Characterising Open Cast Mining from Satellite Data

by

Austin Dibble

In partial fulfilment of the requirements
for the degree of MSc in Advanced Computer Science
with Artificial Intelligence

Department of
Computer and Information Sciences

August, 2023

Abstract

Open-pit mining, a prevalent surface mining method, is responsible for extracting a significant portion of the world's minerals due to its economic advantages. However, this technique poses considerable environmental and social challenges. The need for effective mapping and monitoring of global mining areas is becoming increasingly critical to assess the associated environmental and anthropological risks.

This research leverages remote sensing and deep learning technologies to monitor and detect spatio-temporal changes in open-pit mining regions. A comprehensive review of the literature is undertaken, examining recent studies on the identification of mining regions, the assessment of environmental impacts, the application of deep learning in change detection, and the surveillance of activity in surface mining areas. The research methodology details the creation of a unique dataset, OMS2CD, comprised of Sentinel-2 satellite images of mining regions which have been annotated for change detection. Deep learning models, including TinyCD, LSNet, and DDPM-CD, are trained on this dataset. To quantify the levels of activity, a novel metric, the Normalized Difference Temporal Change Index (NDTCI), is introduced.

The performance of the models, the influence of cloud cover and seasonal variations, and the impact of utilizing masks for areas of interest is discussed. The application of the trained models in detecting changes and mapping activity levels in unseen mining sites is demonstrated through case studies.

The key findings include the strong correlation between the median model prediction and ground truth data, and the utility of the NDTCI as a normalized measure of activity. In essence, this research proposes a methodology that utilizes freely accessible data and deep learning models to monitor surface mining regions by detecting changes in satellite imagery, serving as a proxy for activity. The NDTCI offers a quantitative approach to mapping the activity levels of mining sites over time.

Declaration

This dissertation is submitted in part fulfilment of the requirements for the degree of MSc in Advanced Computer Science with Artificial Intelligence of the University of Strathclyde.

I declare that this dissertation embodies the results of my own work and that it has been composed by myself.

Following normal academic conventions, I have made due acknowledgement to the work of others.

I declare that I have sought, and received, ethics approval via the Departmental Ethics Committee as appropriate to my research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to provide copies of the dissertation, at cost, to those who may in the future request a copy of the dissertation for private study or research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to place a copy of the dissertation in a publicly available archive.

(please tick) Yes [X] No [ ]

I declare that the word count for this dissertation (excluding title page, declaration, abstract, acknowledgements, table of contents, list of illustrations, references and appendices) is 15551.

I confirm that I wish this to be assessed as a Type 1 2 3 4 (5) Dissertation (please circle).

Signature:

Date: 14/08/2023

Acknowledgements

First and foremost, I'd like to extend my gratitude to Dr. Marc Roper. He not only laid out the blueprint for this project but also has been a source of guidance and expertise.

To my wife, Alicia: your continuous support, patience, and meticulous editing have been instrumental in the completion of this thesis. Your unwavering belief in me has been a constant source of motivation.

To my family: even from afar, your love, understanding, and encouragement have been an anchor during this journey. Being in another country, away from home, presented its own challenges, but your constant belief in my abilities and goals made it easier. Your support has been instrumental in every phase of this adventure.

I am fortunate to have a supportive circle of friends who embarked on this academic adventure alongside me. Together, we navigated challenges, celebrated successes, and supported each other. A special mention to the "Food Amigos"—I am deeply grateful for our collective friendship and our shared experiences from this past year.

Furthermore, I'd like to express my appreciation to Andrea Codegoni for granting me explicit permission to incorporate their code into my research. Their generosity has enriched this work and underscores the collaborative spirit of the academic community.

Lastly, I want to acknowledge all the researchers referenced in this dissertation who have generously released their work under open-source licences. Open-access publications and open-source code are foundational to academic progress, making projects like mine feasible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The mining of mineral resources is crucial for many modern industries, yet it can lead to significant environmental impacts if not properly managed. Open-pit mining (also known as open-cast mining), which focuses on near-surface deposits, now accounts for over 70% of global mineral production due to its economic advantages over underground mining. However, open-pit mines can cause major disturbances to landscapes and ecosystems. Continuous monitoring of active mining sites is therefore critical for environmental protection and sustainable development.

Recent advances in Earth observation satellites and deep learning (DL) algorithms provide new opportunities for automated, large-scale monitoring of mining activity. In particular, change detection techniques applied to multi-temporal satellite imagery can identify where surface changes have occurred within mining areas over time. By training deep neural networks on before-and-after image pairs, surface changes related to mining operations like excavation, road construction, deforestation, and waste dumping may be detected.

This dissertation investigates state-of-the-art deep learning methods for change detection using freely available Sentinel-2 multispectral imagery. Focusing on the novel application of monitoring open-pit mining activity, custom models are trained and tested on a new dataset tailored to mining land cover changes. Quantitative activity metrics are proposed based on the predicted surface changes. Case studies demonstrate how these tech-

niques can map mining expansion and changing activity levels at sites worldwide.

The ability to automatically monitor mining activity supports improved governance and environmental impact assessment. This research explores how recent advances in deep learning and open satellite data can be applied to this real-world challenge. The approaches developed provide new capabilities for transparent monitoring and sustainable mining practices.

1.1 Background

1.1.1 Open-pit Mining

Open-pit mining is a method of surface mining that focuses on extracting ore deposits at or near the earth's surface. It uses a series of horizontal 'benches' that are blasted into the surface. Open-pit mining is commonly used to extract a variety of metallic and non-metallic elements from the surface, including aluminium, bauxite, copper, iron, coal, uranium, and phosphate (Altiti et al. 2021). Surface mining methods such as open-pit mining have a number of comparative advantages over other forms of mineral extraction. Namely, surface mining is in general more economical than underground methods. Because of this, surface mining is responsible for approximately 70% of global mineral production (Haldar 2018, p. 5). Unfortunately, open surface mines can cause a wide range of environmental and ecological impacts. These include the destruction of natural ecosystems, soil, water, and air pollution, and impacts on human settlements. Additionally, open-pit mine areas include not only the mine itself (where the minerals are extracted from the exposed orebody), but also the waste (tailings) dump areas and waste rock piles. For these reasons, mapping and monitoring global mining areas has become increasingly important for assessing the environmental and anthropological risks of mining activity (Maus et al. 2022).

1.1.2 Remote Sensing and Monitoring

Very generally, remote sensing (RS) can be defined as "acquiring knowledge from a distance" (D. Kumar, Singh, and Kaur 2019). One of the most common platforms used for re-

Figure 1.1: Sentinel-2 aerial image of an open-pit mine facility. Some sites, like this one, can be of substantial size. This image covers an area of over 36 km^2.

mote sensing in the 21st century is some type of observational satellite orbiting the earth. Satellites are considered to be "spaceborne" remote sensing platforms. In the context of remote sensing, Earth Observation (EO) is a common term which is focused specifically on observing and studying Earth's systems. As such, EO satellite technologies can include those designed for weather and atmosphere monitoring, as well as surface observation.

Since the launch of the first EO satellite Vanguard 2 by the United States Naval Research Laboratory, the number of EO satellites has grown rapidly. As of 2020, there were 845 imaging instruments currently operating or planned for launch within the next 15 years. Such a significant increase in the number of available EO missions has provided countless opportunities for EO research in areas such as climate and ecological monitoring, agriculture, weather forecasting, and more (Q. Zhao et al. 2022). Here, we review the most common types of remote sensing data available from these EO missions, and how they can be applied to the task of change detection.

1.1.3 Types of Remote Sensing Data

There are three broad categories of remote sensing satellite data for EO which are most commonly used with machine and deep learning applications: Synthetic Aperture Radar (SAR), multi-spectral, hyper-spectral, and panchromatic (W. Shi et al. 2020, pp. 6–7). In this section, each of these are defined in more detail.

1.1.3.1 Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) is an active, rather than passive, EO technology. It uses radar transmissions to detect information about the earth's surface terrain. Because of its active transmission, SAR data is less susceptible to weather conditions and can operate completely independent of surface lighting from the sun or other sources. Using the differences in signal phase between multiple radar acquisitions allows for highly sensitive assessment of terrain changes on the surface. SAR is therefore invaluable for damage assessment after natural events such as earthquakes (Plank 2014, pp. 3–6). Launched by the ESA (European Space Agency) in 2014, Sentinel-1 has become one of the most vital SAR satellite missions[1] due to its high revisit frequency and the ease of access to its mission data (*Sentinel-1 - Missions - Sentinel Online* 2023).

SAR interferometry (InSAR) is a derivative imaging technology of SAR, which uses the phase difference between two SAR acquisitions to create a coherence (or correlation) image. The coherence is a complex phase signal which can be used to identify terrain elevation and terrain movement (differential SAR interferometry) (Plank 2014, pp. 7–8). Such coherence images lend themselves readily to surface change detection tasks and have been used with statistical and deep learning techniques to detect surface changes over time (Jia and Z. Zhao 2021; Qu et al. 2022; Ebel, Saha, and X. X. Zhu 2021; Canty et al. 2020).

While SAR has the aforementioned advantages, it is difficult to use in machine and deep learning applications due to the high amounts of 'speckle noise' present in SAR images. In InSAR, this is largely due to the lack of phase correlation between surfaces such as

[1] Note that the term 'satellite mission' rather than 'satellite' is used to refer to the full mission objectives and goals of a single satellite or a constellation (multiple satellites) designed for the same task.

vegetation. That is one of the greatest challenge that must be overcome in order to use InSAR data in DL systems (Jia and Z. Zhao 2021).

1.1.3.2 Multi-spectral

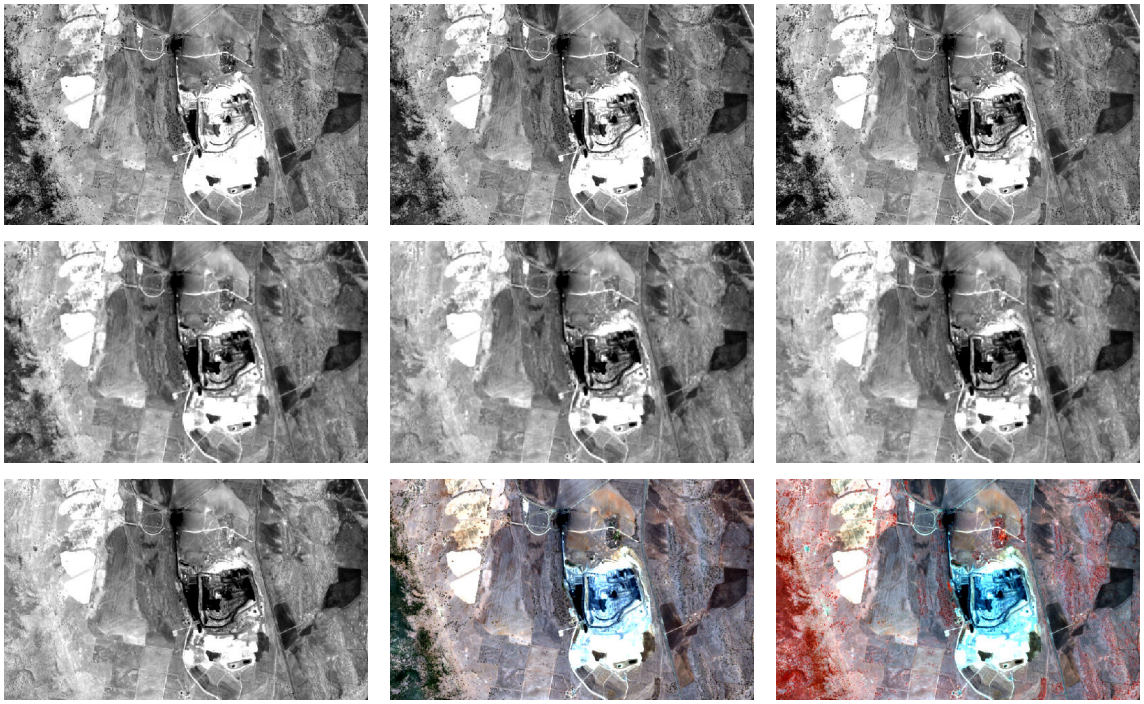


Figure 1.2: Sentinel-2 image with bands 2-7, a true colour image, and a false colour image (substituting red with the NIR band), from top-left to bottom-right.

The term 'multi-spectral' in the context of EO broadly refers to a collection of satellite optical imagery which is collected passively in an array of discrete wavelength bands on the electromagnetic spectrum (usually ranging from infrared to visible). Two multi-spectral satellite missions commonly used for land use/land cover mapping are Sentinel-2 and Landsat-8. Sentinel-2 was launched by the ESA in 2015 as part of the Copernicus initiative, and it is equipped with a multispectral instrument (referred to as MSI). Landsat-8 was launched in 2013 as a continuation of the USGS Landsat Earth observation programme, and it is equipped with the Operational Land Imager (OLI) (Mandanici and Bitelli 2016).

Sentinel-2's data is provided as freely accessible by the ESA. At its peak, it has had a temporal resolution (or surface region revisit frequency) of only 5 days. Additionally, it provides 13 bands of optical data, 4 of which (red, green, blue, and near-infrared) are

at a spatial resolution of 10 metres per pixel[2] (example given in Figure 1.2). Because of its high revisit frequency, it is estimated that it can provide, on average, at least one cloud-free image per month for all areas that it transits (Mandanici and Bitelli 2016; Phiri et al. 2020).

1.1.3.3 Hyper-spectral

Hyper-spectral imaging is based on the same technology as multi-spectral, and typically in a similar wavelength range (visible to infrared). The primary distinction is that hyper-spectral imagery provides hundreds or thousands of narrow, distinct bands in a single acquisition. Some hyper-spectral EO missions include Hyperion, HISUI, and PRISMA which each have over 100 spectral bands in their data (Transon et al. 2018).

While hyper-spectral imagery provides an abundance of spectral information due to its large number of bands, its abundance of data can make it unwieldy for machine and deep learning applications. Redundancy in the data and a lack of interpretability of the high dimensionality of the hyper-spectral data has meant that it has seen less application in these research areas than the aforementioned multispectral data (W. Shi et al. 2020).

1.1.4 Pixel-based Vegetation and Soil Difference Indices

To assist in remote sensing using satellite data, a number of numerical indices have been proposed to assist in spatial classification of satellite imagery. One such index, the Normalised Difference Vegetation Index (NDVI), as shown in Equation 1.1 is commonly used in EO literature. It was first introduced as a metric in 1974 (Vorovencii 2021, p. 6).

$$\text{NDVI} = R_{\text{NIR}} - \frac{R_{\text{Red}}}{R_{\text{NIR}} + R_{\text{Red}}} \tag{1.1}$$

Other such indices include Normalised Difference Water Index (NDWI) (S. Kumar and Arya 2021, p. 1909), and Soil Adjusted Vegetation Index (SAVI) (Vorovencii 2021, p. 7) among others. The different indices range from -1 to 1, and provide a normalised metric of comparison on a pixel-by-pixel basis of measure. These index-based metrics are a

[2]Metres per pixel (often only given as metres) is commonly used to express the image resolution of satellite or aerial imagery.

Figure 1.3: Example image demonstrating an NDVI matrix. Pixels which are more white represent a higher NDVI, and vice-versa.

type of pixel-based classification of image regions that is calculated using the intensity of two or more spectral bands. They have the advantages of being easy to interpret, and fast to calculate. Additionally, they are completely unsupervised and require no training, unlike machine learning and deep learning methods of classification.

1.1.5 Surface Change Detection

Change detection (CD), in the realm of remote sensing, can be understood as a process by which certain features are identified through analysis over different time intervals. It necessitates the use of Geographical Information System (GIS) or EO data to observe changes that can happen due to human activities or natural events. The identification of these alterations is achieved by studying images of specific spatial zones over time. It is an important element in satellite mapping, monitoring environmental shifts, and analysing changes in land use and land cover (LU/LC).

Change detection can be distilled down to a few distinct approaches: algebraic, transform-based, classification-based, and 'advanced' techniques (Afaq and Manocha 2021; S. Kumar and Arya 2021). Alongside these classification approaches, there are two primary types of change which have typically been examined: binary and from-to. Binary CD is concerned only with whether a change has occurred or not, while from-to is interested in the type of change (e.g. from water to grass) (Bai et al. 2022).

While CD as a field of study has been developing for many years, deep learning-based

techniques for change detection (DLCD) have been studied mostly within the last decade. Even so, DLCD has given rise to significant improvements in accuracy over traditional CD methods (Bai et al. 2022).

1.2 Objectives and Summary

This dissertation project aims to approach open-pit mine activity monitoring from a combined perspective of the aforementioned state-of-the-art methods within a resource-constrained environment. To achieve this goal, our work primarily focuses on deep learning methods of change detection which are accessible on freely available computing resources. Besides assessing the methods best suited for open-pit mine monitoring, we desire to establish a functional approach to activity detection in open-pit mine areas that can be broadly replicated.

1.2.1 Research Questions

To achieve the aforementioned goals, our research is informed by a collection of research questions:

1. How can the size, activity level, and changes over time of open-pit mines be measured using state-of-the-art techniques?
2. What is 'change' in the context of activity in open-pit mine areas?
3. What existing DLCD architectures are best suited for this task?
4. Can existing DLCD architectures be adapted to a new data domain using fine-tuning?
5. Can DLCD be used to monitor and detect the activity level of open-pit mines over time?

To answer the proposed research questions, this project focuses on deep learning architectures for binary change detection across bi-temporal[3] image pairs. We utilise freely

[3] In the context of this dissertation, bi-temporal generally refers to a pair of images which represent the same spatial location at two distinct times.

available Sentinel-2 data obtained using the Google Earth Engine (GEE) (Gorelick et al. 2017), following a process similar to the one described by (Balaniuk, Isupova, and Reece 2020). Using such data, we produce the first publicly available open-pit mine Sentinel-2 change detection (OMS2CD) dataset, which is used to train and validate a selection of state-of-the-art change detection architectures.

As a measure of activity for change detection, we propose the Normalised Difference Temporal Change Index (NDTCI) and demonstrate how it can be used to monitor activity in surface mine areas. The trained models are then employed in specific case study areas to measure the NDTCI and evaluate the feasibility of temporal change mapping of open-pit mines.

For this project, it's essential to understand that the training and evaluation of the DLCD model are steps towards our primary objective. Our main aim is to establish a method for monitoring open-pit mines using Sentinel-2 data. While the DLCD models play a crucial role in this, the resulting change prediction is just one component of the broader activity detection process.

Chapter 2

Literature Review

This literature review investigates the state of the art in open-pit/open-cast mine activity monitoring using remote monitoring technologies. In particular, this is a review of recent papers in this area, primarily from the last four years. Although monitoring surface mines for environmental reasons has been a concern for a long time, the amount of literature in adjacent research areas has grown rapidly with the popularity of deep learning (Bai et al. 2022, p. 3).

Numerous research domains intersect with the objectives of this dissertation. Specifically, these include identifying illicit mining practices, pinpointing and charting surface mining zones, detecting changes on the Earth's surface through satellite imagery, discerning the environmental repercussions of mining, and tracking surface mining operations. An exploration of existing literature suggests minimal studies directly align with this dissertation's focus: real-time surveillance of mining sites to gauge their ongoing activity level. Nevertheless, several contemporary research methods can be adapted to address this issue.

2.1 Identifying Surface Mining Areas

The automated detection of mining zones holds significant interest for entities focused on illicit mining activities that are otherwise challenging to oversee and track. For regulatory bodies with limited resources, several studies offer valuable insights, including (Balaniuk,

Isupova, and Reece 2020; Kozińska and Górniak-Zimroz 2021; Nava et al. 2022; Gallwey et al. 2020). These works primarily aim to pinpoint mining operations and compare them against acknowledged legal activities, facilitating the automated identification of potential illegal mining. Research in this domain often leverages cost-effective or complimentary tools and datasets, accommodating resource constraints. This report finds (Balaniuk, Isupova, and Reece 2020) particularly relevant, as it delves into the utilization of freely available technologies, such as Google Colaboratory (or "Colab") (*Google Colaboratory* 2023) and Sentinel-2 data from the Google Earth Engine. (Kozińska and Górniak-Zimroz 2021) highlights the lack of research on illicit mining activity monitoring but reviews existing methodologies employing a variety of technologies, including visual spectrum and InSAR data. Another notable mention is (T. Chen et al. 2022), which adeptly identifies mining zones using an adapted U-Net framework. However, they rely on China's Gaofen-2 VHR (very-high resolution)[1] satellite, which isn't available for our purposes.

C. Wang et al. use a modified Mask R-CNN to identify mining areas and monitor them as they change over time. Their approach is successful on the dataset they collected of a particular region in China. Whether their approach would be applicable to a more diverse range of geographical areas is not explored (Chunsheng Wang et al. 2020). Zhao et al. propose a state-of-the-art segmentation algorithm for identifying open-pit mining areas, and compare the performance of their model over several datasets and against other segmentation models such as Mask R-CNN (L. Zhao et al. 2022). Their novel proposed segmentation models consistently outperform Mask-RCNN on their chosen dataset (as measured using the F1-score). They focus on mining area segmentation, but not temporal monitoring. Such segmentation approaches have the benefit of being able to identify previously unknown mining areas and separate them from the surrounding terrain, but identifying changes over time would require more processing steps.

Shi et al. develop a custom multitask learning framework for land cover classification in open-pit mining areas, combining CBE-04 and Sentinel-2 satellite data. Their results are impressive in pinpointing mining zones, but they have not shared their labelled classifica-

[1] Various terms are used in literature to describe optical satellite data with a resolution approximating 5 metres per pixel or finer. "VHR" is among the most prevalent. It stands out from sources like Sentinel-2 due to its enhanced depiction of surface details. Some privately accessible satellite data offer resolutions of 1 metre per pixel or finer.

tion data (J. Shi et al. 2022). In a similar vein, (Chen Wang, T. Chen, and Plaza 2023) employ multiscale segmentation and classification on VHR images, generating detailed land use classification maps for specific mining areas. However, they use private software for segmentation and keep their data sources confidential. Although both studies aim to identify extensive open-pit mining zones and their land use, the unavailability of their data makes it challenging to incorporate their methods into our research.

Two works previously mentioned, (Nava et al. 2022; Gallwey et al. 2020), are primarily focused on small-scale artisanal mine activity. This activity exists in a unique context and relies heavily on the availability of VHR data. Both studies utilize U-Net architectures to identify artisanal mining areas, a methodology that is adopted by numerous works in this area (Jiang et al. 2022; L. Zhao et al. 2022; Shafique et al. 2022; T. Chen et al. 2022; Huang et al. 2022). Nava et al. achieved a satisfactory F1-score with their dataset. However, they do not provide comparative studies due to the novelty of the research domain. In contrast, Gallwey et al. modify their approach slightly, employing their classifier to identify land use (comprising mining, vegetation, and 'built' zones) and assess land use alterations over time. Their methodology seems promising and aligns with our objectives of activity detection. However, that have not made their labelled dataset publicly available.

A number of studies also investigate the idea of identifying illegal mine activity using InSAR data. One such example uses a novel InSAR processing technique to identify active mine areas (S. Wang et al. 2020). Their satellite observations are then followed up by ground-based and UAV missions to confirm their findings. Such an approach is promising because it utilises an easily replicable statistical process. However, it requires access to legal mine site survey data in order to properly classify the discovered mining areas.

Unique in the area of identification is the work by Maus et al., which is a comprehensive dataset of global mining sites covering an area of 101,583 km^2. Unlike the other works mentioned in this section which use various automatic methods for discovering and categorising mining areas, they provide a novel dataset which is manually constructed and delineated using information about known mining sites. The dataset also provides polygon definitions for all mining areas, making it useful for monitoring the exact area occu-

pied by the mine (Maus et al. 2022). While the Maus et al. dataset is spatially useful for identifying mining site locations, the database provided by Jasansky et al. goes into more depth. It provides, where possible, information about the type and quantity of commodity produced at the facility, dates for the start and end of production, and waste and processing capacity, among other fields (Jasansky et al. 2023). For our research goals, both of these datasets are useful for identifying targets sites for EO data gathering and observation case studies.

2.2 Surface Mine Environmental Impact Assessments

With regard to assessing the extent of environmental impacts, Zhu et al. utilises a novel PCA algorithm using a moving window to calculate the index of multiple environmental indicators around open-pit mining sites. They propose a novel environmental index (MW-RSEI) which demonstrates that the environmental effects of open-pit mining activity extends far beyond the immediate vicinity of the damaged surface area of the mine (D. Zhu et al. 2020). Nascimiento et al. are also primarily concerned with the extent of deforestation as surface mining areas expand. They adopt an algorithmic technique which combines VHR satellite imagery with captured LiDAR (laser imaging, detection, and ranging) in order to segment and classify ground changes (deforestation, re-vegetation, etc.). Using this technique, they successfully assess the exact land area changes observed during their investigation period (Nascimento et al. 2020).

Most of the literature that we have found that is concerned with monitoring the environmental impact of mining areas is focused on measuring the vegetation changes over time. There are a number of case studies such as (Vorovencii 2021; V. Kumar and Yarrakula 2022; Ruifeng et al. 2022; Guo et al. 2022) which focus on vegetative or soil change indices such as the NDVI. Using vegetation indices has the advantage of being a fairly robust and proven proxy metric for environmental changes. Combined with the digital elevation model TanDEM-X, (Wu et al. 2020) achieve excellent change detection performance using a number of RS features such as NDVI and NDWI. Kumar and Arya review algebraic (methods using indices such as NDVI) and DL methods for change detection and make comparisons in their use-cases and accuracy. They observe that the true accu-

racy of each method is hard to compare broadly, and that they may be more appropriate depending on certain factors of the required use-case (S. Kumar and Arya 2021). From the perspective of environmental monitoring specifically, such algebraic indices are easy to interpret, though they are still only a proxy measure for environmental activity. Additionally, such indices are very sensitive to seasonal changes such as vegetative growth and snow cover.

