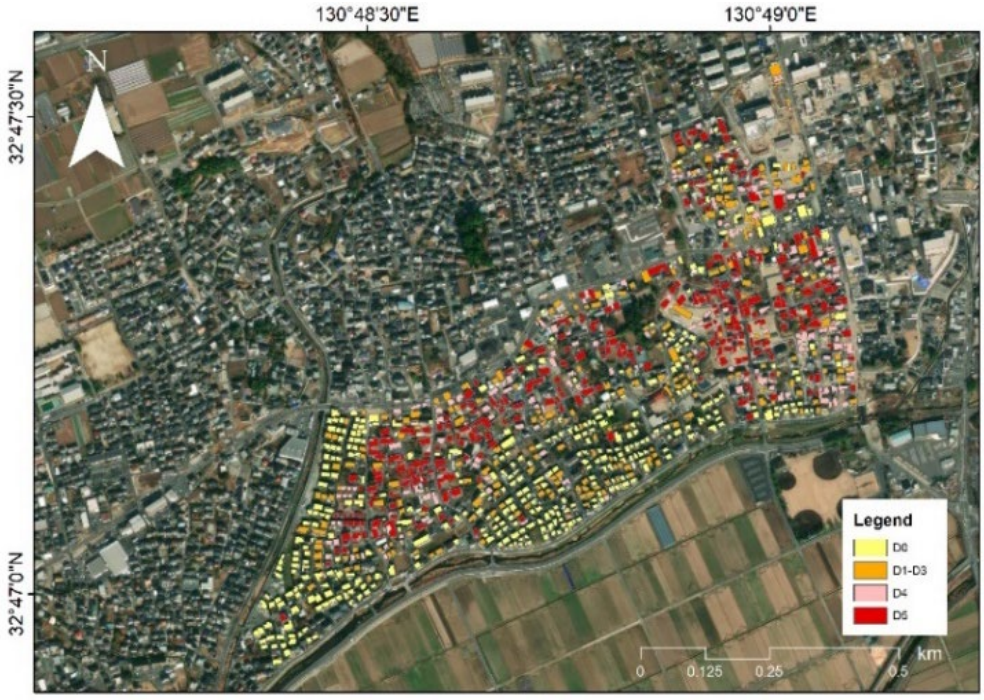



Figure 7-3. Building damage level based on in-situ observation for the 2016 Kumamoto Earthquake

Fourthly, the labelled building footprints shown in Figure 7-1 were fused into Source 4 - post-earthquake non-damage-level point clouds. The result of the fusion was the post-earthquake point clouds with building damage levels. Source 3 was clipped from the post-earthquake airborne Lidar point cloud dataset published by Asia Air Survey Co. (2018) for this earthquake with a point density of 4.47 points/m$^2$. The selected area is shown in Figure

7-4. This original dataset only contains labels of ground and three types of vegetation (low/medium/high) without any colour information.

In order to fuse colour information with the Lidar data, post-earthquake HRSI (Source 5) was applied. This post-earthquake HRSI was collected from K3 on 20/04/2016, with only RGB bands utilized in this research, as shown in Figure 7-2 (b). Before the fusion, it is necessary to reproject the coordinate system of Source 5 to the same system of the point clouds achieved from the fourth step, which is the fifth step.

After the reprojection, colours with RGB bands from the reprojected post-earthquake HRSI were fused with the post-earthquake point clouds processed from the fourth step, which was the sixth step.

The final data pre-processing result was a colourised post-earthquake point cloud dataset labelled four building damage levels. It should be noted that only the labels of the building class were kept in this study, because other classes labelled in the original point clouds were not the research target.

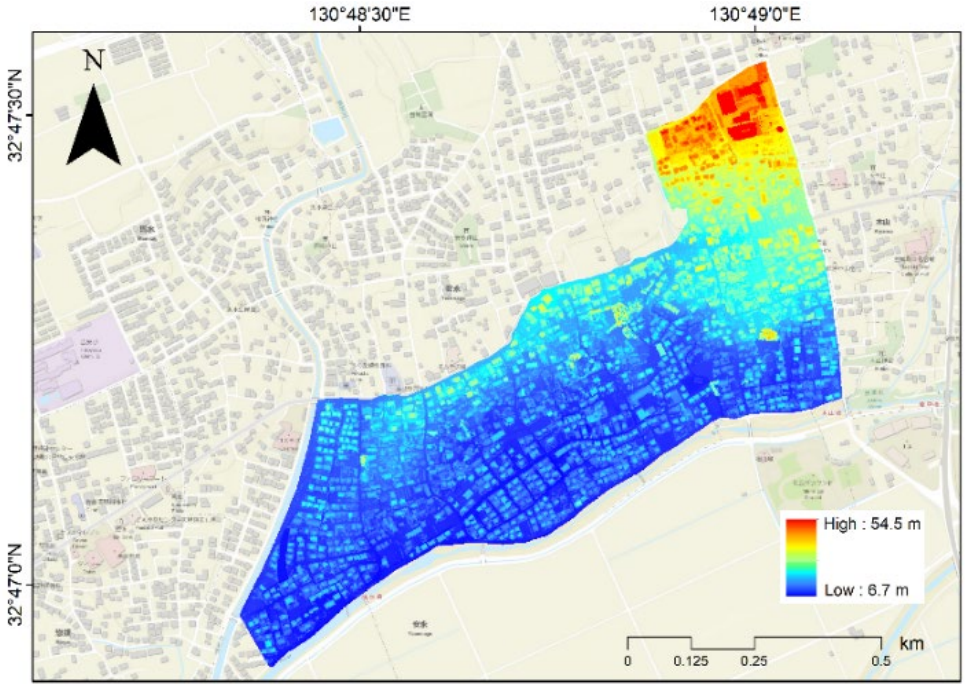



Figure 7-4. Post-earthquake point cloud

## 7.4 Methods

### 7.4.1 Architecture of the deep learning network

The backbone DL network in this study was RandLA-Net (Hu et al., 2020). While various DL networks have been published, most of them were designed for small or indoor scenarios. RandLA-Net, on the other hand, was explicitly proposed for extensive outdoor Lidar semantic segmentation. Therefore, it was adopted as the backbone due to the study's focus on large-scale outdoor scenarios. The details of the architecture of RandLA-Net are provided in its original publication paper (Hu et al., 2020), so the introduction of RandLA-Net is omitted in this study.

This study revised RandLA-Net by adding surface normal vectors. Surface properties have been applied in numerous studies for building damage assessment, such as vertical zenith information, planarity, surface normals, and the angles between surface normals. For instance, Axel and Van Aardt (2017) applied normal vectors and curvature of post-disaster Lidar points for assessing building damage. As a result, this chapter added surface normal information as an input. The developed architecture formulated for this purpose is illustrated in Figure 7-5.

There are four types of information selected as inputs, as shown in the left bottom of Figure 7-5, including colour information, coordinates of points, normal vectors of points, and labelled building damage levels. The surface normal vectors were calculated by the encapsulated 'estimate_normals' function in Open3D, an open-source Python library. This function in this chapter was set to find 30 nearest neighbour points of a point within a 0.1 m radius and to estimate a plane using the least square method with these points. After that, a vertical line of the plane went through that point, which is its normal vector.

The details of these inputs were subsequently stored, as depicted in the light green sections of Figure 7-5. This encompasses the information from point clouds, KDTree files containing the index of neighbouring points for each point, and projection files for the up-sampling step in the architecture. This light green information was retained as a part of the original RandLA-Net architecture. After that, these inputs underwent down-sampling and up-sampling processes, culminating in the production of four building damage levels as the final outputs. Consequently, the output of this network consists of building footprints with four damage levels, as shown in Figure 7-5.

