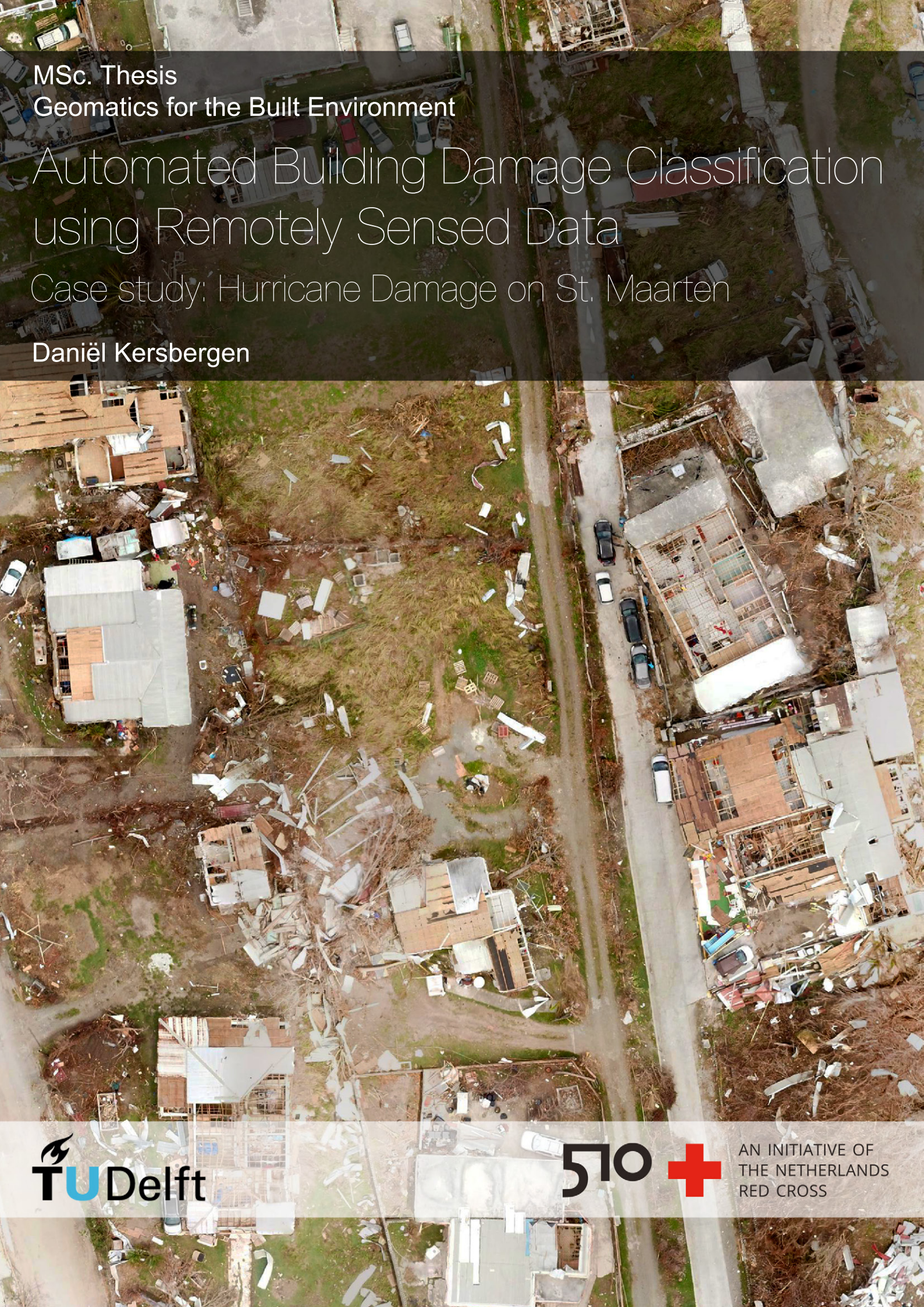


MSc. Thesis
Geomatics for the Built Environment

Automated Building Damage Classification using Remotely Sensed Data

Case study: Hurricane Damage on St. Maarten

Daniël Kersbergen

TU Delft

510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

COVER IMAGE:
UAV imagery from Cay Bay on St. Maarten after Hurricane Irma. (Source: Netherlands Red Cross (11 Sept. 2017), Cole Bay - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org)

AUTOMATED BUILDING DAMAGE CLASSIFICATION USING REMOTELY SENSED DATA

Case study: Hurricane damage on St. Maarten

A thesis submitted to the Delft University of Technology in partial fulfilment of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Daniël Kersbergen

July 2018

The work in this MSc. thesis was made in collaboration with:

Environmental Technology and Design
Department of Urbanism
Faculty of Architecture & the Built Environment
Delft University of Technology

Department of Geoscience and Remote Sensing
Faculty of Civil Engineering and Geosciences
Delft University of Technology

510
An initiative of
The Netherlands Red Cross

Supervisors:	Dr. Jorge Lopes Gil
	Dr. Stef L. M. Lhermitte
Co-reader:	Agung Indrajit MSc.
Board of Examiners:	Luc Willekens MSc.
Company Supervisor:	Dr. Stefania Giodini

ABSTRACT

In the second half of the 20th and beginning of the 21st century the amount of natural disasters has increased rapidly. Due to this rise in occurrences, more people are affected. An important indicator for people affected is the amount of damage to buildings. To gather this information aid workers now have to go into the field to gather data on the amount of destruction. In response to the possible dangers these people encounter in the field, remote sensing and analysis techniques have been developed for automated damage detection. However, due to various limitations on the implementation, these techniques are not yet widely adopted in emergency response and humanitarian aid.

This work compares two methods and two data sources for the detection of building damage. The methods are evaluated on their accuracy and implementability within humanitarian aid in disaster situations. The main methods considered are equalisation of histograms of pre-event and post-event imagery, followed by Univariate Image Differencing; and a convolutional neural network on features withdrawn from post-event imagery, using OpenStreetMap data. Remotely sensed data sources considered are synthetic aperture radar and very high resolution optical imagery. All results are analysed and compared to current standards in damage detection.

From the results it can be concluded that more research is required for a practical implementation of deep learning techniques. The constraint posed by the requirement of large datasets, make these methods impracticable without sufficient preparation and resources. More simpler methods, like Univariate Image Differencing, can be validated on smaller ground-truth datasets, and are therefore easier in implementation when resources are limited. The possible accuracy increase of deep learning methods does, at this moment, not outweigh the ease of an elementary differencing approach.

PREFACE

This work is the culmination of many months of hard work, with many opportunities to learn and experience. Over the past year I have been fortunate enough to get acquainted with humanitarian aid processes and the role geographic information has within it. These insights have been an inspiration for this research and allowed me to put the Master's degree Geomatics for the Built Environment in context of real world applications. At the conclusion of this Master's programme, I can say that I have found my personal drivers and my little place from where I can contribute to a better world.

The process of research and study however, is a path of cooperation, support and learning from others. I would like to thank those who have inspired and supported me throughout. First of all I would to thank my two mentors guiding the thesis process. Jorge and Stef, thank you very much for the all the feedback in the various meetings we had every two weeks. The ideas and concepts you were able to elaborate to me guided me through the process and allowed me to gain purpose in the research. Thank you for steering me back to the main topic on moments my ideas where to broad for a Master's thesis, and explaining that others would be able to start their research from the questions I left unanswered. Besides my main mentors I would like to thank Luc Willekens and Agung Indrajit for their feedback in the official process; and Sisi Zlatanova for her inspiring work on Geomatics for disasters and commencing support for the project.

Secondly, thank you very much to the 510 team from the Netherlands Red Cross. The work done with regards to information within humanitarian aid is an inspiration to me. A special thanks to Stefania for the support of the project and for reading all materials I made for this thesis. Furthermore I was considerably helped by Andy Thean, his expertise and technical understanding allowed me to interpret existing methods and apply those on new datasets.

I would also like to thank my co-student from the Geomatics program. Thank you all very much for the great two years with many experiences. Unfortunately we saw each other less often in the last year but we had so much fun, and shared so much of the stress. And Wicked Climbers, we will climb some more!

Lastly I would like to thank my family. Sorry parents, for moving back in after my Bachelor's. However, I am very grateful as it allowed me to focus on my studies and research. Nicoline, thank you for listening to all the information I had regarding my thesis, even if it was boring, unstructured, or just plain incomprehensible. Thank you all very much for all the support and guidance throughout the process.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

BAR Baker, Achkar, Raymond method

CNN Convolutional Neural Network

CNO CNN approach for Optical UAV data

CNS CNN approach for Satellite SAR data

DEM Digital Elevation Model

DRM Disaster Risk Management

EM Emergency Mapping

EMS Emergency Mapping Service

ESO Equalisation and Subtraction on Optical UAV data

ESS Equalisation and Subtraction on Satellite SAR data

ESA European Space Agency

GIS Geographic Informations Systems

HSL Hue, Saturation, and Lightness

HSV Hue, Saturation, and Value

InSAR Interferometric Synthetic-Aperture Radar

NASA National Aeronautics and Space Administration

NGOs Non-governmental Organisations

NLRC Netherlands Red Cross

RGB Red, Green, and Blue

SAR Synthetic Aperture Radar

SEM Satellite-based Emergency Mapping

SRTM Shuttle Radar Topography Mission

STOA State of the Art

UAV Unmanned Aerial Vehicle

VHR Very High Resolution

WSL Windows Subsystem for Linux

1 INTRODUCTION

Weather related natural disasters cost the world economy around 100 billion dollars every year [Kousky, 2014]. According to the Centre for Research on the Epidemiology of Disaster [2015] 69.800 deaths per year are inflicted as the result of these disasters and earthquakes. The effects are felt around the world, however most deaths occur in low or middle income areas. Advances in technology and preparedness have decreased the amount of deaths caused by natural disasters since the second part of the previous century [United Nations, 2004]. However, due to an increase in the frequency of disasters [figure 1.1] more people are affected and more damages occur; with the most economic damage recorded in 2011 [Coppola, 2015; Kerle, 2015]. 2017 was no exceptions to both trends, as it was the year with the second most economic damage but with less people killed [Munich RE, 2018].

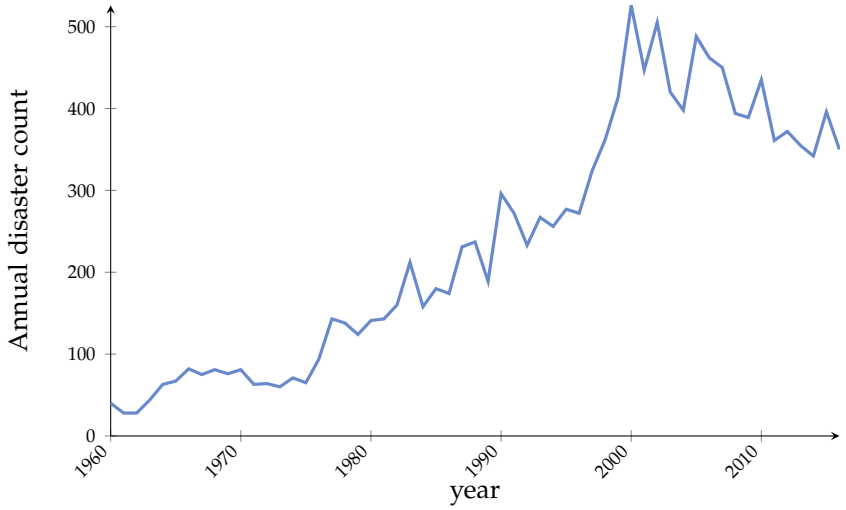


Figure 1.1: Natural disaster per year from 1965 to 2016 [From: EM-DAT: The Emergency Events Database - Universite catholique de Louvain (UCL) - CRED, D. Guha-Sapir - www.emdat.be, Brussels, Belgium]

One of the major disasters of 2017 was hurricane Irma, being labelled the worst storm in the Caribbean in recordARed history [Daniell et al., 2017]. In the first week of September this hurricane raged over multiple islands causing billions worth of damage, affecting millions in its path [Phipps, 2017; Daniell et al., 2017]. One of the islands affected is St. Maarten, part of the Kingdom of the Netherlands. It was hit by the eye of the hurricane on the 6th of September with winds up to 185 miles per hour [Wilts, 2017]. Two other major hurricanes passed over the area in the weeks following hurricane Irma, however, fortunately these did little to no extra damage on the island [Gray, 2017; Bijnsdorp, 2017]. First damage estimates show that 70 - 90% of the island may be affected by the storm [Rode Kruis, 2017; UNOSAT, 2017]. The indication and location of the damage caused is a leading planning tool for organisations like the Netherlands Red Cross (NLRC), providing a first indication of the most vulnerable people in an affected area. Indications of damage have allowed the NLRC to help 18.881 individual people since the landfall of hurricane Irma and subsequent relief operation; and long term operations are being started right now to facilitate the rebuilding of the island [Rode Kruis, 2017].

Around the world, in both wealthy and impoverished regions, individuals and organisations, both governmental and Non-governmental Organisations (NGOs), are motivated to reduce and manage the impact of disasters [Coppola, 2015]. Within this disaster management the decision making process and risk management process are bounded by three key characteristics [Zlatanova and Li, 2008]. [1.] Rapid action needs to be taken, [2.] aware of the situation and context [3.] with a connected overview of the data available. The goal of the Disaster Risk

Management (DRM) is to minimize the impact from a disaster [Piero and FabioGiulio, 2012]. Information on the extent of the damage is therefore paramount, as demonstrated by relief operations from the NLRC. Building damage is an essential indicator for this [Schweier and Markus, 2006]; but can be hard to establish as it requires a lot of manual field labour; a dangerous and timely ordeal for aid workers involved [Kerle, 2010]. Automated detection of building damages based on remotely sensed data could be the solution, allowing for faster and more efficient response [Vetrivel et al., 2016b].

Remote sensing has long been part of the DRM cycle. According to Kerle [2015] this started at the beginning of space-based remote sensing around the 1960s and 1970s and brought about the increase in information within the DRM. From here the development of remote sensing techniques, both in space (optical and Synthetic Aperture Radar (SAR)) and within the atmosphere (aerial and Unmanned Aerial Vehicle (UAV) based), accelerated over the past 50 years and increasingly allows for higher resolution information in a more timely manner. The performance increase in remote sensing solutions make it applicable for the automated classification of building damage [Dell'Acqua and Gamba, 2012; Dong and Shan, 2013]. Many solutions for automated damage detection or classification have been developed over the past years in academia, based on several remote sensing techniques. Dong and Shan [2013] provides a clear overview of the solutions up until 2013 and several more have been developed since [Dominici et al., 2017; Sharma et al., 2017; Kakooei and Baleghi, 2017; Vetrivel et al., 2016b; Menderes et al., 2015]. However, in practice, services from the International Charter, like Copernicus and UNOSAT, are mostly used for Satellite-based Emergency Mapping (SEM) [Voigt et al., 2016] as they produce usable results [Kerle, 2010]. The method for damage classification used by these services is manual visual interpretation of remotely sensed data, as is indicated by the disclaimers or map information of products from these services and a program specialist at UNOSAT [Copernicus EMS, 2017; UNDAC, 2017]. While this approach is safer for the aid workers in the field, it remains a laboursome task.

1.1 PROBLEM STATEMENT

It seems remarkable that the extensive academic research on automated damage detection is not implemented in the disaster relief sector, as the building damage is a fundamental indicator used in DRM and relief operations [Schweier and Markus, 2006]. The lack of implementation of automated methods can be seen as an indicator of the absence of support from the humanitarian agencies. Several considerations could be the cause of this; [1.] there is too little communication between humanitarian agencies and academics, resulting in too complex methods or inadequate solutions. [2.] the methods proposed do not deliver the expected outcomes concerning effectiveness or accuracy. [3.] there are no resources to implement the new methods in existing procedures. An example of this can be found in Ajmar et al. [2011]. This paper mentions the lack of predictable results and time involved as impediments for implementation of automated approaches. Disaster situations require fast implementations as lives might be at stake, while reliable results are necessary for fair distribution of aid.

The stringent requirements from humanitarian organisations in disaster situations results in the recurrence of visual interpretation within damage detection from remotely sensed data. In this process a group of professionals has to visually inspect pre-event and post-event imagery to conclude the extent of damage to areas. Even though this is common practice, it is only viable on smaller areas, as the task of human comparison is labour intensive and requires specialist knowledge. The return to this method is aided by the understanding that automated approaches are not as accurate as human interpretation from satellite imagery. However, this is difficult to corroborate as academic literature seems to insinuate the opposite.

This research explores methods for the accurate classification of building damage after a natural disaster. Existing academic methods will be taken into consideration and tested on the available -real world- data from St. Maarten. Furthermore, the research will be conducted in cooperation with NLRC to cope with some of considerations that might be causing the lack of implementation. Advances in remote sensing techniques, machine learning and Geographic Informations Systems (GIS) are recognised as upcoming and supportive technologies within the organisation, as it established a new data team [510] in 2016. The data team and NLRC, as well as other humanitarian organisations, could benefit from the research into the automated classification of building damage after a disaster, as it would allow for more efficient delivery of aid and humanitarian relief. The academic field working on remote sensing for disaster situations could also benefit from this research as it will provide a comparison between methods in a scenario different from the academic examples.

An viable method would enable humanitarian organisations to accurately plan their interventions and relief operations on the basis of damage inflicted in a disaster struck areas. This would allow for an increase in effectiveness of the activities, and therefore help more people affected by a natural disaster.

