Change Detection of Plant Populations in a Post-Mining Environment

Brook Queree
School of Information Technology and Electrical Engineering
The University of Queensland, Qld., 4072, Australia

Abstract

Plant growth uplift in post-mining environments is an integral part of post-operational mine site restoration. As such, we present an exploration of multiple change-detection methods applied to the problem of detecting and tracking change in plant growth using drone imagery from a coal mine in the Bowen Basin, Queensland, Australia.

The results show that leveraging finetuned foundational object detection models can significantly uplift change detection in scenarios where the objects of interest are of a much smaller scale than traditional change detection datasets. The results also indicate that existing hierarchical structures in Change Detection (CD) models show improved detection of small changes vs. non-hierarchical network architectures, and that further utilising these structures may have benefits for generalising CD models to the plant change detection problem.

1 Introduction

Environmental rehabilitation of mine sites post-operation is an important component of the mining lifecycle, where globally operators are legally obligated to either rehabilitate land directly or contribute to a fund dedicated to land rehabilitation upon closure of mining operations [1] [2].

Traditionally, monitoring of plant growth in rehabilitation scenarios has been performed with a combination of manual site visits [3] as well as in-depth human analyses of remote imagery [4]. Most commonly used is the Normalized Differenced Vegetation Index (NDVI) [4], which is a measure that infers the level of green vegetation growth through the strength of light band measurements.

The use of remote drone imaging and deep learning Change Detection (CD) models represent an opportunity to measure plant growth during rehabilitation at a higher resolution while requiring less human input when generating analyses.

CD models commonly leverage a siamese network architecture. This is an approach where input the images taken at two times T_1 and T_2 are fed through two identical shared-weight networks before being fed through a final decode network where the siamese outputs are used to predict change. The most common current CD dataset for comparison of results is LEVIR-CD [5], which encompasses change detection of building development in remote sensing images.

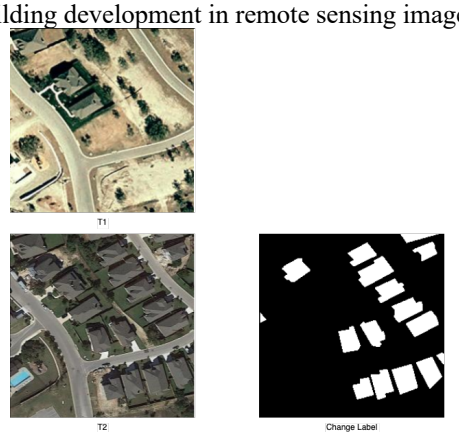

Figure 1: Example LEVIR-CD Bitemporal Image Pair and Change Label

Three of the most-promising CD neural network architectures were trained and evaluated on the full plant population change detection problem. These models were HANet [6], ChangeFormer [7] , and BIT [8].

HANet [6] (Hierarchical Attention Network) is a siamese Convolutional Neural Network (CNN) architecture designed for change detection encompassing a series of pooling and attention layers that serve to allow the network to detect change across a wide range of object scales inside images.

ChangeFormer [7] is a transformer-based siamese neural network architecture that also incorporates the concept of hierarchy and detection at multiple scales through a series of downsampling layers to achieve greater change detection performance.

BIT [8] (Bitemporal Image Transformer) leverages an initial CNN layer that is fed into a siamese tokenising layer which acts to arrange the pixels into groups reflecting a vocabulary of concepts (tokens). The tokens are processed by a transformer-based encoder layer that learns relationships between the input tokens and a decode layer that transforms the tokens back into the pixel space for final change prediction.

These direct CD model architectures were compared to a bounding box object-detection and segmentation approach leveraging the foundational models YOLOv11 [9] (You Only Look Once) and SAM [10] (Segment Anything Model).

YOLO is a CNN based model that performs object detection and classification to output predicted bounding boxes surrounding all instances of an object class within an image.

SAM is a task-specific foundation model created to efficiently generate image segments. It consists of a Vision Transformer [11] based image encoder, followed by a transformer mask prediction network layer.