2.3 General Deep Learning Change Detection

The field of Deep Learning Change Detection (DLCD) aims to automatically detect surface changes in satellite imagery by comparing temporally spaced (before and after) images of the same geolocation. Bai et al. provide a recent and comprehensive review of the state of the art in this area. Overall, they find that deep learning methods are more robust and accurate than traditional statistical and machine learning methods. The two different branches in the field of change detection look at binary (if the surface has changed) or from-to (e.g. from vegetation to mine). They find that the two major challenges facing the area of DLCD methods are 1) insufficient training data and 2) the need for high computational power. After a thorough comparison of the advantages and disadvantages of different DL architectures in change detection, they find that 'coupled' DLNN (deep learning neural network) architectures sustain a performance advantage over 'separate' DLNN architectures. The trade-off is that coupled architectures tend to be more complex to build and understand (Bai et al. 2022). Following their work, we find that the direct classification method (DCM) and differencing neural network method (DNNM) to be the most promising for this project because of their superior performance characteristics.

Another extensive review of the field of DLCD from the same time period is given in (Jiang et al. 2022). They find that CNN architectures are the most popular DNN architectures for DLCD, and that these are most often applied to multi-spectral optical data. While they do not make a performance comparison between the reviewed model architectures, their description of promising model architectures is thorough and compliments the work of (Bai et al. 2022). We find again that coupled DLNN architectures such as Siamese networks show great promise and relevance to our work in this project. While they also

give a comprehensive list of VHR change detection datasets, we find that these datasets would not be useful to us without significant downsampling.

While the area of DLCD is still a small subfield, the last few years have seen rapid growth in the diversity of approaches, catalysed by the growth of DL architectures as a whole. There are a myriad of different architectural designs, each with their own trade-offs. Here we review a few of the most recent. Zhang et al. introduce a deep Siamese network that takes advantage of image transformers for change detection (DSNCoT) in satellite imagery. They achieve state-of-the-art results on the multi-season change detection dataset for general change detection (M. Zhang et al. 2022). Yin et al. also achieve state-of-the-art performance using a self-attention transformer network (SAGNet) on VHR imagery obtained from Google Earth. They provide models of various sizes to allow for performance trade-off selection (Yin et al. 2023). While such transformer-based architectures have yielded groundbreaking performance on DLCD datasets, it is important to note that they are generally less efficient than other approaches due to the increased computational complexity of self-attention blocks (Keles, Wijewardena, and Hegde 2022).

Two promising DLCD papers that we reviewed use SAR image data rather than visual spectrum data with their DL models. Qu et al. propose DDNet to perform multi-modal comparison in the spatial and frequency domain of SAR data. They achieve superior results on the Ottawa and Yellow River SAR datasets compared with other model architectures (Qu et al. 2022). They also provide their PyTorch implementation and Google Colab demo open-source in their GitHub repository. Jia and Zhao propose a novel deep belief network (gΓ-DBN) to find the differences between bi-temporal SAR data. Also using the Yellow River dataset, they demonstrate robust performance in binary change detection (Jia and Z. Zhao 2021).

Because of the lack of available training data in the area of change detection for open-pit mine surfaces, we are also very interested in recent works which focus on reducing hardware and training data requirements via efficient architecture designs. TinyCD is a promising lightweight network based on Siamese U-Net for change detection. It achieves state-of-the-art accuracy on the WHU-CD and LEVIR-CD datasets while having 13 to 140 times fewer parameters than other models of comparable performance (Codegoni,

Lombardi, and Ferrari 2022). LSNet also achieves high accuracy on the CDD benchmark dataset while being over 90% more computationally efficient than other recent models (Liu, Huaixin Chen, and Z. Wang 2022). While focused less on computational efficiency, SemiCD is also a more recent model architecture that operates in a semi-supervised manner and has been shown to achieve comparable benchmark performance while requiring only 10% as much labelled data as other models. Such a drastic reduction in the requirements for labelled training data lends itself well to the area of DLCD because of the difficulty of preparing the ground-truth labelled datasets (W. G. C. Bandara and Patel 2022).

Concluding the discussions in this domain, DDPM-CD stands out as a unique approach to binary DLCD. It leverages a readily available diffusion model, pre-trained on an extensive dataset of over one million unlabelled Sentinel-2 images. The diffusion features are then fed through a lightweight differencing module to produce a binary change map. It achieves state-of-the-art performance on a number of CD datasets, including WHU-CD and LEVIR-CD, and could potentially reduce training requirements compared to other models because of its pre-trained diffusion backbone (W. G. C. Bandara, Nair, and Patel 2022). We are interested in comparing this approach to some others mentioned in our review, although it is worth noting that its diffusion model backbone is quite large and may be difficult to train with limited resources.

2.3.1 Change Detection Datasets

It is noteworthy in this overview to highlight the presence of diverse datasets suitable for training change detection models. An up-to-date list of available databases and open-source repositories for change detection using DL is available on GitHub (wenhwu 2023). Specifically, we find “Sentinel-2 Multitemporal Cities Pairs” (Leenstra et al. 2020), “Onera Satellite Change Detection” (OSCD) (Daudt et al. 2018) (©2011 IEEE), LEVIR-CD (Hao Chen and Z. Shi 2020), “Open-pit Mine Change Detection” (OMCD) (Li et al. 2023) and “Ordos OM Change Detection” (Du et al. 2022) to be particularly relevant datasets to this dissertation. If fine-tuning is necessary, Sentinel data will be preferable since it is readily available through online platforms such as Google Earth Engine. (Li et al. 2023) (©2011 IEEE) and (Du et al. 2022) are also of specific interest since they are the only

publicly accessible open-pit mine change detection dataset that we have identified in our review, even though they are using VHR Gaofen satellite and not Sentinel-2 imagery. As mentioned in the previous model review section, there are many other CD datasets available for supervised training. However, we have found the difference in the training domain (the scale and type of image features) of those datasets makes it difficult to train networks for our specific task of change detection in open-pit mine areas. That challenge is reviewed in Appendix G.

2.4 Open-pit and Surface Mine Change/Activity Monitoring

This subsection delves into literature closely aligned with the core theme of this dissertation. Our exploration yielded only a handful of papers focusing specifically on the longitudinal monitoring and analysis of individual mines. While related to the discussions in the preceding subsections, the literature examined here shifts from broad detection to *change* detection—concentrating on small alterations over time within the mining area.

Moon and Lee use InSAR data to map the surface changes of dumping sites to a novel indicator of activity, the "Normalized Difference Activity Index" (NDAI). The study area is a single iron mine in North Korea, and they utilise the freely available Sentinel-1 SAR data to perform their research. Most noteworthy is that using this method, they map specific mine activity over time, and project it onto a digital elevation model (DEM) to show exactly which areas of the mine surface had changed. Of interest to this dissertation is also the challenges that they mention, notably that vegetation growth can cause decoherence in the InSAR data, which can be easily mistaken for mine activity (Moon and Lee 2021).

L. Wang et al. employ a technique similar to that of Moon and Lee, harnessing InSAR data from Sentinel-1 to observe the NDAI value of a vast mining zone in China over an extended period of time. By applying linear regression to the time-averaged InSAR coherence values, they create a map delineating zones of low and high mining activity. Moreover, they plot the week-by-week NDAI values to study exact spatial transformations in the mine surface. To validate their NDAI observations, they supplement their findings with Sentinel-2 visual spectrum data (L. Wang et al. 2021). Their investigation extends that of Moon and Lee not only by tracking the changes over time, but also circumventing

the need for a DEM to understand the visual changes. We find both works informative to this dissertation, as they offer insightful examples of analytical methodologies and spatio-temporal change mapping.

In our exploration, we identified four papers emphasising the surveillance of open-pit mine change detection via visual spectrum data.

Du et al. introduce a novel deep learning architecture, DA-UNet++, tailored for identifying surface changes in VHR imagery. They also curate a novel dataset featuring masked VHR satellite imagery for change detection to enhance the training of their architecture. In the end, they demonstrate state-of-the-art performance in change detection for both open-pit mine and non-mine surface areas. However, while in their paper they advocate for automated change detection of open-pit mining zones, they refrain from offering a detailed, quantifiable assessment of activity level based on the identified changes (Du et al. 2022).

Echoing from our earlier discussions, C. Wang et al. propose a novel framework based on Mask R-CNN; termed IMRT. Harnessing VHR image data from the Gaofen-1 and Gaofen-2 satellites, they were able to create high resolution segmentation masks of the target mine areas. The segmentation masks are then monitored for changes by looking at the change in mine area over time. Such landscape transformations are mapped to the DROTL index (damages rate of the topographical landscape), offering a quantitative comparison of surface mine environmental impact. Their research specifically underscores the importance of transfer learning in their technique, deploying a Mask R-CNN model which was pre-trained on the COCO image dataset (Chunsheng Wang et al. 2020).

Li et al. propose a Siamese architecture for change detection: SMCDNet. Via their novel methodology, they achieve state-of-the-art results for open-pit mine change detection (OMCD) using VHR imagery of mines in China. We note that although they achieve superior results, the parameter size of their proposed model (54M parameters) makes its inference slower than some smaller, less accurate methods (Li et al. 2023).

2.5 Comparison

In this review, we have provided a broad overview of the relevant literature for monitoring mine activity. In this field, there are a number of major approaches that reflect the specific and relevant research questions in each case. To give a brief comparison, the two most common reasons for identifying and monitoring surface mines are to 1) perform an environmental impact assessment (D. Zhu et al. 2020; Nascimento et al. 2020) or 2) identify mine locations for comparison against legal, known operations. With regard to source data types, they can be broadly categorised into two types that are typically studied: HR or VHR visual spectrum, and SAR/InSAR data.

Visual spectrum data has two advantages over SAR:

- It is easier to interpret and can be displayed in true/false colour.
- It may be less noisy than SAR data (SAR coherence suffers from high speckle noise) (Jia and Z. Zhao 2021).
- There are VHR satellites that provide much higher spatial resolution than most available SAR satellites.

However, it also has the following disadvantages:

- It can be interrupted by cloud cover (L. Wang et al. 2021).
- It can only be captured during the day. Additionally, images may differ in their appearance depending on the position of the sun relative to the satellite.
- Seasonality may change the appearance of the surface features (M. Zhang et al. 2022).
- VHR data generally needs to be obtained from a private source and is not publicly available.

Although InSAR coherence data is precise enough to detect subtle changes in surface height (even as small as only a few centimetres), its sensitivity to factors such as vegetation and surface moisture poses a challenge. Moon and Lee delve into the details of such challenges, describing how they overcame them in their detailed case study (Moon

and Lee 2021, p. 22). One factor adding more complexity is the technical difficulty of data pre-processing and integration with DEMs.

In the area of monitoring and assessing surface activity of open-pit mines, there are a diverse set of techniques that have been engaged. Two papers that we have reviewed (L. Wang et al. 2021; Moon and Lee 2021) compare bi-temporal InSAR images in order to assess land change in terms of the novel NDAI. Using VHR visual spectrum data paired with DL techniques, other studies have been successful at mapping binary surface changes (Du et al. 2022; Chunsheng Wang et al. 2020; Li et al. 2023). None of the visual spectrum studies make a quantitative measure of change besides C. Wang et al., who use the DROTL index to measure the rate of change in the extent of the mine surface. We find both of these proposed indices (NDAI and DROTL) to be useful descriptors of activity and change in open-pit mine areas.

For general change detection in satellite imagery, there is a large and growing body of recent research. While our review highlights five noteworthy sources, numerous other methodologies and architectures have been explored in the review by Bai et al. The key distinctions among these studies typically revolve around the type of data (i.e., whether SAR or visual), the specific datasets employed for model training and evaluation, the unique model architectures introduced, and whether the researchers have made their code publicly available (Bai et al. 2022).

2.6 Summary

In our exploration of open-pit mine change monitoring literature, it is evident that there is a substantial body of work is committed to identifying mining areas and devising strategies for tracking surface changes. Yet, apart from the two references that employ the NDAI for quantitative assessment of change, few studies aim to directly and objectively measure mining activity levels based on observed changes. Moreover, much of the literature relies on the availability of VHR visual spectrum imagery. For the purposes of this dissertation, obtaining access to sufficient quantities of high resolution visual data over open-pit mines is not feasible. Studies emphasising algebraic strategies, such as vegetative indices, provide valuable insights into proxy methods for tracking long-term

changes in mining areas. Hu et al. stand out by illustrating how regression analysis can be performed on these indices to depict vegetation changes (Hu et al. 2022). Nevertheless, this is approach is distinct from direct change monitoring such as the kind seen in other DLCD references already discussed. It is also worth noting that pixel-based indices are particularly vulnerable to seasonable surface variations, while DLCD methods have demonstrated considerable robustness against such disturbances.

For researchers in this domain with limited resources, there is a notable example which exclusively employs open-source and freely available tools (Balaniuk, Isupova, and Reece 2020). Yet, the task of performing optical image change detection in open-pit mines using freely available sources like Sentinel-2 is still largely unexplored. Moreover, there is a lack of literature that aims to correlate the identified changes to meaningful environmental and economic impact metrics.

Chapter 3

Research Methods

Our main research task for this project was to find a practical manner of assessing and monitoring open-pit mine areas using the resources available to us. As a guide, our research questions (given in Section 1.2.1) were intended to reflect what information (and what experimental processes) were required for us to establish a system for accomplishing that task. The experimental processes that we developed are given below, in detail.

3.1 Models

As discovered in our literature review, deep-learning change detection is a growing subfield focused on identifying spatio-temporal changes in EO observation data. Through the utilisation of novel DL architectures, the accuracy of state-of-the-art binary DLCD models has reached a level where they can be practically useful for tasks such as open-pit mine change detection. We have proposed in this project to establish binary CD as a proxy metric for surface activity in open-pit mine areas. In this section, the model selection criteria are reviewed. The selected model architectures were trained and evaluated, and that is discussed in further sections.

3.1.1 Model Selection Criteria from DLCD Literature

The following primary questions were considered when selecting models to train and evaluate for open-pit mine CD:

- Does the model appear to accomplish state-of-the-art (or comparable) performance on established CD benchmarks?
- Would training (or at least fine-tuning) the model be feasible on freely available hardware resources (such as the Google Colab environment)?
- Would training the data on a limited labelled dataset (less than 5k training samples) be sufficient to achieve a desirable level of accuracy?
- Is there public source code available to assist in the implementation of the model in our experiments?

Via the above questions/criteria, the following models were found to be promising candidates for identifying and monitoring changes in open-pit mine areas.

3.1.2 TinyCD

TinyCD was selected on the basis of its state-of-the-art performance, its efficient training size, and the availability of its source code in a public repository (Codegoni, Lombardi, and Ferrari 2022). The original implementation is hosted on GitHub (AndreaCodegoni 2023). The TinyCD architecture features a Siamese U-Net structure with a novel Mix and Attention Mask Block (MAMB) and a pixel-level output classifier (Figure 3.1). As the backbone, TinyCD uses EfficientNet pre-trained on the ImageNet dataset (Codegoni, Lombardi, and Ferrari 2022, pp. 4–5).

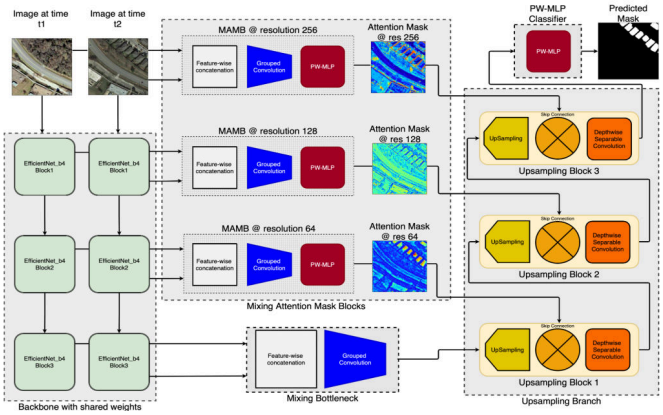


Figure 3.1: TinyCD network architecture. From (Codegoni, Lombardi, and Ferrari 2022, p. 4). Reproduced under CC BY-NC-SA 4.0.

3.1.3 LSNet

Although it is about 2× larger than TinyCD (in terms of the number of parameters), LSNet is a lightweight DLCD architecture which boasts much higher computational efficiency than many other state-of-the-art models (Liu, Huaixin Chen, and Z. Wang 2022). Additionally, its initial implementation has also been published in a public code repository. The LSNet architecture features a light Siamese backbone with a novel differential feature pyramid network (diffFPN) (Figure 3.2).

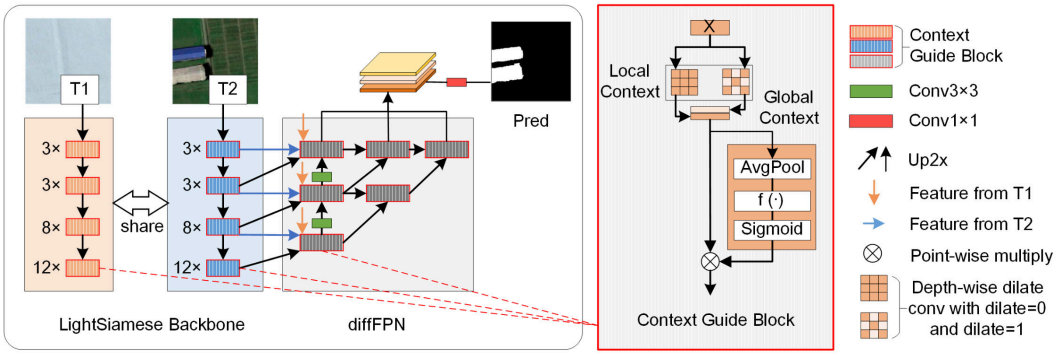


Figure 3.2: LSNet network architecture. From (Liu, Huaixin Chen, and Z. Wang 2022, p. 2). Reproduced under CC BY-NC-SA 4.0.

3.1.4 DDPM-CD

DDPM-CD is quite different from the previous two models, as it relies on a large diffusion model backbone which was pre-trained on Sentinel-2 imagery. Its implementation and trained weights are available publicly online (C. Bandara 2023). The published paper (W. G. C. Bandara, Nair, and Patel 2022) does not evaluate the model size or computational efficiency, so its hardware requirements for training were initially underestimated. The DDPM-CD architecture makes novel use of the diffusion model backbone, which are trained in an unsupervised manner. Following that, a lightweight Siamese difference classifier is used to produce change predictions (Figure 3.3).

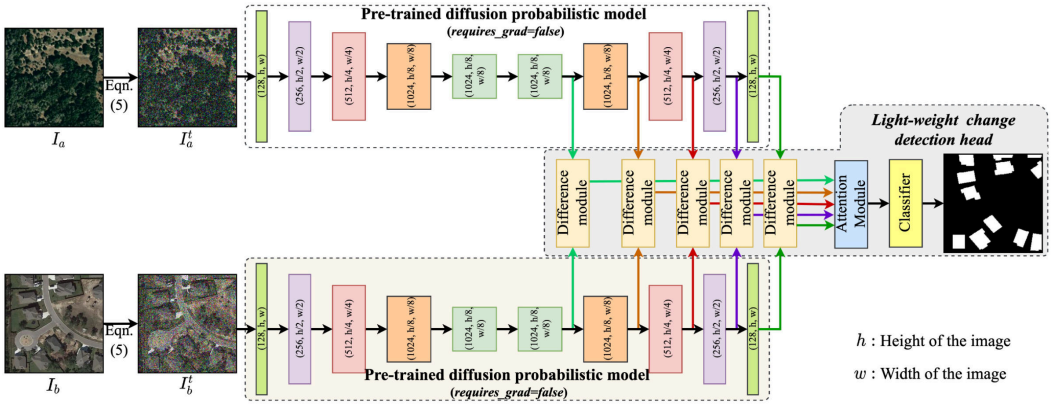


Figure 3.3: DDPM-CD network architecture. From (W. G. C. Bandara, Nair, and Patel 2022, p. 6). Reproduced under CC BY 4.0.

3.2 Data

The selection of a robust training and validation set is potentially more important than the model selected. As discussed in the literature review, there are a collection of benchmark datasets for change detection that are freely available online such as LEVIR-CD (Hao Chen and Z. Shi 2020), OSCD (Daudt et al. 2018), and OMCD (Li et al. 2023). With regard to selecting which online dataset would be the best fit for training our models for open-pit mine change detection, it is worth discussing which data was determined to be the most useful for completing our goal of open-pit mine activity detection. We did not only want to train a model against a benchmark; we also wanted to utilise the trained model to monitor long-term changes in case study areas. In terms of turning the trained models into a tool that is usable for monitoring long-term changes, it was determined that the Sentinel-2 satellite data provided the best opportunity. The reasons are as follows:

1. Sentinel-2 data is freely available from the ESA through online platforms such as Google Earth Engine, which is easily integrated into Google Colab.

2. Compared to other freely available satellite image sources such as Landsat, Sentinel-2 has a higher spatial resolution (10 m vs 30 m).

3. The Sentinel-2 online archives contain data for at least as far back as 2018 for most areas on Earth, and will continue to be added to for the foreseeable future (the launch of Sentinel-2C is planned for 2024 (*Gearing up for third Sentinel-2 satellite*

2023)).

While OSCD was derived from Sentinel-2 multi-spectral imagery, the other benchmark datasets available are mostly from miscellaneous VHR satellite or aerial imagery sources. During experimentation, training the models on the OSCD and OMCD datasets was initially attempted. It was expected that such training would improve change detection ability on Sentinel-2 imagery over open-pit mine areas. However, the domain transfer from the OSCD and OMCD source to the Sentinel-2 target areas appeared to be very poor (see Appendix G for further discussion). We believe that there are likely to be two primary reasons:

1. While OSCD is sourced from Sentinel-2 imagery, it does not feature any open-pit mine areas.
2. OMCD features open-pit mine areas, but it uses VHR imagery sourced from the Gaofen satellites and therefore has different optical characteristics and feature scale.

Additionally, the definition of what counts as 'change' is different in each dataset and benchmark. The reason for the difference in definition is primarily because each dataset is established with a specific goal, namely: what changes should matter to our model? In the context of OSCD, only 'urban' changes are labelled as part of the change class. Such a goal is somewhat at odds with our own priority of identifying human activity within open-pit mining areas, which may not necessarily appear as urban change. As mentioned in (Ebel, Saha, and X. X. Zhu 2021, p. 5) who added Sentinel-1 data on top of the OSCD dataset, there is a principal question about what should count as a change in the ground truth.

3.2.1 Open-pit Mine Sentinel-2 Change Detection Dataset

For the aforementioned reasons, we found that it was necessary to establish a specific definition of relevant change for open-pit mine change and activity detection. That definition was then used to build a custom Open-pit Mine Sentinel-2 Change Detection (OMS2CD) dataset. The definition of relevant change (or activity) includes all surface areas that appear to include the following activities:

- Excavation
- Waste tailings and overburden dumping
- Soil and rock build-up
- Re-filling of site areas for the purposes of reclamation
- Construction or deconstruction of structures or buildings in or near a known open-pit mine
- Visible movement of vehicles in or around a known open-pit mine
- Signs of deforestation
- Construction, deconstruction, or modification of roadways in or near a known open-pit mine
- Establishment of new waterways or the destruction of existing waterways
- Establishment or destruction of tailings ponds

The definition of change for the context of this project was established with consideration for the purposes of activity monitoring for open-pit mines. As reviewed in our introduction, surface mining activities have significant environmental impacts. Such impacts are evident in significant local changes to the surface environment. As found by (D. Zhu et al. 2020), even the surface temperature near surface mining sites can be significantly affected by continued development. As such, any signs of significant human activity in and around surface mines is considered by us to be worth monitoring.

It is worth mentioning that while the above items met the definition of change for open-pit mines established for this context, it was still difficult to identify relevant changes in all the Sentinel-2 images. Several factors, including the angle of the sun on the surface, the presence of clouds, and seasonal changes (such as snow cover), make it difficult to distinguish human activity from natural causes. The implication is that in the annotated OMS2CD dataset that was used for training and validation of the selected models, there are certain areas that were of low confidence to the human annotator. Even so, that was deemed to be an acceptable risk. In the best case, if a model could do as well at

annotation as the human (even making the same mistakes), the performance would still be beyond acceptable.

3.3 Building the OMS2CD Dataset

The OMS2CD dataset which was built and manually annotated was constructed through numerous steps of site selection and image processing to prepare it for use in training our selected network architectures. Below, the process is described in detail.

3.3.1 Data Source Selection

To build the OMS2CD dataset, the first step was mining site selection. The open mining database provided by (Jasansky et al. 2023) proved to be an invaluable resource. Using the QGIS software (*QGIS* 2023), we filtered only known 'active' mining sites, and exported the data as a GeoJSON file. The GeoJSON data was then processed to create a bounding box over the known geolocation area of the mining site. If a geolocation was not present in the database, we used the GADM (*Global Administrative Areas* 2022) entry to find the nearest administrative division (primarily in China), if possible. After manual visual inspection of each site using geojson.io (Mapbox 2023), a total of 16 unique sites across seven distinct geographical areas were selected.

3.3.2 Data Filtering and Acquisition

After the mining site bounding boxes were established, the Google Earth Engine (GEE) through Google Colab was used to download and preprocess the Sentinel-2 Level-2A surface reflectance imagery data for the areas selected. The date range for download for each area was from 2016-01-01 to 2020-12-31, with a filter for images with 20% clouds or less. For each month of data in each mining site area, the least cloudy day was selected for download. Sentinel-2 Level-2A surface reflectance is already atmospherically corrected and images for a given region are automatically registered together. While the Level-1C data from GEE is available from earlier time periods (2015-06-23 as compared to 2017-03-28 for Level-2A) it was simpler to download the Level-2A data without developing an extensive image processing pipeline.

3.3.3 Processing and Annotation Criteria

From the full 13-band images downloaded using Google Earth Engine, only the 3 RGB bands (B4, B3, and B2 for Sentinel-2 data) were exported and scaled down from the standard Sentinel-2 range of 0-10000, to 0-255. Finally, manual visual inspection was used to remove images that were incomplete or had too much cloud cover. This left a total of 255 RGB GeoTIFF images over the aforementioned mining sites. These were then filtered further and organised into 238 bi-temporal image pairs prepared for mask annotation.