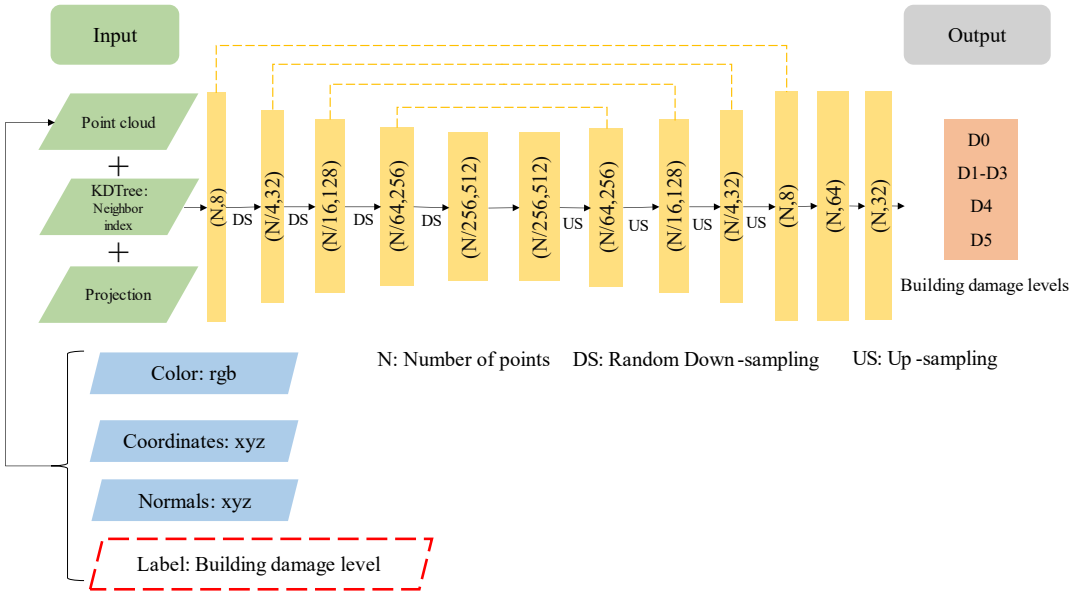



Figure 7-5. The architecture of the deep learning network applied in Chapter 7

### 7.4.2 Experiment information

Considering the total number of buildings, 200 and 100 buildings were randomly chosen for DL network validation and testing, respectively. Ten point cloud samples of tested buildings are shown in Figure 7-6. Data augmentation was applied, such as rotation and random scaling of each input during training with GeForce RTX 2080 Ti GPU. The maximum epoch during training is 50. Surface normal information was added according to the experiment results from the experiments implemented by Liu (2023). IoU of each level was calculated for each test building. The level with the highest IoU was recognized as the tested level of this building.

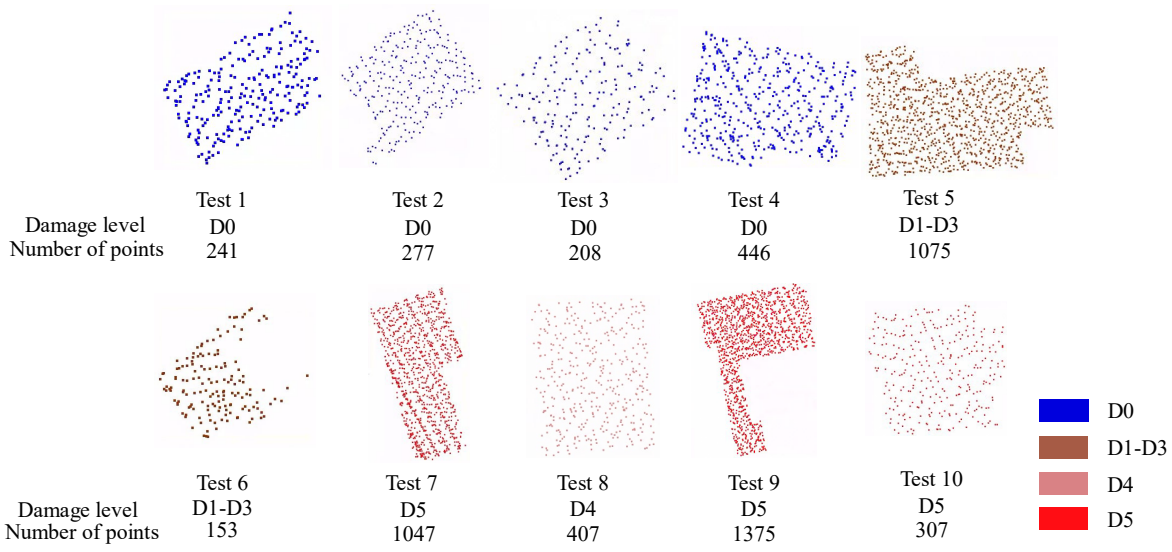



Figure 7-6. Damage levels of 10 sample tested point clouds

The IoU of each building damage level is the result of the number of TP points divided by the number of TP, FP, and FN points in that level. The meanings of each damage level are shown in Table 7-3. For example, In the context of the evaluation, TP indicates that a point was categorised as damage level D0, and its true level is also D0.

Table 7-3. TP, TN FP, and FN of each point in the test dataset

| Result truth \ Ground | **True** | **False** |
|---|---|---|
| **True** | TP | FP |
| **False** | FN | TN |

After getting the classification outputs according to the IoU of each building, the accuracy of each level was calculated, which is the result of the truly categorised number divided by the total number of that class. The mean accuracies of all levels are calculated considering

the weight of each level. The results of the proposed method and the original RandLA-Net backbone.

## 7.5 Results and discussion

### 7.5.1 Results for the proposed method

As mentioned in Section 7.4.2, 100 samples were randomly selected and tested. The locations of the randomly chosen point clouds are shown in Figure 7-7. The classification results are listed in Table 7-4. Different than TP of each point shown in Table 7-3 for calculating IoUs, the TP value in Table 7-4 represents that the classification result of each building sample is the same as its ground truth level. The classification result of each building was decided by the highest IoU result among all damage level results. For instance, if the D0 level had the highest IoU of a tested building, the building would be labelled as D0. If its ground truth is also D0, this building would be considered as TP in Table 7-3.

As introduced in Section 7.4.2, accuracy was applied for the evaluation, which is the result of the TP of each class divided by the total number of buildings at that damage level. It can be found that the accuracies of D0 and D5 are much higher than D1-D3 and D4 in Table 7-4. The two levels in the middle, i.e., D0-D3 and D4, were hard to detect. Partial points of one building in those levels were detected correctly in those two levels, but most points were categorised as D0 or D5. Therefore, the classification results of them were incorrect. A possible reason for this could be the feature difference between the middle two levels was not as obvious as the other two. Even the visual interpretation method finds it difficult to distinguish D1-D3 and D4 between these point clouds building damage levels. The

number of points in each test building is listed in Table 7-5 with its ground truth label and the classification result.