2 Methodology

2.1 Data Preparation

An initial set of 9,635 drone images taken in 2021 and 8,501 drone images taken in 2023 were processed to produce a set of matching image pairs + associated change detection labels for training, testing, and evaluation of a collection of CD model architectures.

First, the images were imported to Agisoft Metashape, a 3D photogrammetric software suite, into projects of <1,000 images, grouped by geolocation and image capture-time. Following this, each project was built into a geographic model using Metashape's model-building functionality. These models were then best-effort aligned by placing alignment markers mapped to source of truth GPS coordinates. With the alignment markers placed, a corresponding orthomosaic was built from each project.

With the orthomosaic generated for each project, the next step was to export the orthomosaic sections as matched image pairs. To do this a script was created to both generate the image pair boundaries, as well as automate the UI actions (menu navigation and form completion). Using this script the total geographic area was split into 3,000 equal-size sections, and where each project had photos overlapping with any of the 3,000 sections, an orthomosaic section was exported as an image.

After exporting the matched sections in all projects, data cleansing was undertaken. Firstly, where the section overlapped the edge of an orthomosaic in such a way that <50% of the exported section contained drone imagery data (i.e. more than half of the image was blank), these images were removed from the final set. Secondly, where more than one project in the same time period (2021 or 2023) had produced an export for the same section, the section with the highest image quality was selected for de-duplication. From this process, a collection of 258 source images (129 matching image pairs) were produced.

After generating matching image pairs, associated change detection labels had to be produced. In order to generate comparable labels between the two change detection techniques (CD models and object detection bounding box inferences), the same set of plant bounding box locations used for object detection were used to generate the segment masks for the CD labels.

This has the effect of allowing both techniques to be trained on the same set of input data in the different required training formats. The complete labeling pipeline is as follows:

The source images were labeled with bounding box classifications using the Roboflow labelling tool. These boxes were placed surrounding every plant in the image, this allows for finetuning of the YOLOv11 bounding box object detection model

The coordinates of the plant bounding boxes were fed into SAM to generate the change-detection output image labels for training the CD models.

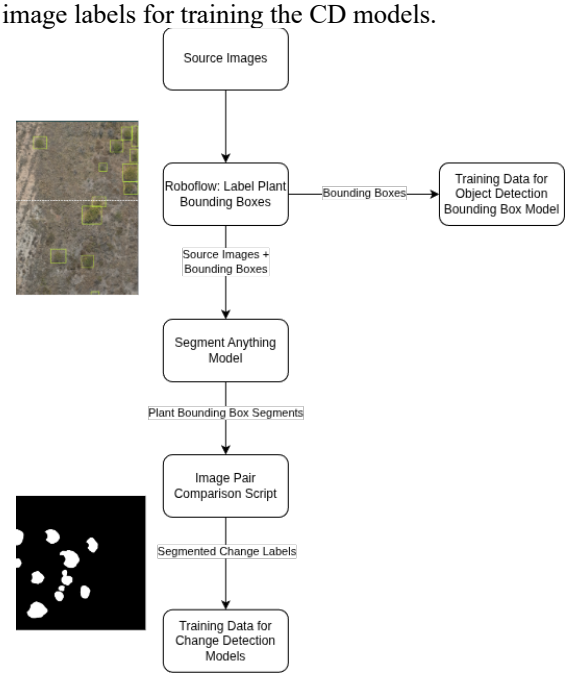

Figure 2: Data Labelling Pipeline

Two segmented change detection label sets were created:

- Positive labels: These represent image plant segments present in the 2023 set that were not present in the 2021 set.
- Negative labels: These represent image plant segments present in the 2021 set that are not present in the 2023 set.

To generate these sets, the plant segment masks from 2021 and 2023 are subtracted from each other (on a per-pixel basis) to remove plant segments that are present in both images.

To generate the final red-green change representations, the positive and negative labels were combined into a union image where the white pixels of the positive labels are represented as green pixels and the corresponding transformation applied to the negative labels becoming red pixels.