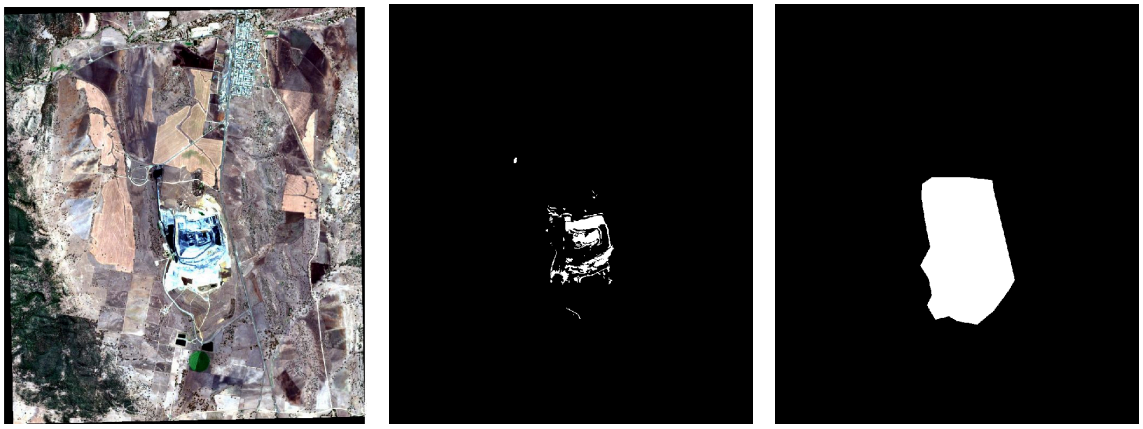


Figure 3.4: From left to right: an RGB Sentinel-2 image tile, the corresponding change mask label, and the area mask polygon.

To assist in annotation and model training, mining site *area mask* polygons were created and exported using QGIS. These were intended to separate approximate mine areas from the surrounding non-mine areas in each of the 16 sites. See Figure 3.4 for an example of one of the dataset images and its corresponding masks. Note that the RGB image tile is a composite raster image[1] created by combining the Sentinel-2 B2-B4 bands as mentioned earlier.

Following the established 'change' criteria described in Section 3.2.1, 70 of the previously selected 238 bi-temporal image pairs were manually annotated using GIMP imaging editing software (The GIMP Development Team 2023). There was unfortunately not sufficient time during the project to annotate all 238 bi-temporal image pairs. However, 70 image

[1] A raster is a matrix of data commonly used in EO. Each element of the matrix (which may represent multiple data bands) corresponds to a specific spatial position, and the data for that position. The data represented could be surface reflectivity, elevation, radar coherence, etc.

pairs is a collection still larger than the aforementioned OSCD benchmark dataset which, between train and test, only contains 50 annotated image pairs.

After manual annotation was completed, files without a completed change mask were removed from the dataset. The remainder were then split into separate train, validation, and test locations. Using different locations for each split of the dataset was intended to reduce spatio-temporal dependencies arising that would have incorrectly given the models a high validation and/or test performance.

Additionally, using an image patch size of 256$\times$256 pixels (the input size used in most DLCD architectures) with a stride of 128 pixels, the locations for each split were selected to approximate a split of 70%, 20%, and 10% for training, validation, and testing respectively. Effort was also made to guarantee that the validation and testing sets would represent as unique geographical areas as possible. As such, it was considered that the models were generalising well only if both validation and test performance appeared to be increasing. In practice, the trained models performed much better on the validation split, but this will be discussed more in our analysis.

3.3.4 Data Preparation and Augmentation

Due to the limited size of the proposed OMS2CD dataset, strategies needed to be employed to increase dataset diversity, enhance model generalisation, and reduce training set over-fitting. Following other works in this area, (Codegoni, Lombardi, and Ferrari 2022; W. G. C. Bandara and Patel 2022; Li et al. 2023) data augmentation methods were employed. The data augmentation included both random geometric (flip, rotation, scale) and non-geometric (random Gaussian noise) operations:

1. Random horizontal and vertical flip.
2. Random Gaussian noise: $\mathcal{N}(0, 0.1)$.
3. Random affine transforms (rotation up to 30 degrees, translation up to 10% on each axis, scaling between 80% and 120%).

Horizontal and vertical flipping and rotation encourage the model to generalise in a way that is not dependent on the specific direction of the visual features. Similarly, scaling

encourages learning scale-invariant features. Gaussian noise is added to increase visual diversity and enhance the model's robustness against speckle noise that may be present in the input images.

In addition to the above augmentations, distribution standardisation was applied to the images to bring the input images into a target distribution of $\mathcal{N}(0, 1)$. The mean and standard deviation of the dataset for standardisation were calculated using only the training data, separately for all image channels. All of the above augmentations and standardisation were performed using the Kornia library (Riba et al. 2019).

As a final method of increasing the total size of the training dataset, input images of size 256$\times$256 were pulled from the larger source images using a sliding window with a stride 100 pixels. This means that each training image (or patch) overlapped with a selection of other nearby patches, but it increased the model's exposure to each area. Additionally, an advantage of providing overlapping patches means that change areas near the edge of one patch (not showing the entire change area) may be presented in the centre of another patch. Combined with the aforementioned augmentation transforms, such a change area may be presented to the model differently than in nearby patches.

3.3.5 Data Representation

Split	Size	Percentage (%)
Train	2130	72.6
Val	463	15.8
Test	342	11.7
Total	**2935**	**100.00**

Table 3.1: Distribution of dataset splits.

The final, labelled, and pre-processed OMS2CD dataset contains 102 Sentinel-2 image tiles of 15 different surface mining sites. The image tiles selected or multi-seasonal and cover different relative sunlight angles. In addition, the dataset has 17 area masks (as defined earlier) and 70 hand-annotated mask labels which cover the 70 bi-temporal image pairs. Using a patch size of 256$\times$256, a stride of 100, and the aforementioned splits, Table 3.1 gives the dataset sizes which we used for training.

All image files in the dataset are presented as .tif files, which are in the GeoTIFF image

format. The binary area and change label masks they are single-channel (or band) with a value range of $[0 - 255]$ (0 for no change, 255 for change). All the other GeoTIFF files are RGB (three-channel) with a value range of $[0 - 255]$. The naming scheme of the mining site images is as follows:

```
s2_<facility_name>_<coordinates>_<YYYY-MM-DD>.tif
```

To organise the dataset into training, validating, and testing splits, there are three .csv files which are the locations for each split. A 'mapping.csv' file is also contained within, which describes which images make up each bi-temporal image pair. We have provided in the project source code a Python class 'OMS2CD' which was used to work with the dataset and prepare it for use with DLCD models. More details are available in Appendix H and Appendix I.

3.4 Model Training

For training the selected models that were described in Section 3.1.1, the primary limitation was hardware resource availability. Hardware limitations, in turn, limited the amount of time with which could be focused on performance tuning with hyperparameter adjustments. The training setup was a bit different for each model, generally following the default setup used by the model's authors as made available in their source repositories. In general (for all models trained), training was approached with the following guidelines:

- Batch size was maximised based on the amount of GPU memory available for the model during training.
- An 'early stopping' algorithm based on validation loss was used to prevent overfitting and to make best use of our available time limitations.
- 16-bit precision was used, where possible, to make training more efficient.
- The original author's model source code was modified as little as possible to achieve training using the PyTorch Lightning library (Falcon and The PyTorch Lightning team 2019).
- Training data were shuffled (randomised in order).

- The same augmentation steps and probabilities were used.
- The pseudorandom systems were seeded with the same values for each training run of each model.
- All training was limited to 50 epochs (though due to early stopping, it rarely reached that point).

3.4.1 Transfer Learning

In the domain of surface mine identification and monitoring, (Chunsheng Wang et al. 2020) serves as a demonstrable example in their approach using transfer learning. Defining transfer learning as "transfer and reuse of knowledge" between a 'source' and 'target' domain, they demonstrate how a model that was pre-trained on the COCO image dataset (Lin et al. 2015) can be adapted to surface mine detection. Such a method of domain transfer provides a potential type of model training that can reduce the amount of training data which is required for convergence. The term fine-tuning then refers to the process of training just a portion of a model on the new domain. The model weights which are not being trained are said to be 'frozen' because they are not updated through gradient backpropagation. Because less of the model weights are being updated, less hardware resources and time are required for training, and the model's performance relies more on the pre-training that it has experienced.

To take advantage of the potential benefits of transfer learning, each model was fine-tuned from the weights provided in the original source. The details of this fine-tuning are discussed below in Section 3.4.3 and in Section 4.1.1.

3.4.2 Focusing on Areas of Interest (AOI)

Open-pit mine CD is unique from change detection in other contexts because, as was established in Section 3.2.1, there is a broad category of change that we desire to identify. Unlike other change detection tasks which may be only interested in changes to buildings or roads, the relevant changes present in open-pit mine areas that reflect *activity* can be substantially more broad. Additionally, there are spatio-temporal changes present in the OMS2CD dataset tiles that do not appear to be related to mining activity. In those

cases, would it be desirable for the model to identify them as change? Would such an identification be considered to be correct or incorrect?

In review of the relevant literature, we did not find such a discussion about the identification of relevant change. Even so, it was experimentally found that these questions pose challenges to model performance in real-world tasks, and this is discussed this further in the analysis in Section 4.4. To evaluate the impact of training and validating the selected models against potentially irrelevant changes that can appear very visually similar to relevant changes, *area masks* were added to the OMS2CD dataset to mark the mine AOI. Each model was then trained using the full Sentinel-2 image tiles or just the tiles that overlap with the AOI mask polygon, and the performance was compared.

3.4.3 Training and Validation Experimental Setup

As briefly mentioned in Section 3.4, there was insufficient time available during the project to tune the hyperparameters of the models. As such, the hyperparameters values that were used can be given here as explicit, hardcoded values. Unless otherwise stated, the values selected were those left as the recommended or default values by the model's original authors. For LSNet and DDPM-CD, these values were derived from the training configuration files created by their authors. In the case of TinyCD, the value was derived from the training source code.

3.4.3.1 TinyCD

For training TinyCD, AdamW optimisation (Loshchilov and Hutter 2019) was used with a learning rate of 0.003, weight decay of 0.009, and not using the AMSGrad variant. Unlike the source paper, no cosine annealing algorithm was implemented on the learning rate, as it was experimentally found that convergence was more stable without it. The target loss function was binary cross-entropy loss, and the batch size used was 16. The total trainable parameters in the TinyCD model numbered 285,000.

For fine-tuning the model, numerous approaches to optimise the models' validation and testing performance on our dataset were attempted. Such approaches included:

- Training the full model from randomised weights.

- Loading the pre-trained backbone model and freezing its weights (just training the TinyCD head).
- Loading the full pre-trained model from the TinyCD repository and training it without freezing its weights.
- Loading the full pre-trained model from the TinyCD repository and only training the final linear classifier layers.

Additionally, the models were trained with and without the data augmentation described in Section 3.3.4. However, training without data augmentation was discovered early-on to continuously provided poorer validation results. Therefore, all of our recorded experiments used augmentation.

3.4.3.2 LSNet

The original LSNet work provided a few different model options, which offer a trade-off between model size and computational complexity. For that reason, the DiffFPN version of the model was selected for training. Within the context of training and evaluation in this project, 'LSNet' can be considered synonymous with LSNet-DiffFPN.

LSNet was trained using the AdamW optimiser, with an initial learning rate of 0.001, weight decay of 0.01, and no AMSGrad. Mirroring the LSNet paper, the StepLR learning rate scheduling algorithm with a step-size of 8, and a gamma of 0.5 was used. The target loss used by the authors of LSNet was not reviewed in their paper. However, in their source code the default loss function was a hybrid loss which was the sum of the binary cross-entropy and dice losses. Therefore, that is what we used. The batch size used was 6. The total trainable parameters in the LSNet model numbered 1.2 million.

Similarly to our TinyCD training process, LSNet was also trained using different levels of fine-tuning. The results from each type of fine-tuning are discussed in our analysis.

3.4.3.3 DDPM-CD

The DDPM-CD architecture consists of two components: the DDPM diffusion backbone (through which the pre- and post-images are run) (Saharia et al. 2021), and a Siamese

difference network which decodes the diffusion features into a final change map output. The authors of DDPM-CD trained the diffusion model backbone on over 1 million Sentinel-2 images, and provided the trained weights publicly. Because of limited hardware resource availability (and our focus on freely accessible resources), in all training instances the parameters of the diffusion model backbone were frozen to focus only on training the difference network.

For DDPM-CD, the Adam optimiser (Kingma and Ba 2017) was used with an initial learning rate of 0.0001 and a linear learning rate decay with a step-size of 3 and a gamma of 0.1. The loss function used was cross-entropy loss. The batch size used was 3. The total parameters in the DDPM-CD model numbered 435 million, but the highest number that we regarded as trainable was 44 million (achieved when treating the diffusion backbone as frozen).

Due to the time requirements of training the DDPM-CD difference head, we were not able to extensively experiment with fine-tuning DDPM-CD.

3.4.4 Comparison Metrics

In DLCD model architecture literature such as LSNet, TinyCD, and DDPM-CD, the most commonly used metrics for performance comparison are F1, IoU, accuracy (also known as overall accuracy (OA)), precision, and recall. While during training and validation the performance was primarily reported as the model's corresponding loss function value, model-to-model comparisons were performed using the aforementioned metrics. Here is a quick definition of each of them, as given by (Liu, Huaixin Chen, and Z. Wang 2022):

$$
\begin{aligned}
Pr &:= \frac{TP}{TP + FP}, \\
Rc &:= \frac{TP}{TP + FN}, \\
F1 &:= \frac{2}{Pr^{-1} + Rc^{-1}}, \\
IoU &:= \frac{TP}{FN + FP + TP}, \\
OA &:= \frac{TP + TN}{FN + FP + TP + TN},
\end{aligned}
$$

Note that the convention used is TP = True Positive (a positive ground-truth value identified correctly), FP = False Positive (a negative ground-truth value identified incorrectly), TN = True Negative (a negative ground-truth value identified correctly), and FN = False Negative (a positive ground-truth value identified incorrectly). For our experimental analysis, each metric has been calculated at the pixel level by comparing the ground-truth and predicted matrices element-wise. Below, we formally define pixel-wise calculation of the above metric components:

Given a ground-truth matrix G and the prediction matrix P of dimensions $m \times n$, the values of TP, FP, TN, and FN can be calculated as:

$$
\begin{aligned}
TP &= \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = 1 \wedge P_{ij} = 1] \\
FP &= \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = 0 \wedge P_{ij} = 1] \\
TN &= \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = 0 \wedge P_{ij} = 0] \\
FN &= \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = 1 \wedge P_{ij} = 0]
\end{aligned}
$$

Where G_{ij} and P_{ij} represent the elements at the i^{th} row and j^{th} column of the ground-

truth and prediction matrices, respectively. $[\cdot]$ is the Iverson bracket that returns 1 if the condition inside the parentheses is true and 0 otherwise.

Considering an example where element $(0, 0)$ of both matrices were 1, then the pixel represented by element $(0, 0)$ would be considered a TP prediction. In a visual image representation, we could then define TP and TN as follows:

$$TP = \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = \text{white} \wedge P_{ij} = \text{white}]$$
$$TN = \sum_{i=1}^{m} \sum_{j=1}^{n} [G_{ij} = \text{black} \wedge P_{ij} = \text{black}]$$

where matching white pixel predictions are considered TP, and matching black pixel predictions are considered TN. For a visual example, see Figure 3.5 where areas that are white in both would be TP, and black in both would be TN.

Figure 3.5: A pair of labelled change prediction mask images. Left: ground-truth label; right: change prediction by LSNet.

3.4.5 Training Hardware Resources

All training of the LSNet and TinyCD models was performed using the Google Colaboratory free-tier T4 GPU instances with 16 GB VRAM. Such instances are free to use by anyone with a G-Suite account (*Google Colaboratory* 2023). We initially attempted using

the same hardware to train DDPM-CD, but the training time was too slow, and the VM instance would time out before a sufficient number of training epochs had been completed. As such, we utilised the 'Pro' tier plan from Paperspace Gradient, (*Paperspace - Gradient Notebooks* 2023) ($8 USD per month) which gave us access to a Jupyter Notebooks VM instance equipped with an NVIDIA Quadro RTX5000 (or RTX A4000, when available) with 16 GB VRAM. The maximum timeout limit available for the Gradient VM instances is 6 hours, which is much more lenient than those available on Google Colab (often only 1–2 hours, in our experience). While such instances on Gradient are not free, they are still very affordable for projects with a limited budget.

3.5 Normalised Difference Temporal Change Index (NDTCI)

As summarised in Section 1.2, one of the primary objectives for this project was to use DLCD models to predict changes in bi-temporal image pairs of open-pit mine areas. However, it was also found that a succinct way of measuring and expressing the specific levels of activity was needed. As discussed in the literature review, only two previous works that were focused on mapping the spatio-temporal changes in open-pit mines as a measure of mine site activity were identified. The proposed measure, Normalised Difference Activity Index (NDAI) was proposed by (Moon and Lee 2021) and paired with Digital Elevation Models (DEM's) to map changes. L. Wang et al. developed the idea further without requiring DEM data (L. Wang et al. 2021). Following on their work, we wanted to take advantage of a similar measure of activity represented as a normalised index measure. However, the NDAI (see Equation 3.1) was intended to be used with InSAR data and requires InSAR coherence measurements (ρ) as inputs.

$$\text{NDAI} = \frac{\rho^{\text{stable}}_{\text{activity}} - \rho^{\text{target}}_{\text{activity}}}{\rho^{\text{stable}}_{\text{activity}} + \rho^{\text{target}}_{\text{activity}}} \tag{3.1}$$

Although we did not use InSAR coherence images, the trained DLCD models do produce binary change prediction images, which may be treated as a proxy for activity measures. As such, we propose to define our index of activity using change detection as the Normalised Difference Temporal Change Index (NDTCI) in Equation 3.2. δ^{target} is the pixel-

level change predictions for the target pixel. δ^{stable} is the mean of the predicted changes for 'stable' regions in the given change image. ε is simply a small regularising constant value that prevents divide-by-zero. In practice, it was found that δ^{stable} could be calculated as the mean value of the lower quartile of the change image distribution. The lower quartile of change predictions should be near zero, which is comparable to how L. Wang et al. selected ρ^{stable} to have a coherence > 0.9.

$$\text{NDTCI} = \frac{\delta^{\text{target}} - \delta^{\text{stable}}}{\delta^{\text{target}} + \delta^{\text{stable}} + \varepsilon} \tag{3.2}$$

The NDTCI has a theoretical range of $[-1, 1]$, with values closer to one signifying greater activity and values closer to -1 signifying lesser activity. Generally, since $\delta^{\text{stable}} \approx 0$, the typical range is $[0, 1]$. The NDTCI is intended to be calculated for an entire image matrix. Such an NDTCI matrix N could be defined as $N = [\text{NDTCI}ij] \in [-1, 1]^{H \times W}$, where NDTCI$ij$ represents the calculated NDTCI value at the i-th row and j-th column of the matrix.

When the NDTCI is calculated for a set of change prediction images, then the mean of the NDTCI values yields an output which shows which areas of the mining site are more or less active. Effectively, it is an activity map. A formal definition and examples of such activity maps have been provided in Section 4.2.3.

3.6 Positive-class Prediction Thresholds

In the analysis (Chapter 4), performance comparisons and other analytical comparisons are made between the models. When performing comparisons using singular metrics like F1, precision, and recall, a choice needed to be made for the positive-class prediction threshold which would maximise the metric scores across all the models. The positive-class prediction threshold is simply the threshold where each pixel in the predicted change mask is determined to be either positive (contains change) or negative (no change). The value compared against the threshold is the predicted probability/confidence of that pixel. In the reviewed DLCD literature, the threshold is generally set at 0.5, but the optimal value may be anywhere within the range of $[0, 1]$ depending on the model.

To determine the optimal value which would yield the highest F1-score, for each model, we calculated and visualised the Precision-recall curve on the OMS2CD dataset. The results of those calculations are reviewed in Section 4.1. We note that while the threshold value which would optimise F1 was selected, there may be valid reasons to select a different threshold to change the balance between precision and recall.

Chapter 4

Analysis

4.1 Model Performance Comparison on Open-pit Mine Change Detection

As described in detail in Section 3.4, each of the selected model architectures were trained on the OMS2CD dataset described in section 3.2.1. In this section, the performance of each model is reviewed and there is a discussion on the specific training conditions and their observed effects on model performance. The performance metrics identified and defined in 3.4.4 were used for direct comparison of the models. The F1-score was the single metric primarily used because it is a balance between precision and recall. A metric such as accuracy is not informative in situations where there is an extreme class imbalance.

In Table 4.1 we provide a comparison of performance metrics for our models. For the comparison results, the best-performing weights (using F1-score) of each model were selected. A positive-class prediction threshold of 0.3 was used, and was selected by using the precision-recall curves in Figure 4.1, 4.2, and 4.3. It can be observed that a threshold value of 0.3 generally yields a good balance between precision and recall. It is noteworthy that while TinyCD is the smallest model, it achieved the best F1-score by a significant margin on both splits. Additionally, it is easy to see that the dataset performance for each split was vastly different. For example, LSNet had an F1-score

	LSNet		TinyCD		DDPM-CD	
Metric	Val	Test	Val	Test	Val	Test
OA	**0.991512**	0.966767	0.990888	**0.969701**	0.986821	0.954244
F1	0.638726	0.254062	**0.671778**	**0.336884**	0.606581	0.258069
Recall	0.551435	0.315948	0.685350	0.429665	**0.746673**	**0.444258**
Precision	**0.758850**	0.212449	0.658733	**0.277057**	0.510753	0.181853
AP	**0.733086**	0.176493	0.716710	**0.264977**	0.655944	0.202095
IoU	**0.682484**	0.114234	0.453052	**0.143711**	0.418034	0.133165

Table 4.1: Model performance comparison for OMS2CD train and validation splits. Best performance in each metric for validation is highlighted in orange and for testing in blue.

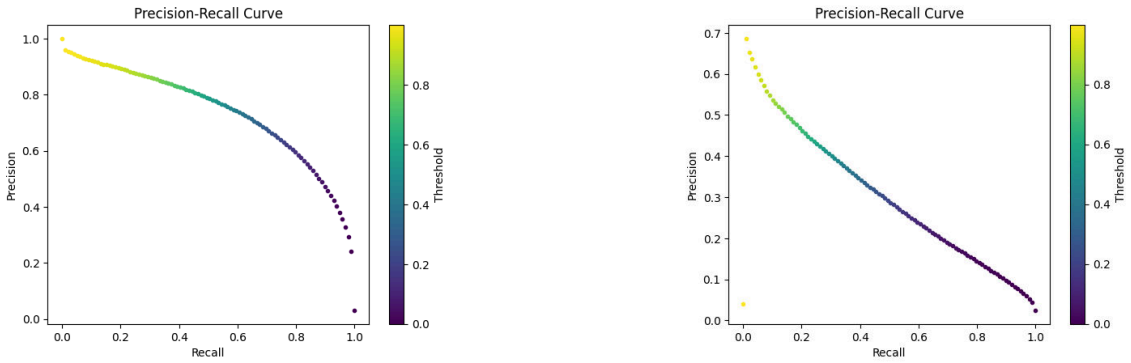


Figure 4.1: Precision-recall curve from TinyCD on the validation (left) and test (right) sets of the OMS2CD dataset.

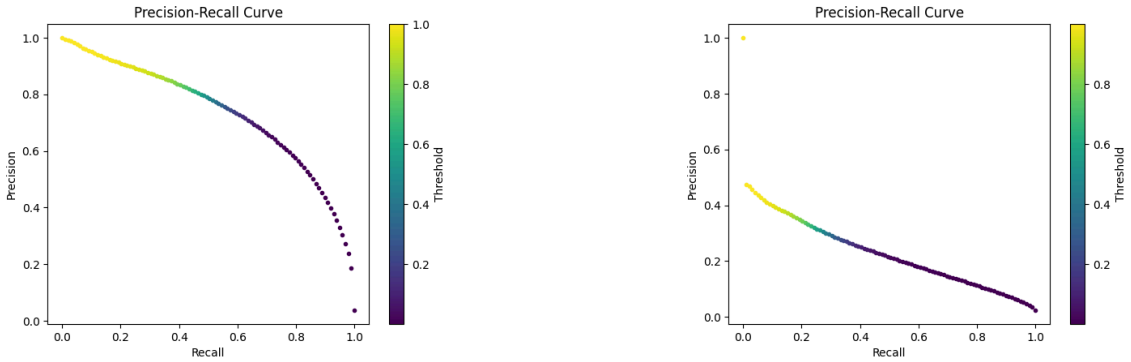


Figure 4.2: Precision-recall curve from LSNet on the validation (left) and test (right) sets of the OMS2CD dataset.

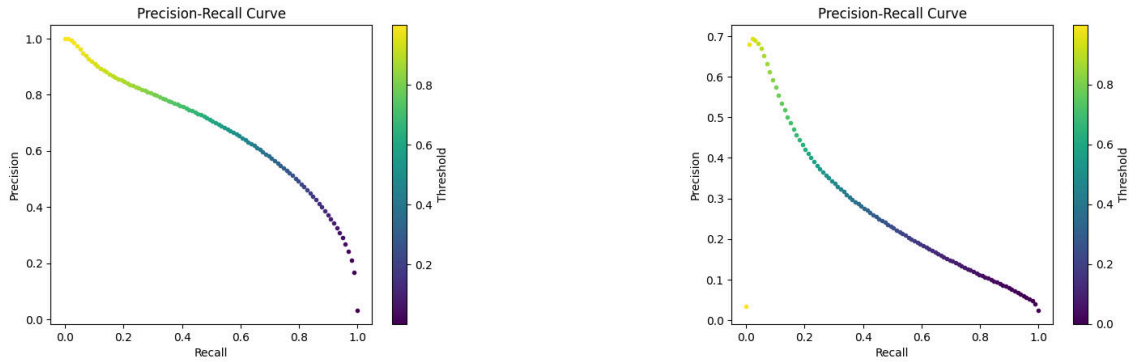


Figure 4.3: Precision-recall curve from DDPM-CD on the validation (left) and test (right) sets of the OMS2CD dataset.

of 0.64 on validation, but only 0.25 on test; but such a difference is reflected in all three models.