1.2 RESEARCH OBJECTIVES

The objective of this research is: *to develop a method for the automatic classification of damage inflicted by hurricanes using remotely sensed data; applied to the case of hurricane Irma on the island of St. Maarten.* This will be based on already existing methods, which will be extended to suit the requirements from humanitarian organisations for improved aid delivery. To achieve this goal the following research question will be answered:

> *Is the use of remotely sensed data a viable option for the automatic classification of hurricane inflicted damage?*

To develop such method various other challenges will need to be handled. These will be dealt with in the literature research and analysis of existing academic methods. To guide this process the following sub-questions have been formulated:

- How is damage determined?
- What criteria are set for damage classification methods?
- Which methods already exist?
- How do these methods perform?
- How does the state of the art compare to these methods?

With answers to these sub-questions, a description of an effective method is developed. This allows for insight in the necessities of such method. The difference between damage detection, classification, and assessment is also considered as all three require different approaches. The restrictions imposed by humanitarian organisations, especially within disaster situations, are also taken into consideration; as these might impose other criteria on methods. As described in this chapter various methods already exist, a subset of these will be considered in this research after the criteria for damage classification have been established. The next scientific challenge lies in the determination of the effectiveness of existing methods. The main question to be answered in this regard is the accuracy comparison between ground truth data from St. Maarten and the results from these methods. An assessment to classify accuracy is necessary to be able to compare the methods. To better map the impact of a disaster an extended overview of the damage is necessary, to achieve this the methods will be examined for extension to allow for damage classification. To ensure the possible solutions and extensions can be applied, these will be evaluated for performance compared to existing methods and ground truth. This will take into account the data derived from Copernicus assessment of the island from September 2017. With answers to all these sub-questions and the main research question, it would be necessary to see the methods and possible extensions on a broader scale and consider their fit within both the academic field and humanitarian field of operation. This will allow for reflection on the results and might allow for discussion on the possibility of implementation in the field.

1.3 RESEARCH METHODOLOGY

In this section the methodology to achieve the main research goal will be defined. As shown in figure 1.2, the methodology for this research is linearly structured with 3 sub-groups, Preparation, Research implementation, and assessments of Results. To answer the first three sub-questions an extensive literature research has been set-up as preparation for the implementation of existing methods. On this bases, methods have been selected to be implemented and possibly extended. Lastly, the results from all implementations will be compared and discussed.

1.3.1 Preparation

The literature review as preparation for the implementation and comparison will form the base for this research. To allow for a thorough, yet workable, approach the following guiding principles have been used. The inclusion criteria for the research were different for the various topics. In general the literature are peer reviewed papers or masters or doctorate theses from the past

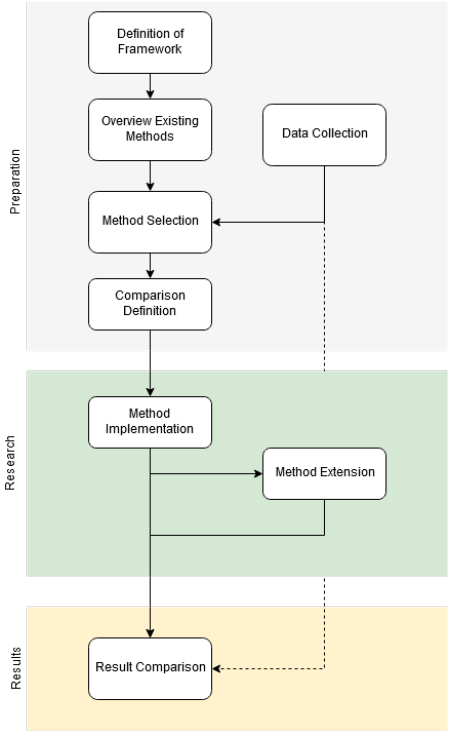


Figure 1.2: Methodology structure, highlighting the various phases [Grey = Preparation, Green = Research implementation, and Yellow = Assessment of Results]

15 years. However, exceptions have been made for explanations of guiding principles. All literature is related to the topic they are describing and have been found through a combination of related key-words. The literature review regarding existing methods have been more strictly limited to the past 15 years with emphasis on newer material. Furthermore, have solutions for all disaster been considered, as long as the damage descriptions were not disaster specific. Particularly design studies have been regarded, in which new methods were developed. This process was guided by key-words and other literature comparing methods. The results from this review can be found in chapter 2.
From the literature review a systematic approach to the testing of methods has been developed. This systematic approach considers both data and method in the comparison. A description of this approach is provided in chapter 3.

1.3.2 Implementation and Comparison

The implementation of methods and the comparison of results is based on the systematic approach defined in the preparation based on the knowledge acquired from the literature review. A selection of methods has been implemented using the description of their respective literature; as well as extended to fit the requirements for classification in a disaster situation. The results from this implementation have been compared to the State of the Art (STOA) approaches in use in humanitarian response to disasters and ground truth data acquired. This comparison provides an insight in the viability of methods to be implemented in disaster situations and possible future research.

1.4 RESEARCH SCOPE

This research will focus on the accurate and automated classification of building damage in a hurricane struck area. Limitations in a masters thesis are unavoidable as time and resources are limited. Spatially the research is restricted to the Dutch part of the island St. Maarten as well as temporally limited to the aftermath of hurricane Irma in 2017. These limitations are set forward by the data available and is also supported by the fact that the author has been to the island in the aftermath of the hurricanes. This on-the-ground experience could be helpful in the understanding of the problem.

The method to be developed will be based on existing methods and will use, for where possible, existing software packages. This will shift the focus from a completely new method, to an adaptation with potential for implementation in various disaster procedures. This allows for more investigation into existing methods, and selecting effective techniques. By choosing this approach, developing a method which would be added to a list of unused, existing methods, is avoided. To further reduce the scope, only two existing methods will be selected for rigorous assessment. These will be chosen through the theoretical framework and literature research, set forward in chapter 2.
For this research, the task of object detection, as used by various other academic methods [Vetrivel et al., 2016b; Kakooei and Baleghi, 2017], will not be used. The datasets available allow for the use of existing building outlines, gathered from recent data. The NLRC uses voluntary cartographers to map disaster areas shortly after an incident and where possible, in case of the hurricane on St. Maarten, shortly before a disaster. Making sure that maps are up to date and can be used for planning. This eliminates the need for object detection from data sources, however requires the data input to be properly matched.
The research is limited to the various datasets available through the 510 team of the NLRC as cleaning, relief and reconstruction activities on the island make new data collection impossible. However, these datasets can be considered a good representation of available data in the wake of large scale disaster. The cooperation with the data team of the NLRC allows for a good balance between the technical approach of the Delft University of Technology and a more societal aspect from a humanitarian organisation.

1.5 READING GUIDE

The following will be discussed in this document:

- Chapter 2 will go into further detail concerning the theoretical framework of this research. Existing methods will be inventoried and two will be selected for further examination. More details on background topics concerning these methods and the methodology will be provided as well.
- Chapter 3 describes the implementation of the existing methods and the extension to allow for classification over detection.
- Chapter 4 showcases the results from the various implementations. All of these are supported by a short discussion with regards to the various requirements and accuracy measures.
- Chapter 5 concludes the research and summarises the findings.
- Chapter 6 lists possible relevant new approaches or angles to problems encountered within the analysis, as well as areas of research which could result in new solutions or findings.

2 THEORETICAL FRAMEWORK

To assess the various remote sensing techniques and damage classification methods, a theoretical framework has to be established. This framework will allow the organisation of methods and technologies and will allow the selection of appropriate approaches in various circumstances. The framework presented here is based on literature research from Dong and Shan [2013] and Kerle et al. [2008] in combination with requirements from the field.

2.1 DISASTER

The International Federation of Red Cross and Red Crescent Societies [2017] defines a disaster as follows:

> *"A disaster is a sudden, calamitous event that seriously disrupts the functioning of a community or society and causes human, material, and economic or environmental losses that exceed the community's or society's ability to cope using its own resources."*

These characteristics in this definition emphasize the need for timely intervention by others outside a community or area to help and support those affected. Ray Shirkhodai notes in Al Achkar et al. [2008, p. i] that a rapid overview of the situation and extent of damage is necessary to achieve this goal, as is corroborated by Okada and Takai [2000] and Schweier and Markus [2006]. This rhetoric is an abstract approach to the problem, in reality every step within the DRM cycle have separate requirements for information regarding to damage. The faster more detailed information is available, the better. However, the first phase of the DRM, Search and Rescue, requires much less detail for teams to be send into the field. This is a limitation set forward by the time necessary for the procurement of data and information. The result is a curve closely linked to time, in which the data detail requirements rise as time passes. This research focusses on the emergency relief and rehabilitation phase, towards recovery [figure 2.1] [Wisner and Adams, 2002; Crutchfield, 2013]. In these first week after a disaster, the needs for information change from a course overview per area affected, to highly detailed information on a block or even building level [Ozisik, 2004].

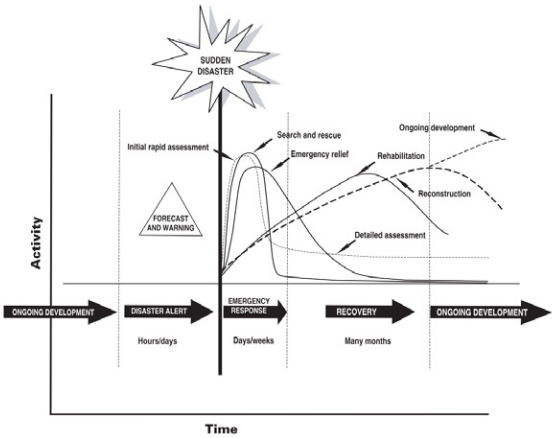


Figure 2.1: DRM activity cycle, including ongoing development. [From: Wisner and Adams [2002, p. 19]]

To gather this data with minimum risk to aid workers, the international community established the so-called International Charter [Bessis et al., 2003]. With resources from various space agencies around the world it became possible to use SEM to provide rapid situation awareness after a disaster. However, as described by Kerle et al. [2008] there are a diverse selection of other remote sensing techniques which can be used for the collection of data, which could make

other forms of Emergency Mapping (EM) possible. This research will take into account this broader approach to remote sensing, looking at advances in UAV and shrinking data collection technologies as new data sources in the immediate aftermath of a disaster.

2.2 DETECTION, CLASSIFICATION, AND ASSESSMENT

Besides spatial resolution scales of information [house, block, neighbourhood], the results of EM can also be categorised in various information densities. For the purpose of this research these have been categorised as follows:

- Damage detection
- Damage classification
- Damage assessment

The lowest information density considered is damage detection [figure 2.2]. In this form of EM the focus is to differentiate between buildings with and without damage. This comparable to the studies that can distinguish between two grades of damage as described by Dong and Shan [2013]. An example of this is Wang and Jin [2012], in which image classification is used to derive the difference between buildings damage and not damaged by a disaster.

One step up from detection is the damage classification as used within this research [figure 2.3]. For classification, various grades of damage need to be considered. Dong and Shan [2013] describe this as three grades or more. They also provide a framework for five damage classes that can be used in achieving damage classification. These classes are derived from Grünthal [1998], however the ambiguity introduced with five classes, especially between the three middle grades, is not beneficial for humanitarian aid organisations. Therefore, the classification method proposed by Al Achkar et al. [2008], known as the Baker, Achkar, Raymond method (BAR), will be used in this research. This framework is focused on damage inflicted to buildings, allows for sufficient variation between damage grades and has example of damage induced by wind. The four classes identified within the BAR method are: [1.] No visible damage, [2.] Minimal Damage, [3.] Significant Damage, and [4.] critical damage, examples of the damage extend are available in figure 2.3 and Al Achkar et al. [2008]. A simplified variation of this is also used by the NLRC and is used within this research due to the availability of data. [Simplified approach available on 510.global].

The last resolution is damage assessment [figure 2.4]. This requires more insight in the damage caused, as well as interior damage which is usually not observable in EM or SEM. Therefore people on the ground are required to do thorough investigation into the damage caused. The exception to this is the use of high resolution oblique imagery, which in some cases might be able to distinguish damage in the vertical plane [Vetrivel et al., 2016a], however most damage assessment will need to be done by humans with guides from governments, like from The Federal Emergency Management Agency [2016].

Classification

Statistical classification is the methods for organising various samples in the right classes [Theodoridis and Koutroumbas, 2009]. Humans are able to discern various features of an object to decide the classification. To achieve the same with the help of machine intelligence, various methods have been described. These range from linear classifiers to Support Vector Machines and Neural Networks. These can be subdivided in Supervised, Semi-Supervised and Unsupervised classification algorithms, based on the amount of pre-existing knowledge [Theodoridis and Koutroumbas, 2009]. In this research two specific approaches to this are researched, an empirical approach, in which Bayes Decision Theory is used as well as a Convolutional Neural Network (CNN). In Bayes Decision Theory an assumption of the distribution of features is made to classify others. A-priori (or existing) knowledge is used to inform the classification process to make decisions [Theodoridis and Koutroumbas, 2009]. With a CNN approach, little a-priori knowledge is used to train an algorithm to classify features.

Figure 2.2: Example of damage detection [Based on: Netherlands Red Cross (11 Sept. 2017), Cole Bay - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

Figure 2.3: Example of damage classification [Based on: Netherlands Red Cross (11 Sept. 2017), Cole Bay - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

Figure 2.4: Example of damage assessment [Based on: Netherlands Red Cross (11 Sept. 2017), Cole Bay - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

2.3 METHOD ASSESSMENT FRAMEWORK

The unexpected nature of a natural disaster, as described in section 2.1, make it hard to prepare; the collection of datasets can therefore, usually, not be planned in advance. However, time is of the essence in a disaster situation, both for response and data acquisition. The first is exemplified by the need for information on the magnitude of disaster. Knowledge regarding the extend of an effected area is paramount for the planning of relief operations [Al Achkar et al., 2008; Schweier and Markus, 2006]. Which requires timely data, this is an conceptualisation of the time it takes to get data to the user while it is, still, or most valuable [Christopher and Doeglas, 2015]. In the case of disaster this can be defined as directly when available. The faster data is available for analysis, the quicker people affected by a disaster can be helped. The latter is described by Christopher and Doeglas [2015] as window of opportunity. This describes the short amount of time certain data can be collected, this may vary per disaster. An example of this is that in case of a hurricane, clouds can obscure the struck area. Furthermore, the accuracy of data is a consideration as well. For the very first relief operations, global data of damage will suffice as these kind of operations require a specific time to set up as well. Very rapidly after that more detailed information is necessary for the planning of future operations and long term relief. Reflecting back on the requirements of time and resolution, it can be concluded that the chosen remote sensing techniques are very dependent on the resolution they can offer as well as the timely manner they can do that in. The methods chosen for implementation will need to reflect this as well.

An extensive analysis and description of remote sensing techniques can be found in Kerle et al. [2008]. The various specifications, capabilities, operation advantages and limitations, and examples are presented per system. This allows for a good overview of all the possibilities. However, humanitarian organisations do rarely have the opportunity to choose out of all options as they require specialist operators or equipment. Those are most of the time neither available in disaster areas which limits the options. Exceptions on this rule is the availability of satellite data that becomes more open for humanitarian organisations around the world through projects like the International Charter [Voigt et al., 2016]. This program allows all involved in relief operations to gain access to satellite data and subsequent information. The advances of portable UAVs and smaller capture technologies also allows humanitarian organisations to quickly gather Very High Resolution (VHR) imagery of areas affected by disasters. They provide an economical substitute for regular aerial surveillance [Nex and Remondino, 2014] and are already more implemented in disaster situations [Lieshout, 2017; Johnson, 2017].

Dong and Shan [2013] give an overview of various building damage research up until 2013. A summery of the various techniques and data types is also presented. The classification of methods is done on the basis of data, collection platform and amount of damage levels that can be detected. This subdivision allows for the selection of appropriate methods in varying situations. However, the clear absence of any indication of accuracy of methodologies only allows for an overview of the field of research and less for selection for implementation in operational procedures.

There is no perfect framework that would classify the various methods on their feasibility in specific disaster situations. This would allow humanitarian data specialist to select the best suitable solution for specific disasters. The main factor from literature on which methods can be classified is accuracy, however for the feasibility of an implementation for humanitarian context other aspect have to be considered as well. Due to the window of opportunity, the timely acquirement of data is critical. Various acquisition techniques described by Kerle et al. [2008] are therefore not available. A similar restriction is set due to monetary restrictions in a disaster situation. The main source of remotely sensed data for humanitarian organisations are: Satellite optical imagery provided through the International Charter, Satellite SAR data provided through various governmental space agencies, and UAV optical imagery collected by delegates in the field. Collection method is therefore part of the framework, combined with acquisition time. Resolution is less of an issue for various of the methods, however the best resolution for humanitarian action would allow identification of damage on a building level without extensive extra data. For an adequate selection of methods a combination of the above, with time indication, resolution, data, and accuracy is necessary. Some of these indicator will be connected and change dependently, however an overview in which all is displayed allows for a quicker overview of methods connected to remote sensing techniques. Table 2.1 shows the various classes which are of importance for the feasibility of methods in humanitarian operations.

Table 2.1: Method assessment framework parameters

Requirement	Description
Accuracy	Percentage of building damage classified correctly.
Acquisition time	Period from disaster to acquisition of data, travel time of delegates not included.
Acquisition method	The technique used for the procurement of the data, mostly limited by financial and time restrictions.
Resolution	The resolution of the data and information retrieved from method.