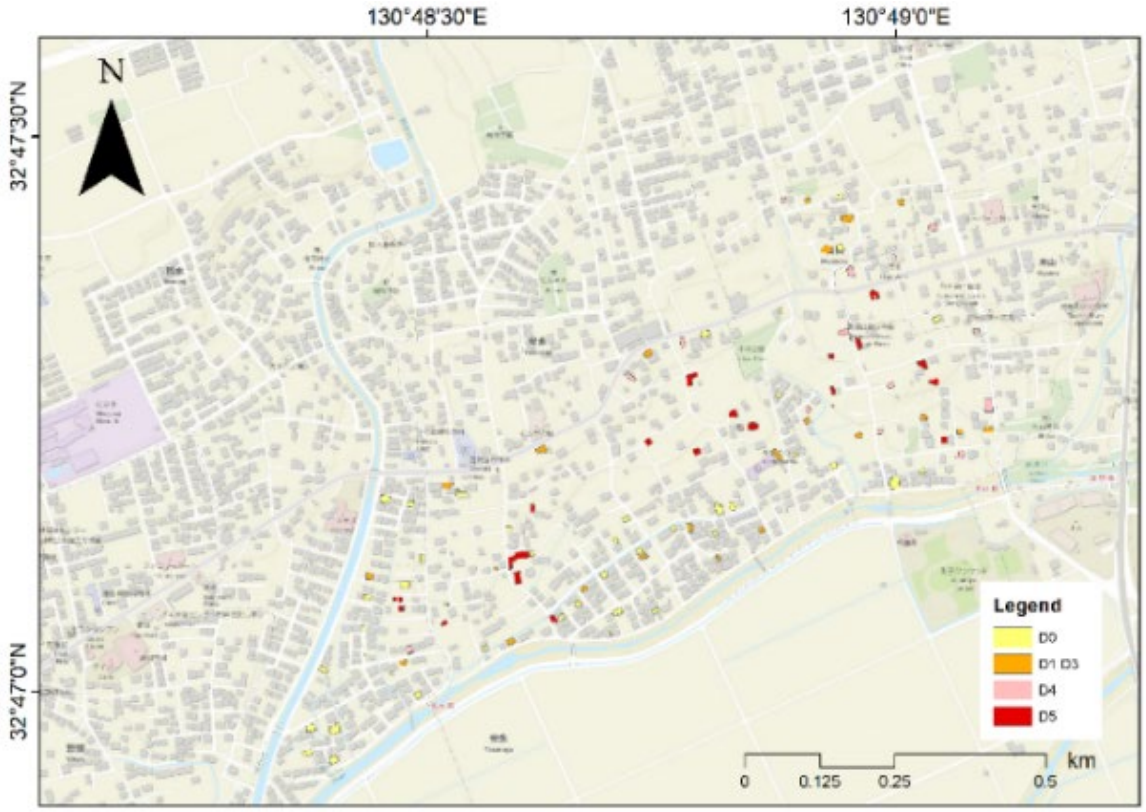



Figure 7-7. The locations of 100 tested buildings

Table 7-4. Building damage level classification results

| Damage level | D0 | D1-D3 | D4 | D5 | Mean |
|---|---|---|---|---|---|
| **Ground truth** | 41 | 23 | 15 | 21 | / |
| **TP of the proposed model** | 31 | 5 | 2 | 13 | / |
| **Accuracy of the proposed model** | 0.76 | 0.22 | 0.13 | 0.62 | 0.51 |
| **Accuracy of RandLA-Net** | 0.75 | 0.19 | 0.11 | 0.61 | 0.50 |

Table 7-5. Details of tested building information

| no. | Point number | True label | Tested | no. | Point number | True label | Tested |
|---|---|---|---|---|---|---|---|
| **1** | 388 | 6 | 9 | **51** | 661 | 7 | 9 |
| **2** | 590 | 6 | 9 | **52** | 122 | 7 | 6 |
| **3** | 795 | 6 | 9 | **53** | 485 | 7 | 6 |
| **4** | 701 | 6 | 9 | **54** | 302 | 7 | 6 |
| **5** | 341 | 6 | 9 | **55** | 160 | 7 | 6 |
| **6** | 241 | 6 | 6 | **56** | 153 | 7 | 6 |
| **7** | 269 | 6 | 6 | **57** | 358 | 7 | 6 |
| **8** | 155 | 6 | 6 | **58** | 378 | 7 | 9 |
| **9** | 151 | 6 | 9 | **59** | 382 | 7 | 9 |
| **10** | 294 | 6 | 9 | **60** | 331 | 7 | 9 |
| **11** | 866 | 6 | 6 | **61** | 1107 | 7 | 9 |
| **12** | 418 | 6 | 6 | **62** | 210 | 7 | 6 |
| **13** | 484 | 6 | 6 | **63** | 241 | 7 | 6 |
| **14** | 96 | 6 | 6 | **64** | 25 | 7 | 6 |
| **15** | 206 | 6 | 6 | **65** | 2114 | 9 | 6 |
| **16** | 277 | 6 | 6 | **66** | 1047 | 9 | 9 |
| **17** | 1018 | 6 | 9 | **67** | 601 | 8 | 9 |
| **18** | 368 | 6 | 9 | **68** | 77 | 8 | 6 |
| **19** | 209 | 6 | 9 | **69** | 659 | 8 | 9 |
| **20** | 759 | 6 | 9 | **70** | 614 | 8 | 9 |
| **21** | 248 | 6 | 9 | **71** | 759 | 8 | 6 |
| **22** | 712 | 6 | 6 | **72** | 649 | 8 | 9 |
| **23** | 687 | 6 | 6 | **73** | 554 | 8 | 9 |
| **24** | 237 | 6 | 6 | **74** | 337 | 8 | 6 |
| **25** | 233 | 6 | 6 | **75** | 255 | 8 | 9 |
| **26** | 208 | 6 | 6 | **76** | 409 | 8 | 9 |
| **27** | 278 | 6 | 6 | **77** | 568 | 8 | 9 |
| **28** | 667 | 6 | 6 | **78** | 350 | 8 | 9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **29** | 171 | 6 | 9 | **79** | 487 | 8 | 6 |
| **30** | 193 | 6 | 6 | **80** | 272 | 8 | 6 |
| **31** | 251 | 6 | 6 | **81** | 124 | 9 | 9 |
| **32** | 429 | 6 | 6 | **82** | 88 | 8 | 6 |
| **33** | 91 | 6 | 6 | **83** | 2224 | 9 | 6 |
| **34** | 76 | 6 | 6 | **84** | 780 | 9 | 6 |
| **35** | 236 | 6 | 6 | **85** | 1019 | 9 | 9 |
| **36** | 446 | 6 | 9 | **86** | 1375 | 9 | 6 |
| **37** | 792 | 6 | 6 | **87** | 1357 | 9 | 6 |
| **38** | 625 | 6 | 6 | **88** | 1496 | 9 | 9 |
| **39** | 443 | 6 | 6 | **89** | 810 | 9 | 6 |
| **40** | 585 | 6 | 6 | **90** | 693 | 9 | 9 |
| **41** | 530 | 6 | 6 | **91** | 287 | 9 | 9 |
| **42** | 1600 | 7 | 9 | **92** | 896 | 9 | 9 |
| **43** | 408 | 7 | 9 | **93** | 536 | 9 | 6 |
| **44** | 255 | 7 | 9 | **94** | 581 | 9 | 9 |
| **45** | 512 | 7 | 9 | **95** | 662 | 9 | 9 |
| **46** | 1075 | 7 | 9 | **96** | 307 | 9 | 6 |
| **47** | 390 | 7 | 9 | **97** | 391 | 9 | 6 |
| **48** | 329 | 7 | 9 | **98** | 331 | 9 | 6 |
| **49** | 227 | 7 | 9 | **99** | 481 | 9 | 9 |
| **50** | 668 | 7 | 9 | **100** | 299 | 9 | 6 |

The results of the proposed method and the RandLA-Net backbone are also listed in Table 7-4. The result of the proposed model is slightly higher than those of the backbone of all damage levels. A possible reason is that the proposed method can provide more features than the backbone, which is the surface normal vectors.

### 7.5.2 Building damage assessment framework

After disaster information analysis, earthquake disaster emergency management is a huge and complex system that has been widely studied in academia. Scholars from different disciplines have different divisions of emergency management stages. One of the most representative is the four-stage crisis model proposed by Fink (1986), including prodromal, acute, chronic, and resolution stages. Pearson and Mitroff (1993) divided crisis management into five stages, namely "signal detection", "probing and prevention", "damage containment", "recovery", and "learning".

Post-earthquake emergency rescue and reconstruction belong to the earthquake disaster emergency management category. In 1970, the United States first put forward the idea of dividing post-earthquake emergency rescue stages. The Federal Emergency Management Agency of the US designed an Incident Command System (ICS) including five functional areas, Planning, Command, Operations, Logistics, and Administration/Finance. Then, other countries also propose their post-earthquake emergency ways.