Why was there such a significant difference between splits? There may be three primary causes:

- Since the dataset (and therefore the splits) are small, each split has a significantly different data distribution.
- Visual inspection of the test split location reveals that it is demonstrably more difficult to identify mining areas and their changes.
- The test data split is considerably smaller than the validation split.

Even so, both evaluation (validation and test) splits are spatially separate from the training areas and from each other. Therefore, both evaluation splits should be used to make a full assessment of model performance. With that said, although TinyCD appears to be significantly better at properly identifying change, it is possible that TinyCD is simply better in the specific distribution represented by the evaluation sets.

In review, it was surprising that the performance of DDPM-CD on the OMS2CD dataset was comparable to that achieved by the other two models. Because the computational requirements of DDPM-CD are considerably larger than the other two models, it does not appear to be an efficient model for the proposed task. However, it is possible that with a larger available training dataset, DDPM-CD may be able to generalise better and outperform the others.

4.1.1 Impact of Fine-tuning

As we defined in Section 3.4.1, fine-tuning is defined as when some part of a pre-trained model is 'frozen' so that only a small subset of the total model weights are updated via gradient backpropagation. Each of the selected models were fine-tuned from the pre-trained weights provided by the original authors. TinyCD and DDPM-CD were pre-trained on the LEVIR-CD dataset from (Hao Chen and Z. Shi 2020), while LSNet was pre-trained on CDD from (C. Zhang et al. 2020).

4.1.1.1 TinyCD

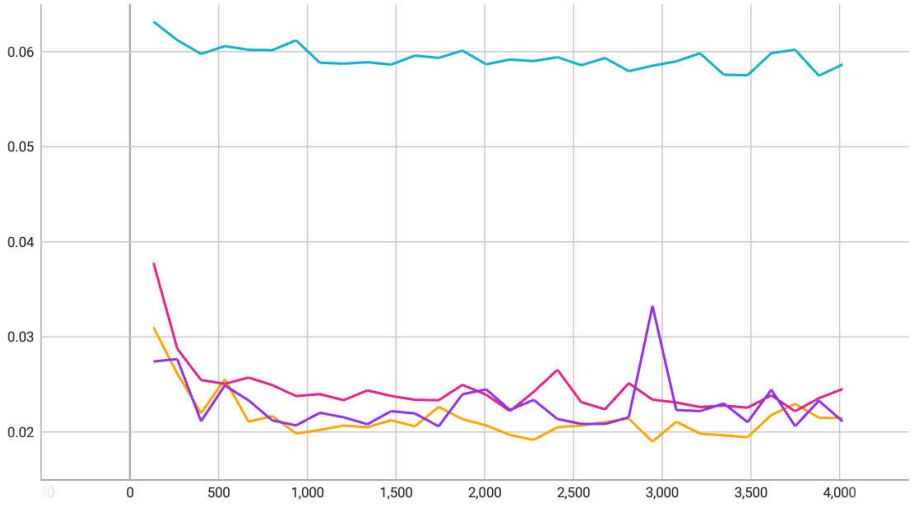


Figure 4.4: Loss vs training iterations for training TinyCD under different fine-tuning scenarios.

On TinyCD, it was discovered that the best F1-score was achieved when the model was fully unfrozen (we were updating all the model's weights). In Figure 4.4 we provide a training comparison using different modes of fine-tuning. The blue line is from fine-tuning only the 'classify' layers of the model (675 parameters), pink is from fine-tuning with the 'backbone' frozen (15.8k parameters), yellow is from training the full network with randomly initialised weights (285k parameters), and purple is from training the full network when initialised with pre-trained weights. Finetuning only 15.8k parameters nearly achieved the same loss as training the full network, which demonstrates how transfer learning can achieve more efficient training performance.

4.1.1.2 LSNet

The LSNet model architecture achieved the highest F1-score using the same strategy as TinyCD; training the full network (no layers frozen), but starting from the pre-trained weights provided by the authors.

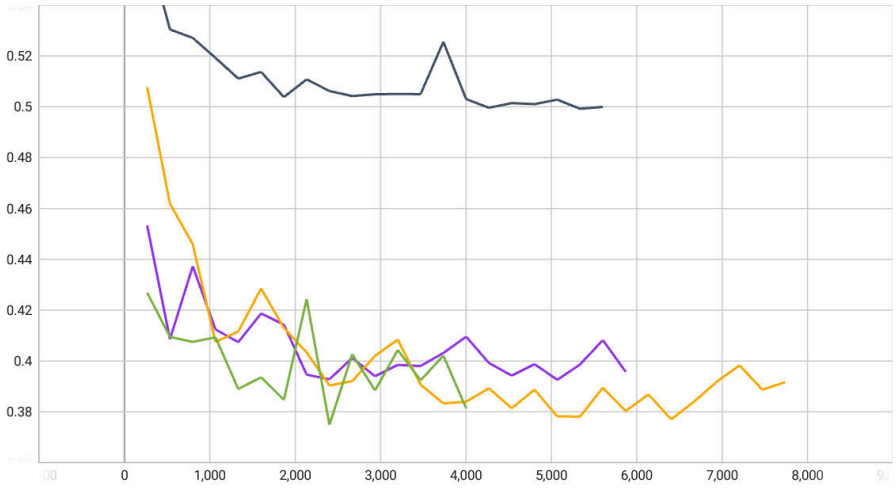


Figure 4.5: Loss vs training iterations for training LSNet. The best validation loss was achieved when training the full network from random weight initialisation (yellow) and when training the full network from pre-trained weights (green).

In Figure 4.5 one can see that the validation performance of LSNet is greatly affected by the number of frozen parameters, and whether the network is initialised with pre-trained weights. The gray line is from fine-tuning only the 'head' of the network unfrozen (3.6k parameters), purple is from fine-tuning with only the 'backbone' frozen (274k parameters), yellow is from training the full network with randomly initialised weights (1.2M parameters), and green is from training the full network initialised with pre-trained weights. It is noteworthy that a similar level of performance is achieved regardless of whether the backbone is frozen, as was seen with TinyCD. Additionally, training the full network from the pre-trained weights seems to achieve better performance in fewer steps than training from randomly initialised weights.

4.1.1.3 DDPM-CD

Due to the training time required for DDPM-CD, there were insufficient resources to perform a full fine-tuning comparison as was completed for LSNet and TinyCD. However, it

was the last model that was trained. To adapt what was learned from the previous models, the Siamese diffusion model training was started with pre-trained weights. By using the pre-trained weights, the goal was to take advantage of the faster convergence that was seen when using the same strategy on LSNet and TinyCD.

4.2 Spatial Change Detection as a Form of Temporal Change Analysis

Following model training and evaluation, the model weights which appeared to optimise the F1-score across the validation and test sets were selected. See Table 4.1 for a cross-metric comparison of the selected models' performances.

The following sections detail the experiments which answer the final research question proposed in the introduction: "Can DLCD be used to monitor and detect the activity level of open-pit mines over time?" In summary: yes. We have experimented with three methods of utilising our trained models to map temporal changes in surface mining areas, with a ground truth comparison. The first method yields a chart of detected surface changes for each bi-temporal image pair and acts as a manner of measuring the amount of 'activity' that occurred in that time delta in terms of the actual surface area change. The second method visually maps the spatial changes over the mining site, with visible areas of change for further investigation. The third method takes inspiration from (L. Wang et al. 2021) and maps a normalized difference index NDTCI as we described in Section 3.5.

4.2.1 Method 1: Plotting Change Area Predictions as an Activity Metric

By taking each bi-temporal image pair and making a change prediction with the models, a timeline of change/activity was created (Figure 4.6). Additionally, the human-labelled ground-truth masks were processed to make a ground-truth comparison timeline to assess the accuracy of the proposed method. The y-axis metric is in square area (m^2), because each pixel of the Sentinel-2 tiles in the selected bands is 10m$\times$10m, or 100 m^2. The calculated area of change is simply the summed area of the pixels which were predicted to contain change. While creating a plot like that in Figure 4.6 does not pro-

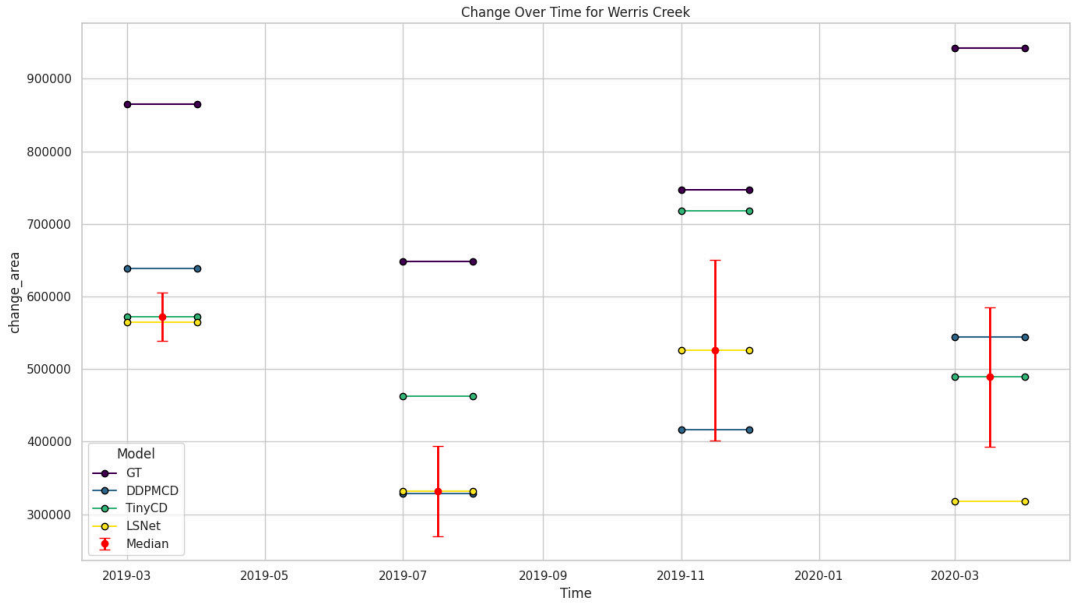


Figure 4.6: Left: A plot of surface area changed (in m^2) over time. Each bar represents the amount of change that was predicted to have occurred over that interval of time. The plot covers the Werris Creek area from our OMS2CD dataset. The red median points are the median change value of the three models for each time interval, while the area bars are the standard deviation of the predictions at that interval.

vide insight into the specific regions of change in the target area, it does create an easy visual estimate of the level of activity over time. See the appendix for plots covering the predicted change in the other evaluation areas.

The change predictions for all areas of the evaluation data were calculated using the three trained models and ground-truth, and those predictions were used to calculate the Pearson correlation coefficient between each pair. In Figure 4.7 the heatmap of correlation between each model pair is shown. It is significant that the highest single-model correlation with the ground-truth is LSNet (with a correlation of 0.73) but that the median value of the model predictions has a correlation of 0.84. It seems to demonstrate that together the three models are able to provide a more informed prediction than when they are separate.

Such a statistically significant correlation result between our model predictions and the ground-truth is a positive sign that a plausible form of activity detection has been proposed.

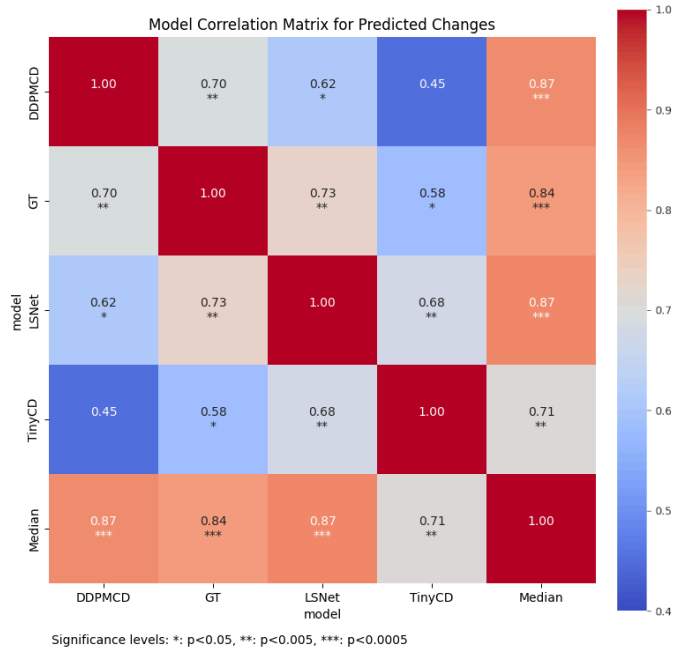


Figure 4.7: A heatmap of the Pearson correlation between the model change area predictions.

4.2.2 Method 2: Visually Mapping Changes over Time

By taking the predicted change images from a model and stacking them (therefore making the later images appear on top of the earlier ones) mine site changes as they occurred over time can be visualised (Figure 4.8). Such a visualisation demonstrates how the proposed methodology provides a valuable tool in activity detection. A map of specific spatio-temporal changes provides easy targets for follow-up investigation in the monitoring of mining sites.

4.2.3 Method 3: NDTCI as a Normalised Activity Measure for Spatio-temporal Changes

The NDTCI as proposed in Section 3.5 is intended to act as a single, normalised measure of activity using bi-temporal, optical image pairs. While it can be used to measure the spatio-temporal changes between two images at a location, it can also be used to create a visual map of regions of high and low activity (Figure 4.9). We have derived this method of taking the average over the NDTCI predictions (which we call μ_{NDTCI}) from the method

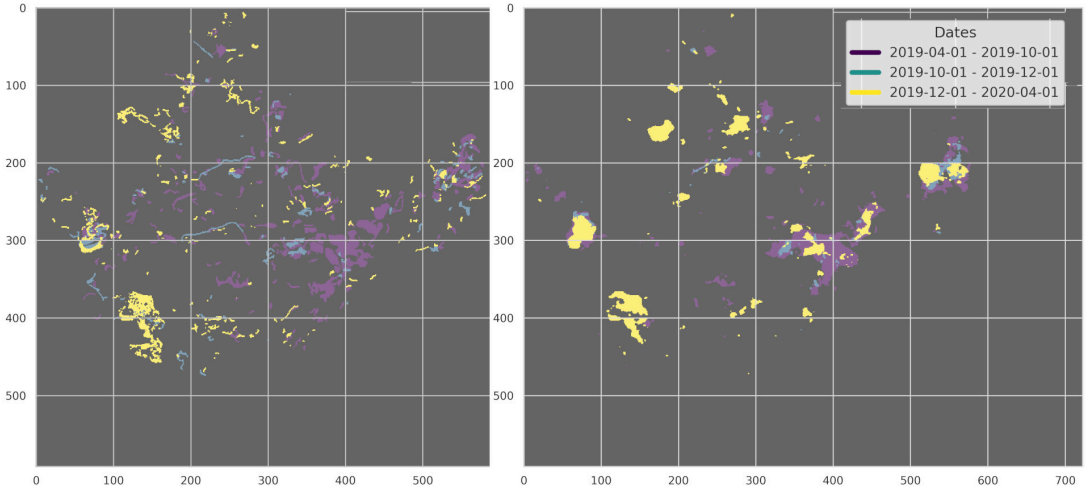


Figure 4.8: A map of changes in the Guizhou mine site area from OMS2CD. Left: ground-truth change labels; right: changes as predicted by DDPM-CD. Each colour band represents a bi-temporal pair, with the dates shown in the legend. Horizontal and vertical axes are measured in pixels.

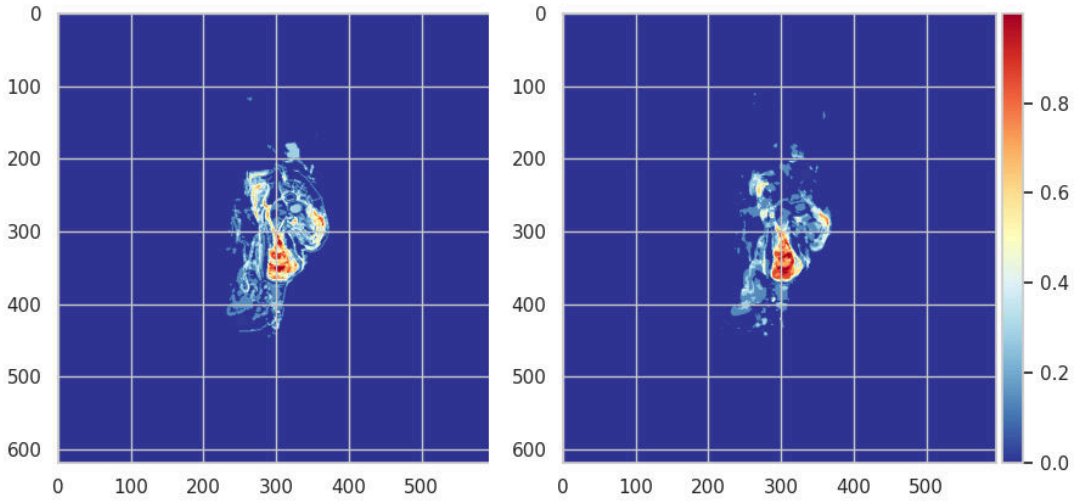


Figure 4.9: A map of μ_{NDTCI} in the Gunnedah Vickery mine site area from OMS2CD. Left: ground-truth change labels; right: changes as predicted by TinyCD. Horizontal and vertical axes are measured in pixels. The colour represents the NDTCI value.

mentioned with NDAI in (L. Wang et al. 2021, p. 9), but we provide a formal definition in Equation 4.1. Therefore, each pixel in this new activity matrix μ_{NDTCI} represents the average level of activity in that location over the provided set of T sequential NDTCI matrices from time $t = t_{\text{start}}$ until time $t = t_{\text{end}}$. Ideally, the set of T NDTCI matrices should be uniformly spaced, but in practice that could be difficult to achieve due to missing observations. A weighted form of the metric which is balanced by the length of each interval could be useful, but we do not define one here.

$$\mu_{\text{NDTCI}} = \frac{1}{T} \sum_{t=t_{\text{start}}}^{t_{\text{end}}} \text{NDTCI}_t \tag{4.1}$$

4.3 Effects of Cloud Cover and Seasonal Changes

Figure 4.10: Prediction example. From left to right: Sentinel-2 'pre' image tile; Sentinel-2 'post' image tile; LSNet model change predictions.

As detailed in Section 2.5, multispectral optical imagery has a number of disadvantages when compared to SAR. Such disadvantages include being inoperable at night, being occluded by cloud cover, and sensitivity to the angle of the sun on the terrain. In our experimentation, we have found that the greatest natural obstacle to accurate change detection is cloud cover. To build the OMS2CD dataset, images with the least cloud cover were selected to be part of the dataset. While such an approach is sensible (and typical), it may have helped to make the DLCD models more robust to cloud occlusion if more cloud cover areas had been included in the training data. While the models appear to be fairly robust to small amounts of cloud occlusion (Figure 4.10), that they are still somewhat susceptible to counting cloud movements as relevant change if cloudy images

are not filtered from the input data (Figure 4.11). But such effects are not consistent between models, nor across spatial areas.

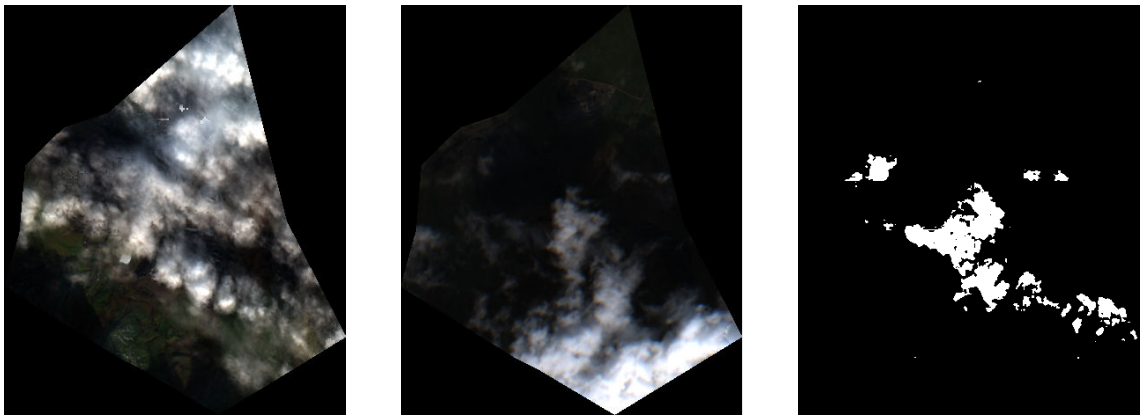


Figure 4.11: Cloud-occluded prediction example. From left to right: Sentinel-2 'pre' image tile; Sentinel-2 'post' image tile; LSNet model change predictions.

Besides cloud occlusion, one of the other major challenges for CD is seasonal changes to the target terrain. One of the most distinct seasonal changes in temporal regions is snow cover, but they may also include agricultural development (planting, harvesting, etc.), tree canopy changes (in deciduous forest areas), and changes to water sources (water level rising or falling), among others. Each of the listed categories of seasonal change pose a unique challenge, as they are represented as entirely unique visual artefacts. Agricultural development, for example, may be easy to confuse with the clearing of vegetation which occurs before excavation. In the Sentinel-2 images at 10 metre resolution, both types of human activity appear quite similar.

In the OMS2CD training dataset, there are a number of images which include snow cover. Following the established change criteria, areas with snow cover were not labelled as 'changed' unless there appeared to be clear human activity (such as the clearing of snow for roadways). As such, the trained models do not seem sensitive to mistakenly labelling snow cover or melt as changed (Figure 4.12). With that said, we have found that all three models tend to make false-positive predictions on agricultural developments which appear similar to mining surface changes. Such a mistake seems sensible, because at 10 m resolution it can be very difficult to tell the difference even for a human annotator. Figure 4.13 is an example of difficult seasonal changes. Due to the likelihood of false-positive

Figure 4.12: Snow-cover seasonal prediction. From left to right: Sentinel-2 ‘pre’ image tile; Sentinel-2 ‘post’ image tile; TinyCD model change predictions.

Figure 4.13: Seasonal prediction with agricultural changes. From left to right: Sentinel-2 ‘pre’ image tile; Sentinel-2 ‘post’ image tile; DDPM-CD model change predictions. In the given mask, white represents TP predictions, yellow represents FP predictions, and blue represents FN predictions.

predictions in areas outside the target mine area, we demonstrate how separating the AOI using a polygon mask can be helpful in obtaining a more accurate change prediction.

4.4 Impact of Predicting within a Defined AOI

As reviewed in the methodology in Section 3.4.2, very little discussion was identified in the relevant literature regarding purposefully bounding the relevant change areas (AOI). One reason for that may be that, among CD tasks, open-pit mines represent a unique challenge. Unlike in urban development, surface mining may blend in with the surrounding natural terrain or agricultural activity. In regions where the mine is near to human activities such as farming and deforestation, true relevant change can be difficult to identify. In such cases, we recommend restricting the prediction area to be as close to the target mine as

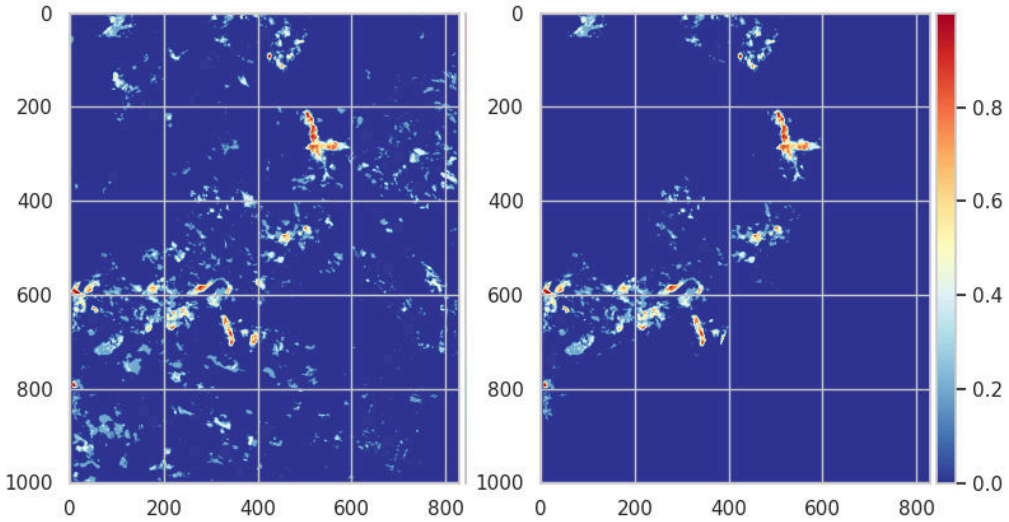


Figure 4.14: A comparison of μ_{NDTCI} maps for the same area based on model predictions. Left is the μ_{NDTCI} plot without utilising an area/AOI mask, while right is with the mask.

possible. To achieve that, AOI 'area mask' polygons as defined in Section 3.3.3 were created.

Let us define the predicted matrix as $P = [p_{ij}] \in [0, 1]^{H \times W}$, where p_{ij} represents the prediction value at the i-th row and j-th column of the matrix. Then the AOI mask as described can be defined as $M = [m_{ij}] \in 0, 1^{H \times W}$, where m_{ij} represents whether the pixel at the i-th row and j-th column is inside the area of interest. The final change predictions within the area of interest can be computed as the element-wise multiplication of the two matrices.