2.4 DATASETS AVAILABLE

As described in section 2.3, the availability of remotely sensed datasets is limited in a disaster situation. However, this is changing with more emerging techniques. An overview of datasets usually available in a disaster can be provided taking operational limitations into account. This section, summarised in table 2.2, will provide an overview of these datasets and their advantages and disadvantages. These are based on operational restrictions from the 510 team at the NLRC. A clear similarity between the datasets is the use of governmental acquiring techniques [Satellite acquired data and Digital Elevation Model (DEM)] or individual acquiring techniques [UAV data]. These datasets are usually available due to the limited resources necessary in collection, as no specialised acquisition companies are involved. The access to satellite datasets is regulated in the international charter, which ensures the availability of these datasets in case of a disaster. Satellite imagery usually provides extensive coverage for an affected area, however it lacks on interpretability and requires specialist knowledge. More NGOs and humanitarian organisations acquire UAVs to quickly collect data after a disaster or in preparation of expected disasters. This allows for focused collection of data with high resolution, however the technique is extremely limited in the geographic coverage. All datasets are a balance between speed and interpretability. This trade-off by design becomes part of the methods for damage detection.

Table 2.2: Datasets usually available after a disaster with resolution per pixel and respective advantages and disadvantages.

Dataset	Resolution	Advantages	Disadvantages
Satellite Optical	2x2m - 0.3x0.3m	• Large coverage • quickly available • centralised approach	• High resolution is expensive • affected by atmospheric conditions • resolution can be limiting in interpretation
Satellite SAR	20x20m - 1x1m	• Large coverage • quickly available • centralised approach • less affected by atmospheric conditions	• High resolution is expensive • difficult to interpret • specialist necessary for information deduction
UAV Optical	0.2x0.2m - 0.02x0.02m	• High resolution relatively cheap • easily interpretable	• Small coverage • slow acquisition • affected by weather conditions
UAV DEM	5x5m - 1x1m	• High resolution relatively cheap	• Small coverage • slow acquisition • affected by weather conditions
Global DEM	30x30m	• Freely available world wide	• Low resolution • without other data sources fruitless in damage detection • only available data from before the disaster

2.5 EXISTING TECHNIQUES

Table 2.3 summarises the findings of various research of the past 15 years into the [semi-] automated detection or classification of damage to buildings. These have been selected by the inclusion criteria as described in section 1.3.1. This is a non-exhaustive summary but allows for a clear overview of the various approaches. Only [semi-] automated solutions are taken into consideration, all information about technique, resolution and accuracy has been taken from the respective paper, while acquisition time has been related to sensing technique or taken from Kerle et al. [2008]. From this table it is clear that most research is already focused on the datasets mostly available in disaster situations. However, most of the research is dedicated to the detection of damage and not classification. This proves a gap in the research which could benefit the humanitarian organisations.

Table 2.3: Overview of existing methods [Alphabetic order]

Method	Description
[Antonietta et al., 2015]	Various semi-automated approaches based on satellite optical data from typhoon Hainan. Methods consist of [1.] Multi-temporal change detection based on various algebraic and image transformations, [2.] and Mono-temporal Segmentation.
[Brunner et al., 2010]	Damage detection using VHR optical and VHR SAR datasets. Method based on image matching and height estimation. Limited to rectangular buildings and tested on limited dataset.
[Li et al., 2017]	Method for building and damage detection using VHR satellite optical imagery. Based on Chinese Restaurant Franchise machine learning, combined with unsupervised approach for earthquake damage detection.
[Martha et al., 2015]	Through the use of stereographic panchromatic images, a 3D overview of the 2011 Sikkim earthquake was formed. The method used block triangulation and stereographic visual change detection for geologic and infrastructural damage detection. A lack of image signatures made damage detection on building level scales not possible.
[Menderes et al., 2015]	A change detection algorithm based on DEM data from pre-event and post-event. The detection is based on a threshold on the difference in coherence between a Digital Surface Model and Digital Terrain Model in both situations.
[Ozisik, 2004]	Damage identification on VHR optical imagery was achieved through: colour indices, edge intensity, and variance of edge intensity. An implementation of the Maximum likelihood classification algorithm allowed for improved damage identification.
[Samadzadegan and Rastiveisi, 2005]	An implementation in which vector features are used for the classification of damage. Genetic Algorithm and Fuzzy Inference System were used to handle optimum feature selection and to eliminate ambiguity in the process.
[Vetrivel et al., 2016b]	Multiple methods are proposed for the use of CNN. The algorithms are used for the identification of features and the detection of damage. Various CNNs are used with varying results however all work with high accuracies in detection. More in section 2.5.2.
[Yun et al., 2015]	This method uses radar [SAR] coherence change detection for the creation of damage proxy maps. These maps highlight damage and are used for visual interpretation of the information. SAR preprocessing and histogram matching are used for the correct identification of damage. More in section 2.5.1.

Methods considered for this research are limited to those that can be used with the data available and that have considerable accuracies in the prediction of damage classes. From table 2.4 it is clear that there are various valid options for the detection of damage and even possibilities for the classification of damage.

The selected methods for this research are Vetrivel et al. [2016b] and Yun et al. [2015]. These showed most promise in the research for the summary represented in table 2.3. Vetrivel et al. [2016b] uses a machine learning approach and achieves a high accuracy. This methods is similar, using more modern techniques to Ozisik [2004] which was able to achieve classification

Table 2.4: Framework analysis of existing methods. * marks classification, and ° marks detection as described in section 2.2. Info scale is the scale of the resulting information; B = Building, BL = Building Block, and N = Neighbourhood. [Alphabetic order]

Method	Technique	Resolution	Acq. time	Accuracy	Info scale
[Antonietta et al., 2015]°	Satellite Optical	0.8x0.8m	6 days	70-80%	B
[Brunner et al., 2010]°	Satellite Optical and Satellite SAR	0.6x0.6m 1.1 x 1.0m	6 days	90%	B
[Li et al., 2017]°	Satellite Optical	0.6x0.6m	6 days	70%	B
[Martha et al., 2015]°	Satellite Optical	0.6x0.6m	6 days	n/a	N
[Menderes et al., 2015]°	Aerial Optical	0.3x0.3m	Days	90%	BL
[Ozisik, 2004]°	UAV Optical	n/a	Hours	70-80%	B
[Samadzadegan and Rastiveisi, 2005]*	Satellite Optical	2.44x2.44m	3 days	74%	B
[Vetrivel et al., 2016b]°	UAV Optical	n/a	Hours	80-90%	B
[Yun et al., 2015]°	Satellite SAR	2.7x22m	6 days	n/a	BL

of datasets available for this research. The implementation of the method will provide insights into the use of machine learning for classification of damage and the transferability of such method. Yun et al. [2015] is the only paper in the list that can cite users of the data and provides analysis using this method for other disasters. Providing a strong base for the possible transferability of method, which would allow it to be used in other disaster situations. The difference in data used for these approaches to the problem also allows for comparison of methods on various sorts of data. Within this comparison some other approaches to existing datasets will be used, these are further explained in chapters 3 and 4.

2.5.1 Equalisation and subtraction - Yun et al. [2015]

In their paper, Yun et al. [2015] develop a method for the detection of damage after a disaster. This methods is tested on the 2015 Gorkha earthquake in Nepal, but their database shows the application on disasters of other origin as well [JPL, 2018]. Their guiding principle is the change in coherence between SAR images. Coherence maps are used to identify noise in images [Ferretti et al., 2007]. This approach by Yun et al. [2015] is based on the notion that the structure within an image changes when destruction occurs, hence a change in noise within the coherence map. This can be visualised in an optical image [figure 2.5]; in which the human brain can identify an building by the cohesion between pixels in an image, while a damaged building can be identified by the incoherence of pixels within the image.

(a) Black and White coherence of image pre-event

(b) Black and White coherence of image post-event

Figure 2.5: Structure within an image, pre-event (Left) and post-event (Right) damage sustained [From: Left: IGN France (16 Feb. 2017), Saint-Martin Orthoimage [georeferenced image] — Right: Netherlands Red Cross (15 Sept. 2017), Quilletor Dr - Sint Maarten [georeferenced image] — both used under CC-BY4.0 as part of Open Imagery Network, retrieved from `www.openaerialmap.org`]

To identify this structural change Yun et al. [2015] use a equalised coherence maps, one pre-event and one post-event. These maps are created from CSK and ALOS-2 datasets with respec-

tive resolutions of 3 meter and 10 meter. Equalisation is achieved through the matching of the histograms of the coherence maps, resulting in statistically equal images in which specific pixel values can vary. Empirically a threshold is chosen to differentiate between noise and change for the detection of damage. Figure 2.6, shows all steps of the process. The results from this approach are used by various agencies around the world, as it allows for a quick visual overview of the damage sustained in a disaster, which can be used for prioritisation of humanitarian aid. The damage sustained is not quantified or projected on [building] features, which limits the process to the interpretation of a user.

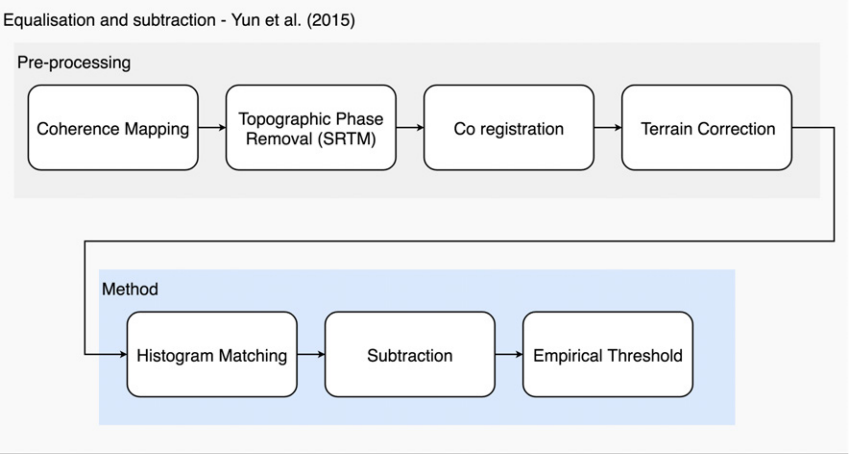


Figure 2.6: Method used by Yun et al. [2015]

InSAR and Coherence

Satellite Interferometric Synthetic-Aperture Radar (InSAR) is a particular use of radar, namely SAR. Radar makes use of electromagnetic pulses in the microwave spectrum, therefore it is considered an active imaging sensor [Hanssen, 2001; Ferretti et al., 2007]. These properties make it extremely valuable for the imaging of the earth in sub-ideal conditions [night-time or under cloud cover]. As described in section 2.3, time is of the essence in a disaster situation, the ability to gather data under sub-ideal conditions is extremely beneficial for a humanitarian mision. InSAR uses two observation of the same part of the earth to calculate the interferometric observations [Ferretti et al., 2007]. These images have a slight offset, which can be achieved by either using two sensors, spatially offset from each other on the same carrier, or the same platform observing the same area twice. This multiplication of the phases of either observation can be used for various remote sensing applications, most notably the generation of a DEM, or the deformation of the sensed area.

A coherence map is a product derived from the interferometric observation of earth in which the cross-correlation of the phases of either observation is estimated over a kernel [Hanssen, 2001; Ferretti et al., 2007]. The normalised value is a indication of coherence between observations, which can be used for the estimation of noise in observations, as well as change detection. Areas with little to no coherence can usually be correlated to vegetation or other natural environments in which change is observed between offset observations. Man-made structures are related to a high coherence as these are likely to remain similar over larger time spans or various angles.

Change detection

For the detection of change between two images various solutions have been developed over the years based on different approaches [Singh, 1989; Tewkesbury et al., 2015]. Yun et al. [2015] have chosen for a pixel based approach. This Univariate image differencing [equation 2.1] subtracts one observation from the other and transforms these to a positive number [Singh, 1989]. It is however noted that the threshold chosen for the detection or classification of change influences the outcome of this approach. Various methods for esthablishing the threshold have been developed, ranging from empirically set thresholds to the assessment of the meta-data

and various specifics of the images used [Singh, 1989]. Yun et al. [2015] use an empirically set threshold, in which ground truth data can be used to adjust this parameter.

$$Dx_{ij}^{k} = x_{ij}^{k}(t_2) - x_{ij}^{k}(t_1) + C \tag{2.1}$$

2.5.2 Convolutional Neural Network - Vetrivel et al. [2016b]

Vetrivel et al. [2016b] describe a broad approach to the detection of building damage through the use of machine learning, CNN and Support Vector approaches. Their method is tested on various earthquake datasets, i.a. 2010 Port-au-Prince Haiti, 2012 Mirabello Italy, 2015 Kathmandu Nepal. Their proposed method is a comprehensive approach using oblique UAV imagery. Various characteristics of the imagery are used in the approach which mainly relate to the use of 3D point-cloud features. Oblique imagery allows for the creation of 3D point clouds through a photogrammetric process. The fusion of this point cloud with the imagery is the bases for a Support Vector Classification of the damage. However, in the design process various other methods using machine learning have been tested with varying results, the most promising is the detection of damage features using CNN.

Three approaches to the use of CNN have been defined by Vetrivel et al. [2016b]. The first approach is the definition of a new CNN and training from scratch. For this method the architecture [figure 2.7] and various settings are provided in the paper. Secondly a pre-trained network from Krizhevsky et al. [2012] has been modified and retrained for the classification of damage. Lastly this retrained network is used for the extraction of features to be used in combination with the 3D point cloud. All CNN approaches have been validated for damage detection on the same datasets with respectively 7130 and 5414 classified samples. No significant accuracy increases where found for any of the approaches, all scored within the 90th percentile. These accuracy scores and model transferability quoted in the paper make this a valid approach for similar datasets, in the case of this research ortho-rectified imagery.

CNN architecture for training from scratch

Layer number	Layer name	Properties
1	Input layer	Input image patch size: $100 \times 100 \times 3$
2	Convolutional	Number of filters: 9; filter size: 11×11
3	RELU	–
4	Maxpooling	Pool size 2×2
5	Convolutional	Number of filters: 21; filter size: 7×7
6	RELU	–
7	Maxpooling	Pool size 2×2
8	Convolutional	Number of filters: 41; filter size: 3×3
9	RELU	–
10	Maxpooling	Pool size 2×2
11	Fully connected	Size: 1×256
12	RELU	–
13	Dropout	Dropout ratio: 0.5
14	Fully connected	Size: 1×100
15	Fully connected	Size: 1×2
16	Softmax	–

Figure 2.7: CNN architecture [From: Vetrivel et al. [2016b]]

Due to the various restrictions of this research only one approach to CNN for damage detection will be considered. As there are no accuracy benefits for choosing any, other circumstances will be taken into consideration. The simple approach to a new CNN, which can be trained from scratch will guarantee the model awareness of the data, increasing the changes of correct classification. A similar sized training set is available and the new approach allows for fine-tuning. The steps necessary to achieve an approach as described by Vetrivel et al. [2016b] is summarised in 2.8.

Neural network

Neural networks, and therefore also CNNs, have come from the field of deep learning [Hardesty, 2017]. This approach to machine learning is roughly modelled on the human brain. A

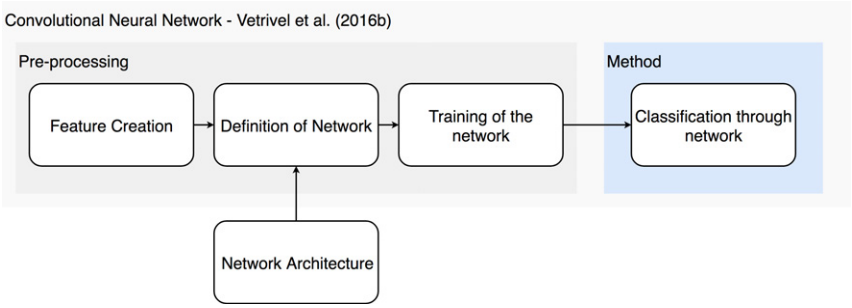


Figure 2.8: CNN implementation from scratch, Vetrivel et al. [2016b]

network of connections is being trained to perform a specific task, for example the classification of data. The first mentions of such an approach to machine learning date back to the second half of the last century [Lawrence et al., 1997; Hardesty, 2017]. CNNs have evolved to become a distinct approach within Neural Networks, in which various layers guide data through filters to reduce data and deduct weights and thresholds. A simple CNN has been shown in figure 2.9, in which the various convolutional layers reduce the images by kernels, activated with a specific algorithm. These so-called pooling algorithms are used to reduce the data. This pyramid structure allows the last layer, usually a regression algorithm, to determine the class assigned to the specific feature [Lawrence et al., 1997].