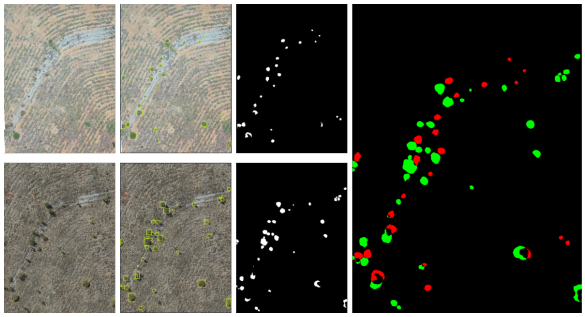

Figure 3: Progression of Bitemporal Image Pairs Through Data Labelling Steps (left to right) - Bitemporal Image Pair, Bounding Box Labeling, Segment Generation, and Red-Green Positive and Negative Change Label Generation

2.2 Candidate Model Architecture Selection

With the help of Open-CD [12], an open source change-detection toolbox based on the MMEngine library, a number of modern published CD models were explorable using a singular platform. These models were run on the same hardware with identical software dependency versions.

A number of candidate models were trained on the positive change labels to identify which change-detection model architectures best fit to the problem of plant change detection. Training on each model was run for 40,000 epochs on the training set. For each trained model, the epoch with the best Intersection over Union (IoU) result on the training set was selected:

TABLE I. PER-PIXEL PERFORMANCE OF CHANGE DETECTION ARCHITECTURES ON UNCHANGED / CHANGED PIXEL LABELS

Model Name	*F1 Score*	*Precision*	*Recall*	*IoU*
BIT [8]	98.8 / 54.1	98.9 / 53.6	98.8 / 54.6	97.7 / 37.1
CGNet [13]	98.9 / 50.6	98.6 / 60.0	99.3 / 43.7	97.9 / 33.9
ChangeFormer [7]	98.9 / 59.0	99.1 / 54.2	98.6 / 64.8	98.6 / 64.8
ChangerEx [14]	97.0 / 34.6	99.0 / 23.9	95.0 / 62.8	94.1 / 20.9
Changestar-farseg [15]	98.8 / 5.7	97.6 / 62.5	99.9 / 2.9	97.6 / 2.9
Changestar-upernet [15]	98.8 / 28.8	98.0 / 56.5	99.6 / 19.4	97.6 / 16.9
HANet [6]	99.0 / 57.5	98.8 / 64.8	99.3 / 51.6	98.1 / 40.3
STANet-BAM [5]	98.5 / 35.0	98.3 / 39.8	98.8 / 31.1	97.1 / 21.1

Here the unchanged labels represent the pixels for which either no plants existed at T_1 or T_2 (and hence no plant change was detected), or a plant existed at that pixel location at both T_1 and T_2. The unchanged labels represent the vast majority of the training data (visualised by the black background in the rightmost image of Figure 3). As such, the process of learning to detect the lack of change is the simpler half of this binary problem. We can see in Table 1. That Changestar-farseg has been unable to escape the local minima represented by the model simply predicting no change for all pixels and still correctly labelling the vast majority of pixels. This is reflected in only 2.9% of change pixels being correctly recalled by the highest-performing Changestar-farseg epoch.

The Three best-performing models (ChangeFormer, HANet, and BIT) were selected to perform full positive and negative training against to produce a two-model positive and negative plant change inference model.

2.3 Two-Model Inference Generation

For the purpose of generating both positive and negative change inferences, separate positive and negative CD model instances were trained for each architecture. The first CD model trained against the T_1 and T_2 images as input and the positive change labels as the output, and the second model trained against the reversed bitemporal images T_2 and T_1 as input and the negative change labels as output.