In our experiments, after a model made its predictions over the entire Sentinel-2 tile, the change prediction image was simply multiplied element-wise by the area mask (Equation 4.2) to get rid of predictions outside the AOI. The resulting matrix $C = [c_{ij}]$ represents the prediction after applying the AOI mask. A value of 1 in C means that both the prediction and the mask were 1 (i.e., change is predicted within the AOI), and a value of 0 means either no change was predicted or the location is outside the AOI.

$$C = P \circ M \tag{4.2}$$

While creating an AOI mask for each target region requires some manual work, it is a

rather simple process that greatly reduces the model false-positives (Figure 4.14). In a real-world production pipeline, an AOI area mask could be defined for each target region as a simple way of improving change prediction accuracy by reducing false-positives. On the evaluation data, using the area masks reduced the median NDTCI timeline prediction RMSE by over 40% when using a positive-class prediction threshold of 0.4 (Figure 4.15).

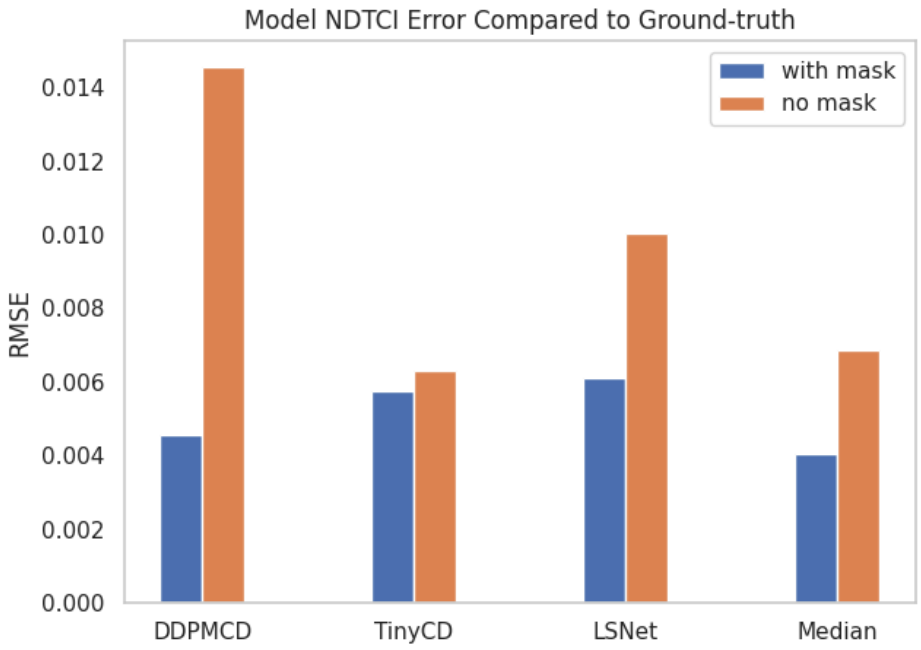


Figure 4.15: A comparison of NDTCI timeline prediction values by model with and without an area mask. Prediction threshold 0.4.

4.5 Selecting a Prediction Threshold for Inference

In Section 4.1 the PR curve from each model was used to determine an optimal prediction threshold for the metric comparison. For the model inference experiments in Section 4.2 through Section 4.4 a threshold of 0.4 was used. While it was found that any threshold value in the range of $(0.3, 0.7)$ tended to give acceptable results, it is worth discussing how the decision of the prediction threshold measurably affects the model inference performance. For example, compare Figure 4.16 to the previous Figure 4.15. The 'with mask' RMSE was significantly higher for all models when the threshold was set to 0.6.

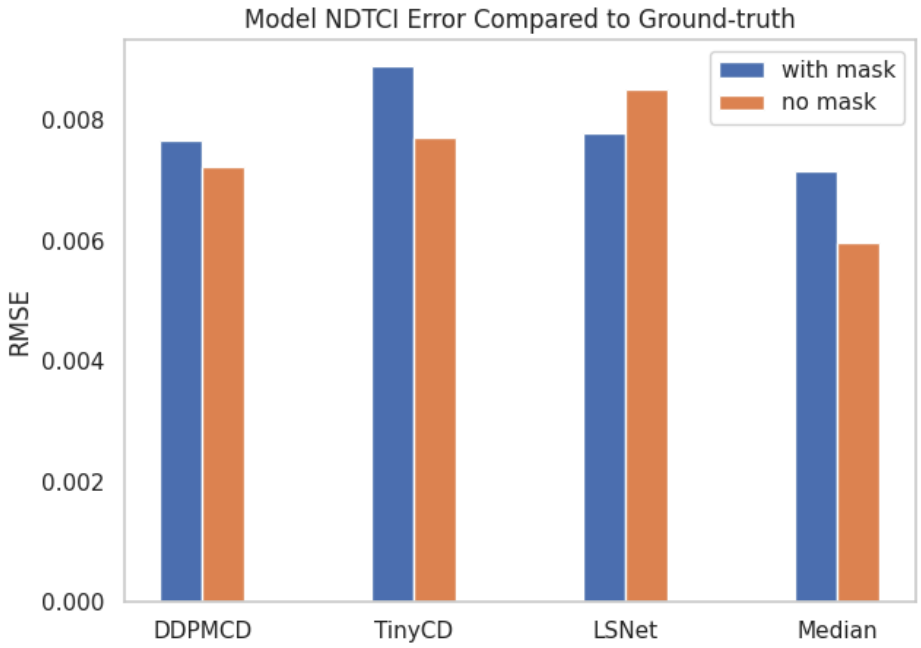


Figure 4.16: A comparison of NDTCI timeline prediction values by model with and without an area mask. Prediction threshold 0.6.

Interestingly, the 'with mask' and 'no mask' RMSE were not very different when using a threshold of 0.6, unlike when the threshold of 0.4 was used. It may be that when the threshold was set to 0.6, then there were many fewer false-positive predictions (as the required confidence level is higher). Since the mask was being used to reduce false-positive predictions, then the effect was less noticeable. We note that the median 'no mask' RMSE was about the same for both thresholds, demonstrating again that the median prediction value was more robust than the individual models. Figure 4.17 represents another example, where the correlation with ground-truth was considerably different depending on the chosen threshold. In our case, the correlation with the ground-truth was higher with the threshold of 0.4, largely due to the models' tendency to underestimate change.

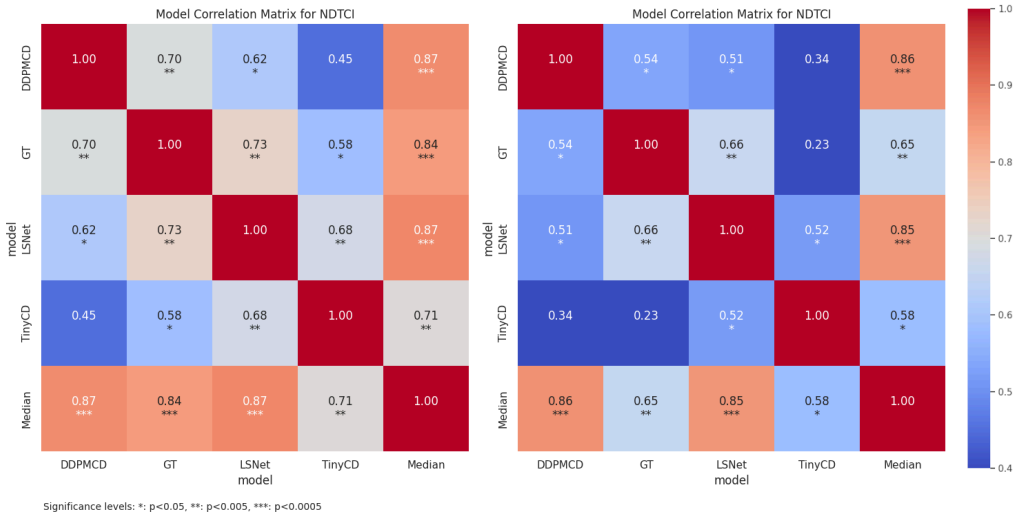


Figure 4.17: A comparison of NDTCI timeline prediction correlation heatmaps by model using a prediction threshold of 0.4 (left) and 0.6 (right).

Facility Name	Country	Production Start Year
Meliadine	Canada	2019
Carrapateena	Australia	2019
Ubuntu	South Africa	2020
MetCoal	Indonesia	2020
Merthyr Tydfil	United Kingdom	2022
Iluka Western Australia	Australia	2019

Table 4.2: Case study site information.

4.6 Change Detection Case Study Areas

Following the experimental analysis of the trained model performance on the evaluation data from the OMS2CD dataset, the models were applied to additional unseen mining sites as case studies for the proposed activity detection pipeline. The case study areas were selected using a methodology similar to that described in Section 3.3.1, but the selection criteria were different. We selected sites that had a production start (key 'production_start' in the Jasansky et al. database) in or after 2019, with the goal of demonstrating clear changes in activity as the mining sites grew. Using the specified filter criteria, six sites were found and their information was exported as GeoJSON data. Geojson.io was again used for manual inspection, and six sites were selected: five from the Jasansky et al. database, and one of additional interest (Table 4.2).

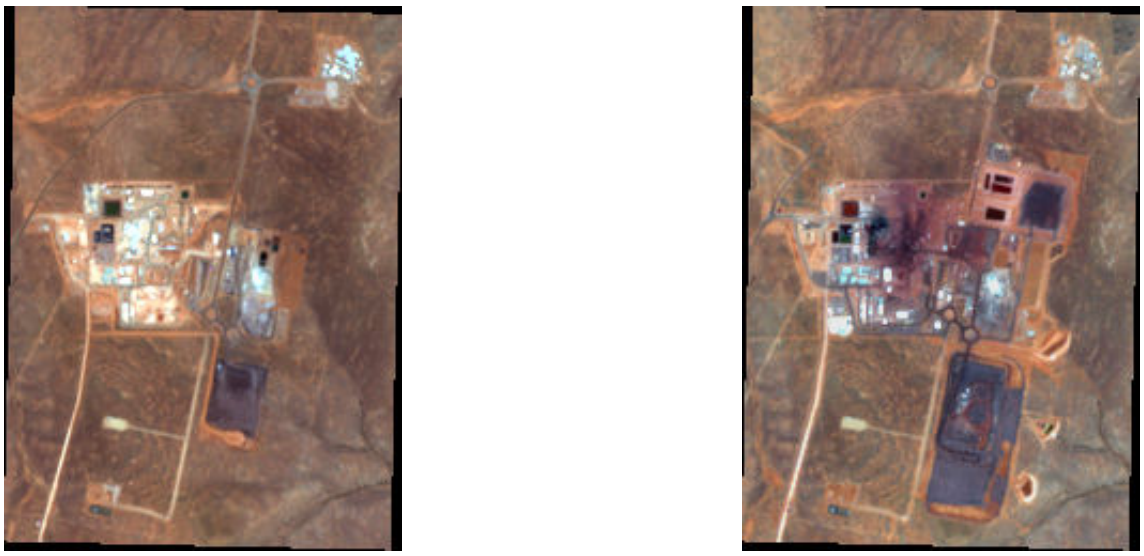


Figure 4.18: Sentinel-2 image tiles of the Carrapateena mining site. The width and height of each tile is 186×255 pixels, or 1.86 km×2.55 km with 10 m resolution. Left: the site at the beginning of the target study period. Right: the site at the end of the target study period. The active surface area of the mine appears to have grown significantly during that time.

After site selection, GEE was again used to download the Sentinel-2 Level-2A tiles for each region. The date range given was ±2 years from the start of production, and we applied a cloud filter of 50%. For every two-month period, the least cloudy image for each area was selected. Only the RGB bands were saved for each tile. After the data was downloaded, images that were incomplete or completely occluded by clouds were removed.

4.7 Case Study Analyses

For each case study location, we performed the same analyses that were used on the evaluation data in Section 4.2. Without a ground-truth comparison, manual assessment was required to determine the activity prediction accuracy. With that said, the achieved activity prediction performance was very good, especially in areas where there was little cloud occlusion and very few seasonal changes. In the appendices, we have provided an extensive set of figures which demonstrate the results of our case study observations for a selection of site areas.

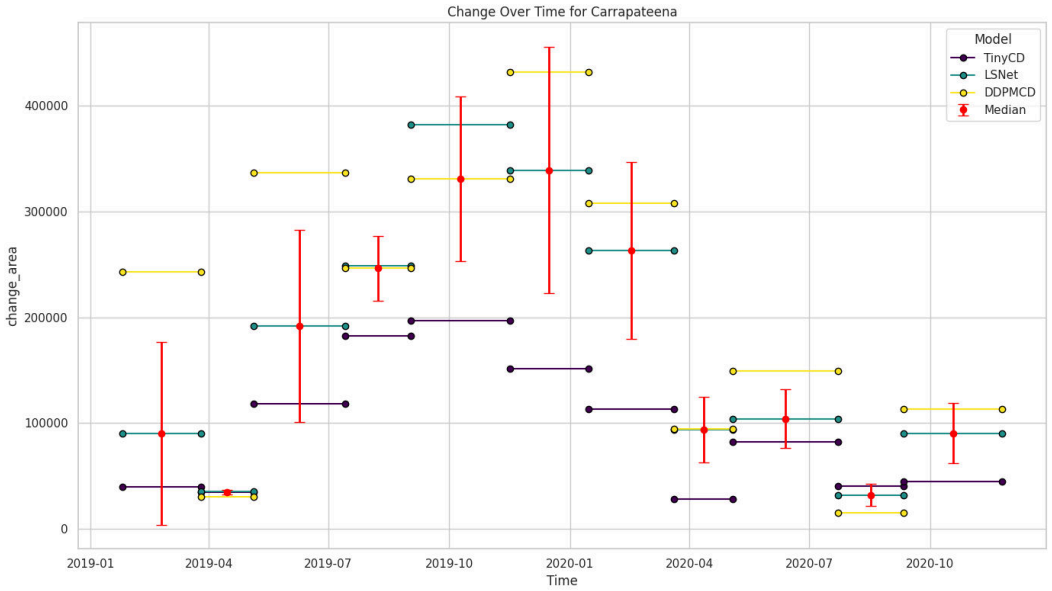


Figure 4.19: Timeline plot demonstrating the amount of surface area predicted to have changed for each time interval (change is in m^2).

For this section, the Carrapateena facility location is reviewed, which started production in 2019. Using the GEE pipeline, a contiguous series of Sentinel-2 tiles covering the dates 2019-01-25 to 2020-11-25 on a bi-monthly basis with no cloud occlusion was obtained. The facility has been listed as an underground gold and copper mine, not open-pit[1]. Even so, the visible surface changes were extensive enough during that time period to still provide a meaningful assessment of mine activity. Figure 4.18 demonstrates the difference in appearance between the start and end of the study period.

Using a prediction threshold of 0.4, the timeline of area change as assessed by the trained DLCD models was plotted (Figure 4.19). The plot demonstrates how the surface activity (measured as surface area change) continued to increase until the beginning of 2020, and then decreased again as expansion slowed. The maximum median surface change occurred between 2019-11-16 and 2020-01-15. Figure 4.20 (left) demonstrates a visualisation of the μ_{NDTCI}, where the highest values are mapped to the most active areas of the facility. Figure 4.20 (right) is a spatio-temporal visualisation which facilitates an

[1]While this facility is not an open-pit mine, we have chosen to review it as a case study using our pipeline because it demonstrates the robustness of our activity detection methodology on many types of mine facilities. As previously stated, see the appendices for further examples of analysis run on open-pit mine facilities

understanding into when specific surface changes occurred in time. Using such a visualisation makes it easy to understand how the lower section of the site area was expanded outward over time.

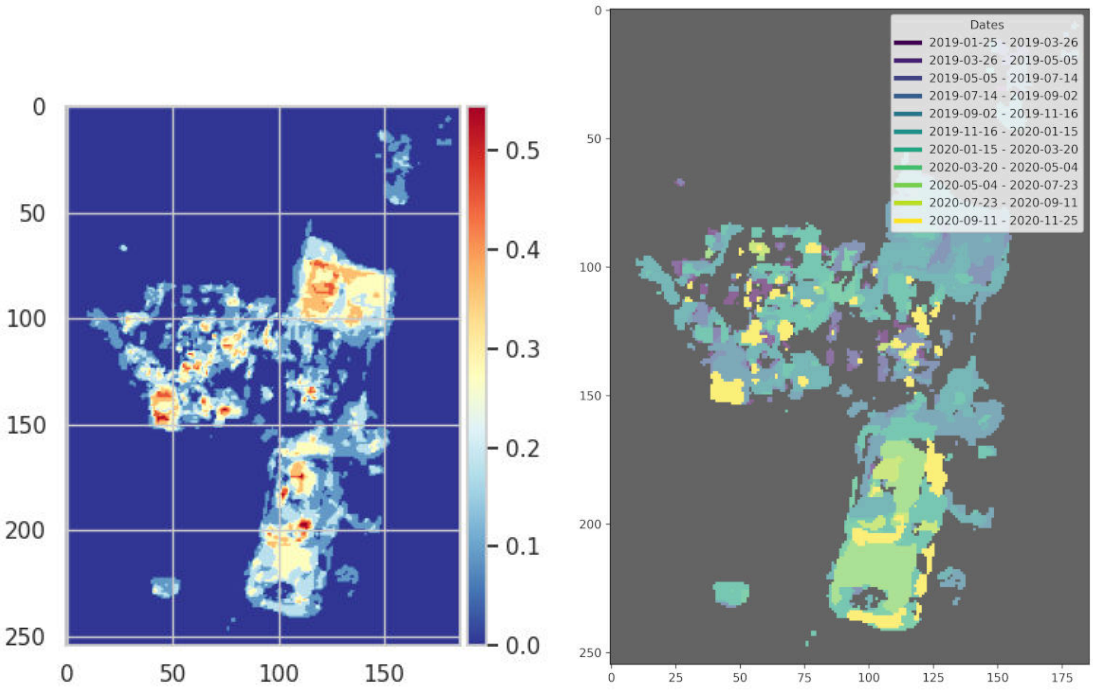


Figure 4.20: Left: Visualisation of the μ_{NDTCI} calculated using the change masks predicted by LSNet over the study period. Right: Visualisation demonstrating the spatio-temporal changes at the facility. Vertical and horizontal axes are measured in pixels.

As a demonstration, the 10 m resolution μ_{NDTCI} image has been stacked on top of Google Earth imagery from the site (Figure 4.21). Even though it is difficult to track exact surface changes using the 10 m resolution Sentinel-2 tiles, the activity estimation provided by the proposed DLCD methodology is sufficient to direct focus to specific areas which can later be better-understood through high-resolution image sources. While such an approach would require manual follow-up, the DLCD models still provide valuable insight into the surface areas which are more active.

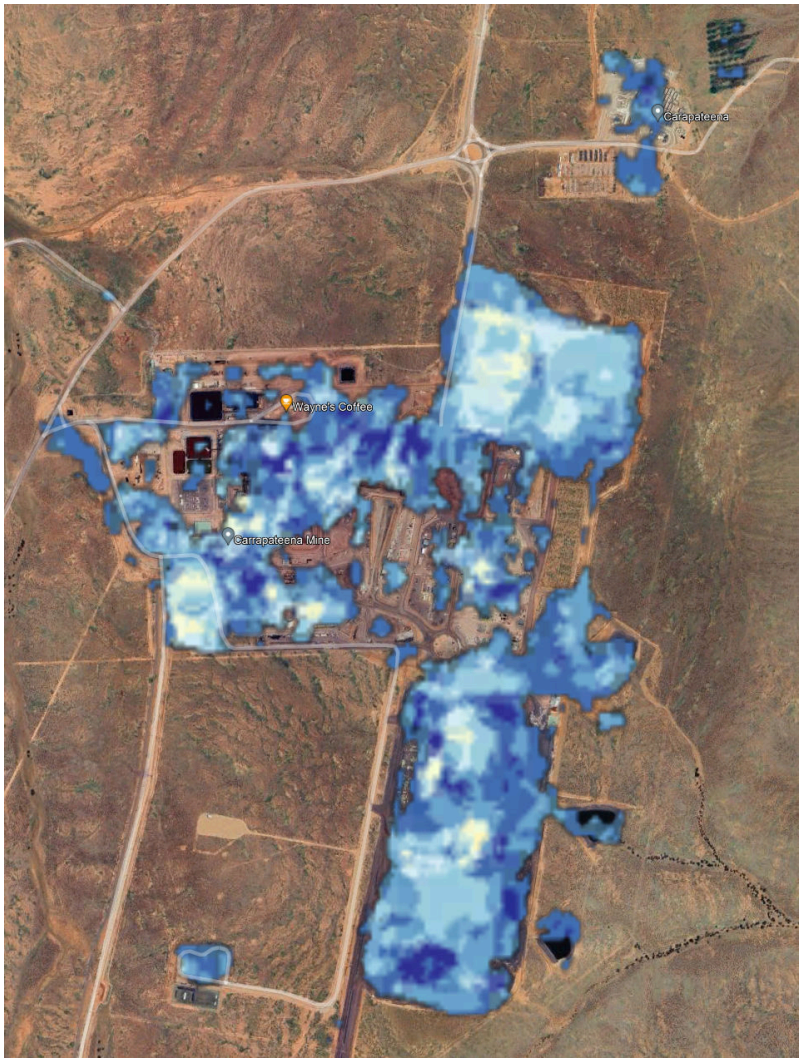


Figure 4.21: Demonstration of overlaying the 10 m resolution μ_{NDTCI} on top of a much higher resolution Google Earth image for observation.

4.8 Observations on the Scale of Changes

Throughout the experimentation performed in this section, one phenomenon which has consistently been observed is the trained models' inability to identify small changes (those which are five pixels or smaller in any one dimension). While some are identified, it appears that generally the models find it difficult to identify such small changes. For examples of such behaviour, see Appendix D and Appendix E. In Figure D.2 there are significantly more small-scale changes in the ground-truth than in any of the model μ_{NDTCI} predictions.

Why do the models miss small changes? It may be that the model architectures which we have chosen and trained on the OMS2CD dataset simply were not intended to detect changes which are less than five pixels across. The convolutional layers in the models may have larger receptive fields, causing them to overlook minute details. Additionally, the training data might not have insufficient examples of small changes, leading the models to generalise them as noise or irrelevant information. The loss functions used during training could also play a role; if they do not penalise missing small changes heavily, the models might not prioritise detecting them.

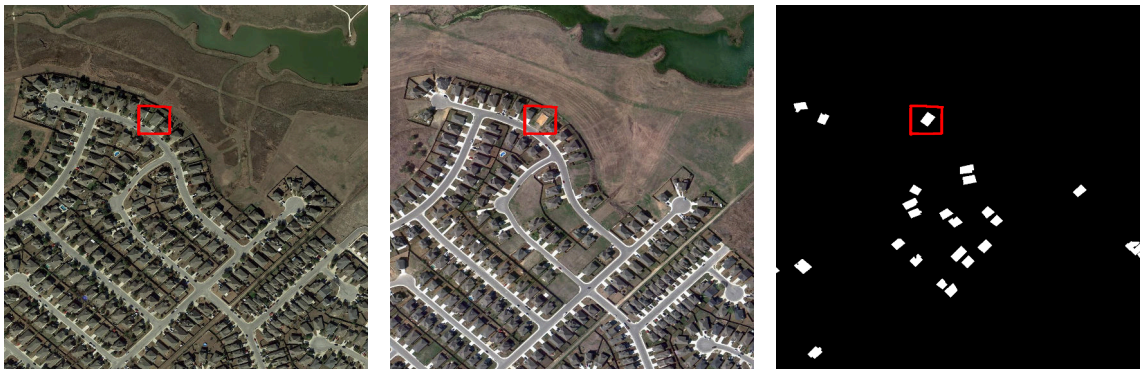


Figure 4.22: A bi-temporal image pair example from the LEVIR-CD dataset. Left to right: earlier image, later image, change mask label. The red box highlights just one of the house changes, which we measured to cover an area of approximately 25×18 pixels.

Something to be considered is that the high-resolution CD datasets normally used as benchmarks feature changes of a much larger scale. For example, in the LEVIR-CD dataset, it is common for a single house to be represented by an area of approximately 20×20 pixels (Figure 4.22). In OMS2CD, that would be an area of 200 m×200 m. As

such, even significant surface changes may present themselves in Sentinel-2 imagery as differences of only a few pixels.

Overall, further investigation is needed to pinpoint the exact reasons for the missed changes, and future work could focus on refining the selection of model architectures or training strategies to address this limitation.

Chapter 5

Conclusions and Recommendations

5.1 Discussion

We have explored the emerging application of deep learning for change detection in open-pit mining areas as a way to monitor and quantify mining activity over time. As highlighted in the literature review, remote sensing and deep learning have been applied in several related domains that provide useful context:

- Identifying and mapping mining areas using satellite data and deep learning models like U-Nets (Balaniuk, Isupova, and Reece 2020; T. Chen et al. 2022). These works focus on delineating mining sites, rather than tracking longitudinal changes within mines. However, their mapping techniques could be combined with change detection to locate active mining zones.
- Assessing environmental impacts through land cover change analysis (D. Zhu et al. 2020; Nascimento et al. 2020). While environmental assessment is not a direct focus of this dissertation, tracking mining activity over time could serve as an important input for quantifying habitat destruction, vegetation loss, pollution levels, and other mining impacts. Change detection provides key activity data to support impact evaluation.

- General deep learning approaches for remote sensing change detection (Bai et al. 2022; Jiang et al. 2022). Our work draws directly upon recent advances in domain-agnostic change detection models like TinyCD and evaluates their effectiveness on the novel application of monitoring mining operations.
- Using radar coherence data and indices like NDAI to measure activity levels in mines (Moon and Lee 2021; L. Wang et al. 2021). We adapted this concept of a normalised activity index for change detection predictions, proposing the NDTCI as a proxy metric for mining intensity over time.