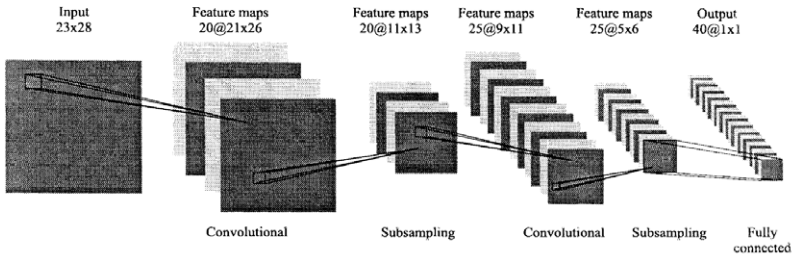


Figure 2.9: A simple CNN. [From: Lawrence et al. [1997]]

2.6 RELATED BACKGROUND INFORMATION

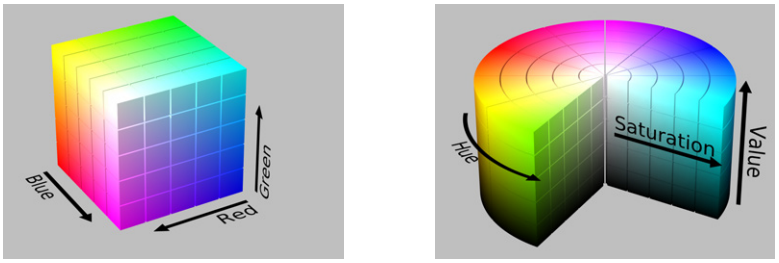
2.6.1 Colour space

For the comparison of approaches as mentioned in section 1.3, various colour spaces are used to gain a simplification of the dataset. Colour spaces are methods used for standardisation of the description of colour. These are all based around the sensations regarding colour perception as described by Hunt and Pointer [2011]; Ford and Roberts [1998]. These are:

- Brightness of a colour, regarding the variance in light
- Hue of a colour, the similarity between colour, usually expressed in Red, Green, and Blue (RGB)
- Colourfulness of a specific area, the amount of hue in a feature
- Lightness, this is a description of brightness referenced to a white area
- Chroma, is the colourfulness referenced to lightness
- Saturation, is the colourfulness relative to the brightness.

Ford and Roberts [1998] describe the various colour spaces available for use. These are all used for various applications and are simplifications of the sensations of colours for humans. These simplifications are necessary for computers to represent colours as data. The colour spaces considered in this research are RGB and Hue, Saturation, and Lightness (HSL), specifically Hue, Saturation, and Value (HSV). RGB is the most commonly used colour space, most cameras and

monitors represent colour in this way [Ford and Roberts, 1998]. It separates the colour into 3 bands described by Red, Green, and Blue, while all other colours are a mix of these 3. It is easy to implement and understand [figure 2.10a], but are non-linear with human sensations of colours. HSL is a group description for colour spaces in which the first two parameters are Hue and Saturation, while the last is on of the other options for sensations of colour as described in the list above. For this research HSV has been chosen. Value is used to describe the colourfulness of an area [figure 2.10b]. As RGB these colour spaces are non linear, and can therefore be linearly transformed both ways using pseudo-code as described by Schwarz et al. [1987] [Code block 2.2]. However, they are more intuitive and comparable to the sensation of colour for humans. This comparability allows for introduction into the classification process of damage. Visual interpretation is the preferred method used for damage detection [chapter 1] and a more comparable approach to colour definition could help the introduction of automated approaches. HSV is more computationally effective than HSL and has therefore been applied in this research [Schwarz et al., 1987].



(a) RGB colour space **(b)** HSV colour space

Figure 2.10: Colour spaces visualised in 3D. [From: Horvath [2011]]

Psuedo code RGB to HSV from Schwarz et al. [1987]

$$F = H * 6 - Floor(H * 6)$$
$$T_1 = V * (1.0 - S)$$
$$T_2 = V * (1.0 - (S * F))$$
$$T_3 = V * (1.0 - (S * (1.0 - F)))$$

$$\begin{aligned} &case\ floor(H * 6)\ mod\ 6\ of \\ &\quad 0 : (R, G, B) = (V, T_3, T_1) \\ &\quad 1 : (R, G, B) = (T_2, V, T_1) \\ &\quad 2 : (R, G, B) = (T_1, V, T_3) \\ &\quad 3 : (R, G, B) = (T_1, T_2, V) \\ &\quad 4 : (R, G, B) = (T_3, T_1, V) \\ &\quad 5 : (R, G, B) = (V, T_1, T_2) \end{aligned} \tag{2.2}$$

2.6.2 Inter-rater statistics

Cross classification is a popular statistic to describe the accuracy of a method for prediction [Warrens, 2011]. The description of these inter-rater statistics have been developed throughout many academic fields. These have originally been developed to determine the accuracy of two human raters in various research fields [Bhowmick et al., 2008]. However, with the growth of machine learning and other predictors, these indices have been used more to describe their performance with regards to ground truth. This more elaborate description of cross classification or inter-rater statistics is used in this field to compensate for the possibility of unbalanced input and outcomes, in which a larger subset of a specific class might affect the accuracy, in which only the correct answers are considered [Equations 2.3].

$$\text{accuracy} = \frac{\text{correctly classfied samples}}{\text{total classified samples}} \tag{2.3}$$

For this research, two methods to describe inter-rater statistics have been chosen. These are the so-called F1-score [Sørensen–Dice coefficient] and the Cohen Kappa coefficient [Dice, 1945;

Sørensen, 1948; Cohen, 1960].Both of these are popular and established methods for the description of inter-rater accuracy for the evaluation of learning algorithms.

Cohen Kappa

The Cohen Kappa is developed in 1960 by Jacob Cohen at New York University [Cohen, 1960]. It's first implementation was used to describe the categorisations [nominal descriptors] of 2 observers in clinical-social-personality studies. The kappa score is described by Cohen [1960] as an indicator of agreement. In the definition of the Cohen Kappa coefficient, it compares the agreement between observers compared to random assignment [Warrens, 2011]. From which the statistic can handle imbalances between classes as well as multiple classes [Kampakis, 2016]. These characteristics make it useful for the description of performance in nominal machine learning techniques, and therefore a suitable descriptor for the performance of the methods considered in this research.

As described by Warrens [2011], the Cohen Kappa coefficient is given in equation 2.4. Which holds true if the categories are in the same order for both observers, such that the row and column totals can be described as in equation 2.5.

$$\kappa = \frac{P - E}{1 - E}$$

Where

$$P = \sum_{j=1}^{n} p_{jj} \quad \text{and} \quad E = \sum_{j=1}^{n} p_j q_j \tag{2.4}$$

$$p_j = \sum_{k=1}^{n} p_{jk} \quad \text{and} \quad q_j = \sum_{k=1}^{n} p_{kj} \tag{2.5}$$

F1-Score

The F1-score or Sørensen–Dice coefficient is based on the work of Dice [1945], in which the association between species is measured, and Sørensen [1948], in which the same is described for plants. This method considers the harmonic mean of both recall and precision [Sasaki, 2007]. In this descriptor of inter-rater statistics, the precision is considered the rightly classified as proportion of the total classified within one category; recall is considered the inverse of precision as described as the proportion of the rightly classified over the ground truth total within one category [Buckland and Gey, 1994]. This description was generalised by Chinchor [1991] in which the classifier could be varied by an β constant. As described by Sasaki [2007], the F1-score reflects to $\beta = 1$, with a resulting equation 2.6, in which P is precision and R is recall.

$$F = \frac{2PR}{P + R} \tag{2.6}$$

The development of this method for cross classification in information retrieval by Chinchor [1991] and the broader statistics of the data with the inclusion of precision and recall per category, allow for an in-depth understanding of the results of machine learning algorithms. These characteristics suit this research, as it provides more understanding of the results per classification category than the generalised Cohen Kappa coefficient. The inclusion of the other precision and recall indicators is, however, paramount.

Critique on inter-rater statistics

Both the F1-Score and Cohen Kappa coefficient are not unscrutinised [Feinstein and Cicchetti, 1990; Pontius and Millones, 2011; Powers, 2015]. The use of Cohen Kappa leaves some room for ambiguity, however is still the most accurate method to measure agreement while factoring in random assignment. It can however be bias in some extremely unbalanced test sets in which the categories are skewed to a particular classification. As described by Powers [2015], the F1-Score should only be used to look at one category at a time, this because it is very sensitive to majority classes. A balance between methods is therefore necessary. The inclusion of recall and precision can be used to further explain the Cohen Kappa coefficient, while taking into account limitations from both methods.

3 IMPLEMENTATION

This chapter describes the implementation of the methods for comparison of the performance in a disaster situation. The implementation is mainly focused on the comparison of accuracy of the models. A discussion on the implementation, including the humanitarian aspects, can be found in section 4.5.

3.1 PREPARATION

In the preparation of this research various background information has been collected, the results of this research have been represented in chapter 2. These results are based on literature studies and practical experience from the field regarding the use of damage data in the wake of a disaster. However, the methods selected for comparison have also been selected for their variety in data sources used. This guarantees the availability of either dataset in case of unseen disaster situations. To thoroughly compare the impact of data and methods, table 3.1 has been developed for easier comparison. The goal of this matrix is the comparison between the influence of data and methods. Further more it allows for the exploration of comparable methods between datasets to achieve new methods and possibilities for classification. As the method by Yun et al. [2015] is developed for SAR data, a comparable method will need to be found or developed for optical data. Vice versa for the method described by Vetrivel et al. [2016b], in this case the methods are written for optical data and should be considered for SAR data. Section 3.3 will elaborate on these comparable approaches to damage detection and how these might apply to damage classification.

Table 3.1: Design matrix for combination of methods and data. Combinations have been abbreviated for continuity in text. These combinations are: Equalisation and Subtraction on Optical UAV data (ESO), Equalisation and Subtraction on Satellite SAR data (ESS), CNN approach for Optical UAV data (CNO), and CNN approach for Satellite SAR data (CNS).

	Optical UAV Data	Satellite SAR Data
Equalisation and subtraction	ESO	ESS
Convolutional Neural Network	CNO	CNS

3.2 TOOLS AND DATASETS USED

Tools and data are necessary to facilitate the implementation of the methods in design matrix [Table 3.1]. the following sections provide a short overview of the tools and data used in the implementation. These relate to the methodology described in section 1.3 and methods selected in section 2.5.

3.2.1 Tools

The tools in table 3.2 are select for the assessment of the existing methods as described in section 2.5. These are based on Open Source or Free to Use software packages as these are also available for deployment in the field by the NLRC. Furthermore allow these package for more customisability for implementation with specific datasets. All code is available on GitHub [1].

1 https://github.com/DKersbergen/AutomatedDamageClassification

Most of the implementation has been done using Python, this universal programming language is compatible with the other tools available for this research. Some steps required specialised algorithms, these steps are the following: [1.] To accommodate for the use of [visual] geo-programming, Qgis has been used. All code was written in Python but based on internal analysis methods from Qgis. Other geo-programming components have been done in Python using GDAL and OSR. [2.] For the specialised pre-processing and analysis performed on SAR imagery, the free to use European Space Agency (ESA) processing tool SNAP has been used. [3.] Lastly, to perform deep learning within the Python language, the module tflearn for Tensorflow has been used. This module has compatibility problems with Windows, therefore a Windows Subsystem for Linux (WSL) has been used to accommodate this tool.

Table 3.2: Tools and modules used in this research [Alphabetic order]

Tool	Used for	Notes
Python	General programming language, tflearn module available as implementation of Tensorflow	Most steps are done through automation via Python to ensure quick overviews of large datasets.
Qgis	For visualisation, change detection, and Geo-referencing	–
SNAP	To prepare SAR datasets for change detection	Free to use [for non-profits] SAR processing software from ESA.
Tensorflow	Machine learning tool, with pre-trained networks available	Used through a WSL install of Ubuntu.

3.2.2 Data

The data available for this research has been summarised in table 3.3. These datasets have been guidance in the selection of methods to be implemented, as put forward in section 1.4.

The specifications of the datasets are provided in table 3.3, including the sources. All datasets are available through the various portals. Ground truth data is excepted as it might contain sensitive data about the people on the island of St. Maarten. The portals for the data are [1.] OpenAerialMap [2] for the UAV and aerial datasets, [2.] Copernicus Scihub [3] for the SAR datasets, [3.] the Shuttle Radar Topography Mission (SRTM) dataset has been acquired via a plug-in to the SNAP tool, [4.] Turbo Overpass [4] has been used to acquire the building vector layer from OpenStreetMap [5], and [5.] lastly the STOA data for damage classification from manual visual interpretation was retrieved from the Copernicus Emergency Mapping Service (EMS) Hub [6]. Furthermore, it must be noted that not all UAV optical data is of the exact same resolution or from the exact date, therefore the average has been indicated in the table for the resolution and a interval for the dates. Lastly, the ground truth data is derived, by the 510 team of the NLRC, from VHR UAV imagery by visual interpretation following the principles discussed in section 2.2. Per dataset an example is provided.

Table 3.3: Datasets with specifications, available for this research [Alphabetic order]

Platform	Technique	Resolution	Acquisition date	Source	fig.
Aerial	Optical	0.2x0.2 m	20-Feb-2017	IGN	3.1
Satellite	SAR	2.7x22 m	11-Aug, 23-Aug, 16-Sept-2017	Sentinel	3.2
UAV	Optical	0.04x0.04 m	11-Sept-2017 – 2-Oct-2017	NLRC – RescUAV	3.3
Satellite	DEM	30x30 m	unknown	SRTM	n/a
–	Building outline	building	2-Oct-2017	OpenStreetMap	3.4
–	Ground truth	building	2-Oct-2017	NLRC	3.4
–	STOA	building	14-Sept-2017	Copernicus EMS	3.4

2 www.openaerialmap.org
3 scihub.copernicus.eu.org
4 https://overpass-turbo.eu/
5 www.openstreetmap.org
6 https://emergency.copernicus.eu/mapping/

Figure 3.1: Example of aerial dataset [From: IGN France (16 Feb. 2017), Saint-Martin Orthoimage [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

Figure 3.2: Example of SAR dataset [From: Copernicus Open Access Hub (2017), ESA [georeferenced image] - used under open access by In-Orbit Commissioning Review, retrieved from scihub.copernicus.eu.org]

Figure 3.3: Example of UAV dataset [RescUAV (13 Sept. 2017), Philipsburg NE - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

Figure 3.4: Example of building dataset [From: OpenStreetMap contributors (2017), Philipsburg - Sint Maarten [georeferenced data], used under ODbL as part of OSMF, retrieved from www.openstreetmap.org]

3.3 RESEARCH IMPLEMENTATION

The research phase can be subdivided in the four subcategories described in table 3.1. This section will go into detail about the approach to each of these combinations of data and methodology, while the results will be presented in chapter 4.

3.3.1 Equalisation and subtraction

SAR data

For the ESS approach, the method is already elaborated by Yun et al. [2015] and in section 2.5.1. This method has two clear phases, preprocessing and method implementation. Due to use restrictions, the software from National Aeronautics and Space Administration (NASA), described by Yun et al. [2015], is not available outside the United States and therefore this research. However, the ESA also provides similar tools in their SNAP package, which is available freely for use by NGOs as explained in section 3.2. The methodology as described by Yun et al. [2015] can therefore still be implemented. The same steps will be followed for the pre-processing of the data, namely:

- Set building through the pairing of the Sentinel 1 SAR datasets. These sets are labelled pre-event [11th of August 2016 paired with 23rd of August 2017] and post-event [23th of August 2016 paired with 16th of September 2017].
- Coherence mapping with a window of size 3 pixels in the azimuth direction, proportionally scaled to the SAR data available.
- Topographic phase removal using SRTM DEM data.
- Co registration of the various swaths using telemetry data.
- Terrain correction for geo-referenced result using SRTM DEM data.

The result of these steps are two coherence maps, pre-event and post-event respectively. In the combination of these coherence maps, the method determines if damage is present. To do this the histogram of the second coherence map is matched to the first, before univariate image differentiation is applied. This method is not described in detail by Yun et al. [2015], so the following approach has been defined:

- The array size of the post-event coherence map is determined
- Both coherence maps are flattened to a single list of pixels
- All unique pixels values are listed for both datasets
- The cumulative distribution function is determined by normalising the cumulative sum of the amount of pixels in each map
- linear interpolation is used to match the statistics of the post-event map to that of the pre-event data

From here the matched coherence maps are subtracted from each other using univariate image differencing, in which each pixel value from the post-event coherence map is subtracted by the value of the pixel with the same X,Y coordinate in the pre-event coherence map [as described in section 2.5.1]. To visualise the change all change is considered absolute, in which an increase in pixel value can also be classified as change. Empirically a threshold will be set to determine the pixels classified as damage. As ground truth data is available this will be used to determine a more optimal threshold for the detection of damage. Yun et al. [2015] describe a masking technique which involved the human settlement index, this is used to focus the change detection on the areas in which people are living. However, as this research is focused on building scale level and building features available are, will this not be implemented and will the results be evaluated on a building scale.

Optical data

For the ESO approach similiar steps are implemented for the detection of damage using histogram matching and univariate image differencing. The optical data from UAVs differs from SAR data in various ways. First of all is the optical data available in products with 3 bands namely RGB. Furthermore, only one dataset is available from before the disaster [aerial optical imagery from the IGN], as well as one after the disaster [UAV optical imagery from the NLRC and RescUAV]. This requires a different approach the the pre-processing of the data before histogram matching and univariate image differencing can be implemented. To achieve a comparable approach, one band simplification of the dataset is applied. the optical imagery will be transformed to a HSV image, of which each band is considered separate. Though this process the image is changed from a mix of colours to a description of features within the colour space, as described in section 2.6. In this new description of the data, three key characteristics of colour are described, particularly the Hue of a pixel, the Saturation, and the Value. For change detection it is important to understand what feature changes in case of damage. As can be seen in figures 3.5 and 2.10b the saturation of an object describes the presence of colour. This clearly does shift in the case of damage sustained to a building. From the same figures, the hue of a building might change not as much, as the colours are roughly the same or of the same family in the pre-event and post-event imagery. Lastly the value, this seem to change quiet a bit during a disaster in which the amount of light reflected from a building clearly decreases as more light is bounced around between the rouble before it is bounced back, hence a decrease in lightness. For this method comparison all three reductions of information however are considered and taken into account in chapter 3. Considerations with all of the simplifications, from RGB to single HSV layers, concern the time shift between images, creating other artefacts or noise. An example of this is the value change due to a different lighting condition. However the histogram matching as used by Yun et al. [2015] does account for these changes in lighting, reducing the amount of artefacts or false positives.