Fortunately, remote sensing techniques can help with the emergency response. According to the building damage level results from this chapter, a framework is proposed for rapid building damage assessment, as shown in Figure 7-8. This framework introduces the building-up stages of an AI-based building damage assessment system using multi-source data. Firstly, multiple types of pre- and post-disaster sources are collected. Secondly, these data can be fused, providing information from different aspects, and be applied in AI-based models, such as CNLNet, to extract ground objects. Building damage levels can also be

detected. The building damage level classification method proposed in this chapter is in the "data processing" stage in this framework. Next, some applications can be realised by these results, including but not limited to disaster information intelligent recognition, multi-source data matching, number of buildings in each damage level, and damage level visualisation. All these applications can help to provide an emergency response system for decision-makers and rescue team members as the last stage.

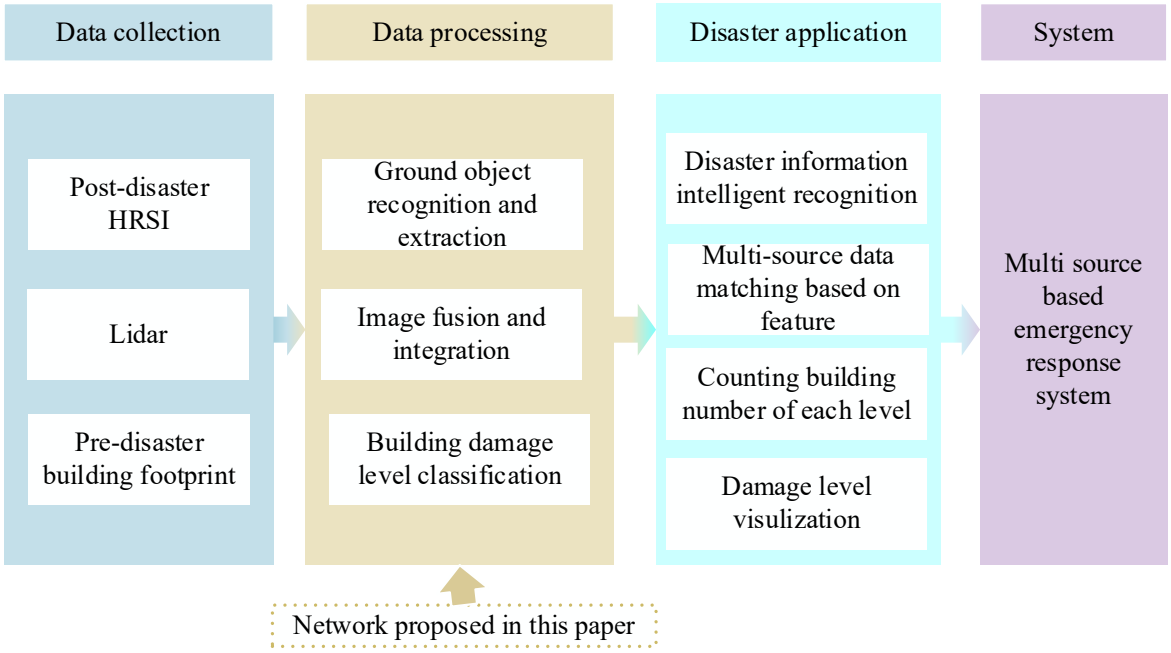



Figure 7-8. The framework of building damage assessment

## 7.6 Conclusion

This study developed an automated DL-based method for classifying post-earthquake building damage into four levels. The case study is the 2016 Kumamoto Earthquake with a created in-house labelled dataset using both post-earthquake Lidar and HRSI data. The building footprints were extracted from pre-earthquake vector shapefiles. The findings

reflect that the first and last levels, which are no/minor damage and story failure, have better classification results than the other two. One benefit of this method is applying multi-source data, including Lidar point clouds and HRSI, which provides more information than only one single type of input. The integration of both height and spectral information can improve the accuracy of the BDLC results. The study underscores the efficiency of satellite and Lidar data in building damage assessment and emphasises the significance of remote sensing technology. Therefore, multi-source remote sensing-based building damage classification is very promising.

Another noteworthy advantage of the workflow design for data preparation in this study is the incorporation of pre-earthquake building footprint vectors to ascertain the precise location of each building. Additionally, in order to identify damaged buildings, surface normal information was calculated and included as an input in the RandLA-Net backbone. The resulting network demonstrates better outcomes compared to the backbone, as evident from the accuracy achieved at each damage level.

Based on the outcomes of the proposed method, a building damage assessment framework was introduced, designed to support post-earthquake emergency response systems. This framework facilitates the identification of buildings significantly affected by earthquakes, thus facilitating the prioritisation of rescue and recovery endeavours by furnishing comprehensive data on the extent of the damage. Equipped with this information, governmental institutions, insurance companies, and other organisations can make informed decisions regarding resource allocation and providing optimal assistance to affected communities. One limitation of this study is its reliance on a single case study

presented in this study. Future research could extend the analysis to include multiple cases and explore the potential of using the developed network with transfer learning for conducting experiments.

# Chapter 8

# Discussion and conclusion

## 8.1 Overview

After a catastrophic earthquake, a high number of damaged buildings always leads to a high number of casualties. Rapid post-earthquake classification of building damage levels is an important part of rescuing human beings because the damage level information is one of the key factors in determining the allocation of rescue personnel and resources.

Considering the rapid development of DLSS in the remote sensing field, the aim of this study was proposed in Chapter 1 as "to propose novel DL models to classify building damage into four levels with large-scale in-house labelled datasets considering both pre- and post-earthquake periods". Four specific objectives were proposed according to the aim.

To achieve the objectives, this study adopted quantitative techniques for analysing and evaluating the results of trained DL methods. In the proposed DL methods of this study for extracting building footprints, which can be applied for pre-earthquake analysis, the possibility of earthquakes was considered when choosing the locations of case studies. The proposed post-earthquake BDLC method was tested with the happened earthquake.

The main chapters of this study, which are Chapters 4, 5, 6, and 7, are designed to achieve these four objectives. The key findings reported in these four main chapters are summarised in the four subsections of Section 8.2 as follows.

## 8.2 Key findings

### 8.2.1 DL-based large-scale BDLC method with 2D satellite images

To fulfil the first objective of this study, an initial exploratory study was undertaken. Chapter 4 explores the possible advantages of 2D optical satellite images for BDLC by proposing a novel DLSS method. The findings illustrate that four-level BDLC performs well based on the proposed DL-based method using pre-earthquake building footprint location information and post-earthquake HRSI. Those categorised building damage levels were hard or impossible to achieve by human eyes. The findings also suggest that the proposed DL-based method with adding SE CA block can achieve higher results than the original HRNet backbone. A larger input size can have better results but use much more computing time. Transfer learning with the pre-trained ImageNet dataset does not have advantages because the dataset does not contain several damaged building images. The block with Sigmoid function has slightly better performance than that with Hard-Sigmoid. Therefore, this chapter successfully provides a quick DL-based method for post-earthquake BDLC with optical satellite images to achieve Objective 1.

### 8.2.2 DL-based pre-earthquake building footprint extraction with 3D Lidar data

To fulfil the second objective of this study, a DLSS network for pre-earthquake building footprint extraction using 3D Lidar point clouds was proposed and tested. Chapter 5 implemented the preparation experiments to test the influences of features on the accuracy of DL-based building footprint extraction approaches. Chapter 6 designed the architecture of the proposed DLSS network according to the results of Chapter 5.