We have made a novel contribution by specialising in tracking changes in open-pit mines using freely available Sentinel-2 multispectral data and state-of-the-art deep learning architectures like TinyCD and LSNet. One of our key contributions is the introduction of the OMS2CD dataset, containing over 2000 manually annotated image patches tailored to mining-related changes like excavation, waste dumping, road construction, deforestation, and equipment activity. Most existing change detection datasets such as OSCD focus solely on generic urban changes, rather than the diverse range of surface changes relevant to mining operations. While many existing datasets highlight urban transformations, this new collection emphasises variations specific to mining, providing essential training material to fine-tune general change detection algorithms for mining scenarios.

Through a series of experiments, we demonstrated how models like TinyCD can be successfully fine-tuned using our mining-specific dataset to identify relevant changes in open-pit mines, even though they were originally developed for urban settings. Among the models tested, TinyCD achieved the highest performance in detecting mining-related changes. This suggests that efficient deep learning architectures can generalise effectively using methods like transfer learning, even with limited training data. Building on indices like NDAI, we also introduced innovative methods to quantify and visualise mining activity over time based on model change predictions. This included the NDTCI activity index and cumulative change map overlays. Our validations results indicate a good correlation between NDTCI activity measurements and human-annotated data from mining sites.

5.2 Recommendations

5.2.1 Enhancing Governance

The potential for transformation in the field of mine monitoring and governance is significant, particularly with the integration of automatic change detection systems that utilise freely available data, such as that from Sentinel-2. The implementation of efficient deep learning (DL) models like TinyCD could enable a shift towards automated, continuous monitoring and analysis. This not only offers a more flexible and comprehensive alternative to manual inspections, but also takes advantage of free, open data.

Furthermore, the establishment of data and tool standards by regulators could create interoperability between different monitoring approaches, enhancing the overall effectiveness of their governance. In a similar vein, industry groups are encouraged to consider voluntary self-monitoring using remote sensing technologies. By publishing their activities using standardised measurements, these groups can demonstrate their commitment to transparency, sustainable development, and climate goals. Such advancements have the potential to revolutionise the way we approach mining governance.

5.2.2 Improving Model Performance

There is great potential for enhancing model performance, particularly with the development of a larger, more comprehensive annotated change detection dataset tailored specifically to mining changes. Such a dataset could significantly improve the accuracy and robustness of the models, thereby enhancing the reliability of the monitoring systems. It is therefore recommended that future research in this area prioritise the publication of more labelled data, ensuring coverage of a diverse range of geographies, mine types, and land cover contexts. This would not only improve the performance of existing models but also contribute to the development of more advanced, accurate monitoring systems in the future. No matter the approach, more public datasets are needed. Many of the promising models and methodologies that were identified in the literature review were impossible to replicate due to the lack of availability of their labelled data.

Looking ahead, the exploration of multi-modal architectures, which combine optical, SAR,

and other types of data, presents a promising avenue for future research. The fusion of different sensor data can provide complementary information, potentially enhancing the resilience of the monitoring systems to common challenges such as cloud cover and speckle noise. Moreover, the use of different data types and spectral bands could offer a more nuanced characterisation of surface mine land cover changes. This could lead to more accurate and detailed monitoring.

5.2.3 Standardising Activity Metrics

Improving and standardising indices like NDTCI and NDAI could offer a better way to track mining activities over time. The development of an open standard would facilitate quantitative cross-comparisons, thereby improving the reliability and consistency of activity assessments. Moreover, an integrated measurement approach that combines various indices (NDTCI, NDAI, MW-RSEI, DROTL, etc.) could be proposed as a unified activity index. The establishment of such a standardised metric could prove invaluable for comparing different mining operations and shaping policy, thereby contributing to more effective and informed governance of mining activities.

5.2.4 Empowering Communities

The advent of remote monitoring offers a powerful toolset for communities impacted by mining activities, enabling them to independently track these activities and provide evidence-based input into decision-making processes. It is crucial for researchers to actively engage with these communities in the development of appropriate monitoring systems and data standards, ensuring that the tools are accessible, user-friendly, and tailored to their needs. Furthermore, providing global access to regular activity reports could significantly enhance participation and accountability, empowering communities and fostering a more transparent and inclusive approach to mining governance.

5.2.5 Summary

In conclusion, the priorities identified in this research—developing automated monitoring systems, exploring multi-modal architectures, and standardising activity metrics for

transparency—represent key areas of focus for the future. The applications proposed are only a glimpse into the many capabilities of artificial intelligence in the domain of remote monitoring. With affordable and accessible remote sensing capabilities, there is great potential to revolutionise mining governance. Increased investment in this area is likely to further amplify its potential impacts, underscoring the need for more research, and active community participation. These elements will be crucial in harnessing the full potential of these technologies to create a more transparent and sustainable mining sector.

5.3 Reflection

Through this work, we have established a proof-of-concept for the use of deep learning and open-access satellite data in monitoring mining activity over time. Venturing into an application domain that has yet to be extensively explored, we have showcased the adaptability of models like TinyCD in detecting relevant mining changes, even with minimal labelled training data. The methods we have proposed for quantifying and visualising activity serve as practical tools for temporal analysis.

A significant achievement of this work is the demonstration of deep learning's potential in this novel application, despite the challenge of limited training data. We overcame this key barrier through the use of transfer learning and efficient models like TinyCD, which generalise well from scarce labels. The creation of a tailored, mining-focused dataset was also a pivotal step, given that most existing benchmarks are centred on urban change detection.

However, our proposed methods are not without limitations. The geographic diversity of our annotated dataset is limited, which restricts the generalisation of our models. The 10-metre resolution limit of Sentinel-2 also poses a challenge to change interpretability, particularly when compared to VHR data, which can provide more visual detail. Seasonal variations and weather events can also confound mining change detection. Furthermore, more robust statistical validation on a diverse range of sites is needed to fully characterise the real-world performance of our methods.

Despite these limitations, we have proposed an affordable, adaptable approach to moni-

toring mining areas over time using freely available data and models. This approach has the potential to support environmental assessment and responsible mining on a global scale. While refinements are necessary, our techniques have demonstrated the potential of deep learning to unlock new value from open-access satellite data. With continued improvements, we believe that our approach could pave the way for greater transparency and sustainability in the mining industry.

Bibliography

Afaq, Yasir and Ankush Manocha (2021-07-01). “Analysis on change detection techniques for remote sensing applications: A review”. In: *Ecological Informatics* 63, p. 101310. ISSN: 1574-9541. DOI: 10.1016/j.ecoinf.2021.101310. URL: https://www.sciencedirect.com/science/article/pii/S1574954121001011 (visited on 2023-03-18).

Altiti, Awwad H. et al. (2021-02-17). *Open Pit Mining*. Publication Title: Mining Techniques - Past, Present and Future. IntechOpen. ISBN: 978-1-83962-369-1. DOI: 10.5772/intechopen.92208. URL: https://www.intechopen.com/chapters/71931 (visited on 2023-03-18).

AndreaCodegoni (2023-06-19). *Tiny_CD*. original-date: 2022-06-15T08:55:31Z. URL: https://github.com/AndreaCodegoni/Tiny_model_4_CD (visited on 2023-07-27).

Bai, Ting et al. (2022-07-01). “Deep learning for change detection in remote sensing: a review”. In: *Geo-spatial Information Science* 0.0. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10095020.2022.2085633, pp. 1–27. ISSN: 1009-5020. DOI: 10.1080/10095020.2022.2085633. URL: https://doi.org/10.1080/10095020.2022.2085633 (visited on 2023-03-18).

Balaniuk, Remis, Olga Isupova, and Steven Reece (2020-01). “Mining and Tailings Dam Detection in Satellite Imagery Using Deep Learning”. In: *Sensors* 20.23. Number: 23 Publisher: Multidisciplinary Digital Publishing Institute, p. 6936. ISSN: 1424-8220. DOI: 10.3390/s20236936. URL: https://www.mdpi.com/1424-8220/20/23/6936 (visited on 2023-03-03).

Bandara, Chaminda (2023-07-27). *DDPM-CD: Remote Sensing Change Detection using Denoising Diffusion Probabilistic Models*. original-date: 2022-04-20T15:27:08Z. URL: https://github.com/wgcban/ddpm-cd (visited on 2023-07-27).

Bandara, Wele Gedara Chaminda, Nithin Gopalakrishnan Nair, and Vishal M. Patel (2022-06-27). *DDPM-CD: Remote Sensing Change Detection using Denoising Diffusion Probabilistic Models*. DOI: 10.48550/arXiv.2206.11892. arXiv: 2206.11892[cs]. URL: http://arxiv.org/abs/2206.11892 (visited on 2023-05-31).

Bandara, Wele Gedara Chaminda and Vishal M. Patel (2022-04-21). *Revisiting Consistency Regularization for Semi-supervised Change Detection in Remote Sensing Images*. DOI: 10.48550/arXiv.2204.08454. arXiv: 2204.08454[cs,eess]. URL: http://arxiv.org/abs/2204.08454 (visited on 2023-05-31).

Canty, Morton J. et al. (2020-01). “Statistical Analysis of Changes in Sentinel-1 Time Series on the Google Earth Engine”. In: *Remote Sensing* 12.1. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 46. ISSN: 2072-4292. DOI: 10.3390/

rs12010046. URL: https://www.mdpi.com/2072-4292/12/1/46 (visited on 2023-05-30).

Chen, Hao and Zhenwei Shi (2020-01). "A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection". In: *Remote Sensing* 12.10. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 1662. ISSN: 2072-4292. DOI: 10.3390/rs12101662. URL: https://www.mdpi.com/2072-4292/12/10/1662 (visited on 2023-06-01).

Chen, Tao et al. (2022). "Open-Pit Mine Area Mapping With Gaofen-2 Satellite Images Using U-Net+". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 3589–3599. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2022.3171290.

Codegoni, Andrea, Gabriele Lombardi, and Alessandro Ferrari (2022-11-07). *TINYCD: A (Not So) Deep Learning Model For Change Detection*. DOI: 10.48550/arXiv.2207.13159. arXiv: 2207.13159[cs,eess]. URL: http://arxiv.org/abs/2207.13159 (visited on 2023-06-01).

Daudt, Rodrigo Caye et al. (2018-07). "Urban Change Detection for Multispectral Earth Observation Using Convolutional Neural Networks". In: *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. ISSN: 2153-7003, pp. 2115–2118. DOI: 10.1109/IGARSS.2018.8518015.

Du, Shouhang et al. (2022-08-09). "Open-pit mine change detection from high resolution remote sensing images using DA-UNet++ and object-based approach". In: *International Journal of Mining, Reclamation and Environment* 36.7. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17480930.2022.2072102, pp. 512–535. ISSN: 1748-0930. DOI: 10.1080/17480930.2022.2072102. URL: https://doi.org/10.1080/17480930.2022.2072102 (visited on 2023-03-18).

Ebel, P., S. Saha, and X. X. Zhu (2021-06-28). "FUSING MULTI-MODAL DATA FOR SUPERVISED CHANGE DETECTION". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B3-2021, pp. 243–249. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLIII-B3-2021-243-2021. URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B3-2021/243/2021/ (visited on 2023-03-28).

Falcon, William and The PyTorch Lightning team (2019-03). *PyTorch lightning*. Version 1.4. DOI: 10.5281/zenodo.3828935. URL: https://github.com/Lightning-AI/lightning.

Global Administrative Areas (2022-07-16). *GADM database of Global Administrative Areas*. Version 4.1. URL: https://gadm.org/.

Gallwey, Jane et al. (2020-10-01). "A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining". In: *Remote Sensing of Environment* 248, p. 111970. ISSN: 0034-4257. DOI: 10.1016/j.rse.2020.111970. URL: https://www.sciencedirect.com/science/article/pii/S0034425720303400 (visited on 2023-05-29).

Gearing up for third Sentinel-2 satellite (2023). URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Gearing_up_for_third_Sentinel-2_satellite (visited on 2023-07-26).

Google Colaboratory (2023). URL: https://colab.research.google.com/ (visited on 2023-07-29).

Gorelick, Noel et al. (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote Sensing of Environment*. Publisher: Elsevier. DOI: 10.1016/j.rse.2017.06.031. URL: https://doi.org/10.1016/j.rse.2017.06.031.

Guo, Junting et al. (2022). "Monitoring of Vegetation Disturbance and Restoration at the Dumping Sites of the Baorixile Open-Pit Mine Based on the LandTrendr Algorithm". In: Num Pages: 9066 Publisher: MDPI AG, p. 9066. ISSN: 1661-7827. DOI: 10.3390/ijerph19159066. URL: https://www.proquest.com/docview/2700621161/abstract/3975F92076BA4F4DPQ/1 (visited on 2023-03-17).

Haldar, Swapan Kumar (2018-01-01). "Chapter 12 - Elements of Mining". In: *Mineral Exploration (Second Edition)*. Ed. by Swapan Kumar Haldar. Elsevier, pp. 229–258. ISBN: 978-0-12-814022-2. DOI: 10.1016/B978-0-12-814022-2.00012-5. URL: https://www.sciencedirect.com/science/article/pii/B9780128140222000125 (visited on 2023-03-03).

Hu, Jiameng et al. (2022). "Remote Sensing Monitoring of Vegetation Reclamation in the Antaibao Open-Pit Mine". In: Num Pages: 5634 Publisher: MDPI AG, p. 5634. DOI: 10.3390/rs14225634. URL: https://www.proquest.com/docview/2739456007/abstract/84BE408E67B140E2PQ/1 (visited on 2023-03-17).

Huang, Jianhua et al. (2022-07). "Remote sensing building change detection based on improved U-Net". In: *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*. 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), pp. 772–775. DOI: 10.1109/ICBAIE56435.2022.9985853.

Jasansky, Simon et al. (2023-01-24). "An open database on global coal and metal mine production". In: *Scientific Data* 10.1. Number: 1 Publisher: Nature Publishing Group, p. 52. ISSN: 2052-4463. DOI: 10.1038/s41597-023-01965-y. URL: https://www.nature.com/articles/s41597-023-01965-y (visited on 2023-03-28).

Jia, Meng and Zhiqiang Zhao (2021-12-11). "Change Detection in Synthetic Aperture Radar Images Based on a Generalized Gamma Deep Belief Networks". In: *Sensors (Basel, Switzerland)* 21.24, p. 8290. ISSN: 1424-8220. DOI: 10.3390/s21248290.

Jiang, Huiwei et al. (2022-01). "A Survey on Deep Learning-Based Change Detection from High-Resolution Remote Sensing Images". In: *Remote Sensing* 14.7. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute, p. 1552. ISSN: 2072-4292. DOI: 10.3390/rs14071552. URL: https://www.mdpi.com/2072-4292/14/7/1552 (visited on 2023-03-18).

Keles, Feyza Duman, Pruthuvi Mahesakya Wijewardena, and Chinmay Hegde (2022-09-11). *On The Computational Complexity of Self-Attention*. DOI: 10.48550/arXiv.2209.04881. arXiv: 2209.04881[cs]. URL: http://arxiv.org/abs/2209.04881 (visited on 2023-08-10).

Kingma, Diederik P. and Jimmy Ba (2017-01-29). *Adam: A Method for Stochastic Optimization*. DOI: 10.48550/arXiv.1412.6980. arXiv: 1412.6980[cs]. URL: http://arxiv.org/abs/1412.6980 (visited on 2023-07-28).

Kozińska, P. and J. Górniak-Zimroz (2021-11). "A review of methods in the field of detecting illegal open-pit mining activities". In: *IOP Conference Series: Earth and Environmental Science* 942.1. Publisher: IOP Publishing, p. 012027. ISSN: 1755-1315. DOI:

10.1088/1755-1315/942/1/012027. URL: https://dx.doi.org/10.1088/1755-1315/942/1/012027 (visited on 2023-03-18).

Kumar, Dilip, R. B. Singh, and Ranjeet Kaur (2019). "Remote-Sensing Technology". In: *Spatial Information Technology for Sustainable Development Goals*. Ed. by Dilip Kumar, R.B. Singh, and Ranjeet Kaur. Sustainable Development Goals Series. Cham: Springer International Publishing, pp. 27–58. ISBN: 978-3-319-58039-5. DOI: 10.1007/978-3-319-58039-5_3. URL: https://doi.org/10.1007/978-3-319-58039-5_3 (visited on 2023-08-04).

Kumar, Saurabh and Shwetank Arya (2021-09-01). "Change Detection Techniques for Land Cover Change Analysis Using Spatial Datasets: a Review". In: *Remote Sensing in Earth Systems Sciences* 4.3, pp. 172–185. ISSN: 2520-8209. DOI: 10.1007/s41976-021-00056-z. URL: https://doi.org/10.1007/s41976-021-00056-z (visited on 2023-03-15).

Kumar, Vignesh and Kiran Yarrakula (2022-09-01). "Environmental impact assessment of limestone quarry using multispectral satellite imagery". In: *Earth Science Informatics* 15. DOI: 10.1007/s12145-022-00845-0.

Leenstra, Marrit et al. (2020-11-19). *Sentinel-2 Multitemporal Cities Pairs*. DOI: 10.5281/zenodo.4280482. URL: https://zenodo.org/record/4280482 (visited on 2023-03-28).

Li, Jun et al. (2023). "Change Detection of Open-Pit Mine Based on Siamese Multiscale Network". In: *IEEE Geoscience and Remote Sensing Letters* 20. Conference Name: IEEE Geoscience and Remote Sensing Letters, pp. 1–5. ISSN: 1558-0571. DOI: 10.1109/LGRS.2022.3232763.

Lin, Tsung-Yi et al. (2015-02-20). *Microsoft COCO: Common Objects in Context*. DOI: 10.48550/arXiv.1405.0312. arXiv: 1405.0312[cs]. URL: http://arxiv.org/abs/1405.0312 (visited on 2023-07-28).

Liu, Biyuan, Huaixin Chen, and Zhixi Wang (2022-01-22). *LSNet: Extremely Light-Weight Siamese Network For Change Detection in Remote Sensing Image*. DOI: 10.48550/arXiv.2201.09156. arXiv: 2201.09156[cs]. URL: http://arxiv.org/abs/2201.09156 (visited on 2023-05-31).

Loshchilov, Ilya and Frank Hutter (2019-01-04). *Decoupled Weight Decay Regularization*. DOI: 10.48550/arXiv.1711.05101. arXiv: 1711.05101[cs,math]. URL: http://arxiv.org/abs/1711.05101 (visited on 2023-07-28).

Mandanici, Emanuele and Gabriele Bitelli (2016-12). "Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use". In: *Remote Sensing* 8.12. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 1014. ISSN: 2072-4292. DOI: 10.3390/rs8121014. URL: https://www.mdpi.com/2072-4292/8/12/1014 (visited on 2023-07-25).

Mapbox (2023). *geojson.io*. URL: https://github.com/mapbox/geojson.io (visited on 2023-08-03).

Maus, Victor et al. (2022-07-22). "An update on global mining land use". In: *Scientific Data* 9, p. 433. ISSN: 2052-4463. DOI: 10.1038/s41597-022-01547-4. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9307859/ (visited on 2023-02-25).

Moon, Jihyun and Hoonyol Lee (2021). "Analysis of Activity in an Open-Pit Mine by Using InSAR Coherence-Based Normalized Difference Activity Index". In: Num Pages: 1861 Publisher: MDPI AG, p. 1861. DOI: 10.3390/rs13091861. URL: https://www.

proquest.com/docview/2530134248/abstract/4D12507C8F4143PQ/1 (visited on 2023-03-18).

Nascimento, Filipe Silveira et al. (2020-01). "Land Cover Changes in Open-Cast Mining Complexes Based on High-Resolution Remote Sensing Data". In: *Remote Sensing* 12.4. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 611. ISSN: 2072-4292. DOI: 10.3390/rs12040611. URL: https://www.mdpi.com/2072-4292/12/4/611 (visited on 2023-03-17).

Nava, Lorenzo et al. (2022). "Artisanal and Small-Scale Mine Detection in Semi-Desertic Areas by Improved U-Net". In: *IEEE Geoscience and Remote Sensing Letters* 19. Conference Name: IEEE Geoscience and Remote Sensing Letters, pp. 1–5. ISSN: 1558-0571. DOI: 10.1109/LGRS.2022.3220487.

Paperspace - Gradient Notebooks (2023). URL: https://www.paperspace.com/gradient/notebooks (visited on 2023-07-29).

Phiri, Darius et al. (2020-01). "Sentinel-2 Data for Land Cover/Use Mapping: A Review". In: *Remote Sensing* 12.14. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute, p. 2291. ISSN: 2072-4292. DOI: 10.3390/rs12142291. URL: https://www.mdpi.com/2072-4292/12/14/2291 (visited on 2023-07-25).

Plank, Simon (2014-06). "Rapid Damage Assessment by Means of Multi-Temporal SAR — A Comprehensive Review and Outlook to Sentinel-1". In: *Remote Sensing* 6.6. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute, pp. 4870–4906. ISSN: 2072-4292. DOI: 10.3390/rs6064870. URL: https://www.mdpi.com/2072-4292/6/6/4870 (visited on 2023-07-23).

QGIS (2023-06-24). *QGIS geographic information system*. Version 3.32 Lima. URL: https://www.qgis.org.

Qu, Xiaofan et al. (2022). "Change Detection in Synthetic Aperture Radar Images Using a Dual-Domain Network". In: *IEEE Geoscience and Remote Sensing Letters* 19, pp. 1–5. ISSN: 1545-598X, 1558-0571. DOI: 10.1109/LGRS.2021.3073900. arXiv: 2104.06699[cs,eess]. URL: http://arxiv.org/abs/2104.06699 (visited on 2023-03-18).

Riba, Edgar et al. (2019-10-09). *Kornia: an Open Source Differentiable Computer Vision Library for PyTorch*. DOI: 10.48550/arXiv.1910.02190. arXiv: 1910.02190[cs]. URL: http://arxiv.org/abs/1910.02190 (visited on 2023-07-27).

Ruifeng, Liu et al. (2022-12-19). "Extraction and spatiotemporal changes of open-pit mines during 1985-2020 using Google Earth Engine: A case study of Qingzhou City, Shandong Province, China". In: *Environmental Monitoring and Assessment* 195.1, p. 209. ISSN: 1573-2959. DOI: 10.1007/s10661-022-10837-8.

Saharia, Chitwan et al. (2021-06-30). *Image Super-Resolution via Iterative Refinement*. DOI: 10.48550/arXiv.2104.07636. arXiv: 2104.07636[cs,eess]. URL: http://arxiv.org/abs/2104.07636 (visited on 2023-07-28).

Sentinel-1 - Missions - Sentinel Online (2023). Sentinel Online. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-1 (visited on 2023-07-23).

Shafique, Ayesha et al. (2022-01). "Deep Learning-Based Change Detection in Remote Sensing Images: A Review". In: *Remote Sensing* 14.4. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 871. ISSN: 2072-4292. DOI: 10.3390/rs14040871. URL: https://www.mdpi.com/2072-4292/14/4/871 (visited on 2023-03-22).

Shi, Jieqing et al. (2022-09). "Intelligent classification of land cover types in open-pit mine area using object-oriented method and multitask learning". In: *Journal of Applied Remote Sensing* 16.3. Publisher: SPIE, p. 038504. ISSN: 1931-3195, 1931-3195. DOI:

10.1117/1.JRS.16.038504. URL: https://remotesensing.spiedigitallibrary.org/journals/journal-of-applied-remote-sensing/volume-16/issue-3/038504/Intelligent-classification-of-land-cover-types-in-open-pit-mine/10.1117/1.JRS.16.038504.full (visited on 2023-05-31).

Shi, Wenzhong et al. (2020-01). "Change Detection Based on Artificial Intelligence: State-of-the-Art and Challenges". In: *Remote Sensing* 12.10. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 1688. ISSN: 2072-4292. DOI: 10.3390/rs12101688. URL: https://www.mdpi.com/2072-4292/12/10/1688 (visited on 2023-03-18).

The GIMP Development Team (2023-07-27). *GIMP*. Version 2.10.34. URL: https://www.gimp.org.

Transon, Julie et al. (2018-02). "Survey of Hyperspectral Earth Observation Applications from Space in the Sentinel-2 Context". In: *Remote Sensing* 10.2. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 157. ISSN: 2072-4292. DOI: 10.3390/rs10020157. URL: https://www.mdpi.com/2072-4292/10/2/157 (visited on 2023-07-25).

Vorovencii, Iosif (2021-01-05). "Changes detected in the extent of surface mining and reclamation using multitemporal Landsat imagery: a case study of Jiu Valley, Romania". In: *Environmental Monitoring and Assessment* 193.1, p. 30. ISSN: 1573-2959. DOI: 10.1007/s10661-020-08834-w. URL: https://doi.org/10.1007/s10661-020-08834-w (visited on 2023-03-24).

Wang, Chen, Tao Chen, and Antonio Plaza (2023-08-01). "MFE-ResNet: A new extraction framework for land cover characterization in mining areas". In: *Future Generation Computer Systems* 145, pp. 550–562. ISSN: 0167-739X. DOI: 10.1016/j.future.2023.04.001. URL: https://www.sciencedirect.com/science/article/pii/S0167739X23001358 (visited on 2023-05-31).

Wang, Chunsheng et al. (2020-10-22). "Automatic Identification and Dynamic Monitoring of Open-Pit Mines Based on Improved Mask R-CNN and Transfer Learning". In: *Remote Sensing* 12, p. 3474. DOI: 10.3390/rs12213474.

Wang, Lili et al. (2021-01). "Monitoring Mining Activities Using Sentinel-1A InSAR Coherence in Open-Pit Coal Mines". In: *Remote Sensing* 13.21. Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, p. 4485. ISSN: 2072-4292. DOI: 10.3390/rs13214485. URL: https://www.mdpi.com/2072-4292/13/21/4485 (visited on 2023-03-24).

Wang, Shunyao et al. (2020). "Evaluating the Feasibility of Illegal Open-Pit Mining Identification Using Insar Coherence". In: Num Pages: 367 Publisher: MDPI AG, p. 367. DOI: 10.3390/rs12030367. URL: https://www.proquest.com/docview/2550249601/abstract/9C9E51FF1DEF4812PQ/1 (visited on 2023-03-17).

wenhwu (2023-03-27). *Awesome Remote Sensing Change Detection*. original-date: 2018-11-03T04:09:01Z. URL: https://github.com/wenhwu/awesome-remote-sensing-change-detection (visited on 2023-03-28).