(a) Colour representation of image pre-event

(b) Colour representation of image post-event

Figure 3.5: Change between pre-event and post-event colour. Structure within an image. [From: a: IGN France (16 Feb. 2017), Saint-Martin Orthoimage [georeferenced image] — b: Netherlands Red Cross (15 Sept. 2017), Quilletor Dr - Sint Maarten [georeferenced image] — both used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

The reduction from a three layered RGB dataset to a single layered map of the data allows for the implementation of histogram matching as described for the SAR data in the previous

subsection. However, as the pre-event and post-event datasets are not of the same resolution [table 3.3] an extra step of pre-processing is necessary before the univariate image differencing can be used. In this step the post-event image will be down-sampled using the algorithm defined by Lanzcos [represented by equations: 3.1], as it performs best along other re-sampling algorithms [Narendra and Madhukar, 2013]. From here an empirically set threshold can be used to determine the damage detection, which can be combined with an informed approach by optimising the threshold using ground truth data.

$$L(x) = \begin{cases} sinc(x).sinc(X/a) & \text{if } -a < x < a \\ 0 & \text{elsewhere} \end{cases} \tag{3.1}$$

Classification

To extent the method described by Yun et al. [2015], to allow for classification, an empirically set thresholds can be used to determine more classes than only detection. However, this is only possible if the groups of damaged features are clearly distinguishable in the datasets acquired. This does not translate to a clear identification by visual interpretation, but clear distribution along the pixels classified for change, with correct density functions. A clear density function can allow for the identification of the damage classes, as described by Theodoridis and Koutroumbas [2009] [figure 3.6]. Aided by the existing ground truth data on building level, the threshold will be set empirically to determine the correct classification for damage.

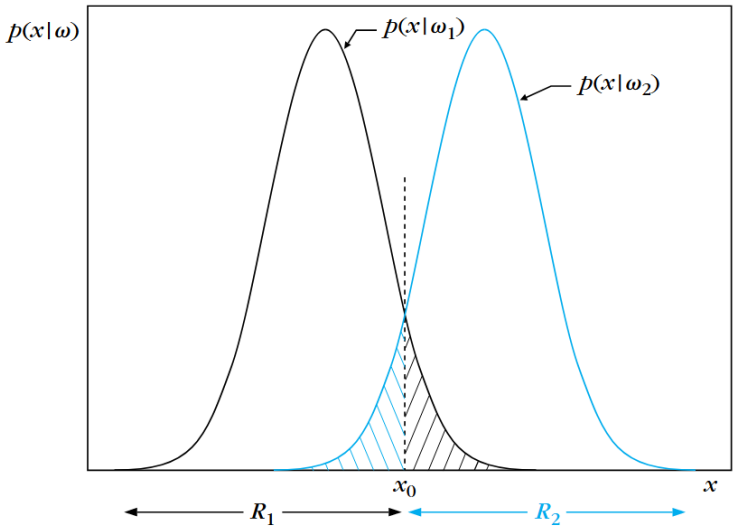


Figure 3.6: Probability density function used for object classification with two classes [From: Theodoridis and Koutroumbas [2009]]

3.3.2 Convolutional Neural Network

Optical data

For the CNO approach, the method described by Vetrivel et al. [2016b] is applicable. The steps remain relatively similar for this research, however as this is a comparison the CNN defined in the paper has been used. To detect damage on the optical dataset using CNN, the following steps will be taken.

- Feature creation
- Neural network training
- Classification of samples

Due to the availability of features from OpenStreetMap, the feature creation step will vary from the approach by Vetrivel et al. [2016b]. In this research a bounding box around the objects will be determined, based on the outline of the buildings acquired from OpenStreetMap via

the Missing Maps [7] initiative. For a neural network, inputs with the same dimensions are beneficial. To ensure all features are of the same size and dimensions, the bounding boxes are taken as squares around the building features. These bounding boxes are used to cut features from the post-event UAV optical imagery. The resulting features are then rescaled to be the same pixel dimensions, namely 100 x 100 pixels. These are introduced to the Neural Network as described in figure 2.7. 70% of this dataset is used for training the CNN, while the other 30 % is used as validations set, this to the parameters described by Vetrivel et al. [2016b]. The learning rate is set to 0.001 to allow for swift learning while ensuring no over-fitting occurs due to a large learning rate. The resulting trained network is saved and used as input for the classification of an unseen dataset. This unseen dataset is a set of features extracted from the main set before training. Geographically this set is defined by the area described in chapter 4. This set will be used in the results phase for comparison.

SAR data

In the CNS approach, the same definition as for the CNO approach can be used with regards to the use of a CNN for damage detection. In this approach, the features can be created from the post-event SAR data in the pre-processing phase, in contrast to the post-event UAV data described in the previous subsection. However, the data used in this research is not compatible with such approach as will be elaborated upon in chapter 3.

Classification

CNN methods are very suitable for extension of the classification classes, in this case from two to four. A network can be trained for four classes instead of two by simply tweaking the deeper layers of the network. However, as networks are usually designed for a specific goal there is no guarantee that CNN architectures available for damage detection will also work for damage classification with multiple classes. For this research the same network will be re-trained for the classification of the images to 4 classes. In this approach the same features will be used, however the training and validation set will be classified differently. This diversification of the dataset results in a decline of features per class, which can negatively impact the performance of the method.

3.4 ASSESSMENT OF RESULTS

To allow for a good comparison between the ground truth data, methods described in this chapter, and visual interpretation as used by Copernicus EMS, various techniques for the rating of inter-rater agreement will be used. These will be the Cohen Kappa Score, the resulting confusion matrices, and the F1 score. All of these methods are used to determine the consensus between the prediction and ground truth data, as described in section 2.6.2. Furthermore, will these principles also be used for the performance rating of the various empirically set thresholds for the method described by Yun et al. [2015]. The guiding principle in the selection of results will be the Cohen Kappa Score. Agreement is seen as good within this context and is therefore used for interpretation of the methods.

7 www.missingmaps.org

4 RESULTS

This chapter show cases the results from the implementation as described in the previous chapter. The evaluations in this chapter are used to answer the main research question; *Is the use of remotely sensed data a viable option for the automatic classification of hurricane inflicted damage?*. This chapter mainly focuses on the following sub-questions: *How do these methods perform?* and *How does the state of the art compare to these methods?*. To answer these questions the results of the methods described in chapter 3 will be discussed as well as compared to each other, ground truth, and STOA methods. The implementation validation is focused on an area called Middle Region in the centre of the island. Around 900 houses are part of this neighbourhood, which was moderately hit [figure 4.1]. This area has been selected to ensure sufficient data was available for all methods to perform analysis and validation on. The extent has been defined by the available UAV datasets of this area. Methods will be tested on this part on the island, while training will be based on data from all the other areas available. This chapter also discusses the results and reflects back on the framework developed in section 2.3.

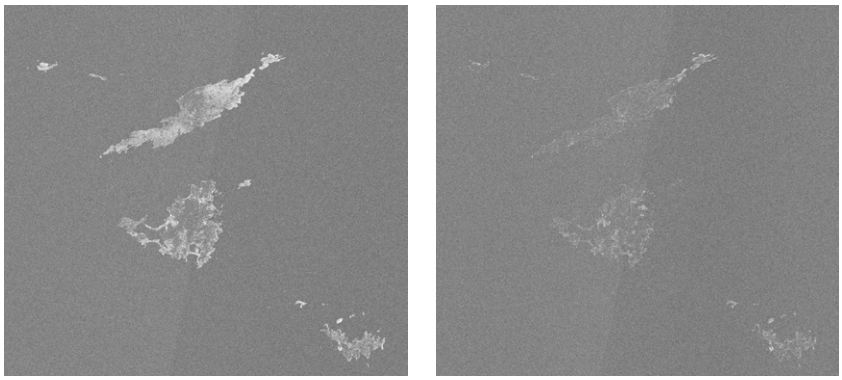
Class	# buildings
none	326
partial	193
significant	186
destroyed	135
unknown	*16*

Figure 4.1: Map of Sint Maarten with Middle Region highlighted [Left]. Ground truth classification of buildings from the NLRC [Right]. These have been derived by visual interpretation from the VHR UAV imagery guided with the principles described in section 2.2. [From: OpenStreetMap contributors (2017), Sint Maarten [georeferenced data], used under ODbL as part of OSMF, retrieved from `www.openstreetmap.org`]]

4.1 EQUALISATION AND SUBTRACTION

4.1.1 SAR data

For the ESS, the method as described in sections 3.3.1 has been applied. From the pre-processing steps, two coherence maps where derived [figure 4.2]. Visual inspection of these datasets already show less cohesion over the whole island after the hurricane. This is to be expected as all vegetation was likewise hit by the hurricanes. These two coherence maps are used as input for the histogram matching algorithm and univariate image differencing to allow for change detection. As described by Yun et al. [2015] a threshold is empirically set for the definition of damage detection, in this case 0.3. The resulting image [figure 4.3], clearly shows the outline of the island and urban areas, in which the most change due to damage is expected. However the paper does not quantify any of the results and focuses on the creation of damage proxy maps from this result, which in this case would highlight various urban areas for damage. These can only be used by visual interpretation for estimation of amount of damage in an area. Section 4.3 will quantify the results of this map and compare these to the results from other approaches.



(a) Coherence map pre-event **(b)** Coherence map post-event

Figure 4.2: Resulting coherence maps from ESS approach as described in section 3.3.1. Showing aggregated coherence per pixel, an increase in lightness correlates to an increase in coherence.

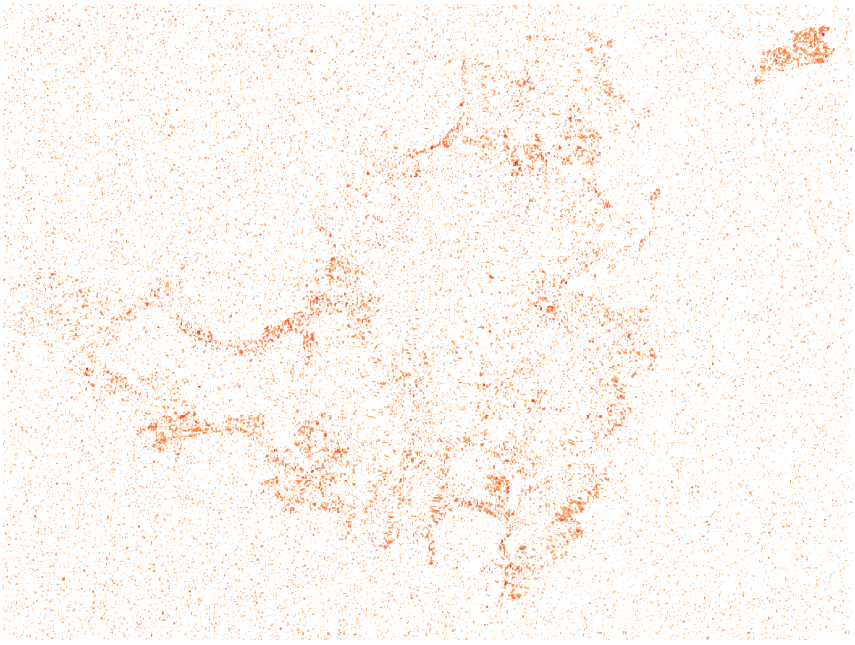


Figure 4.3: Result from empirically set threshold [0.3] after Univariate Image Differencing from the coherence maps. Darker colour indicate more change. All change under the threshold has been discarded and pixels are considered not damaged.

4.1.2 Optical data

The UAV dataset for Middle Region from RescUAV and aerial imagery from IGN have been used for the implementation of the ESO approach. To achieve a satisfactory result, the pre-event imagery has to be of the same geographic dimensions as the post-event imagery both as bounding box and concave hull. This to ensure the equalisation algorithm only affects -real-data and not the masked out areas. Figure 4.5 shows the characteristic non-rectangular outline from stitched UAV data. The correct pre-event dataset was created by using a mask with the outline of the post-event data from the UAV. This was done with the following steps:

- For every pixel in the post-event imagery, it was decide if this pixel had data, if so it was classified as 1, otherwise 0.
- The resulting raster was vectorised to create a mask of the area.
- The polygon with value 0 was discarded.

- Using the polygon with value 1 a comparable pre-event dataset was created from the aerial imagery

However, the first results show an error in the geographic alignment of the pre-event and post-event data [Figure 4.4]. There is a very apparent shift in the datasets, this could be caused by various sources. As can be seen in figure 4.4, the shift is not of the same magnitude throughout the image. The pool on the right hand side of the imagery has moved further than the buildings on the left hand side of the imagery. One explanation would be an error in the geo-referencing algorithm that stitched either the pre-event or post-event data. However, it is more likely that the UAV data has a lower accuracy in positioning than the aerial imagery used. This could cause the uneven shift throughout the image. To continue the research, some [four] of the datasets have been manually geo-referenced, using rubber-sheeting algorithms in Qgis, based on the pre-event aerial imagery available. These areas are Middle Region, Cul du Sac, St. Peters East, and Billy Folly.

Figure 4.4: Result of univariate image differencing before rubber-sheeting for geo-referencing.

From the creation of these dataset the method as described in 3.3.1 could be implemented. For Middle Region all three characteristic bands where calculated and went through the histogram matching and univariate image differencing algorithm. All with varying results. In figure 4.6 the results are visually compared to the original RGB image. From this comparison it can be concluded that the use of lightness, equalised relative to the pre-event situation allows for the best manual interpretation of the data. However, this is due to the capabilities for humans to build connections and recognise objects in abstract representations of data, section 4.3 will quantify the results. The lighter colour implies more change, in the hue change map nearly all buildings are a lighter colour indicating that something else is causing the change. The same amounts for saturation in which non of the buildings seem to have undergone significant change, while this is certainly the case. The value layer is not a perfect one-to-one indication of damage to building, however it does indicate change on the buildings which have sustained damage. This is specifically true in the cohesion of the pixels within the feature, not specifically on their intensity.

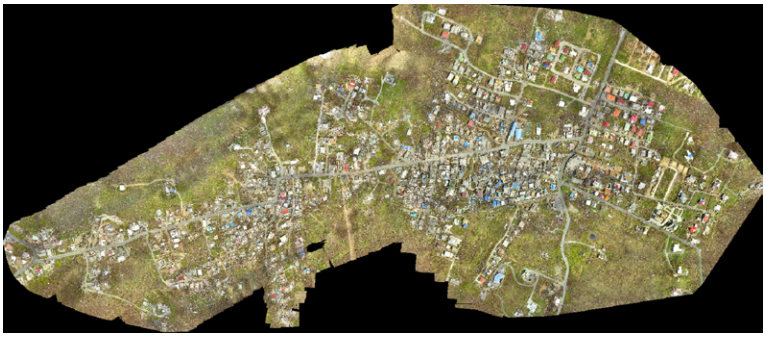
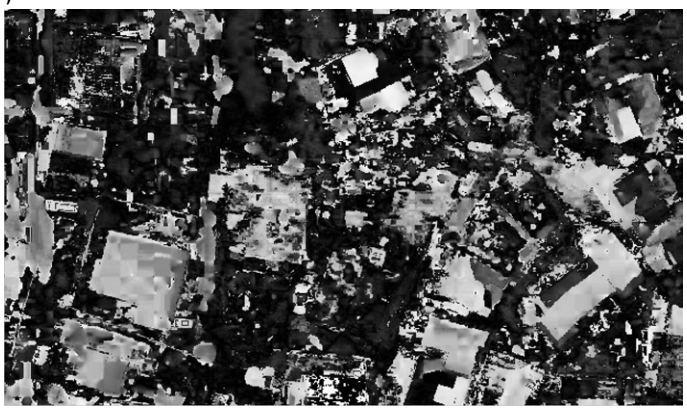


Figure 4.5: Optical UAV imagery Middle Region, with masked extent [From: RescUAV (17 Sept. 2017), Middle Region - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

(a) Raw data



(b) Hue change map

(c) Saturation change map

(d) Value change map

Figure 4.6: Comparison between Dataset Middle Region on a normalised scale. [From: RescUAV (17 Sept. 2017), Middle Region - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

4.2 CONVOLUTIONAL NEURAL NETWORK

4.2.1 Optical data

As described in section 3.3.2 there are three steps necessary for the creation of a CNN method. First of all the features need to be created. In the case of St. Maarten these features can be created from available data. The square bounding boxes necessary for a fruitful implementation of a CNN was done using features, with classification from OpenStreetMap. These bounding boxes where than used to cut features of 100x100 pixels from the various input layers. These layers are the UAV datasets, which have been used for the creation of the ground truth data; which is the foundation of the CNN. A sample of these features have been represented in figure 4.7. From this image it can already be noted that not all of the features have been selected correctly. This is mainly due to the geographical shift between images from before and after the disaster. As described in section 4.1 this shift has a varying magnitude through out the datasets and can only be fixed by rubber-sheeting geo-referencing. Due to the time constraints of this research, these transformations where not possible. Hence an attempt has been done with the available data.