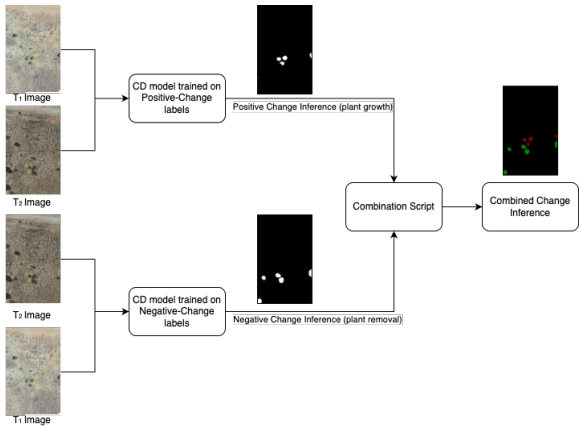

Figure 4: Two-Model Positive Negative Change Detection Inference Generation

2.4 Bounding Box Object Detection Model

During exploration of the plant change detection problem, an alternative approach was devised using bounding box object detection from a foundational model in combination with a secondary image segmentation step.

This approach substitutes training a CD model using the segmented images generated from bounding box labels for finetuning an object detector model (YOLOv11) on the bounding box labels directly. The YOLOv11 model was finetuned on the bounding boxes from the same set of training images that the CD models were trained on for a fair comparison result.

The output of the object detection model is then used as an input to SAM, which generates the corresponding segmented change labels.

Using this approach, the model is not directly learning to detect change. We transform the Change-Detection problem into two simpler subproblems of 1) object-detection and 2) segment generation and combination across the two images.

3 Change Detection Model Results

Both a positive and negative change-detection model were trained for the three selected architectures (BIT, ChangeFormer, and HANet). Using the two-model inference architecture described in 2.3.

TABLE II. TEST SET PER-PIXEL PERFORMANCE OF SELECT ARCHITECTURES ON POSITIVE/NEGATIVE CHANGE LABELS

	F1 Score	*Precision*	*Recall*	*IoU*
BIT				
Positive Change	23.4	23.2	23.6	13.3
Negative Change	13.4	45.4	7.9	7.2
All Change (combined)	19.7	26.4	15.8	11.0
ChangeFormer				
Positive Change	34.0	26.3	48.0	20.5
Negative Change	41.6	57.1	32.8	26.3
All Change (combined)	36.7	33.6	40.4	22.5
HANet				
Positive Change	37.7	32.3	45.4	23.3
Negative Change	38.7	53.11	30.5	24.0
All Change (combined)	38.1	38.3	38.0	23.5

Here we can see that the CD models with hierarchical structures (ChangeFormer and HANet) noticeably outperformed BIT, the model architecture without such a structure.

4 Bounding Box Object Detection Results

To produce the final red-green images, the inferred segments from YOLOv11 and SAM were combined using the same process as the combination of the positive and negative CD model inferences shown in Figure 4. We then evaluated the resulting images using the same per-pixel comparison process used in Table 2 for the CD models.

TABLE III. PER-PIXEL PERFORMANCE OF OBJECT-DETECTION INFERRED CHANGE LABELS

	F1 Score	*Precision*	*Recall*	*IoU*
Training Set				
Positive Change	74.2	64.6	87.1	59.0
Negative Change	70.3	60.9	83.2	54.2
All Change (combined)	72.5	62.9	85.4	56.8
Test Set				
Positive Change	66.0	58.1	76.3	49.3
Negative Change	64.0	64.7	63.3	47.0
All Change (combined)	65.0	60.9	69.8	48.2

These results represent a significant improvement over the equivalent test set results in Table 2, which indicates the YOLOv11 foundational model has a much greater ability to adapt to small-scale change.

5 Discussion

5.1 Study Limitations

One of the main drawbacks in the input data was low resolution and GPS accuracy in the 2021 dataset. As a result, the change-detection models are potentially hampered by both the lack of clarity in the T_1 period images, as well as the potential for the slight geographical drift limiting the ability of the model to detect true change vs. GPS shift.

Additionally, due to hardware and time constraints, not every initial change-detection model architecture was fully optimised from a configuration and hyperparameter standpoint. It is likely that the model performance for architectures that did not produce sufficient results in Table 1 could still be improved somewhat with optimisation of these parameters and further training.