Wu, Qianhan et al. (2020). "Integration of TanDEM-X and SRTM DEMs and Spectral Imagery to Improve the Large-Scale Detection of Opencast Mining Areas". In: Num Pages: 1451 Publisher: MDPI AG, p. 1451. DOI: 10.3390/rs12091451. URL: https://www.proquest.com/docview/2400045657/abstract/89C40EFC3863476BPQ/1 (visited on 2023-03-17).

Yin, Hongyang et al. (2023-03-01). “Attention-guided siamese networks for change detection in high resolution remote sensing images”. In: *International Journal of Applied Earth Observation and Geoinformation* 117, p. 103206. ISSN: 1569-8432. DOI: 10.1016/j.jag.2023.103206. URL: https://www.sciencedirect.com/science/article/pii/S1569843223000286 (visited on 2023-03-18).

Zhang, Chenxiao et al. (2020-08-01). “A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166, pp. 183–200. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2020.06.003. URL: https://www.sciencedirect.com/science/article/pii/S0924271620301532 (visited on 2023-07-29).

Zhang, Mengxuan et al. (2022). “Deep Siamese Network with Contextual Transformer for Remote Sensing Images Change Detection”. In: *Intelligence Science IV*. Ed. by Zhongzhi Shi, Yaochu Jin, and Xiangrong Zhang. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, pp. 193–200. ISBN: 978-3-031-14903-0. DOI: 10.1007/978-3-031-14903-0_21.

Zhao, Lingran et al. (2022). “Application of Improved Instance Segmentation Algorithm Based on VoVNet-v2 in Open-Pit Mines Remote Sensing Pre-Survey”. In: Num Pages: 2626 Publisher: MDPI AG, p. 2626. DOI: 10.3390/rs14112626. URL: https://www.proquest.com/docview/2674398282/abstract/7D7460779B094337PQ/1 (visited on 2023-03-17).

Zhao, Qiang et al. (2022-01). “An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends”. In: *Remote Sensing* 14.8. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, p. 1863. ISSN: 2072-4292. DOI: 10.3390/rs14081863. URL: https://www.mdpi.com/2072-4292/14/8/1863 (visited on 2023-07-25).

Zhu, Dongyu et al. (2020-05-01). “Monitoring the effects of open-pit mining on the eco-environment using a moving window-based remote sensing ecological index”. In: *Environmental Science and Pollution Research* 27.13, pp. 15716–15728. ISSN: 1614-7499. DOI: 10.1007/s11356-020-08054-2. URL: https://doi.org/10.1007/s11356-020-08054-2 (visited on 2023-03-18).

Appendix A

Ideal Case Study Example: Iluka Western Australia Facility

In this appendix, we review another case study area example. At the Iluka Western Australia site we were able to obtain a contiguous series of cloud-free images, which allows for an accurate assessment of spatio-temporal changes. The Iluka site is listed in the Jasansky et al. database to have started production in 2019, and to be producing primarily iron ore products. The time period of observation in our downloaded Sentinel-2 tiles is from 2019-01-27 until 2020-11-07, and during that time there is a substantial amount of visible surface change (Figure A.1). Our quantitative visual analysis results are given in figures A.2, A.3, and A.4. Figure A.4 demonstrates how the facility expanded towards the south-east over time, similar to the visible expansion of the Carrapateena site in Section 4.7.

Figure A.1: Sentinel-2 image tiles of the Iluka Western Australia mining site. The width and height of each tile is 439×455 pixels, 4.39 km×4.55 km with 10 metre resolution. Left: the site at the beginning of the target study period. Right: the site at the end of the target study period. The active surface area visibly grows during the target time period.

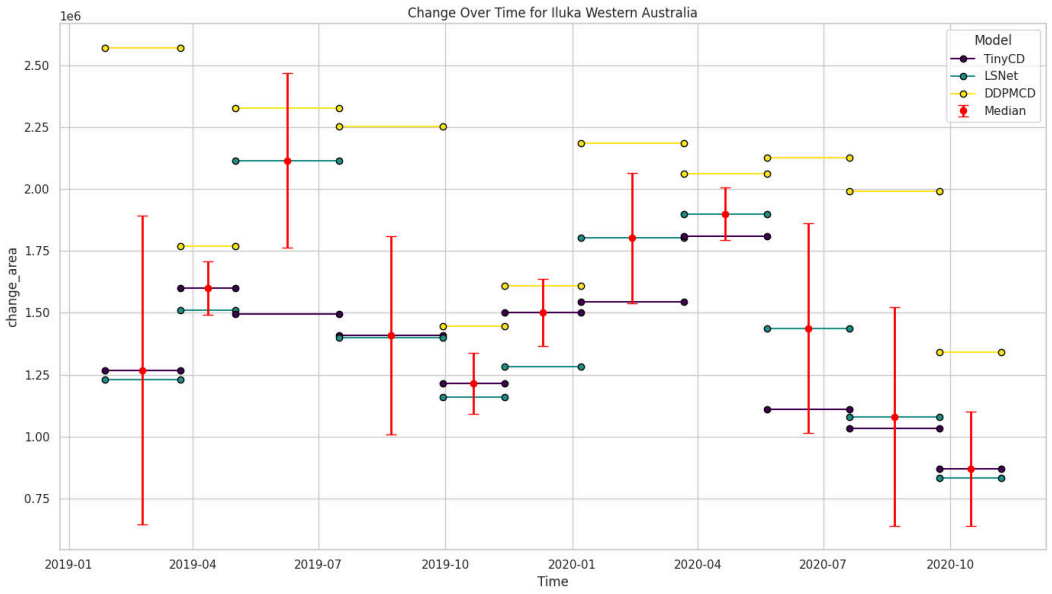


Figure A.2: Timeline plot demonstrating the amount of surface area predicted to have changed for each time interval (change is in m^2) at the Iluka Western Australia facility.

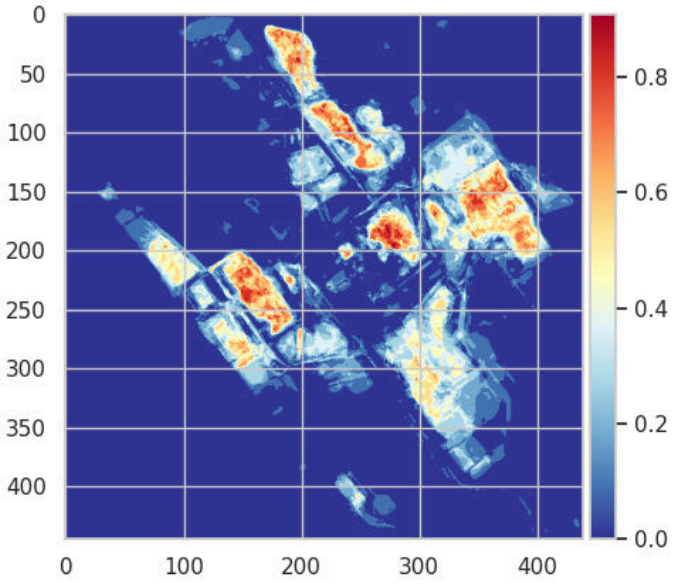


Figure A.3: Visualisation of the μ_{NDTCI} calculated using the change masks predicted by LSNet over the study period for the Iluka Western Australia facility.

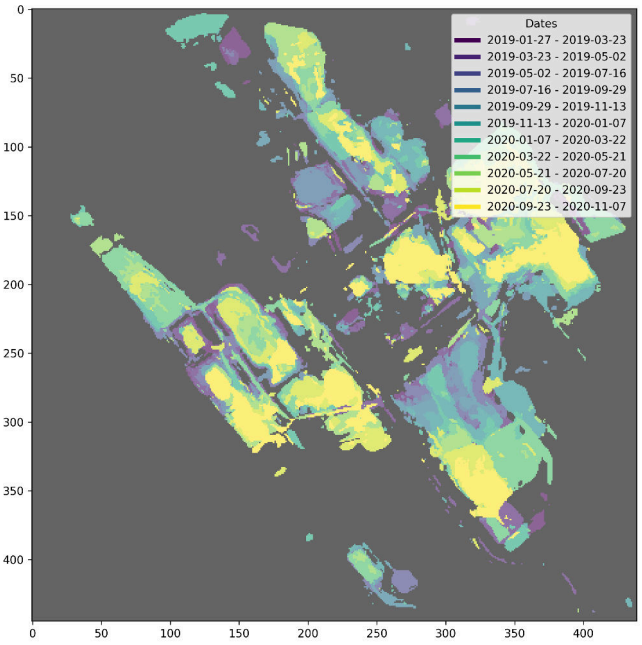


Figure A.4: Visualisation demonstrating the spatio-temporal changes at the Iluka Western Australia facility. Vertical and horizontal axes are measured in pixels.

Appendix B

Non-ideal Case Study Example: MetCoal Facility

In this appendix, we review a case study area example which tends to be occluded by clouds. The site is listed coal mine under the facility name 'MetCoal' in the Jasansky et al. with a production start date of 2019. The time period of observation in our downloaded Sentinel-2 tiles is from 2019-01-06 until 2020-11-16. As the site is in Indonesia, many of the Sentinel-2 tiles include significant visible cloud cover. Figure B.1 shows an example of how such occlusion tricks the LSNet model into making false-positive and false-negative predictions due to the changing surface lighting caused by cloud cover.

In the quantitative change assessments in figures B.2, B.3, and B.4 there are significant artefacts caused by the changing cloud cover which has been falsely predicted as change. In Figure B.2 the false-positive predictions are visible where DDPM-CD and LSNet predict a higher level of change than the other two models in the first and last time period intervals. However, we see that using the median of the three model predictions helps to ignore the outlier changes caused by cloud cover movements. In Figure B.3 and Figure B.4 there are spatio-temporal changes throughout the image that have arisen from cloud cover movements.

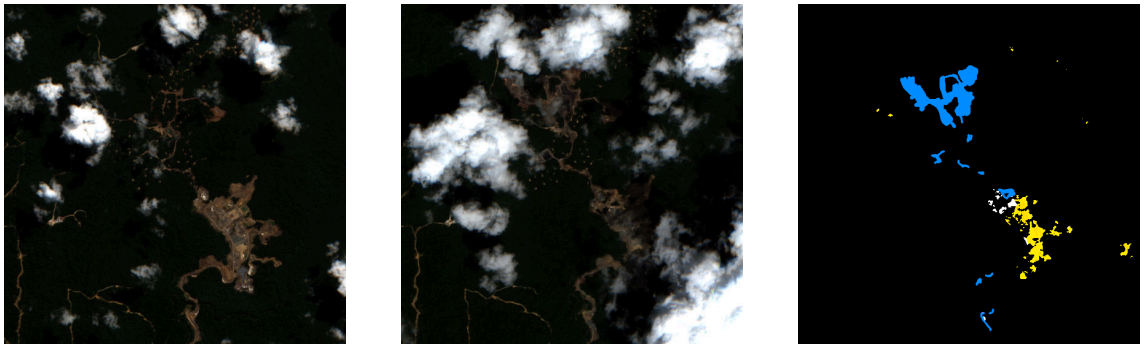


Figure B.1: Cloud-occluded prediction example from MetCoal site. From left to right: Sentinel-2 'pre' image tile; Sentinel-2 'post' image tile; LSNet model change predictions. In the prediction mask, white represents TP changes, yellow represents FP changes, and blue represents FN changes.

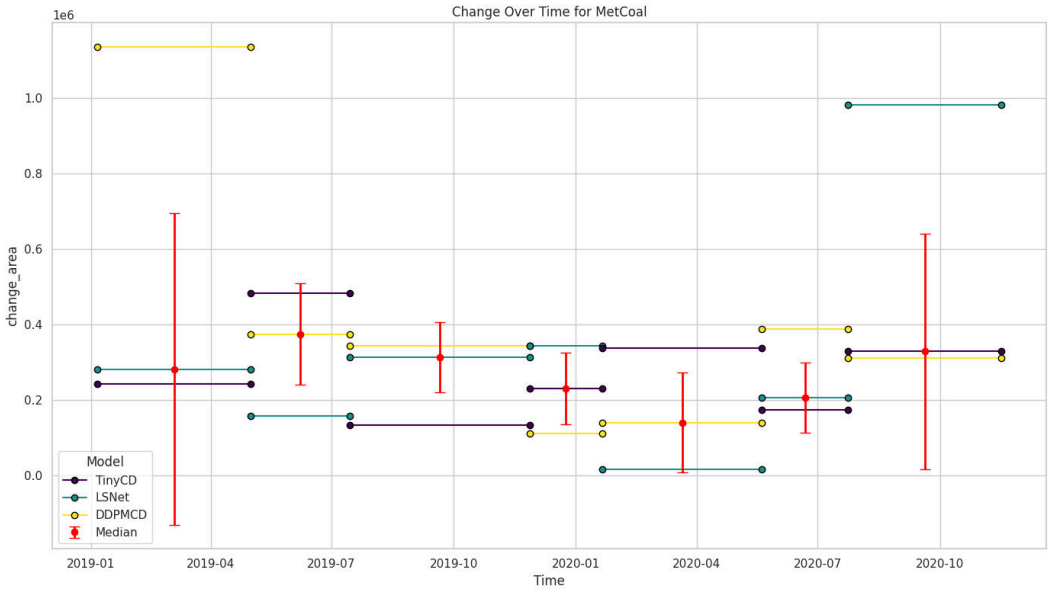


Figure B.2: Timeline plot demonstrating the amount of surface area predicted to have changed for each time interval (change is in m^2) at the MetCoal facility.

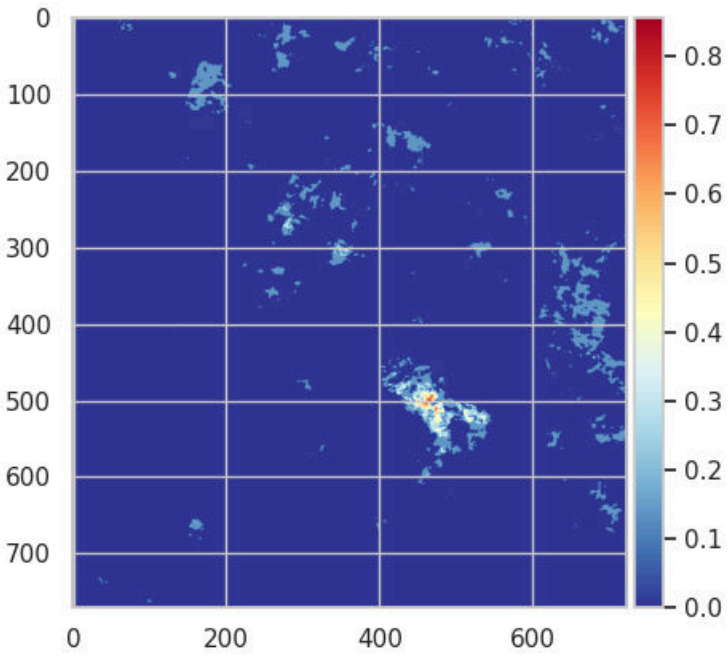


Figure B.3: Visualisation of the μ_{NDTCI} calculated using the change masks predicted by LSNet over the study period for the MetCoal facility.

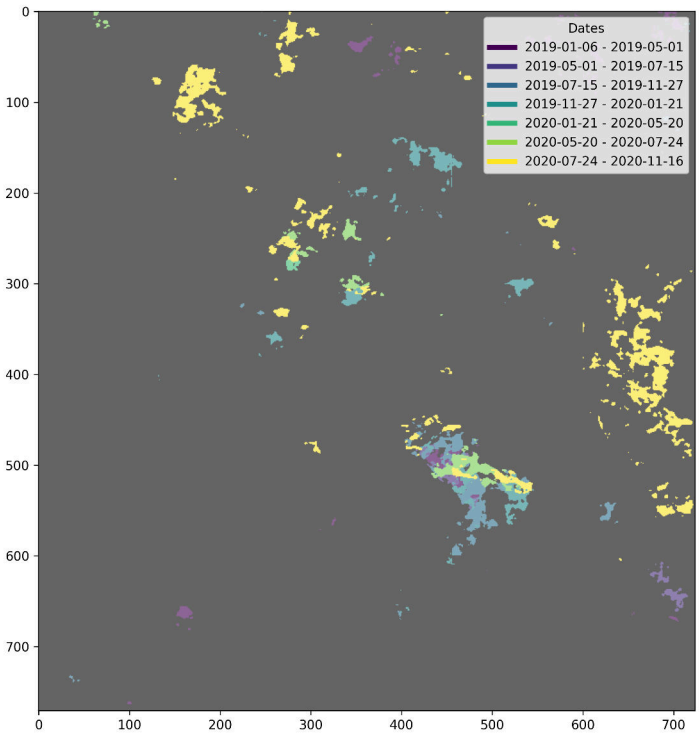


Figure B.4: Visualisation demonstrating the spatio-temporal changes at the MetCoal facility. Vertical and horizontal axes are measured in pixels.

Appendix C

Model Prediction Comparison: Ubuntu Facility Mask Predictions

We present in this appendix a side-by-side comparison of predicted changes for the same bi-temporal image pair from different models (Figure C.2). The changes were predicted by each of the three models using two different positive-class thresholds (0.4 and 0.6). We selected this particular pair of images to compare because it is particularly challenging due to visible seasonal changes such as vegetation growth (Figure C.1).

It is difficult to say which model performs the 'best' on this specific image pair, because some changed features are quite small and we have not prepared a ground-truth for comparison. With that said, it is easy to see how the number of change-class predictions decreases when the positive-class threshold is increased. Additionally, we can note that at the given thresholds, LSNet and DDPM-CD seem to be much more sensitive to the visible changes. None of the models incorrectly predict change in the areas where there is only vegetative growth, which demonstrates that they are all robust against that type of image noise.

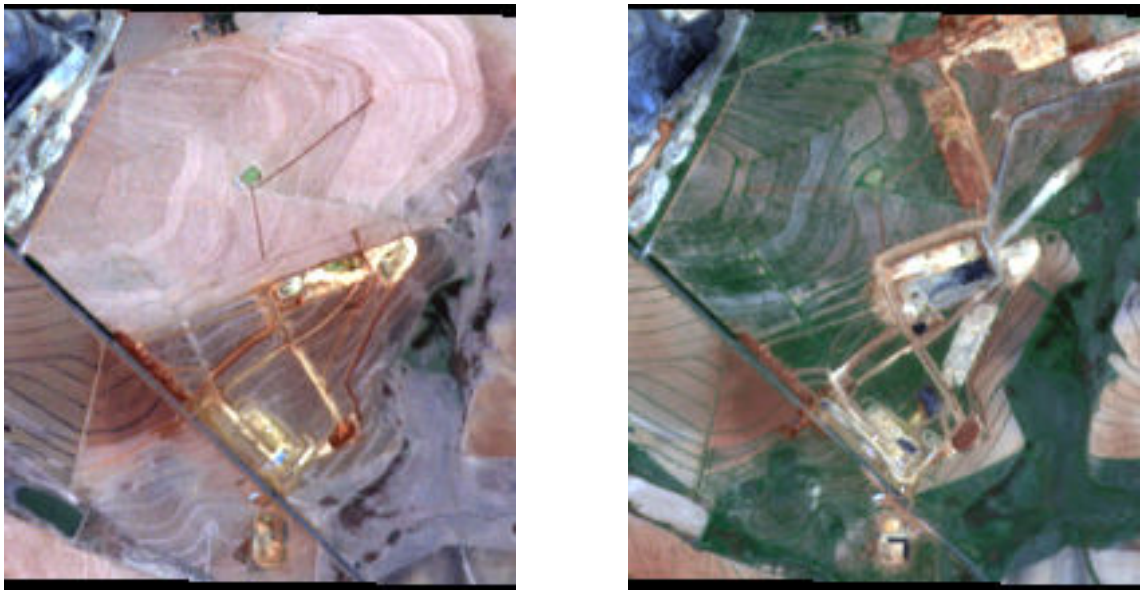
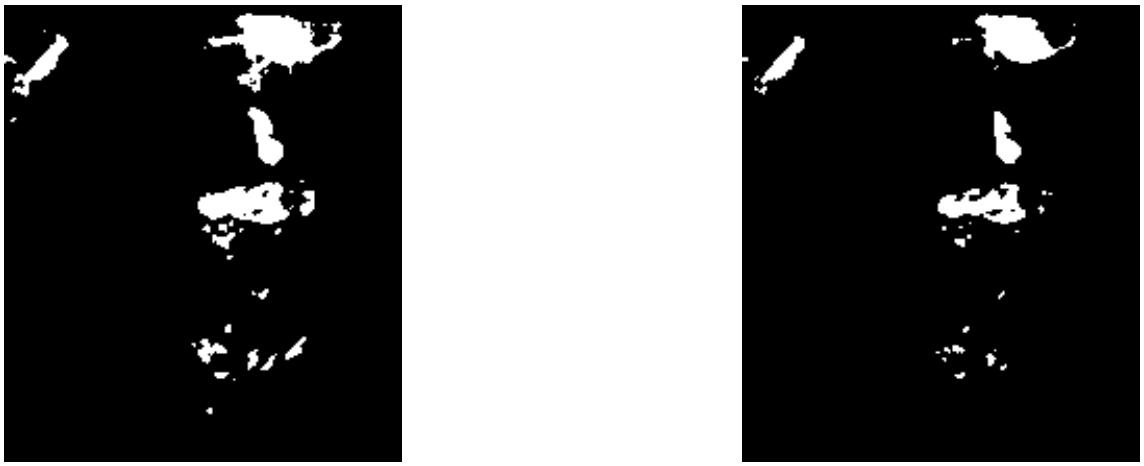


Figure C.1: Ubuntu case study facility Sentinel-2 bi-temporal image pair. Left: facility at 2019-09-19. Right: facility at 2019-11-28. Both images have been processed with histogram equalisation to make the terrain features more easily visible.

(a) TinyCD change predictions.



(b) LSNet change predictions.

(c) DDPM-CD change predictions.

Figure C.2: A comparison of the predicted masks for the bi-temporal image pair given in Figure C.1. The left column predictions were all produced using a positive-class threshold of 0.4, while the right column predictions were produced with a threshold of 0.6.

Appendix D

Model Prediction Comparison: NDTCI for Guizhou Facility

We present in this appendix a side-by-side comparison of the predicted μ_{NDTCI} for the Guizhou facility from the OMS2CD dataset. The changes were predicted by each of the three models using two different positive-class thresholds (0.4 and 0.6). Figure D.1 shows a bi-temporal pair for the first and last Sentinel-2 image tiles in our downloaded series, which were used for the μ_{NDTCI} calculations. The ground-truth μ_{NDTCI} is given in Figure D.2, and the predictions for each model are shown in Figure D.3. Note that the AOI mask discussed in Section 4.4 was used to focus the predictions of the models and reduce false-positives. By comparing the predicted μ_{NDTCI} maps to the ground-truth, it appears that one of the greatest differences across all models is the absence of small changes.

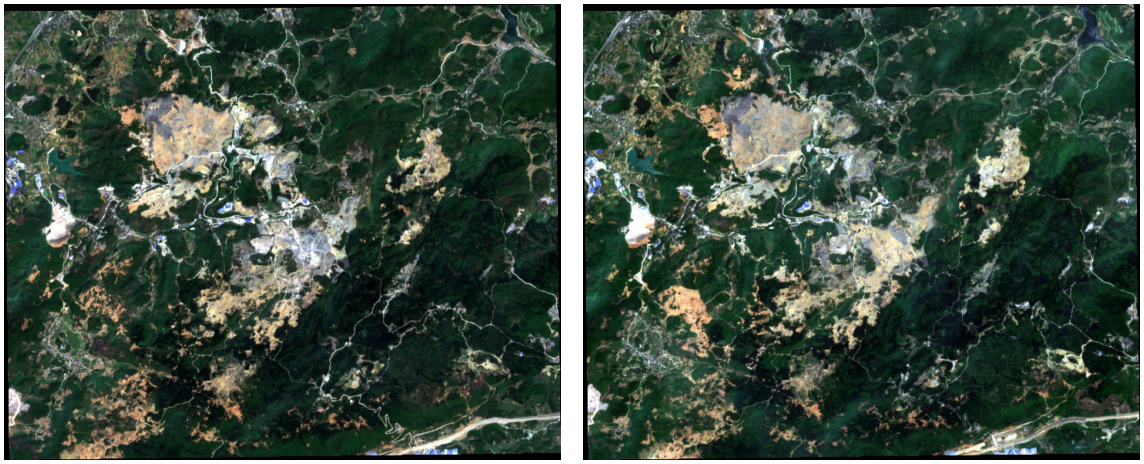


Figure D.1: Guizhou facility Sentinel-2 bi-temporal image pair for the beginning and end of the dataset time period. Left: facility at 2019-04-01. Right: facility at 2020-04-01. Both images have been processed with histogram equalisation to make the terrain features more easily visible.