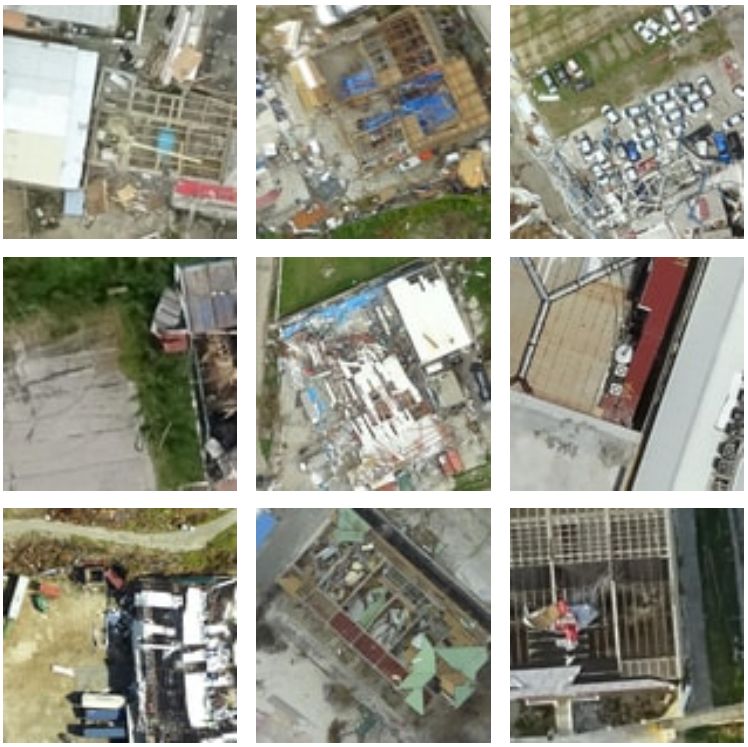


Figure 4.7: Features generated from bounding boxes, classified as destroyed

The CNN architecture described in 3.3.1 has been implemented in TFlearn for Python and run on the available dataset. This run is based on the settings from Vetrivel et al. [2016b] and implemented for damage detection. The first training was based on a learning rate of 0.1, 5 epochs, and no normalisation. Unfortunately where the first runs unsuccessful with over-fitting of the regression layer. This resulted in a varying loss indicator, but a stable validation accuracy. To avoid over-fitting more epoch where run with the same settings. A slow increase to 5, 10, 15, 25, and 35 epoch had no influence on the results. Over-fitting was the norm and no to little increases in the loss function or accuracy where noted. A similar approach was used for the increase of the learning rate from 0.1 to 0.01, with similar results. The addition of normalisation should allow for a better distribution of the features, resulting in better predictability. However, after 35 epochs [figure 4.8] the neural network still over-fits. The resting validation accuracy can be interpreted as the regression layer continually predicting the same class, the validation accuracy in this case therefore arises from the fact that guessing this option gives the best outcome most of the time. This is the result of a skewed input regarding the classes, or skewed input regarding the actual features.

```
raining Step: 4059  | total loss: 0.52611 | time: 293.533s
 Adam | epoch: 034 | loss: 0.52611 - acc: 0.7810 | val_loss: 0.54905 - val_acc: 0.7619 -- iter: 6336/7662
-
raining Step: 4080  | total loss: 0.56647 | time: 420.401s
 Adam | epoch: 034 | loss: 0.56647 - acc: 0.7474 | val_loss: 0.54894 - val_acc: 0.7619 -- iter: 7662/7662
-
raining Step: 4158  | total loss: 0.54208 | time: 247.685s
 Adam | epoch: 035 | loss: 0.54208 - acc: 0.7678 | val_loss: 0.54896 - val_acc: 0.7619 -- iter: 4992/7662
-
raining Step: 4200  | total loss: 0.54314 | time: 421.486s
 Adam | epoch: 035 | loss: 0.54314 - acc: 0.7669 | val_loss: 0.54891 - val_acc: 0.7619 -- iter: 7662/7662
-
```

Figure 4.8: Training of CNN, last 2 epochs, full training in appendix A.1

4.2.2 SAR data

The characteristics of the SAR data do not allow for the use of a CNN method. As described in section 2.5.2 for a CNN to work it needs to do convolutions on the datasets. As shown in figure 4.9 the pixel size of the SAR dataset is of the same magnitude as the building outlines. This results in less than 1 pixel per structure, thus per feature. A CNN can not perform any convolutions or other transformations on a feature that starts with only one pixel value. This would not allow for the detection of borders or other characteristics which might encompass change or damage. A machine learning approach could still be feasible, however this would now be focussed on a 1D feature vector, comparable to those of support vector machines or least squares adjustment. As these methods differ from a close comparison to the method proposed by Vetrivel et al. [2016b] these have not been considered in this research.

Figure 4.9: SAR coherence map superimposed on buildings [Based on: RescUAV (17 Sept. 2017), Middle Region - Sint Maarten [georeferenced image], used under CC-BY4.0 as part of Open Imagery Network, retrieved from www.openaerialmap.org]

4.3 COMPARISON

From sections 4.1 and 4.2 the matrix from section 3.1 can be filled. Matrix 4.1 indicates which methods have been tested and which was not possible to execute. However, some implementations so far have been focused on visual interpretation as a last step in determining damage. This section will quantify the results for a direct comparison on accuracy performance for damage detection as described in section 2.2. The extension and comparison of the methods is described in section 4.4.

Table 4.1: Matrix used for comparison methods and data, Green = Implemented, Orange = Not implemented.

	Optical UAV Data	Satellite SAR Data
Equalisation and subtraction	ESO	ESS
Convolutional Neural Network	CNO	CNS

The quantification for the ESO and ESS is done by empirically checking various possible combinations of thresholds and feature statistic and the calculation of the Cohen Kappa Coefficient. Only a summary of these results with the highest Cohen Kappa coefficient will be compared, these are fully represented in appendix A.2. For the CNO approach, the network as described in sections 2.5.2 and 3.3.2 is trained on all data available expect Middle Region. This pre-trained network is then used for the detection of damage in the features from Middle Region. Lastly the STOA method, implemented by Copernicus EMS is considered and compared. All of these results are available in table 4.6.

The ground truth data consist out of vector representations of the buildings on the island and their respective classification as described in section 2.2. To allow fair comparison between this data and the other methods for damage detection this datasets will need to be translated to match the granularity of the damage description. In this case, four categories have been reduced to two. These classes have been used in the comparison between methods described in this section. Table 4.2 shows this reduction translation between classes.

Table 4.2: Translation from damage classification provided in ground truth to damage detection, as defined in section 2.2

Damage classification	Damage Detection
No Damage Minimal Damage	No Damage
Significant Damage Critical Damage	Damage

For the ESO and ESS approaches four change maps, namely the interferometry and HSV change sets from the UAV optical data, are considered. To do so these pixel values have been aggregated over the various buildings outlines. These vector representations of the buildings are derived from the ground truth data. By subsequently taking the median of intersecting or enclosed pixels an aggregated value is derived per building. To determine whether a building is damaged a threshold needs to be empirically set to differentiate between damaged buildings and non damaged buildings. Such a threshold is determined for all four changes maps systematically. These have been generated for every 0.01 of change within the change maps, resulting in one hundred thresholds per image. All quantifications are then compared to the ground truth data obtained by visual interpretation by the NLRC. Table 4.6 summarise the best findings, based on their Cohen Kappa Scores. In appendix A.2 the full F1 matrices, and confusion matrices from the four change maps are represented. From that table 4.6 it can be concluded that although the method is working, it is not on the data for which it is intended. The interferometry data has a low Kappa score and F1-score of 0.45, this is about comparable to estimated guessing with a-priori knowledge of the damage distribution. For the optical approaches, namely Saturation and Value, the empirical threshold does allow for better results especially considering regarding the considerations saturation and value, even though these might not have seem a direct hit on visual interpretation, they do provide knowledge about the distribution of damage, with both Cohen Kappa scores around the 0.5 and F1-scores close to 0.7. Furthermore, both perform well on the detection of damage, for which the F1-scores are 0.74 and 0.73 respectively.

As mentioned in section 4.2.1 the CNO approach fails to converge on the loss indicator. This is due to the skewed input as a result of geographic misalignment. Therefore, all features have been classified as not damaged, as shown in the confusion matrix in table 4.3. This is reflected in the indices for both F1-score and Cohen Kappa coefficient [Table 4.2].

Table 4.3: Confusion matrix [Vertical Ground Truth, Horizontal predicted] for the CNO based method.

	No	Damage	unknown
No	519.0	0.0	0.0
Damage	321.0	0.0	0.0
unknown	16.0	0.0	0.0

Lastly, the STOA method implemented by Copernicus EMS was compared to the ground truth. As with the ground truth, the STOA data are vector representations of the buildings with classifications attached. The STOA data has been spatially matched to the vector representations of the ground truth data, to ensure the vector representations corresponded to the right classifications in both methods. Like with the ground truth data, the STOA classifications need

to be translated to damage detection, table 4.4 shows the corresponding categories. From the comparison and corresponding confusion matrix [Table 4.5], using the same accuracy measurements for agreement as the other methods, it can be concluded that the STOA approach heavily overestimates the damage. It is also indicated that the method does not significantly outperform other methods implemented in this research.

Table 4.4: Translation from damage classification provided in the STOA method to damage detection, the latter as defined in section 2.2

Damage classification	Damage Detection
Not affected Negligible to slight Damage	No Damage
Moderately Damaged Highly Damaged Completely Destroyed	Damage

Table 4.5: Confusion matrix [Vertical Ground Truth, Horizontal predicted] for the STOA approach.

	No	Damage	unknown
No	142.0	377.0	0.0
Damage	51.0	270.0	0.0
unknown	4.0	12.0	0.0

Table 4.6: Accuracy comparison between methods described in chapter 3, quantified using thresholds and ground truth data.

Technique	Threshold	Kappa coeff.	Avg. F1-Score
ESO Hue	0.11	0.070	0.47
ESO Saturation	0.07	0.429	0.71
ESO Value	0.21	0.389	0.69
ESS	0.30	0.059	0.54
CNO	n/a	0.000	0.46
STOA	n/a	0.093	0.45

4.4 EXTENSION TO CLASSIFICATION

To answer the main research question fully, an evaluation of the extended methods for classification is indispensable. The implementation for extension require little change to the existing methods and is based on the same principles. The definition for classification of damage as described in section 2.2 is used. Therefore, the ground truth data is used without translation as it is already based on this definition. To facilitate the extension of the ESO and ESS, multiple thresholds are required. As in section 4.3, these threshold will be systematically empirically checked. With the four classes defined for damage classification this translates to three thresholds. The extension of the CNO approach requires little tweaking of the deepest, regression layer of the CNN, in which the parameters are tuned from two to four classes. The STOA classification still differentiates from the defined classification and requires a translations schema. The summary of the results is presented in table 4.9.

For the systematic approach to the three thresholds for the ESO and ESS methods, all valid combinations of thresholds is examined. Intervals between thresholds are defined by 0.025 steps between 0 and 1, a combination of thresholds is considered valid when the third threshold is larger than the second threshold, which has to be larger than the first threshold. From all valid thresholds, those with the highest Cohen Kappa score are represented in table 4.7.

The CNO approach has similar steps as in section 4.3. Through these steps, the network is retrained with the deepest layer altered for four classification categories, and this trained network is applied to the unseen dataset. However, as in the comparison for damage detection, the CNN does not converge on the loss indicator [Appendix A.3 provides an overview of the

training] and all features classified as not damaged.

As the STOA approach to classification is defined differently than in this research, the data needs to be translated for comparison. The same vector outlines as in section 4.3 have been used but with the translation provided in table 4.7. The resulting accuracy indices indicate a bleak implementation. The STOA approach heavily overestimates the damage, corroborated by the confusion matrix [Table 4.8], as in the damage detection.

Table 4.7: Translation from damage classification provided in the STOA method to damage classification as defined in section 2.2

Damage classification	Damage Detection
Not affected Negligible to slight Damage	No Damage
Moderately Damaged	Partial Damage
Highly Damaged	Significant Damage
Completely Destroyed	Critical Damage

Table 4.8: Confusion matrix [Vertical Ground Truth, Horizontal predicted] for the STOA approach for classification.

	No	Damage	unknown
No	142.0	377.0	0.0
Damage	51.0	270.0	0.0
unknown	4.0	12.0	0.0

Table 4.9: Accuracy comparison between methods described in chapter 3, quantified using thresholds and ground truth data.

Technique	Threshold	Kappa coeff.	Avg. F1-Score
ESO Hue	0.08, 0.11, 0.88	0.054	0.23
ESO Saturation	0.08, 0.08, 0.31	0.250	0.37
ESO Value	0.13, 0.18, 0.26	0.188	0.40
ESS	0.23, 0.31, 0.34	0.0511	0.30
CNO	n/a	0.00	0.21
STOA	n/a	0.078	0.24

4.5 DISCUSSION

This section provides discussion on the results from the comparison for damage detection and damage classification, as well as an interpretation of the framework for method assessment defined in section 2.3. The conclusions and recommendations from this discussion are presented in chapters 5 and 6 respectively.

The results gathered in sections 4.3 and 4.4 leave room for discussion and interpretation. First of all it must be noted that, as described in section 2.6.2, the inter-rater statistics provide a tinted look on accuracy. However, agreement is considered a better description for accuracy in this case study. Although the extension of methods to damage classification is possible, they perform considerably less than the methods for damage detection. The main negative outliers in both approaches are the CNO, STOA, and ESS approaches. As the results of the CNO approach are faulted by design, these are not considered for discussion. However, a lesson learned from this faulty approach is the need for properly geo-referenced imagery in the use of multiple datasets for damage determination, as corroborated in section 4.1. The bad performance in both damage classification as well as damage detection by the STOA approach is unexpected. However, the speed with which these datasets are provided by UNOSAT or Copernicus EMS is nearly unbeatable as the first damage maps were available one week after the hurricane struck the island. Furthermore, this approach overestimates the damage in a disaster struck

area, which is desirable for humanitarian organisations to ensure proficient aid is supplied. Lastly,the ESS approach is not suited for damage determination on a building scale level, however, this is also not claimed by Yun et al. [2015]. Therefore, this method might be unsuitable for building damage determination but it might still be useful in larger disaster where prioritisation of areas is paramount to an operation.

All methods considered, the ESO approach was the most surprising, as a deducted method from Yun et al. [2015], it performs reasonable on a building scale level, mainly due to the increase in resolution. However, the empirically set threshold is not a robust approach for other disasters, as variations in camera systems or external influences might impact the performance. Ground truth data is therefore indispensable to make sure a threshold can be set systematically with a machine learning minimisation approach.

The implementations of the various methods and STOA approach, have varying degrees of accuracy. However, accuracy is only one of the parameters defined in section 2.3 for the assessments of methods for damage detection or classification. The others are acquisition time, acquisition method, and resolution for input data and output information. Acquisition time and acquisition method are highly depended and pros and cons have been summarised in table 2.2. The datasets used in this research are complementary, especially concerning the spatial coverage and resolution of input data. Most methods described in this research are able to provide output on a building level resolution, without regards to accuracy, and this is the required level of information for the phase from the DRM activity cycle on which this research focused. The SAR data is highly regular and can be provided within 6 days of a disaster indifferent of the atmospheric conditions, however the larger coverage area from this method is balance by the lower input resolution. Paired with the specialised technical know-how, and disappointing results in accuracy, this would not be the best implementation for the St. Maarten hurricane. In this specific disaster, the higher resolution input data from UAV imagery, provides more information than the larger extent, mainly due to the small size of the island. The methods used for optical imagery provide sufficient accuracy to be a better indication than STOA approaches, and manual visual interpretation could provide ground truth information about the damage extent.

5 CONCLUSIONS

The main goal of this research was to find a method for the automatic classification of damage inflicted by hurricanes on the island of St. Maarten using remotely sensed data to support the operations of the NLRC and other humanitarian NGOs. To achieve this goal the following question was defined, *Is the use of remotely sensed data a viable option for the automatic classification of hurricane inflicted damage?* The various sub-questions all contributed to a stream-lined method for answering this question.

5.1 HOW IS DAMAGE DETERMINED?

To determine the damage to buildings, various information densities can be used. Section 2.2 provides an exemplified overview of the three nominal gradations used in this research. These are the detection, classification, and assessment of damage. The difference between these methods is in the detail to which the damage is described. The increase in inclusion of detail is proportional to the ongoing developments within the DRM activity cycle of a humanitarian organisation after a disaster; the further into this development, the more detail is necessary for the correct planning of humanitarian aid. Limited standardised methods exist for all three nominal gradations of damage determination, this research therefore proposes the use of a classification diagram, based on efforts from the Harvard Humanitarian Initiative and 510 from the NLRC. This classification diagram is applicable to both automated and manual approaches for damage classification and allows for clear distinction between damage classes while avoiding ambiguity.