As visualised in Figure 5 below, the CD models struggle to detect change of small scale within the entire image. As such, higher resolution T_1 images would better allow for exporting a smaller physical area per bitemporal pair while maintaining clarity, making the plants (and hence the change) of a larger scale within the image. With these larger-scale plant images, it is possible that the CD models would perform significantly better due to its closer resemblance to standard reference CD datasets.

5.2 Results

From the exploration of multiple change detection architectures during initial training, it appears the problem of plant change detection greatly favours architectures that can encode a concept of hierarchy (HANet and ChangeFormer), in particular models that focus on detection of change and learning at multiple object scales through the use of pooling and downsampling layers.

From training a number of modern change-detection model architectures in direct comparison with bounding-box object detection in combination with post-detection segmentation, it was seen that the bounding-box approach produces significantly higher F1 Score and IoU results when applied to change-detection of plants in post-mining environment rehabilitation. As such when combined with a higher quality dataset this approach has a high feasibility of allowing for accurate change detection in real-world post mining environments.

Acknowledgment

We would like to express appreciation for the UQ Sustainable Minerals Institute for providing access to

the source data upon which this research was conducted and thank the team Ben Seligmann, Nikodem Rybak, Peter Erskine and Steven Micklethwaite for sharing their industry subject matter expertise. Special thanks to Alina Bialkowski for her computer vision and machine learning expertise and guidance.

References

[1] J. Skousen and C. E. Zipper, "Post-mining policies and practices in the Eastern USA coal region," *International Journal of Coal Science & Technology,* vol. 1, pp. 135-151, 2014.

[2] Government of Western Australia, *Mining Rehabilitation Fund Act,* Perth, Western Australia, 2012.

[3] S. F. Gould, "Comparison of Post-mining Rehabilitation withReference Ecosystems in Monsoonal EucalyptWoodlands, Northern Australia," *Restoration Ecology,* vol. 20, no. 2, pp. 250-259, 2012.

[4] P. B. McKenna, A. M. Lechner, S. Phinn and P. D. Erskine, "Remote Sensing of Mine Site Rehabilitation for Ecological Outcomes: A Global Systematic Review," *Remote Sensing,* vol. 12, no. 21, p. 3535, 2020.

[5] H. Chen and Z. Shi, "A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection," *Remote Sensing,* vol. 12, no. 10, p. 1662, 2020.

[6] C. Han, C. Wu, H. Guo, M. Hu and H. Chen, "HANet: A Hierarchical Attention Network for Change Detection With Bitemporal Very-High-Resolution Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 16, pp. 3867-3878, 2023.

[7] W. G. C. Bandara and V. M. Patel, "A Transformer-Based Siamese Network for Change Detection," in *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*, Kuala Lumpur, 2022.

[8] H. Chen, Z. Q. and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 60, pp. 1-14, 2021.

[9] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," *arXiv preprint arXiv:2410.17725,* 2024.

[10] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar and R. Girshick, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929,* 2020.

[12] K. Li, J. Jiang, A. Codegoni, C. Han, Y. Deng, K. Chen, Z. Zheng, H. Chen, Z. Liu, Y. Gu, Z. Zou, Z. Shi, S. Fang, D. Meng, Z. Wang and X. Cao, "Open-CD: A Comprehensive Toolbox for Change Detection," *arXiv preprint arXiv:2407.15317,* 2024.

[13] C. Han, C. Wu, H. Guo, M. Hu, J. Li and H. Chen, "Change Guiding Network: Incorporating Change Prior to Guide Change Detection in Remote Sensing Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 16, pp. 8395-8407, 2023.

[14] S. Fang, K. Li and Z. Li, "Changer: Feature Interaction Is What You Need for Change Detection," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 61, pp. 1-11, 2023.

[15] Z. Zheng, A. Ma, L. Zhang and Y. Zhong, "Change is everywhere: Single-temporal supervised object change detection in remote sensing imagery," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15193-15202.

