Hybrid Foundation Model for Change Detection in Remote Sensing Data

1st Dheeraj NVS
Department of CISE
University of Florida
Gainesville, Florida, USA
vnaganaboina@ufl.edu

2nd Subhash Vadlamani
Department of CISE
University of Florida
Gainesville, Florida, USA
v.vadlamani@ufl.edu

***Abstract*—Change detection in remote sensing imagery plays an important role in monitoring dynamic environments. This paper proposes a novel approach that uses a UNET like architecture to build a Siamese Segmentation Network to capture changes in Google Earth images using the LEVIR-CD dataset. Our primary objectives are to assess the foundational potential of this model, evaluate its computational feasibility, and determine its ability to generate change maps from pre-change and post-change images.**

Building upon traditional methods like Multivariate Alteration Detection (MAD) and iterative MAD, as well as Trilateral change detection networks, our approach focuses on the creation of a foundational model. Unlike previous methods that address specific use cases, our model is designed to serve as a versatile foundation for diverse applications in change detection.

I. Introduction

Change detection in satellite imagery, particularly in the context of Remote sensing images, is an activity with far-reaching societal implications. The ability to discern alterations in geographical features over time has significant applications in urban planning, environmental monitoring, disaster management, and beyond. As our world undergoes dynamic transformations, the need for advanced and adaptable models for change detection becomes increasingly evident.

Previous methodologies in this domain, including Multivariate Alteration Detection (MAD) [1], iterative MAD, and Trilateral Change Detection Networks, have made significant strides in capturing temporal shifts within images. However, a critical gap remains in the lack of a foundational model that serves as a versatile framework, capable of evolving to meet specific use-case requirements. The existing paradigms, while effective for specific tasks, often fall short when confronted with the demand for a unified approach that can serve as a base for various other applications.

MAD and its iterative variants operate on the principle of statistical analysis, attempting to identify changes by comparing multi-temporal pixel values. While successful in certain scenarios, these methods often struggle with adaptability, hindering their scalability across diverse applications. Trilateral Change Detection Networks, as explored in the paper titled ”TCDNet: Trilateral Change Detection Network for Google Earth Image,” [2] introduce a neural network-based approach, showcasing the potential of deep learning in addressing the intricacies of change detection. However, even these advancements lack a broader foundational perspective.

In response to these challenges, this paper proposes an innovative approach that combines Siamese neural networks, Convolutional Neural Networks (CNNs), and a UNET-based segmentation architecture. By integrating these components, we aim to craft a foundational model that not only captures fine changes in Google Earth images but also lays the groundwork for a more adaptive and robust change detection framework. Our method seeks to go beyond the confines of specific applications, envisioning a model that can be fine-tuned for diverse use cases.

The computational feasibility of our approach is a paramount consideration, recognizing the practical challenges associated with processing vast amounts of satellite imagery data. As we delve into the intricacies of our proposed methodology, we concurrently explore the broader societal impacts of accurate change detection. Urban development, environmental conservation, disaster response, and other domains stand to benefit significantly from a foundational model that offers both accuracy and adaptability.

In the subsequent sections, we will delve into the specifics of our proposed approach, combining it with established methodologies, and presenting empirical evidence from the analysis of the LEVIR-CD dataset [3]. Through this exploration, we aim to determine the potential of our model to serve as a cornerstone in the field of change detection, addressing the dual challenges of computational feasibility and applicability across diverse scenarios.

II. Problem Definition

A. Problem Overview

The fundamental challenge addressed in this paper revolves around the development of a hybrid foundation model leveraging Convolutional Neural Networks (CNNs) and Siamese architecture [4] to effectively capture temporal dynamics and changes within Earth images. The primary objective is to create a versatile model that can serve as a foundational framework for diverse applications requiring the observation and analysis of changes in data.

B. Conceptual Framework

The model is conceptualized as a fusion of segmentation and Siamese architecture, capitalizing on their unique strengths. The CNN component of the Siamese network is designed to comprehend the spatial relationships within Earth images, allowing the model to discern patterns, structures, and changes in different geographical features. Simultaneously, the upsampling part is integrated to capture the temporal aspects of the image data, enabling the model to analyze the progression of changes over time.

C. Input

The input to the model consists of pairs of pre-change and post-change Earth images, where each image is represented as a matrix of pixel values. The pre-change and post-change images are processed parallelly through the encoder and the absolute difference of features from each of the layers of the encoder is processed through the decoder of the segmentation architectures.

D. Output

The output of the model is a change map that highlights the areas where alterations have occurred between the pre-change and post-change images. The change map provides a visual representation of the detected changes, aiding in the interpretation and analysis of temporal dynamics within the Earth images.