5.2 WHAT CRITERIA ARE SET FOR DAMAGE CLASSIFICATION METHODS?

Operation criteria for the classification of damage are set by humanitarian organisations, like the NLRC. These are based on the information needs in a disaster situation. As described in section 2.1, this research focusses on the emergency relief and rehabilitation phases. The detail requirements are set to block or building level, in which classification, as described in section 2.2, suffices. To be able to select a method for rapid assessment using remotely sensed data, a framework has been established with 4 parameters: [1] Accuracy, [2] Acquisition time, [3] Acquisition method, and [4] Resolution. These are the four parameters that should be taken into account in the development or assessment of a method for use in a humanitarian context. In this context, accuracy describes the percentage of correct classification, which can be defined in various manners, as described in section 2.6.2. A clear guideline to the accuracy is to include all people in need of help, while still differentiating to ensure effective distribution of means. Acquisition time is an operation requirement that relates to the time constraints in a DRM activity cycle. For a method to be implemented in a specific part of the humanitarian response to a disaster, time constraints the possibilities for data collection. This is closely linked with the acquisition method, which describes how and what data can be collected. This would be limited by sources of data, as well as the specialities within a humanitarian team. Resolution is a two-fold parameter to describe methods for damage detection or classification. It describes both the resolution of the input data and output information, which are constraint by time and resources. In the various phases of the DRM activity cycle, the requirements for the information vary. The definition of sufficient detail in the output information, is proportionally linked with the activity of ongoing development. Furthermore, are humanitarian mission restricted by resources in the various phases to process high resolution data.

For most of these parameters no specific guidelines for values are provided, as every situation in humanitarian action is different and all require slightly different datasets or implementations. In the selection of method for damage classification, it is therefore inevitable that guidelines to these parameters are set by the specific humanitarian mission and will vary as a result.

5.3 WHICH METHODS ALREADY EXIST?

A broad selection of methods are already available in academic literature. Section 2.5 provides a non-exhaustive overview. While all approaches differ in exact techniques used to determine damage, the range of remotely sensed data they use is limited to satellite optical and SAR data or UAV optical data. This is only a small subset of the possibilities described by Kerle et al. [2008]. Most quoted accuracies range between 70% - 90%, with acquisition times under a week. This indicates suitability for use in the DRM activity cycle after a disaster. Vetrivel et al. [2016b] and Yun et al. [2015] describe methods with most potential. The CNN approach by Vetrivel et al. [2016b] yields the highest accuracy claims, with over 90% accuracy. The equalisation and univariate differencing technique described by Yun et al. [2015] has found the most traction in the humanitarian sector with uses by various NGOs.

5.4 HOW DO THESE METHODS PERFORM?

To indicate performance, a pure definition of accuracy is not sufficient. In case of unbalanced data this might lead to a false indication of accuracy. Other inter-rater statistics are therefore necessary. In this research the Cohen Kappa coefficient and F1-score have been used for their inclusion of randomness in the accuracy measurement, and inclusion of precision and recall, respectively. These approaches to accuracy are used to provide a more complete overview of the performance of the methods.

In this research the methods by Vetrivel et al. [2016b] and Yun et al. [2015] have been adapted in compared. Both methods have been implemented with the data they were intended for UAV optical and satellite SAR, respectively. However, these methods have also been extended to allow for comparison between the SAR data and optical data, this translate to implementations for the other datasets as well. Due to characteristics of the SAR data, it is not possible to include an implementation based on a CNN approach. Furthermore, all implementations have been extended to allow for smaller granularity in the output information, which translates to an implementation for damage detection and damage classification.

Quantified inspection of the methods indicate that the adapted approached of Equalisation and Subtraction and a CNN approach are not sufficiently suitable for direct implementation in damage classification, with low values for both the Cohen Kappa coefficient and average F1-score. The methods based on differencing perform all roughly equal, while the CNN approach is hindered by inappropriate input features. For the detection of damage these methods show suitable results for the use by humanitarian organisations. The methods based on optical image differencing work adequately, specifically on those features which indicate damage. The same approach based on the SAR data performs not as well, especially with regards to the identification of damage. The CNN approach suffers the same problems as in the classification approach and perform not up to par.

5.5 HOW DOES THE STATE OF THE ART COMPARE TO THESE METHODS?

The state of the art for damage determination in humanitarian aid is the use of visual inspection by Copernicus and UNOSAT. The comparable information for this research is based upon satellite optical data. Compared to the ground truth data this approach over-estimates the damage in an area. This is in line with the approach of humanitarian organisations which prefer to supply more people with aid than have people left out. Comparatively this approach is quantifiable less accurate, however has some major implementation advantages. While the damage is overestimated, other humanitarian organisations are not burdened with the task of collecting and processing data. This set-up has a clear operational advantage as less specialised personnel is necessary for the implementation of new automated measures. It offers an all-inclusive solution with fast results. In comparison a subtraction approach based on UAV optical imagery would take more investment to obtain sufficient workable coverage of a larger scale disaster, even if it provides better accuracy. A similar problem arises with the use of CNN approaches. In that case specialised people are necessary to train and maintain a CNN to allow for fast implementation after a disaster, which can furthermore be hindered by the lack of granular data collection.

5.6 SUMMARY

Is the use of remotely sensed data a viable option for the automatic classification of hurricane inflicted damage? Yes, there are many possibilities for the use of remotely sensed data to be used for the classification of damage. However, the practical implementations are not available yet. Various possibilities for future research, to make this possible, have been illustrated in chapter 6. Methods for damage detection exist and can be introduced in emergency situations, however they are not user friendly for the humanitarian delegates in the field, without technical know-how. The result of this is the return to practical solutions offered by other NGOs, based on visual interpretation. A more human approach to the problem would allow non-technicians to get acquainted with the possibilities of remotely sensed data and the benefits for disaster response. This could pave the path for willingness to invest in new automated approaches to make humanitarian aid more effective and efficient.

6 RECOMMENDATIONS

Due to the broad nature of this research various recommendations can be made regarding new research topics and practical implementations. These could improve existing methods, extent methods for the classification as described in section 2.2, or be possibly developed into new solutions.

6.1 OPTICAL DATA

Regarding the use of univariate image differencing for the detection of change based on RGB or HSV values could be improved in various ways. First of all, an approach as described by Yun et al. [2015], in which the coherence between datasets is used, could also be advisable for applications based on multi-layered optical images. In that case multiple datasets should be pre-emptively collected, to allow for the creation of coherence maps. This requires pre-emptive resource investments. However, these would allow for description of the coherence between pixels, which could be used as an indicator of damage. Practically, this approach might incur more opposition due to the investment necessary. However, especially regarding repetitive and predictable disasters, like hurricanes, this could be a benefit for all involved in the humanitarian operation later. Secondly, the datasets could be subtracted in multiple bands, this would allow for more data to be used in the detection of change and or damage. To achieve this, a new method would need to be developed capturing various features of the date to capture the coherence. A CNN or other convolution algorithm would allow for the description of the pixel coherence to be used for classification. Lastly, the change indicator layers as presented in section 4.1, figure 4.6, could provide extra indication of damage in case visual interpretation is used. This as it highlights possible change, which would make manual classification more accessible. Which is especially useful in the possible crowd-sourcing of this kind of information using approaches similar to Missing Maps [1].

In the use of CNN for damage classification, various improvements could be made. A single implementation architecture for damage detection or classification is hard to establish. This is due to data variability caused by variance in the sensors, conditions, or other exterior influences. To achieve a result for all datasets, a library with various datasets within would need to be created. This approach, similar to ImageNet [2], would allow NGOs to develop tools based on data available from a diverse range of disaster, all oriented on the classification of damage. With this database, new networks could be designed to cope with various inputs and variations in other exterior influence. An existing database with known damage identified on post-event imagery could also inspire other technologies for damage classification.

Furthermore, would a CNN approach be helped by improved creation of the features. This could also be achieved through more machine learning. If it would be possible for a Neural Network approach to identify buildings on a pre-disaster image, it would be possible to create clear outlines for use on OpenStreetMap as well as within operations using CNNs for the classification or detection of damage. This also reflects back to the geo-referencing of imagery. Newer techniques allow for faster and more accurate data collection, however the smaller packages and cheaper sensors do not allow for perfect geo-referencing. An improvement in the alignment of imagery and the geo-referencing of optical imagery would be beneficial for the results of various methods.

6.2 SAR DATA

The use of SAR in both approaches is limited by the resolution of the input. Higher resolution inputs, or lower resolution expected output, could allow the use of CNN for all datasets. How-

1 http://www.missingmaps.org/
2 http://www.image-net.org/

ever it must be noted that it can be hard to distinguish features on SAR data. This property would transfer to the use of machine learning, specifically CNNs, as the method is based on the use of distinguishable features for classification. A further deep-dive into the methodology devised by Yun et al. [2015] could also uncover possible areas of improvement. Various research around the world have been able to detect millimetre changes in static objects using satellite SAR data [Sousa et al., 2010].

Another solution for the use of lower resolution data is an aggregation of damage after the pixels have been classified. In this case, if a building is covered by 4 pixels partly, an aggregate of these pixels would be subscribed to the building. If in that case one of the pixels is classified as damaged and overlaps the building by 25%, this building would be considered damaged for 25%. A similar approach would be valid for use with the optical data sets in which it would be the amount of pixels classified as damaged relative to the total amount of pixels within the building bounds.

6.3 DATA COMBINATION

While this research only considered datasets separately, the combination of datasets could provide new possibilities for the determination of damage. As mentioned in section 4.5, the advantages of satellite based SAR, in particular the operational agility and large coverage area, are complementary to the VHR UAV imagery, which are limited in coverage area but provide more information about damage. In the fusion of these datasets, and the possible addition of others, combination of methods could be used to fully- or semi- automatically determine the damage extent after a disaster.

6.4 DISASTER SPECIFIC PROPERTIES

This research considered general approaches to the detection and classification of damage. This is based on the premise that building damage is generally similar between disasters, however, it can be noted that various disasters might have specific damage patterns around buildings. The manual method proposed by Okada and Takai [2000] is based on the premise of building damage patterns to determine the extend of the area affected. An automated approach could use similar indicators for remotely sensed data. An example of damage patterns surrounding buildings could be the debris spread observed after a hurricane, where debris is scattered around a building in the direction of the prevailing winds; compared to earthquake damage, where debris is usually less scattered and more concentrated on the building footprint.

6.5 METHOD ASSESSMENT FRAMEWORK

The pre-existing conditions regarding damage classification and information needs within humanitarian aid changes constantly. A more social approach to the problem would allow for clearer definitions of needs and possible holes in the information provision with regards to remotely sensed data. An implementation review of the approach used by Copernicus and UNOSAT, which would focus on the use of the information, would allow for understanding of their products. This would further allow the introduction of new data sources in the programmes of NGOs, which in turn would favourably influence the research with regards to implementation possibilities.

In this research the focus was placed on a part of the DRM activity cycle. However, information needs vary within this cycle as well. Other methods, including predicting approaches, would be useful in the various stages of the DRM activity cycle for humanitarian organisations. Examples of this include the work done by Lint [2016] and Bulte [2017], in which predictive models estimate the damage after typhoons and earthquakes. Similarly an approach in which verified data from the field can be used to train new methods for more accurate damage classification in later stages of a disaster, allowing for better accuracy on a higher resolution.

Other methods identified in this research, could transfer better to the datasets used. More research regarding their capabilities would allow for the development of new practicable methods. These methods in combination with the knowledge from this research, might be capable of the classification of damage, automatically or semi-automatically. Lastly, the use of semi-automated methods could also be a practical approach. As disasters vary from location, time,

and intensity, a flexible human approach could benefit the method. In such an approach some steps are identified by humans, so a-priori knowledge can be supplied to the automatic classification tools. This synergy would allow for the best of both approaches, the flexible insights of a human and the meticulous repetitive approach of machine learning, allowing for faster creation of information. This would not be a troublesome practical approach as most disaster datasets have to be scrambled from various sources; human involvement is - for now - indispensable.

6.6 COMPARISON OF RESULTS

The use of the Cohen Kappa Coefficient, F1 Score, and Confusion Matrix are non optimal indicators for the use in humanitarian circumstances. It would benefit the humanitarian organisations if the risk of false negatives, regarding building damage would be minimised. This would ensurethat everybody in need of assistance receives aid. Another minimisation technique might therefore be beneficial for these problems. This could be achieved by the addition of weights to certain classes, in which it is more important that these are classified correctly. A certain distinction is still necessary and the balance should not be moved to classifying as many buildings as possible as damaged, as this would defy the goal of a damage classification method for disaster impact. This could be done with the F1-Score mentioned in section 2.6.2 as well as other classification techniques.

REFERENCES

Ajmar, A., Boccccardo, P., and Giulio Tonolo, F. (2011). Earthquake damage assessment based on remote sensing data. The Haiti case study. *Italian Journal of Remote Sensing*, pages 123–128.

Al Achkar, Z., Baker, I. L., and Raymond, N. A. (2008). Assessing Wind Disaster Damage to Structures. *Harvard Humanitarian Initiative*, (viii).

Antonietta, F., Boccardo, P., Tonolo, F. G., Vassileva, M., and Dist, T. (2015). Damage assessment exploiting remote sensing imagery : review of the typhoon haiyan case study. pages 3556–3559.

Bessis, J.-L., Béquignon, J., and Mahmood, A. (2003). The International Charter Space and Major Disasters initiative. *Acta Astronautica*, 54:183–190.

Bhowmick, P. K., Mitra, P., and Basu, A. (2008). An agreement measure for determining inter-annotator reliability of human judgements on affective text. *Proceedings of the Workshop on Human Judgements in Computational Linguistics - HumanJudge '08*, (August):58–65.

Bijnsdorp, L. (2017). Hurricane Maria passes South of St. Maarten.

Brunner, D., Lemoine, G., and Bruzzone, L. (2010). Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery. 48(5):2403–2420.

Buckland, M. and Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*, 45(1):12–19.

Bulte, E. (2017). *Estimating Post-Earthquake Aid Priority Areas*. PhD thesis, Univerity of Utrecht.

Centre for Research on the Epidemiology of Disaster (2015). Poverty & Death: Disaster Mortality 1996-2015. pages 1–20.

Chinchor, N. (1991). MUC-4 evaluation metrics. *Proceedings of the 3rd conference on Message understanding - MUC3 '91*, pages 22–29.

Christopher, D. and Doeglas, A. (2015). *Time-Sensitive Remote Sensing*.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, XX(1):37–46.

Copernicus EMS (2017). Philipsburg - SINT MAARTEN.

Coppola, D. P. (2015). *The Management of Disasters*.

Crutchfield, M. (2013). Phases of Disaster Recovery: Emergency Response for the Long Term.

Daniell, J., Mühr, B., Pomonis, A., Schäfer, A., and Mohr, S. (2017). Center for Disaster Management and Risk Reduction Technology CEDIM Forensic Disaster Analysis Group (FDA) Focus on Caribbean up until 8 th September 2017. Technical report.

Dell'Acqua, F. and Gamba, P. (2012). Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE*, 100(10):2876–2890.

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302.

Dominici, D., Alicandro, M., and Massimi, V. (2017). UAV photogrammetry in the post-earthquake scenario: case studies in L'Aquila. *Geomatics, Natural Hazards and Risk*, 8(1):87–103.

Dong, L. and Shan, J. (2013). A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84:85–99.

Feinstein, A. R. and Cicchetti, D. V. (1990). High agreement but low Kappa: I. the problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6):543–549.

Ferretti, A., Monti-guarnieri, A., Prati, C., and Rocca, F. (2007). *InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation*. Number February. European Space Agency, Noordwijk, The Netherlands.

Ford, A. and Roberts, A. (1998). Colour space conversions. *Westminster University, London*, 1998:1–31.

Gray, M. (2017). Demolished by Irma, Barbuda spared from Hurricane Jose.

Grünthal, G. (1998). Cahiers du centre européen de géodynamique et de séismologie luxembourg. *European Macroseismic Scale 1998*, 15:1–99.

Hanssen, R. F. (2001). *Radar Interferometry*, volume 276.

Hardesty, L. (2017). Explained: Neural networks.

Horvath, M. (2011). ColorSolid.

Hunt, R. W. G. and Pointer, M. (2011). Measuring colour.

IFRC (2017). What is a disaster?

Johnson, E. M. (2017). Red Cross launches first U.S. drone program for disasters.

JPL (2018). Aria Database.

Kakooei, M. and Baleghi, Y. (2017). Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment. *International Journal of Remote Sensing*, 38(8-10):2511–2534.

Kampakis, S. (2016). Performance Measures: Cohen's Kappa Statistic.

Kerle, N. (2010). Satellite-based damage mapping following the 2006 Indonesia earthquake-How accurate was it? *International Journal of Applied Earth Observation and Geoinformation*, 12(6):466–476.

Kerle, N. (2015). Disasters: risk assessment, management, and post - disaster studies using remote sensing. *Remote sensing of water resources, disasters, and urban studies*, (October):455–482.

Kerle, N., Heuel, S., and Pfeifer, N. (2008). Real-time data collection and information generation using airborne sensors. In Zlatanova, S. and Li, J., editors, *Geospatial Information Technology for Emergency Response*, chapter 2, pages 43–74. Taylor - Francis Group, London, 1 edition.

Kousky, C. (2014). Informing climate adaptation: A review of the economic costs of natural disasters. *Energy Economics*, 46:576–592.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. volume 2, pages 1097–1105. cited By 10330.

Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113.

Li, S., Tang, H., Huang, X., Mao, T., and Niu, X. (2017). Automated Detection of Buildings from Heterogeneous VHR Satellite Images for Rapid Response to Natural Disasters. *Remote Sensing*, 9(11):1177.