The findings from Chapters 5 and 6 show that the proposed method can be applied for pre-earthquake building footprint extraction. The findings also suggest that the proposed method with adding surface normal information and CA mechanism has higher accuracy than the original RandLA-Net backbone for classifying the building class with improving mIoU around 1% to 2%. Moreover, adding either surface normal or CA can achieve better results than the backbone. Therefore, the findings reveal that either surface normal or CA can help to improve accuracy for building footprint extraction. Since the proposed method can not only extract the building class but also vegetation, it is also illustrated that the proposed method performs better for building footprint extraction than vegetation classification. The generalisability of the proposed network for the building class is the best.

### 8.2.3 DL-based post-earthquake BDLC with 3D Lidar data

The experiments for achieving the third objective were implemented in Chapter 7. Building damages with four levels were detected by the proposed DL-based method using Lidar data. The proposed method added surface information to the network and was compared with the original RandLA-Net backbone. The findings show that the proposed BDLC method

outperforms the original RandLA-Net backbone, underscoring the advantage of incorporating surface normal information. The findings also show promising outcomes, particularly in accurately categorising no/minor damage and story failure levels. Consequently, this chapter not only demonstrates the practical utility of DL networks in assessing building damage after disasters under realistic operational conditions but also emphasises the significance of the proposed method and framework in bolstering public safety and guiding decision-making during the critical phases of post-earthquake recovery and reconstruction.

### 8.2.4 2D satellite and 3D Lidar labelled dataset creations of pre-earthquake building footprints and post-earthquake multi-level damaged building information

To achieve Objective 4, Chapters 4 to 7 all generated their own labelled large-scale datasets. Specifically, Chapter 4 created a four-level building damage dataset using satellite images from the 2010 Haiti Earthquake. Chapters 5 and 6 constructed labelled land cover Lidar datasets, and the labelled classes include ground, low vegetation, medium vegetation, high vegetation, and buildings. These Lidar datasets were collected from New Zealand and Japan. Chapter 7 generated a four-level building damage Lidar dataset from the 2016 Kumamoto Earthquake. Consequently, Objective 4 was completed as both satellite and Lidar datasets of pre-earthquake building footprints and post-earthquake multi-level damaged building information were created.

## 8.3 Theoretical contribution

This study proposed three DLSS networks for multi-level BDLC after earthquakes. The proposed methods include one 2D image-based four-level BDLC method, one Lidar-based pre-earthquake building footprint extraction method, and one Lidar-based post-earthquake four-level BDLC method. To bridge the gap that lacks the discussion of large-scale outdoor scenarios, all scenarios applied in this study are large-scale. Through extensive experimentation, this study demonstrated the feasible applications of the proposed methods for multi-level BDLC in large-scale scenarios. Moreover, compared to existing networks, this study demonstrated that normal information could provide more feature information, and channel attention can emphasise key information in channels so they can improve the accuracy of the building information extraction results.

## 8.4 Practical contribution

This study underscores the practical contributions that DL and semantic segmentation can make to enhance earthquake management. This study has led to the creation of DL networks capable of analysing seismic data, satellite imagery, and Lidar point clouds in near real-time. These networks provide early warnings of impending disasters, enabling authorities to initiate evacuation procedures and resource mobilisation promptly. Following an earthquake, this study employs computer vision and remote sensing techniques to assess the extent of damage. By analysing high-resolution satellite imagery and Lidar data, we can quickly identify affected areas and damaged levels, facilitating targeted response efforts. Therefore, the proposed methods in this study provide possible solutions in remote sensing-

based earthquake-related research, even natural disaster research, which is a road map for future research. As mentioned in Section 8.2.4, this study also generated three own labelled imagery and Lidar datasets to provide more datasets including building damage level information in the relevant research field.

## 8.5 Implications

This study related to disaster management is expected to have a significant implication for humanitarian-related purposes closely aligned with the United Nations (UN) Sustainable Development Goal (SDG) 11 'Make cities and human settlements inclusive, safe, resilient and sustainable'. After an earthquake, this study employs remote sensing to estimate building damage. It enables emergency responders and policymakers to rapidly and efficiently allocate resources for rescue and recovery operations, especially for low or middle-income countries due to limited resources, such as Pacific Island countries.

Several Pacific Island countries are prone to volcanic eruptions, often leading to severe earthquakes. It is expected that this study work will be capable of providing the degree of building damage in each area, including no damage, minor damage, major damage, or complete collapse. The building damage levels may serve as an indicator of the dangerousness of the rescue operation. This ensures an accurate and rapid allocation of resources for post-earthquake rescue and recovery in these communities. The results can also be applied for predisaster urban visualisation information storage, update, and management.

In conclusion, the main impact of this study is to benefit the community after an earthquake by informed decision-making. The information provided by this study is crucial for saving resources and lives in the affected communities by the disaster, not only for developed countries or regions but also for developing regions such as Pacific Island countries. The impact of this study will be visible in reduced casualties and enhanced community resilience. Additionally, this study will promote collaboration between scientists and policymakers, leading to the development of a more coordinated and effective response to earthquakes.

## 8.6 Recommendations for future work

As evident from the aforementioned results and findings, this study has frequently opted for a single earthquake event as the case study for each main chapter. Given the substantial variations in the aftermath of different earthquakes, the potential for these proposed methods to be widely applicable or generalised effectively may be limited.

Considering the results of this study, further research is suggested to find an approach to improve the segmentation accuracy of separating low, medium, and high vegetation. The quality of the training set could also be improved in future studies, such as increasing point numbers or densities of point clouds. Moreover, as the literature has shown in Chapter 2, many building damage evaluation codes and standards are developed from the structural engineering perspective. More unified and official codes and standards intended from the remote sensing perspective are suggested. Besides that, as shown in Section 6.5.3, a higher or lower resolution of 2D satellite images has little influence 3D Lidar data. Therefore, the

performance improvements that can be achieved using both the 2D satellite imagery and 3D Lidar data should be further tested and discussed in future.

# References

Akhlaghi, MM, Bose, S, Mohammadi, ME, Moaveni, B, Stavridis, A & Wood, RL 2021, 'Post-earthquake damage identification of an RC school building in Nepal using ambient vibration and point cloud data', *Engineering Structures,* vol.227, pp.111413.

Akkar, S, Ilki, A, Goksu, C & Erdik, M 2021, *Advances in assessment and modeling of earthquake loss,* Cham, Switzerland, Springer.

Asia Air Survey Co. Ltd. 2018, 'Post-Kumamoto earthquake rupture Lidar scan'. https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.052018.2444.1

Axel, C & Van Aardt, J 2017, 'Building damage assessment using airborne Lidar', *Journal of Applied Remote Sensing,* vol.11, no.4, pp.046024-046024.

Berman, M, Triki, AR & Blaschko, MB 2018, 'The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4413-4421.

Bhatta, B 2013, *Research methods in remote sensing*, Springer.

Chen, J 2018, *Automated indoor mapping with point clouds*, University of California, Santa Barbara.

China Earthquake Administration 2018, 'DB/T 75-2018 remote sensing assessment of earthquake disaster'.

Cloudeo Company 2023, *Brief report - kahramanmaraş earthquake: Damage assessment,* https://www.cloudeo.group/?utm_source=EarthquakeReportTurkey&utm_medium=report [Accessed on 17 March 2023].

Cotrufo, S, Sandu, C, Giulio Tonolo, F & Boccardo, P 2018, 'Building damage assessment scale tailored to remote sensing vertical imagery', *European Journal of Remote Sensing,* vol.51, no.1, pp.991-1005.

Courbariaux, M, Bengio, Y & David, JP 2015, 'Binaryconnect: Training deep neural networks with binary weights during propagations', *Advances in neural information processing systems,* vol.28, pp.3123-3131.

Deng, J, Dong, W, Socher, R, Li, LJ, Li, K & L, FF 2009, 'Imagenet: A large-scale hierarchical image database', *2009 IEEE conference on computer vision and pattern recognition*, pp.248-255, 10.1109/CVPR.2009.5206848.