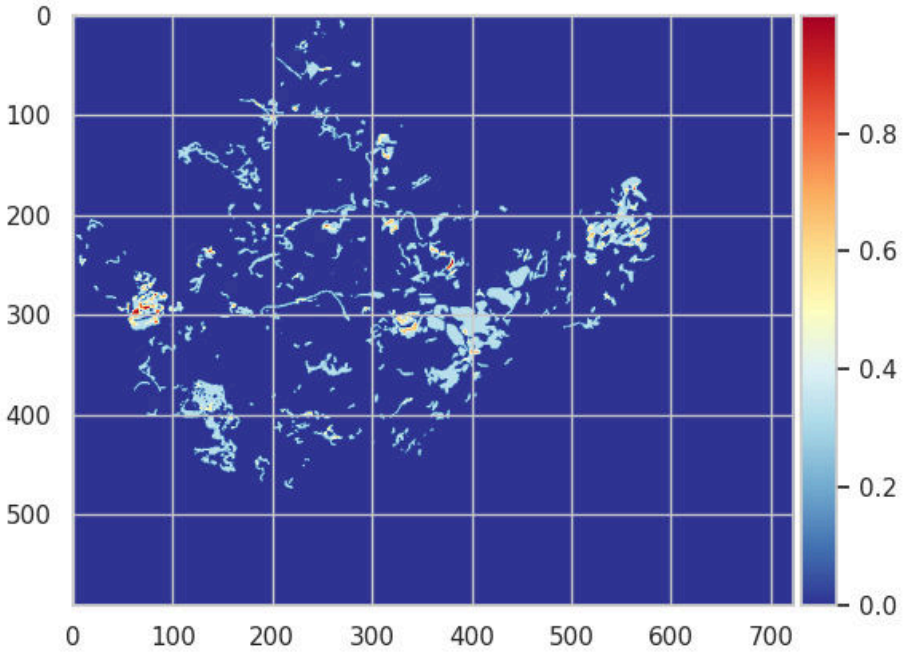
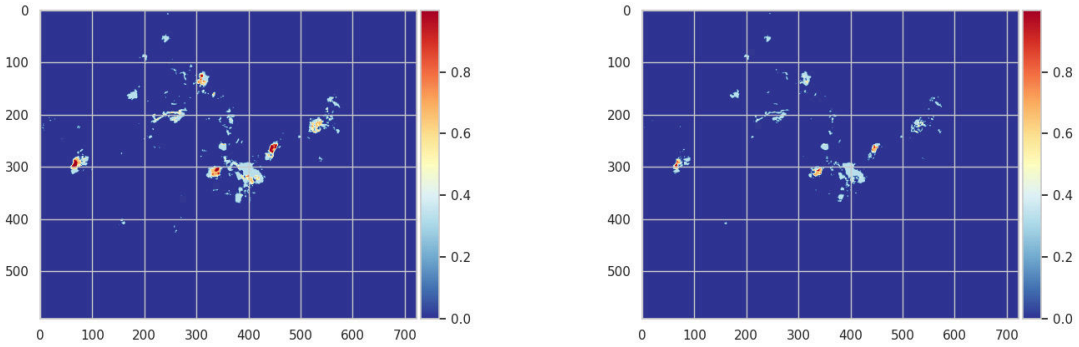
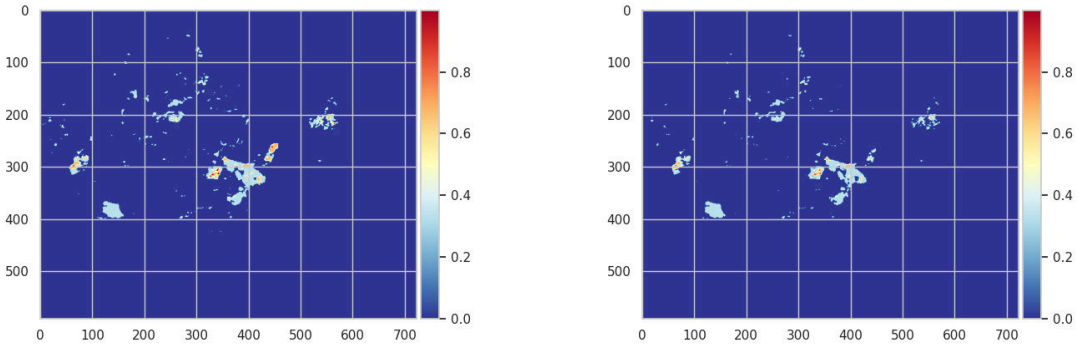


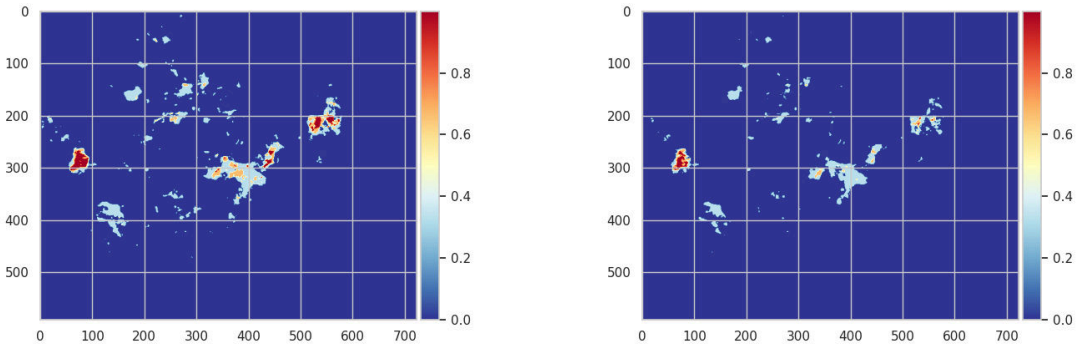
Figure D.2: Ground-truth μ_{NDTCI} visualisation for the Guizhou facility. Calculated using the annotated ground-truth masks.



(a) TinyCD μ_{NDTCI} predictions.



(b) LSNet μ_{NDTCI} predictions.



(c) DDPM-CD μ_{NDTCI} predictions.

Figure D.3: A comparison of the predicted μ_{NDTCI} maps for the Guizhou facility in the time period represented in Figure D.1. The left column predictions were all produced using a positive-class threshold of 0.4, while the right column predictions were produced with a threshold of 0.6.

Appendix E

Model Prediction Comparison: NDTCI for Mianchi Facility

We present in this appendix a side-by-side comparison of the predicted μ_{NDTCI} for the Mianchi facility from the OMS2CD dataset. The changes were predicted by each of the three models using two different positive-class thresholds (0.4 and 0.6). Figure E.1 shows a bi-temporal pair for the first and last Sentinel-2 image tiles in our downloaded series, which were used for the μ_{NDTCI} calculations. The ground-truth μ_{NDTCI} is given in Figure E.2, and the predictions for each model are shown in Figure E.3. Note that the AOI mask discussed in Section 4.4 was used.

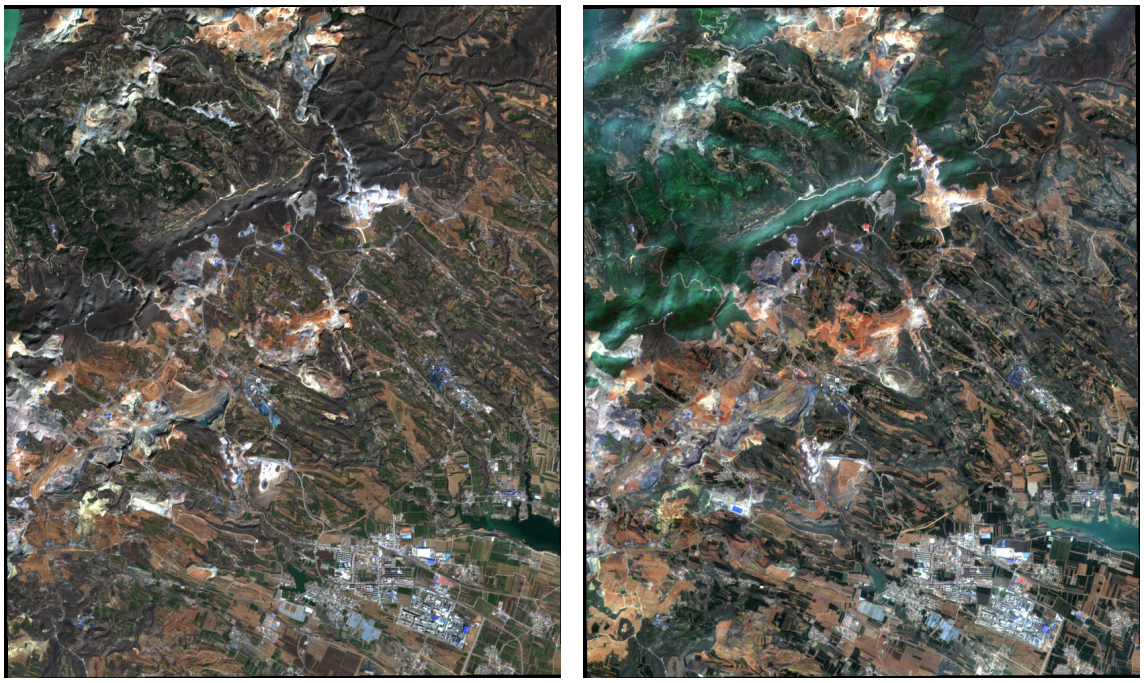


Figure E.1: Mianchi facility Sentinel-2 bi-temporal image pair for the beginning and end of the dataset time period. Left: facility at 2019-03-01. Right: facility at 2020-04-01. Both images have been processed with histogram equalisation to make the terrain features more easily visible.

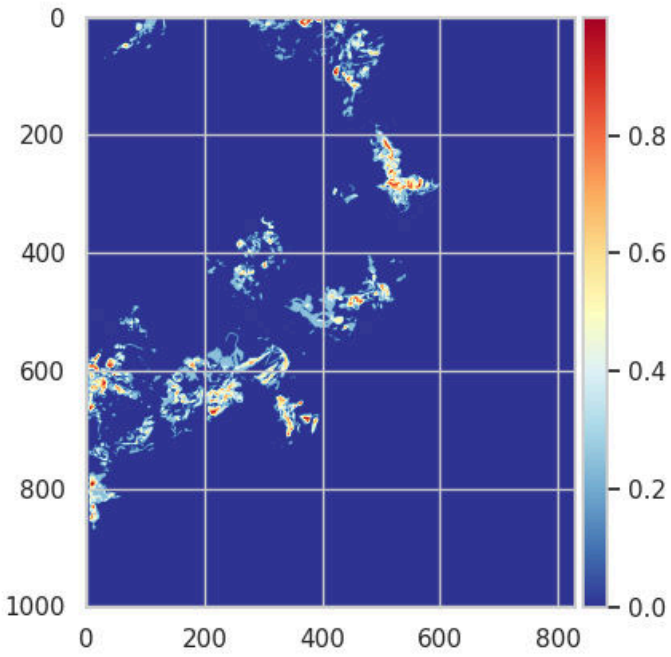
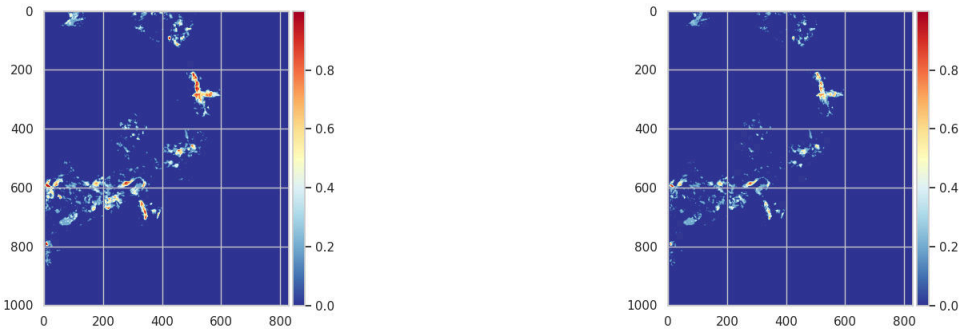
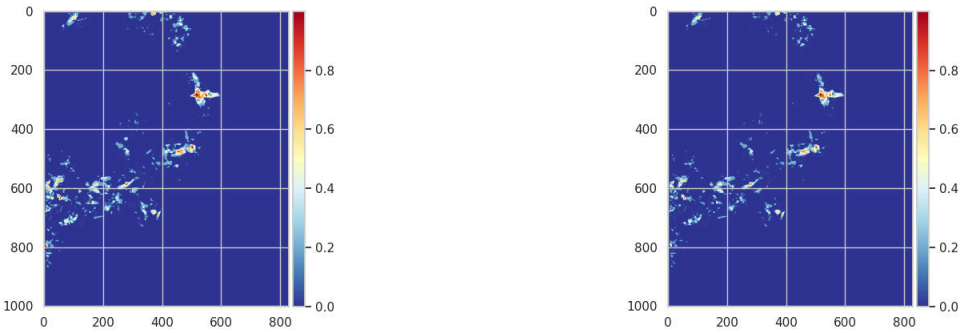


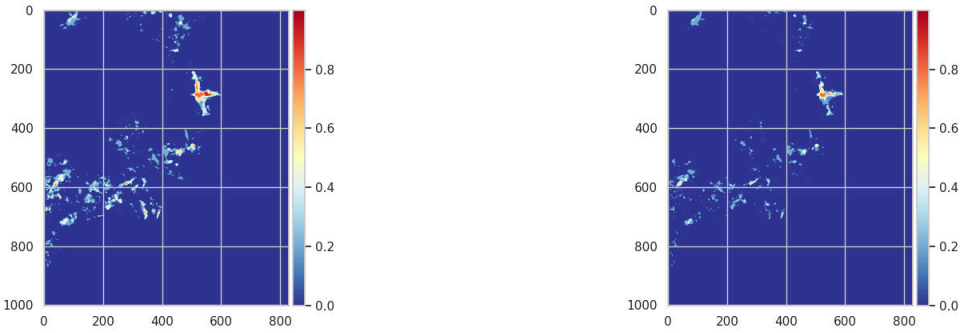
Figure E.2: Ground-truth μ_{NDTCI} visualisation for the Mianchi facility. Calculated using the annotated ground-truth masks.



(a) TinyCD μ_{NDTCI} predictions.



(b) LSNet μ_{NDTCI} predictions.



(c) DDPM-CD μ_{NDTCI} predictions.

Figure E.3: A comparison of the predicted μ_{NDTCI} maps for the Mianchi facility in the time period represented in Figure E.1. The left column predictions were all produced using a positive-class threshold of 0.4, while the right column predictions were produced with a threshold of 0.6.

Appendix F

Model Prediction Comparison: NDTCI and Spatio-Temporal Change for Ubuntu Facility

We present in this appendix a side-by-side comparison of the predicted μ_{NDTCI} and spatio-temporal area changes for the Ubuntu facility. The changes were predicted by each of the three models using a prediction threshold of 0.4. Figure F.1 shows a bi-temporal pair for the first and last Sentinel-2 image tiles in our downloaded series which were used for the μ_{NDTCI} calculations. The predicted μ_{NDTCI} and spatio-temporal change maps for each model are shown in Figure F.2.

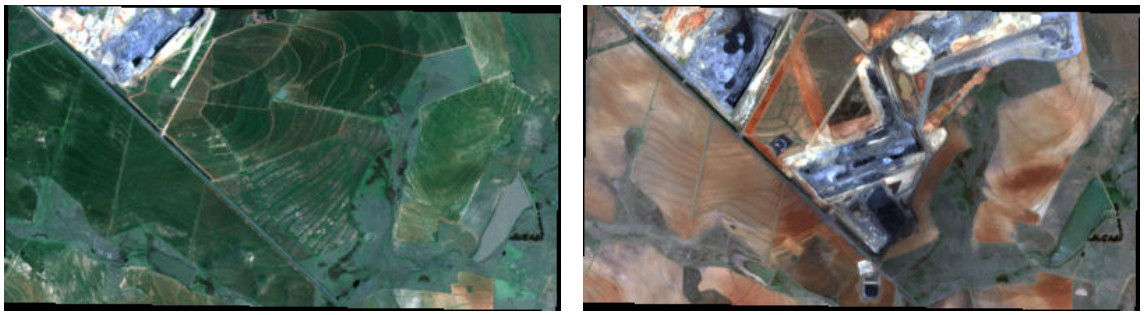
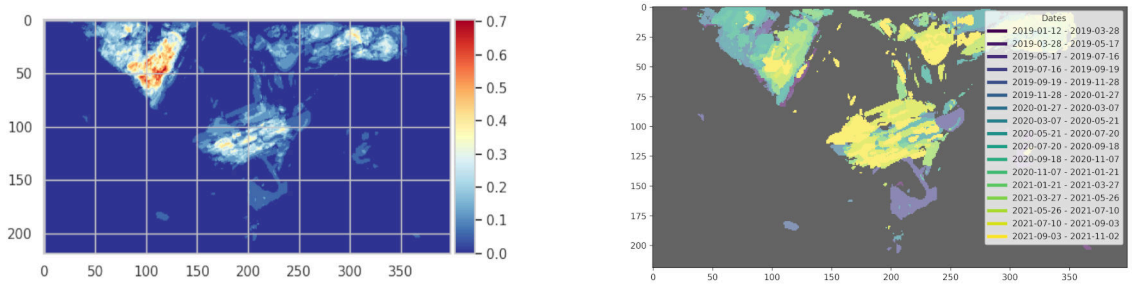
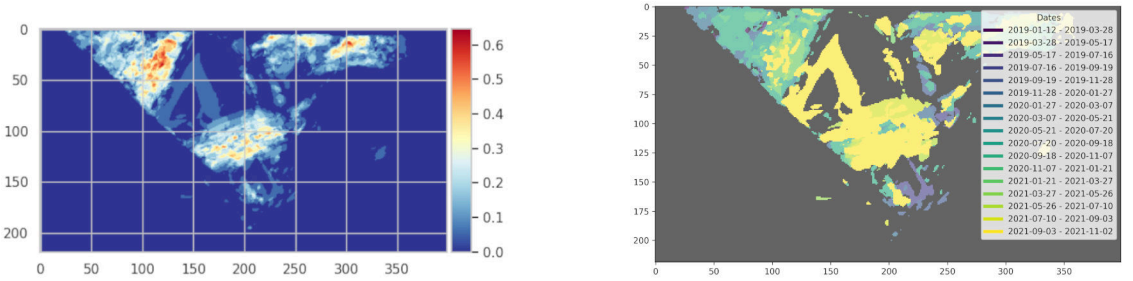


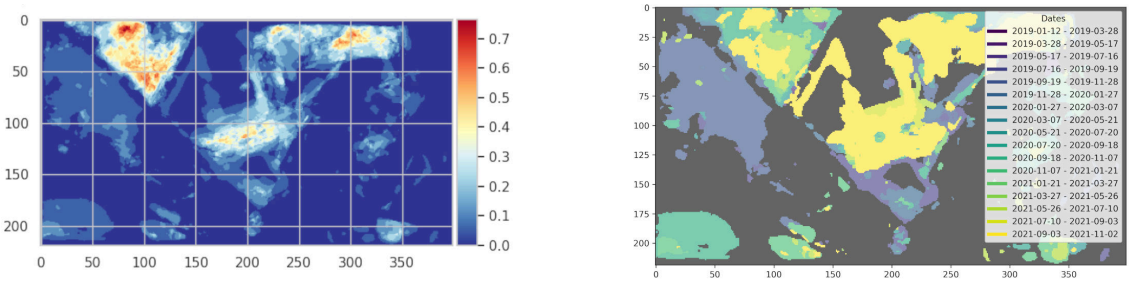
Figure F.1: Ubuntu facility Sentinel-2 bi-temporal image pair for the beginning and end of the dataset time period. Left: facility at 2019-01-12. Right: facility at 2021-11-02. Both images have been processed with histogram equalisation to make the terrain features more easily visible.



(a) TinyCD μ_{NDTCI} predictions.



(b) LSNet μ_{NDTCI} predictions.



(c) DDPM-CD μ_{NDTCI} predictions.

Figure F.2: A comparison of the predicted μ_{NDTCI} and spatio-temporal change maps for the Ubuntu facility in the time period represented in Figure F.1. The left column is the μ_{NDTCI} map for each model architecture, while the right column is the spatio-temporal change map. All were produced with a threshold of 0.4.

Appendix G

Data Domain Differences

As highlighted in Section 2.3.1, our target domain, which comprises Sentinel-2 tiles of mining areas, exhibits marked distinctions when compared to benchmark datasets like OSCD, OMCD, and more. These disparities are evident in aspects such as the magnitude and nature of surface alterations and object attributes. For a visual comparison drawn from each dataset, refer to Figure G.1. Section 4.8 delves deeper into the significance of these scale differences.

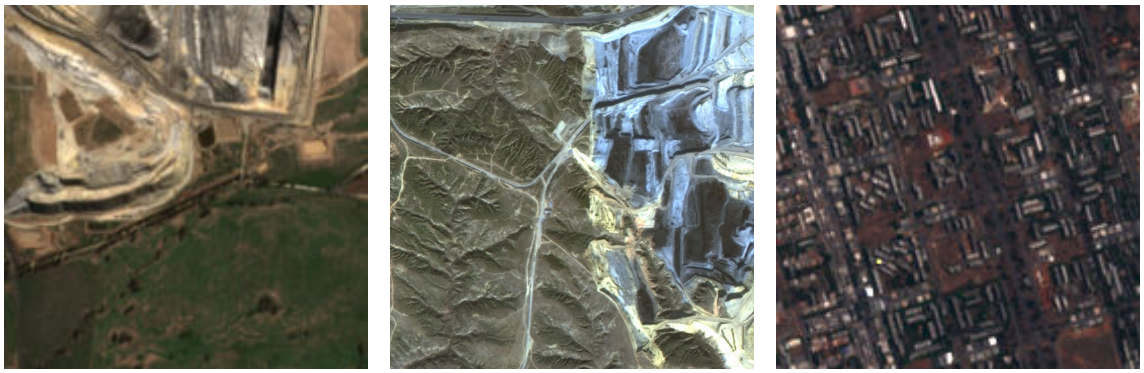


Figure G.1: Scale and feature comparison on image patches across three datasets. Left to right: OMS2CD, OMCD (Li et al. 2023) (©2023 IEEE), OSCD (Daudt et al. 2018) (modified under CC BY 4.0). Each image covers a surface area of 1.8 km$\times$1.8 km. In the OSCD and OMS2CD datasets, spatial resolution is 10 m. In OMCD, the spatial resolution is much higher at 2 m, which is demonstrated by the visibility of finer surface details. Note also that OSCD is focused on urban change, so the image tiles cover developed urban areas.

In our experimental analysis, it became evident that the dataset distributions' discrepan-

Model	Training Dataset	Val F1-score	Test F1-score
TinyCD	LEVIR-CD	0.00036	0.0
TinyCD	OSCD	0.14	0.16
TinyCD	OMCD	0.002	0.0
LSNet	CDD	0.0	0.0
DDPM-CD	LEVIR-CD	0.00021	0.00049

Table G.1: Summary of Model Performances

cies made it impractical to employ DLCD inference on Sentinel-2 tiles with models trained on conventional CD benchmark datasets. Table G.1 showcases the performance metrics of TinyCD, LSNet, and DDPM-CD as evaluated on OMS2CD, post their training on these benchmark datasets. The results, without further fine-tuning on OMS2CD data, are evidently not satisfactory. However, it's noteworthy that a fine-tuning process on OSCD yielded the most promising F1-scores, peaking at $0.14/0.16$. This might be attributed to OSCD being the sole Sentinel-2 dataset we explored.

Appendix H

OMS2CD Dataset Structure and Description

The dataset, designed for use in Deep Learning Change Detection (DLCD) models, is organized into a hierarchical structure. This appendix provides a comprehensive breakdown of the dataset's components and their respective locations.

H.1 Directory Structure

```
Dataset Root Directory
|-- GeoTIFF RGB: s2_<facility_name>_<coordinates>_<YYYY-MM-DD>.tif
|-- CSVs: mapping.csv, train.csv, val.csv, test.csv
|-- area_mask/ - Binary Area .tif Files
|-- mask/ - Change Label .tif Files
```

H.2 Detailed Description

To provide further detail on the structure of the dataset:

- **GeoTIFF RGB Images**: Located directly in the root directory, these images adhere to a specific naming convention:

`s2_<facility_name>_<coordinates>_<YYYY-MM-DD>.tif`.

Each image consists of three channels (RGB) with pixel values ranging from 0 to 255.

- **CSV Files**: The ‘mapping.csv‘ in the root directory details which images constitute each bi-temporal image pair. Additionally, ‘train.csv‘, ‘val.csv‘, and ‘test.csv‘ specify the facilities of the respective dataset splits.
- **Subdirectories**:
  - ‘area_mask‘: Houses single-channel GeoTIFF files representing binary area masks. These images possess pixel values in the range of [0-255]. Each area mask file follows the file naming convention of `<facility>_.tif`
  - ‘mask‘: Contains single-channel GeoTIFF files denoting change label masks, with pixel values also in the [0-255] range. Each mask file follows the naming convention of `<facility>_<####>.tif` where ‘####’ is a four-digit integer.

H.3 ‘mapping.csv’

Each row of the ‘mapping.csv’ file has four columns: ‘id’, ‘imageA’, ‘imageB’, and ‘mask’. For each row, the corresponding labelled file in the ‘mask/‘ directory can be found by combining that row’s facility name and that row’s ‘id’. The AOI mask can be found using the filename given in the ‘mask’ column. Please note that this file was created before the dataset was hand-labelled, which means that there are more rows than there are existing image-mask sets. Our provided Python utility class (discussed below) takes this into account by skipping row entries that contain a non-existent imageA, imageB, or mask file.

H.4 Python Utility Class

For streamlining interactions with this dataset and its preparation for DLCD models, a Python class named ‘OMS2CD‘ is provided in the project’s source code. Further details about this class can be found in Appendix I.

Appendix I

Source Code Repository

The source code which was developed for the purposes of carrying out the experiments described in this dissertation is available in an online repository[1]. Please refer to the README for information about specific files and the licensing of the code therein.

The OMS2CD and OMS2CDDataModule classes contained within the source repository provide an example for how to load and the provided OMS2CD dataset for training and inference. They utilise the `rasterio` and `tiler` Python libraries to load the dataset GeoTIFF files and break them into small patches which serve as model inputs. For extensive examples demonstrating how to use the OMS2CD classes, see the provided notebook files. For a simple case, see the code below. It creates an instance of the dataset class, with the root directory at 'OMS2CD/', uses the default patch size (256 pixels) and stride (128 pixels).

```
# Prepare the OMS2CD dataset for use
dataset = OMS2CD(root='OMS2CD', split='val', bands='rgb')
sample = dataset[0]

# Or, if using with PyTorch Lightning.
datamodule = OMS2CDDataModule(root='OMS2CD', bands='rgb',
    batch_size=16, tile_mode="constant", stride=100)
```

[1] https://github.com/Dibz15/OpenMineChangeDetection