Biography

Brook Queree is a part time Master of Computer Science student at UQ working as a software engineer in the energy industry. Brook completed his Bachelor of Engineering (Hons) in Computer Engineering at the University of Canterbury in New Zealand.

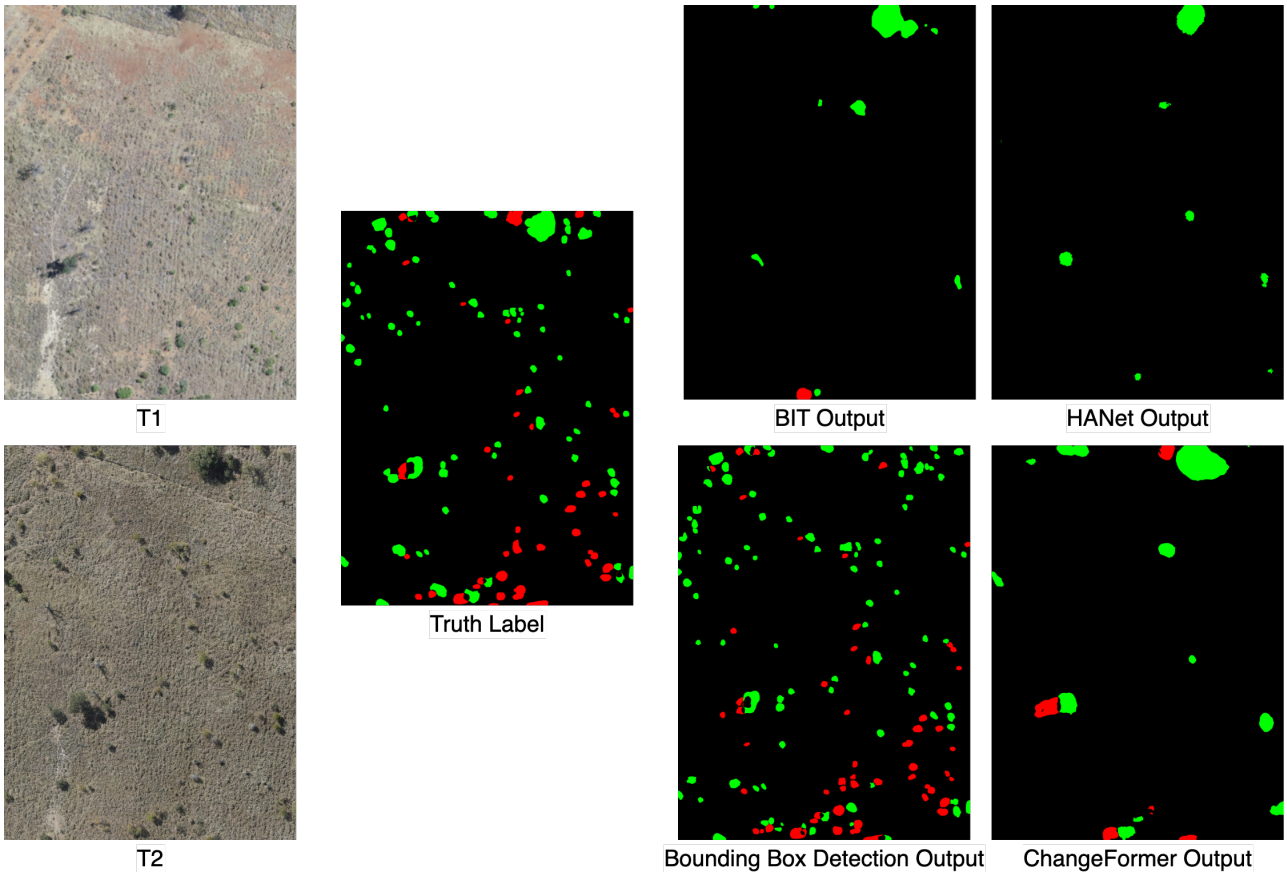

Figure 5: Test Set Image Pair Example Model Outputs