Desroches, R, Comerio, M, Eberhard, M, Mooney, W & Rix, GJ 2011, 'Overview of the 2010 Haiti Earthquake', *Earthquake Spectra,* vol.27, no.1_suppl1, pp.1-21, 10.1193/1.3630129.

DIUX 2019, 'Xview2 scoring'.

Dos Santos, RC, Galo, M & Carrilho, AC 2019, 'Extraction of building roof boundaries from Lidar data using an adaptive alpha-shape algorithm', *IEEE Geoscience and Remote Sensing Letters,* vol.16, no.8, pp.1289-1293.

Dosovitskiy, A, Beyer, L, Kolesnikov, A, Weissenborn, D, Zhai, X, Unterthiner, T, Dehghani, M, Minderer, M, Heigold, G & Gelly, S 2020, 'An image is worth 16×16 words: Transformers for image recognition at scale', *arXiv preprint*, 10.48550/arXiv.2010.11929.

Dumoulin, V & Visin, F 2016, 'A guide to convolution arithmetic for deep learning', *arXiv preprint arXiv:1603.07285*, 10.48550/arXiv.1603.07285.

Eslamizade, F, Rastiveis, H, Zahraee, N K, Jouybari, A & Shams, A 2021, 'Decision-level fusion of satellite imagery and lidar data for post-earthquake damage map generation in Haiti, *Arabian Journal of Geosciences,* vol.14, pp.1-16.

Federal Emergency Management Agency 2016, 'Damage assessment operations manual: A guide to assessing damage and impact'. Washington, D.C., U.S.

Fink, S 1986, *Crisis management: Planning for the inevitable*, American Management Association, Amacom.

Geiger, A, Lenz, P & Urtasun, R 2012, 'Are we ready for autonomous driving? The Kitti vision benchmark suite, 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp. 3354-3361.

General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China 2009, 'GB/T 24335-2009: Code for classification of earthquake damage to buildings and special structures'. Beijing: Standards Press of China.

Geng, X, Ji, S, Lu, M & Zhao, L 2021, 'Multi-scale attentive aggregation for Lidar point cloud segmentation', *Remote Sensing,* vol.13, no.4, pp.691.

Gerke, M & Kerle, N 2011, 'Automatic structural seismic damage assessment with airborne oblique pictometry© imagery', *Photogrammetric Engineering & Remote Sensing,* vol.77, no.9, pp.885-898.

Google 2005, *Google Maps* [Online]. Available: https://developers.google.com/maps/documentation/static-maps/ [Accessed on 04 August 2023].

Grünthal, G 1998, 'European macroseismic scale 1998'. European Seismological Commission (ESC).

Guo, Y, Wang, H, Hu, Q, Liu, H, Liu, L & Bennamoun, M 2020, 'Deep learning for 3d point clouds: A survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.43, no.12, pp.4338-4364.

Gupta, A, Watson, S & Yin, H 2021, 'Deep learning-based aerial image segmentation with open data for disaster impact assessment', *Neurocomputing,* vol.439, pp.22-33.

Gupta, R, Goodman, B, Patel, N, Hosfelt, R, Sajeev, S, Heim, E, Doshi, J, Lucas, K, Choset, H & Gaston, M 2019a, 'Creating xBD: A dataset for assessing building damage from satellite imagery', *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp.10-17.

Gupta, R, Hosfelt, R, Sajeev, S, Patel, N, Goodman, B, Doshi, J, Heim, E, Choset, H & Gaston, M 2019b, 'XBD: A dataset for assessing building damage from satellite imagery', *arXiv preprint arXiv:1911.09296*.

Hackel, T, Savinov, N, Ladicky, L, Wegner, JD, Schindler, K & Pollefeys, M 2017, 'Semantic3D.Net: A new large-scale point cloud classification benchmark', *arXiv preprint arXiv:1704.03847*.

Hanks, TC & Kanamori, H 1979, 'A moment magnitude scale', *Journal of Geophysical Research: Solid Earth,* vol.84, no.B5, pp.2348-2350.

He, K, Zhang, X, Ren, S & Sun, J 2016a, 'Deep residual learning for image recognition', *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770-778.

He, M, Zhu, Q, Du, Z, Hu, H, Ding, Y & Chen, M 2016b, 'A 3D shape descriptor based on contour clusters for damaged roof detection using airborne lidar point clouds', *Remote Sensing,* vol.8, no.3, pp.189.

Helber, P, Bischke, B, Dengel, A & Borth, D 2019, 'EuroSat: A novel dataset and deep learning benchmark for land use and land cover classification', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol.12, no.7, pp.2217-2226.

Hu, J, Shen, L & Sun, G 2018, 'Squeeze-and-excitation networks', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141.

Hu, Q, Yang, B, Khalid, S, Xiao, W, Trigoni, N & Markham, A 2021, 'Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges', Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4977-4987.

Hu, Q, Yang, B, Xie, L, Rosa, S, Guo, Y, Wang, Z, Trigoni, N & Markham, A 2020, 'RandLA-Net: Efficient semantic segmentation of large-scale point clouds', Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11108-11117.

Huang, C 2022, *RandLA-Net-enhanced online codes* [Online]. Available: https://github.com/HuangCongQing/RandLA-Net-Enhanced [Accessed on 10 February 2022].

Huang, J, Zhang, X, Xin, Q, Sun, Y & Zhang, P 2019, 'Automatic building extraction from high-resolution aerial images and Lidar data using gated residual refinement

network', *ISPRS Journal of Photogrammetry and Remote Sensing,* vol.151, pp.91-105.

Ioffe, S & Szegedy, C 2015, 'Batch normalization: Accelerating deep network training by reducing internal covariate shift', *Proceedings of the 32nd International Conference on Machine Learning,* vol.37, pp.448-456.

Ji, M, Liu, L & Buchroithner, M 2018a, 'Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 Haiti Earthquake', *Remote Sensing,* vol.10, no.11, pp.1689, 10.3390/rs10111689.

Ji, S, Wei, S & Lu, M 2018b, 'Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set', *IEEE Transactions on Geoscience and Remote Sensing,* vol.57, no.1, pp.574-586.

Jiang, M, Wu, Y, Zhao, T, Zhao, Z & Lu, C 2018, 'Pointsift: A sift-like network module for 3D point cloud semantic segmentation', *arXiv preprint arXiv:1807.00652*.

Kalantar, B, Ueda, N, Al-Najjar, HA & Halin, A A 2020, 'Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre-and post-event orthophoto images', *Remote Sensing,* vol.12, no.21, pp.3529.

Kapiti Coast District Council 2022, *Kapiti's natural hazards* [Online]. Available: https://www.kapiticoast.govt.nz/our-district/cdem/kapitis-natural-hazards/ [Accessed on 10 July 2022].

Khodaverdi, N, Rastiveis, H & Jouybari, A 2019, 'Combination of post-earthquake lidar data and satellite imagery for buildings damage detection', *Earth Observation and Geomatics Engineering,* vol.3, no.1, pp.12-20.

Kolar, Z, Chen, H & Luo, X 2018, 'Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images', *Automation in Construction,* vol.89, pp.58-70, 10.1016/j.autcon.2018.01.003.

Koo, J, Seo, J, Yoon, K & Jeon, T 2020, 'Dual-HRNet for building localization and damage classification', SI Analytics (ed.). Republic of Korea. https://github.com/DIUx-xView/xView2_fifth_place/blob/master/figures/xView2_White_Paper_SI_Analytics.pdf

Krizhevsky, A, Sutskever, I & Hinton, GE 2012, 'Imagenet classification with deep convolutional neural networks', *Advances in neural information processing systems,* vol.25, pp.1097-1105.

Krupiński, M, Lewiński, S & Malinowski, R 2019, 'One class SVM for building detection on Sentinel-2 images', *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019,* vol.11176, pp.1117635, 10.1117/12.2535547.