Lieshout, I. v. (2017). Big data en drones helpen de hulpverlening.

Lint, S. V. (2016). *Sense-making of The Netherlands red cross priority index model case: Typhoon Haiyan, Philippines*. PhD thesis, Wageningen UR.

Martha, T. R., Babu Govindharaj, K., and Vinod Kumar, K. (2015). Damage and geological assessment of the 18 September 2011 Mw 6.9 earthquake in Sikkim, India using very high resolution satellite data. *Geoscience Frontiers*, 6(6):793–805.

Menderes, A., Erener, A., and Sarp, G. (2015). Automatic Detection of Damaged Buildings after Earthquake Hazard by Using Remote Sensing and Information Technologies. *Procedia Earth and Planetary Science*, 15:257–262.

Munich RE (2018). Natural catastrophe review: Series of hurricanes makes 2017 year of highest insured losses ever.

Narendra, B. N. and Madhukar, R. (2013). *Lanczos Resampling for the Digital Processing of Remotely Sensed Images*. Banglore.

Nex, F. and Remondino, F. (2014). UAV for 3D mapping applications: A review. *Applied Geomatics*, 6(1):1–15.

Okada, S. and Takai, N. (2000). Classifications of Structural Types and Damage Patterns of Buildings for Earthquake Field Investigation. *12th World Conference on Earthquake Engineering*, pages 1–8.

Ozisik, D. (2004). Post-earthquake Damage Assessment Using Satellite and Aerial Video Imagery. *Geo-Information Science*.

Phipps, C. (2017). Irma's destruction: island by island — World news — The Guardian.

Piero, B. and FabioGiulio, T. (2012). Remote-sensing techniques for natural disaster impact assessment. *Advances in Mapping from Remote Sensor Imagery*, pages 387–414.

Pontius, R. G. and Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15):4407–4429.

Powers, D. M. W. (2015). What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. Technical report, Flinders University, Adelaide.

Rode Kruis (2017). Eerste publieksterugkoppeling Nationale Actie Nederland helpt Sint-Maarten. Technical report.

Samadzadegan, F. and Rastiveisi, H. (2005). Automatic Detection and Classification of Damaged Buildings , Using High Resolution Satellite Imagery and Vector Data. pages 415–420.

Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor mater*, pages 1–5.

Schwarz, M. W., Cowan, W. B., and Beatty, J. C. (1987). An experimental comparison of rgb, yiq, lab, hsv, and opponent color models. *ACM Transactions on Graphics (TOG)*, 6(2):123–158.

Schweier, C. and Markus, M. (2006). Classification of collapsed buildings for fast damage and loss assessment. *Bulletin of Earthquake Engineering*, 4(2):177–192.

Sharma, R. C., Tateishi, R., Hara, K., Nguyen, H. T., Gharechelou, S., and Nguyen, L. V. (2017). Earthquake damage visualization (EDV) technique for the rapid detection of earthquake-induced damages using SAR data. *Sensors (Switzerland)*, 17(2).

Singh, A. (1989). Review Articlel: Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, 10(6):989–1003.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.*, 5:1–34.

Sousa, J. J., Ruiz, A. M., Hanssen, R. F., Bastos, L., Gil, A. J., Galindo-Zaldívar, J., and de Galdeano, C. S. (2010). Ps-insar processing methodologies in the detection of field surface deformation—study of the granada basin (central betic cordilleras, southern spain). *Journal of Geodynamics*, 49(3):181 – 189. WEGENER 2008 - Proceedings of the 14th General Assembly of Wegener.

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., and Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160:1–14.

The Federal Emergency Management Agency (2016). Damage assessment operations manual - A guide to assessing damage and impact.

Theodoridis, S. and Koutroumbas, K. (2009). *Pattern Recognition*. Elsevier Ltd, Burlington, 4 edition.

UNDAC (2017). UNDAC Talks Episode 2 - The Power of Remote Sensing.

United Nations (2004). *Living with risk: a global review of disaster reduction initiatives*, volume 1.

UNOSAT (2017). Irma-17: comprehensive satellite detected building damage assessment overview as of 21 September 2017. Technical report, UNITAR, Geneva.

Vetrivel, A., Duarte, D., Nex, F., Gerke, M., Kerle, N., and Vosselman, G. (2016a). Potential of Multi-Temporal Oblique Airborne Imagery for Structural Damage Assessment. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3(September):355–362.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., and Vosselman, G. (2016b). Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, (March).

Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., Platzeck, G., Kaku, K., Kumar Hazarika, M., Czaran, L., Li, S., Pedersen, W., Kadiri James, G., and Proy, C. (2016). Global trends in satellite-based emergency mapping. *Science*, 353(6296):247–252.

Wang, T.-L. and Jin, Y.-Q. (2012). Postearthquake Building Damage Assessment Using Multi-Mutual Information From Pre-Event Optical Image and Postevent SAR Image. *IEEE Geoscience and Remote Sensing Letters*, 9(3):452–456.

Warrens, M. J. (2011). Cohen's kappa is a weighted average. *Statistical Methodology*, 8(6):473–484.

Wilts, A. (2017). Irma: Category 5 hurricane flattens 'most solid buildings' on Saint Martin island — The Independent.

Wisner, B. and Adams, J. (2002). Environmental Health in Emergencies and Disasters. Technical report, World Health Organization, Geneva, Switzerland.

Yun, S.-H., Hudnut, K., Owen, S., Webb, F., Simons, M., Sacco, P., Gurrola, E., Manipon, G., Liang, C., Fielding, E., Milillo, P., Hua, H., and Coletta, A. (2015). Rapid Damage Mapping for the 2015 M _w 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO–SkyMed and ALOS-2 Satellites. *Seismological Research Letters*, 86(6):1549–1556.

Zlatanova, S. and Li, J. (2008). Introduction. In Zlatanova, S. and Li, J., editors, *Geospatial Information Technology for Emergency Response*, pages Xi –Xii. Taylor - Francis Group, London, 1 edition.

A APPENDICES

A.1 NEURAL NETWORK TRAINING FOR DAMAGE DETECTION

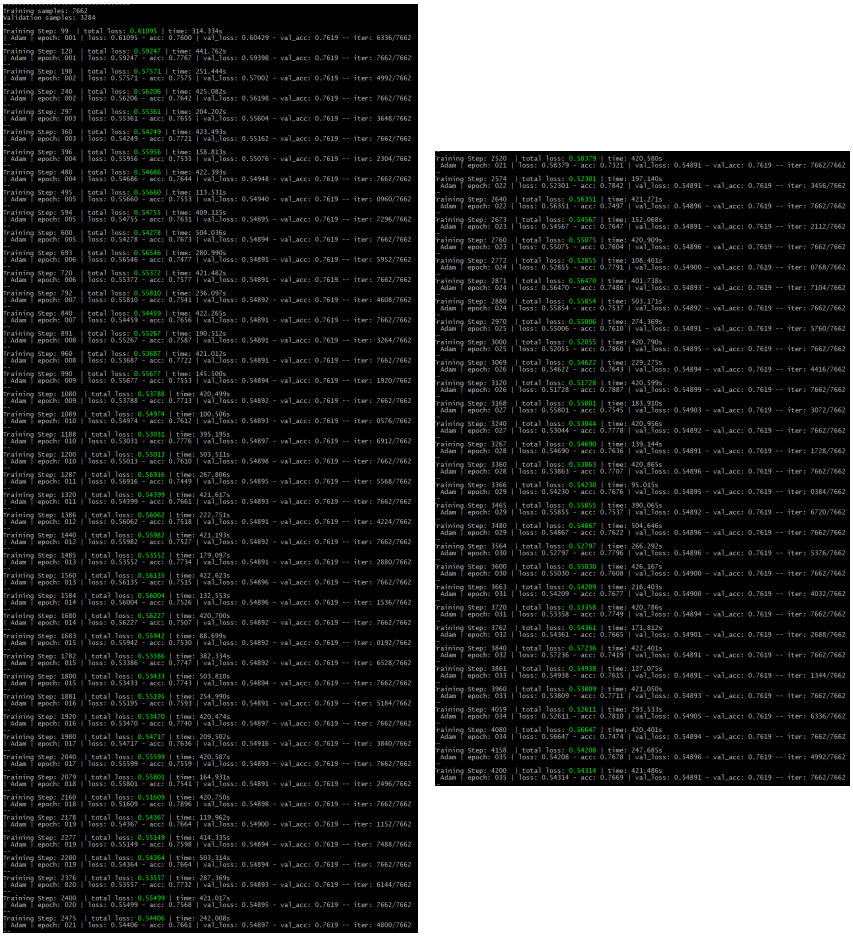


Figure A.1: Full results training CNN with 35 epochs.

A.2 RESULTS EMPIRICAL APPROACH TO DAMAGE DETECTION

Table A.1: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for interferometry univariate change detection with a threshold of 0.30. Cohen Kappa Score: 0.05896

	No	Damage	unknown
No	439.0	80.0	0.0
Damage	254.0	67.0	0.0
unknown	12.0	4.0	0.0

	precision	recall	f1-score	support
No	0.44	0.21	0.28	321
Damage	0.62	0.85	0.72	519
unknown	0.00	0.00	0.00	16
avg / total	0.54	0.59	0.54	856

Table A.2: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Hue based univariate change detection with a threshold of 0.11. Cohen Kappa Score: 0.07099

	No	Damage	unknown
No	179.0	340.0	0.0
Damage	83.0	238.0	0.0
unknown	8.0	8.0	0.0

	precision	recall	f1-score	support
No	0.41	0.74	0.52	321
Damage	0.66	0.34	0.45	519
unknown	0.00	0.00	0.00	16
avg / total	0.55	0.49	0.47	856

Table A.3: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Saturation based univariate change detection with a threshold of 0.07. Cohen Kappa Score: 0.42963

	No	Damage	unknown
No	349.0	170.0	0.0
Damage	64.0	257.0	0.0
unknown	5.0	11.0	0.0

	precision	recall	f1-score	support
No	0.59	0.80	0.68	321
Damage	0.83	0.67	0.74	519
unknown	0.00	0.00	0.00	16
avg / total	0.73	0.71	0.71	856

Table A.4: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Value based univariate change detection with a threshold of 0.21. Cohen Kappa Score: 0.38926

	No	Damage	unknown
No	347.0	172.0	0.0
Damage	78.0	243.0	0.0
unknown	5.0	11.0	0.0

	precision	recall	f1-score	support
No	0.57	0.76	0.65	321
Damage	0.81	0.67	0.73	519
unknown	0.00	0.00	0.00	16
avg / total	0.70	0.69	0.69	856

Table A.5: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for the CNN based method. Cohen Kappa Score: 0.0

	No	Damage	unknown
No	519.0	0.0	0.0
Damage	321.0	0.0	0.0
unknown	16.0	0.0	0.0

	precision	recall	f1-score	support
No	0.00	0.00	0.00	321
Damage	0.61	1.00	0.75	519
unknown	0.00	0.00	0.00	16
avg / total	0.37	0.61	0.46	856

Table A.6: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for the Copernicus classification. Cohen Kappa Score: 0.09283

	No	Damage	unknown
No	142.0	377.0	0.0
Damage	51.0	270.0	0.0
unknown	4.0	12.0	0.0

	precision	recall	f1-score	support
No	0.41	0.84	0.55	321
Damage	0.72	0.27	0.40	519
unknown	0.00	0.00	0.00	16
avg / total	0.59	0.48	0.45	856

A.3 NEURAL NETWORK TRAINING FOR DAMAGE CLASSIFICATION

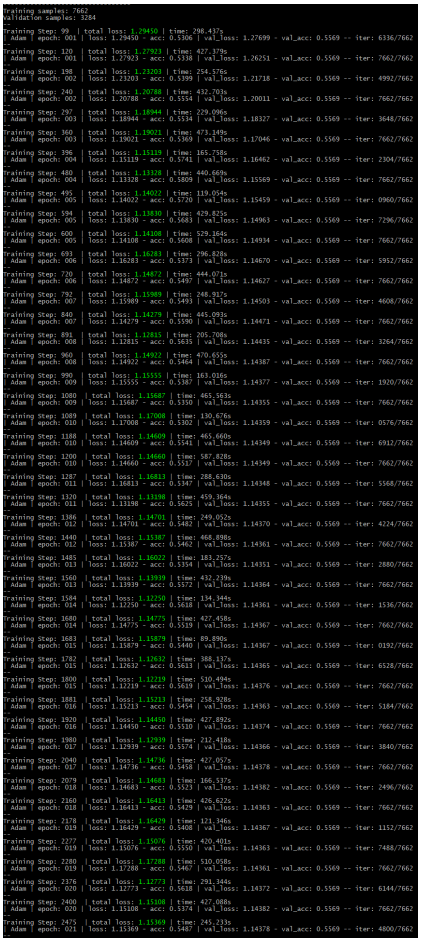


Figure A.2: Full results training CNN with 35 epochs.

A.4 RESULTS EMPIRICAL APPROACH TO DAMAGE DETECTION

Table A.7: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for interferometry univariate change detection with a threshold of 0.23, 0.31, 0.34. Cohen Kappa Score: 0.05110

	none	partial	significant	destroyed	unknown
none	241.0	42.0	8.0	35.0	0.0
partial	128.0	34.0	8.0	23.0	0.0
significant	136.0	18.0	11.0	21.0	0.0
destroyed	91.0	14.0	7.0	23.0	0.0
unknown	9.0	3.0	0.0	4.0	0.0

	precision	recall	f1-score	support
none	0.22	0.17	0.19	135
partial	0.40	0.74	0.52	326
significant	0.31	0.18	0.22	193
destroyed	0.32	0.06	0.10	186
unknown	0.00	0.00	0.00	16
avg / total	0.33	0.36	0.30	856

Table A.8: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Hue based univariate change detection with thresholds of 0.08, 0.11, 0.88. Cohen Kappa Score: 0.0543

	none	partial	significant	destroyed	unknown
none	79.0	41.0	205.0	1.0	0.0
partial	31.0	28.0	132.0	2.0	0.0
significant	25.0	20.0	141.0	0.0	0.0
destroyed	23.0	15.0	97.0	0.0	0.0
unknown	3.0	5.0	8.0	0.0	0.0

	precision	recall	f1-score	support
none	0.00	0.00	0.00	135
partial	0.49	0.24	0.32	326
significant	0.26	0.15	0.19	193
destroyed	0.24	0.76	0.37	186
unknown	0.00	0.00	0.00	16
avg / total	0.30	0.29	0.25	856

Table A.9: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Saturation based univariate change detection with a threshold of 0.08, 0.08, 0.31. Cohen Kappa Score: 0.25029

	none	partial	significant	destroyed	unknown
none	280.0	0.0	46.0	0.0	0.0
partial	104.0	0.0	88.0	1.0	0.0
significant	53.0	0.0	133.0	0.0	0.0
destroyed	41.0	0.0	94.0	0.0	0.0
unknown	6.0	0.0	10.0	0.0	0.0

	precision	recall	f1-score	support
none	0.00	0.00	0.00	135
partial	0.58	0.86	0.69	326
significant	0.00	0.00	0.00	193
destroyed	0.36	0.72	0.48	186
unknown	0.00	0.00	0.00	16
avg / total	0.30	0.48	0.37	856

Table A.10: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for Value based univariate change detection with a threshold of 0.13, 0.18, 0.26. Cohen Kappa Score: 0.18838

	none	partial	significant	destroyed	unknown
none	141.0	93.0	57.0	35.0	0.0
partial	44.0	69.0	54.0	26.0	0.0
significant	12.0	54.0	85.0	35.0	0.0
destroyed	5.0	27.0	64.0	39.0	0.0
unknown	0.0	4.0	9.0	3.0	0.0

	precision	recall	f1-score	support
none	0.28	0.29	0.29	135
partial	0.70	0.43	0.53	326
significant	0.28	0.36	0.31	193
destroyed	0.32	0.46	0.37	186
unknown	0.00	0.00	0.00	16
avg / total	0.44	0.39	0.40	856

Table A.11: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for the CNN based method. Cohen Kappa Score: 0.0

	none	partial	significant	destroyed	unknown
none	101.0	33.0	6.0	186.0	0.0
partial	41.0	24.0	5.0	123.0	0.0
significant	40.0	10.0	2.0	134.0	0.0
destroyed	11.0	10.0	3.0	111.0	0.0
unknown	4.0	1.0	0.0	11.0	0.0

	precision	recall	f1-score	support
none	0.00	0.00	0.00	135
partial	0.38	1.00	0.55	326
significant	0.00	0.00	0.00	193
destroyed	0.00	0.00	0.00	186
unknown	0.00	0.00	0.00	16
avg / total	0.15	0.38	0.21	856

Table A.12: Confusion matrix [Vertical Ground Truth, Horizontal predicted] and F1 classification for the Copernicus classification. Cohen Kappa Score: 0.07871

	none	partial	significant	destroyed	unknown
none	101.0	33.0	6.0	186.0	0.0
partial	41.0	24.0	5.0	123.0	0.0
significant	40.0	10.0	2.0	134.0	0.0
destroyed	11.0	10.0	3.0	111.0	0.0
unknown	4.0	1.0	0.0	11.0	0.0

	precision	recall	f1-score	support
none	0.20	0.82	0.32	135
partial	0.51	0.31	0.39	326
significant	0.31	0.12	0.18	193
destroyed	0.12	0.01	0.02	186
unknown	0.00	0.00	0.00	16
avg / total	0.32	0.28	0.24	856

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.