E. Objective

The overarching goal is to develop a hybrid foundation model that incorporates the Siamese segmentation architecture along with Convolutional LSTM layers capable of addressing the temporal nuances of Earth image data, thereby facilitating applications such as wildfire spread estimation, flood impact assessment, urban development monitoring, and forest cover analysis. Additionally, the model is designed to be fine-tuned for specific use cases with minimal data, ensuring adaptability and efficiency in scenarios with limited labeled information.

F. Example

Consider a scenario where the input comprises satellite images of a forested region before and after a significant weather event. The hybrid foundation model processes these images, identifying changes indicative of altered forest cover. The output, in the form of a change map, precisely delineates areas where vegetation has been affected. This map serves as a valuable tool for environmental monitoring, aiding authorities in assessing the impact of the weather event on the forested area.

III. Proposed Solution

A. Model Overview

The continual evolution of Earth's landscapes necessitates robust methods for monitoring and understanding changes, prompting advancements in satellite imagery analysis. In this paper, we propose a solution that leverages the strengths of various architectures. The SiameseSegmentationNet is a specialized neural network tailored for change detection in geospatial imagery. Utilizing a Siamese architecture [5] and inspired by UNET architecture [6], it processes pairs of pre- and post-event images, focusing on identifying and delineating changes with high precision. The model's capability to analyze bi-temporal images is essential for applications in environmental monitoring, urban development, and disaster assessment.

B. Architecture Design

This model is structured into two main components: the encoder and the decoder. The encoder, through its convolutional layers, efficiently extracts and compresses features from the input images. Batch normalization and dropout layers are incorporated to enhance the model's generalization abilities. In the decoder, upsampling layers meticulously reconstruct the spatial dimensions, allowing for precise localization of changes. The design is optimized to handle the complexities and nuances of geospatial data. The pre - and post - change images are processed parallelly through the encoder architecture and the decoder architecture uses upsampling techniques to generate a change map using the absolute difference between the features generated by the layers in the encoder architecture. Going into the details of implementation, the encoder block consists of 10 convolutional layers, where each layer is followed by a batch normalization layer, ReLu activation function and a dropout layer for regularization. The decoder block consists of 14 deconvolutional layers each followed by a batch normalization layer and a dropout layer except for the last deconvolutional layer. ADAM optimizer with a learning rate of 1e-3 is used for training and Binary Cross Entropy is used as the loss function.

C. Innovation and Novelty

The SiameseSegmentationNet represents a significant innovation in leveraging Siamese networks for change detection in satellite imagery. Its novelty lies in the simultaneous processing of pre- and post-event images, enabling a more nuanced and detailed understanding of changes. This approach is particularly groundbreaking for geospatial applications, as it allows for the precise detection of changes in landscapes, urban areas, and natural environments. The architecture's ability to handle complex spatial features and temporal dynamics offers a new perspective in remote sensing and change detection, setting a precedent for future research and applications in this field.

D. Major Steps:

- **Architecture Design:** Implementation of the SiameseSegmentationNet, a novel Siamese network architecture tailored for change detection in geospatial imagery. This involves constructing convolutional layers and upsampling mechanisms to process and compare pre- and post-event satellite images.
- **Data Preparation:** Utilization of the LEVIR-CD dataset, which comprises pre- and post-change satellite images, requiring preprocessing for optimal network training.

Breaking each image in LEVIR-CD Datset of size 1024*1024 into patches of size 256*256.

- **Training and Validation Process:** Employing a TPU-based environment for efficient training, the model undergoes rigorous training and validation phases, leveraging PyTorch's XLA (Accelerated Linear Algebra) for parallel processing across TPU cores. This significantly speeds up the training process and enhances the model's performance.
- **Regularization and Loss Function:** Introduction of L1 regularization in the training loop, alongside the Binary Cross-Entropy with Logits Loss function, to mitigate overfitting and improve the model's generalization capabilities.
- **Model Evaluation and Visualization:** Post-training, the model's performance is evaluated using various metrics, and outputs are visualized to assess the effectiveness of the change detection. This step is crucial for understanding the model's practical utility in real-world scenarios.

.

E. Dataset

The LEVIR-CD dataset [3], pivotal to this project, is a specialized collection of high-resolution aerial imagery designed specifically for the purpose of change detection in geospatial analysis. It consists of pairs of satellite images, each capturing the same geographical area before and after an event, such as urban development or natural disasters. These images are of significant resolution, each being 1024x1024 pixels, and are captured in RGB format, providing detailed color information critical for accurate change detection.