Li, J, Zhang, X, Li, J, Liu, Y & Wang, J 2020a, 'Building and optimization of 3d semantic map based on lidar and camera fusion', *Neurocomputing,* vol.409, pp.394-407.

Li, L, Tian, T, Li, H & Wang, L 2020b, 'SE-HRNet: A deep high-resolution network with attention for remote sensing scene classification', *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pp.533-536, 10.1109/IGARSS39084.2020.9324633.

Li, W, He, C, Fang, J, Zheng, J, Fu, H & Yu, L 2019, 'Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source gis data', *Remote Sensing,* vol.11, no.4, pp.403.

Li, Y, Wang, C, Cao, Y, Liu, B, Luo, Y & Zhang, H 2020c, 'A-HRNet: Attention based high resolution network for human pose estimation', *2020 Second International Conference on Transdisciplinary AI (TransAI)*, pp.75-79, 10.1109/TransAI49837.2020.00016.

Liang, S & Wang, J 2020, *Advanced remote sensing: Terrestrial information extraction and applications*, Academic Press.

Liu, C, Sepasgozar, S M, Shirowzhan, S & Mohammadi, G 2021, 'Applications of object detection in modular construction based on a comparative evaluation of deep learning algorithms', *Construction Innovation,* vol.22, no.1, pp.141-159, 10.1108/CI-02-2020-0017.

Liu, C, Sepasgozar, SM, Zhang, Q & Ge, L 2022, 'A novel attention-based deep learning method for post-disaster building damage classification', *Expert Systems with Applications,* vol.202, pp.117268.

Liu, C, Zhang, Q, Shirowzhan, S, Bai, T, Sheng, Z, Wu, Y, Kuang, J & Ge, L 2023, 'The influence of changing features on the accuracy of deep learning-based large-scale outdoor Lidar semantic segmentation', The 2023 International Geoscience and Remote Sensing Symposium (IGARSS), Pasadena, California. IEEE, pp. 4443-4446.

Liu, P, Liu, X, Liu, M, Shi, Q, Yang, J, Xu, X & Zhang, Y 2019, 'Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network', *Remote Sensing,* vol.11, no.7, pp.830.

Liu, Y, Fan, B, Wang, L, Bai, J, Xiang, S & Pan, C 2018, 'Semantic labeling in very high resolution images via a self-cascaded convolutional neural network', *ISPRS Journal of Photogrammetry and Remote Sensing,* vol.145, pp.78-95.

Ma, H, Liu, Y, Ren, Y, Wang, D, Yu, L & Yu, J 2020, 'Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images', *Remote Sensing,* vol.12, no.2, pp.260.

Majd, RD, Momeni, M & Moallem, P 2019, 'Transferable object-based framework based on deep convolutional neural networks for building extraction', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol.12, no.8, pp.2627-2635, 10.1109/JSTARS.2019.2924582.

Maxar 2010, 'Open data program: Haiti earthquake'.

Mnih, V 2013, *Machine learning for aerial image labeling*, University of Toronto (Canada).

Nakano, Y, Maeda, M, Kuramoto, H & Myrakali, M 2004, 'Guideline for post-earthquake damage evaluation and rehabilitation of RC buildings in japan', *13th World Conference on Earthquake Engineering* Vancouver, B.C., Canada.

National Administration of Surveying, 2011, *Map world (in chinese)* [Online]. Available: http://lbs.tianditu.gov.cn/staticapi/static.html [Accessed on 01 August 2023].

Nelson Government 2023, *Weather event August 2022 - recovery progress report march 2023* [Online]. Available: https://www.nelson.govt.nz/services/weather-event-august-2022/ [Accessed on 01 August 2023].

Open3D Company, *Open3D geometry estimating normal function* [Online]. Available: http://www.open3d.org/docs/0.7.0/python_api/open3d.geometry.estimate_normals.html [Accessed on 12 January 2022].

Opentopography 2022, *Lidar of Kapiti Coast, Wellington, New Zealand 2021* [Online]. https://portal.opentopography.org/lidarDataset?jobId=pc1644298041020 [Accessed on 18 August 2022].

Opentopography 2023, *Top of the south flood, Nelson and Tasman, New Zealand 2022* [Online].https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.032023.2193.1 [Accessed on 15 May 2023].

Paszke, A, Gross, S, Chintala, S & Chanan, G 2021, *Pytorch-batchnorm2D* [Online]: Facebook. https://pytorch.org/docs/stable/generated/torch.nn.BatchNorm2d.html [Accessed on June 30 2021].

Pearson, CM & Mitroff, II 1993, 'From crisis prone to crisis prepared: A framework for crisis management', *Academy of Management Perspectives,* vol.7, no.1, pp.48-59.

Qi, C R, Su, H, Mo, K & Guibas, LJ 2017a, 'Pointnet: Deep learning on point sets for 3D classification and segmentation', *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE,* vol.1, no.2, pp.652-660.

Qi, C R, Yi, L, Su, H & Guibas, LJ 2017b, 'Pointnet++: Deep hierarchical feature learning on point sets in a metric space', *Advances in neural information processing systems,* vol.30.

Reichstein, M, Camps-Valls, G, Stevens, B, Jung, M, Denzler, J, Carvalhais, N & Prabhat 2019, 'Deep learning and process understanding for data-driven earth system science', *Nature,* vol.566, no.7743, pp.195-204, 10.1038/s41586-019-0912-1.

Richter, CF 1935, 'An instrumental earthquake magnitude scale', *Bulletin of the Seismological Society of America,* vol.25, no.1, pp.1-32.

Ronneberger, O, Fischer, P & Brox, T 2015, 'U-net: Convolutional networks for biomedical image segmentation', *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, pp.234-241, 10.1007/978-3-319-24574-4_28.

Ross, J 2021, *Earthquakes don't kill people; buildings do. And those lovely decorative bits are the first to fall* [Online]. https://theconversation.com/earthquakes-dont-kill-people-buildings-do-and-those-lovely-decorative-bits-are-the-first-to-fall-168476 [Accessed on 05 August 2023].

Rutzinger, M, Rottensteiner, F & Pfeifer, N 2009, 'A comparison of evaluation techniques for building extraction from airborne laser scanning', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol.2, no.1, pp.11-20.

Safe Software 2022, *Introduction of FME desktop* [Online]. Available: https://www.safe.com/fme/fme-desktop/ [Accessed on 12 August 2022]

Schweier, C & Markus, M 2006, 'Classification of collapsed buildings for fast damage and loss assessment', *Bulletin of earthquake engineering,* vol.4, no.2, pp.177-192.

Shan, J & Toth, CK 2018, *Topographic laser ranging and scanning: Principles and processing*, CRC press.

Shao, Z, Tang, P, Wang, Z, Saleem, N, Yam, S & Sommai, C 2020, 'BRRNet: A fully convolutional neural network for automatic building extraction from high-resolution remote sensing images', *Remote Sensing,* vol.12, no.6, pp.1050.

Shearer, PM 2019, *Introduction to seismology*, Cambridge university press.

Simonyan, K & Zisserman, A 2014, 'Very deep convolutional networks for large-scale image recognition', *arXiv preprint*, 10.48550/arXiv.1409.1556.

Steve Coast 2004, *Openstreetmap* [Online]. Available: https://www.openstreetmap.org/ [Accessed on 04 August 2023].

Su, J, Bai, Y, Wang, X, Lu, D, Zhao, B, Yang, H, Mas, E & Koshimura, S 2020, 'Technical solution discussion for key challenges of operational convolutional neural network-based building-damage assessment from satellite imagery: Perspective from benchmark xBD dataset', *Remote Sensing,* vol.12, no.22, pp.3808, 10.3390/rs12223808.

Sun, K, Xiao, B, Liu, D & Wang, J 2019a, 'Deep high-resolution representation learning for human pose estimation', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5693-5703.

Sun, K, Zhao, Y, Jiang, B, Cheng, T, Xiao, B, Liu, D, Mu, Y, Wang, X, Liu, W & Wang, J 2019b, 'High-resolution representations for labeling pixels and regions', *arXiv preprint*, 10.48550/arXiv.1904.04514.

Suppasri, A, Mas, E, Charvet, I, Gunasekera, R, Imai, K, Fukutani, Y, Abe, Y & Imamura, F 2013, 'Building damage characteristics based on surveyed data and fragility curves of the 2011 great east japan tsunami', *Natural Hazards,* vol.66, pp.319-341.

Tan, W, Qin, N, Ma, L, Li, Y, Du, J, Cai, G, Yang, K & Li, J 2020, 'Toronto-3D: A large-scale mobile lidar dataset for semantic segmentation of urban roadways', Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 202-203.

Tanjung, J, Sanada, Y, Nugroho, F & Wardi, S 2020, 'Seismic analysis of damaged buildings based on postearthquake investigation of the 2018 Palu Earthquake', *International Journal of GEOMATE,* vol.18, no.70, pp.116-122, 10.21660/2020.70.9490.

Tensorflow 2022, *Tensorflow GPU installation* [Online]. https://www.tensorflow.org/install/pip [Accessed on 19 Jule 2022].

UNITAR/UNOSAT/EC/JRC/WB 2010, *Port-au-prince Atlas of building damage assessment* [Online]. http://www.unitar.org/unosat/node/44/1425 [Accessed on 05 March 2021]

United States Geological Survey 2020, *Map of the December 29, 2020 Petrinja, Croatia earthquake* [Online]. https://www.usgs.gov/news/featured-story/magnitude-64-earthquake-croatia#:~:text=The%20USGS%20has%20up%2Dto%2Ddate%20details%20on%20the%20December,%3A20%20pm%20local%20time) [Accessed on 23 June, 2023].

United States Geological Survey 2023, *The science of earthquakes* [Online]. https://www.usgs.gov/programs/earthquake-hazards/science-earthquakes [Accessed on 09 July 2023]

Valentijn, T, Margutti, J, Van Den Homberg, M & Laaksonen, J 2020, 'Multi-hazard and spatial transferability of a cnn for automated building damage assessment', *Remote Sensing,* vol.12, no.17, pp.2839.

Van Etten, A, Lindenbaum, D & Bacastow, TM 2018, 'Spacenet: A remote sensing dataset and challenge series', *arXiv preprint arXiv:1807.01232*.

Verma, V, Kumar, R & Hsu, S 2006, '3D building detection and modeling from aerial Lidar data', 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, pp. 2213-2220.

Wang, J, Sun, K, Cheng, T, Jiang, B, Deng, C, Zhao, Y, Liu, D, Mu, Y, Tan, M & Wang, X 2020, 'Deep high-resolution representation learning for visual recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.43, no.10, pp.3349-3364, 10.1109/TPAMI.2020.2983686.

Wang, Y, Jing, X, Cui, L, Zhang, C, Xu, Y, Yuan, J & Zhang, Q 2023, 'Geometric consistency enhanced deep convolutional encoder-decoder for urban seismic damage assessment by UAV images', *Engineering Structures,* vol.286, pp.116132.

Wei, S, Ji, S & Lu, M 2019, 'Toward automatic building footprint delineation from aerial images using cnn and regularization', *IEEE Transactions on Geoscience and Remote Sensing,* vol.58, no.3, pp.2178-2189.

Wei, X, Li, X, Liu, W, Zhang, L, Cheng, D, Ji, H, Zhang, W & Yuan, K 2021, 'Building outline extraction directly using the U2-net semantic segmentation model from high-resolution aerial images and a comparison study', *Remote Sensing,* vol.13, no.16, pp.3187.

Wheeler, BJ & Karimi, HA 2020, 'Deep learning-enabled semantic inference of individual building damage magnitude from satellite images', *Algorithms,* vol.13, no.8, pp.195, 10.3390/a13080195.

Wierzbicki, D, Matuk, O & Bielecka, E 2021, 'Polish cadastre modernization with remotely extracted buildings from high-resolution aerial orthoimagery and airborne lidar', *Remote Sensing,* vol.13, no.4, pp.611.

Xie, S, Duan, J, Liu, S, Dai, Q, Liu, W, Ma, Y, Guo, R & Ma, C 2016, 'Crowdsourcing rapid assessment of collapsed buildings early after the earthquake based on aerial remote sensing image: A case study of Yushu earthquake', *Remote Sensing,* vol.8, no.9, pp.759.

Xiong, C, Li, Q & Lu, X 2020, 'Automated regional seismic damage assessment of buildings using an unmanned aerial vehicle and a convolutional neural network', *Automation in Construction,* vol.109, pp.102994.

Xiu, H, Shinohara, T, Matsuoka, M, Inoguchi, M, Kawabe, K & Horie, K 2020, 'Collapsed building detection using 3D point clouds and deep learning', *Remote Sensing,* vol.12, no.24, pp.4057.

Yamada, M, Ohmura, J & Goto, H 2017, 'Wooden building damage analysis in Mashiki town for the 2016 Kumamoto earthquakes on April 14 and 16', *Earthquake Spectra,* vol.33, no.4, pp.1555-1572.

Yang, W, Zhang, X & Luo, P 2021, 'Transferability of convolutional neural network models for identifying damaged buildings due to earthquake', *Remote Sensing,* vol.13, no.3, pp.504, 10.3390/rs13030504.

Yang, X, Wang, C, Xi, X, Wang, P, Lei, Z, Ma, W & Nie, S 2019, 'Extraction of multiple building heights using ICESat/GLAS full-waveform data assisted by optical imagery', *IEEE Geoscience and Remote Sensing Letters,* vol.16, no.12, pp.1914-1918.

Yuan, W, Li, J, Bhatta, M, Shi, Y, Baenziger, PS & Ge, Y 2018, 'Wheat height estimation using Lidar in comparison to ultrasonic sensor and UAS', *Sensors,* vol.18, no.11, pp.3731.

Zhan, Y, Liu, W & Maruyama, Y 2022, 'Damaged building extraction using modified mask r-cnn model using post-event aerial images of the 2016 Kumamoto earthquake', *Remote Sensing,* vol.14, no.4, pp.1002.

Zhang, C, Sargent, I, Pan, X, Li, H, Gardiner, A, Hare, J & Atkinson, PM 2019a, 'Joint deep learning for land cover and land use classification', *Remote Sensing of Environment,* vol.221, pp.173-187.

Zhang, L, Wu, J, Fan, Y, Gao, H & Shao, Y 2020, 'An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN', *Sensors,* vol.20, no.5, pp.1465.

Zhang, Q, Ge, L, Hensley, S, Metternicht, GI, Liu, C & Zhang, R 2022, 'PolGAN: A deep-learning-based unsupervised forest height estimation based on the synergy of PolInSAR and Lidar data', *ISPRS Journal of Photogrammetry and Remote Sensing,* vol.186, pp.123-139.

Zhang, Z, Hua, BS & Yeung, SK 2019b, 'Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics', Proceedings of the IEEE/CVF International Conference on computer vision. pp. 1607-1616.

Zhao, H, Jiang, L, Fu, CW & Jia, J 2019a, 'Pointweb: Enhancing local neighborhood features for point cloud processing', Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5565-5573.

Zhao, Z, Wang, H, Wang, C, Wang, S & Li, Y 2019b, 'Fusing Lidar data and aerial imagery for building detection using a vegetation-mask-based connected filter', *IEEE Geoscience and Remote Sensing Letters,* vol.16, no.8, pp.1299-1303.

Zheng, C, Abd-Elrahman, A & Whitaker, V 2021, 'Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming', *Remote Sensing,* vol.13, no.3, pp.531.