Each image pair in the dataset is accompanied by a binary label image of the same resolution. These labels are meticulously annotated to indicate change at the pixel level, with '1' representing a change and '0' signifying no change. This binary labeling is crucial for training the model to differentiate between changed and unchanged areas effectively.

In preparation for training, the dataset underwent a crucial preprocessing step. Each high-resolution image, along with its corresponding label, was segmented into smaller, more manageable patches of 256x256 pixels. Each image in the dataset is broken down into 16 patches. This process not only made the data compatible with the network's input size requirements but also allowed for more efficient processing and a greater focus on localized changes. The segmentation of images into patches also facilitated the handling of large-scale data, optimizing the training process for better performance and accuracy.

The final dataset contains 7280 train images, 1024 validation images, and 2048 test images. This comprehensive approach to dataset preparation underscores the meticulous attention to detail and the focus on precision, which are paramount in the field of geospatial change detection.

F. Experiments

- **VAE architecture [7] to pre-train the CNN block of the Siamese network:** We built a VAE architecture using CNN layers and trained it on Sentinel 2 Surface Reflectance Images. We have chosen few regions of interest that cover different types of landscapes and retrieved the scenes from the Sentinel 2 dataset using Google Earth Engine API. Each scene is broken down into 1024*1024 patches and these patches are used to train the VAE architecture. This model was unable to generalize the remote sensing images so we had to drop this idea.
- **Siamese Segmentation architecture with Bi-directional LSTM layer:** The segmentation architecture is inspired from the UNET model which has encoder and decoder blocks. For the encoder that captures spatial features from the images, we used a pre-trained ResNet18 [8] model without the fully connected layers and concatenated the output for pre-change and post-change images to process through a bi-directional LSTM layer [9]. The images were flattened before passing through the LSTM layers and this caused the images to lose their spatial representation and the final output was jitter.
- **Siamese Segmentation architecture with CNN layers:** This time we removed the LSTM layer from our previous architecture and added deconvolutional layers for upsampling and change map reconstruction. While training this model, it is found the the model is overfitting to the training data as the ResNet architecture is very deep and also we were unable to access feature outputs of intermediate layers to use in the upsampling part of the decoder that reconstructs the change map. So, we finally built the architecture that is presented in this paper which has custom CNN layers in both encoder and decoder blocks.

IV. Evaluation

In the quest to develop a robust change detection model using the SiameseSegmentationNet architecture, our journey involved numerous experimental iterations. This section delves into the evaluation of these models, where we first present the outcomes of our initial trials, highlighting the challenges and learning points. These early experiments, though not meeting our final objectives, were instrumental in guiding our approach towards the successful model.

Subsequently, we showcase the results of our final model, which emerged from these iterative refinements. Here, we detail its performance metrics, visual outputs, and the effectiveness in accurately detecting changes in geospatial imagery. This comparative analysis not only illustrates the evolution of our model but also underscores the significance of each experimental phase in achieving our final successful implementation.

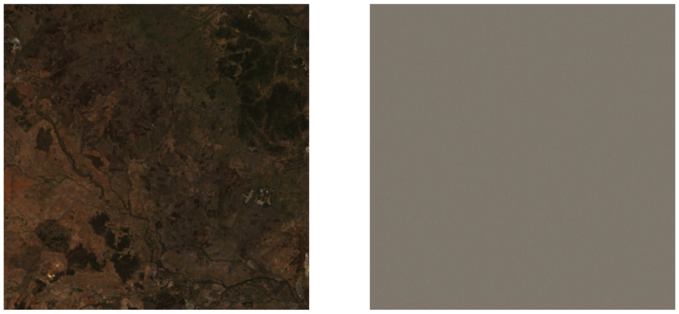

Fig. 1. VAE implementation output

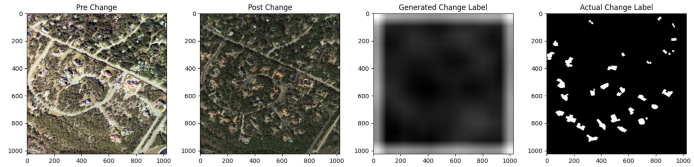

Fig. 2. SiameseLSTM Output

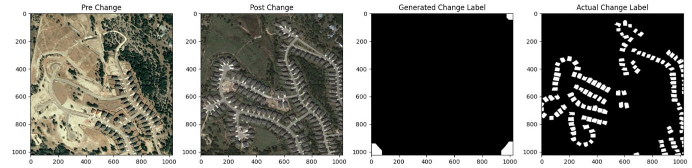

Fig. 3. SiameseCNN Output

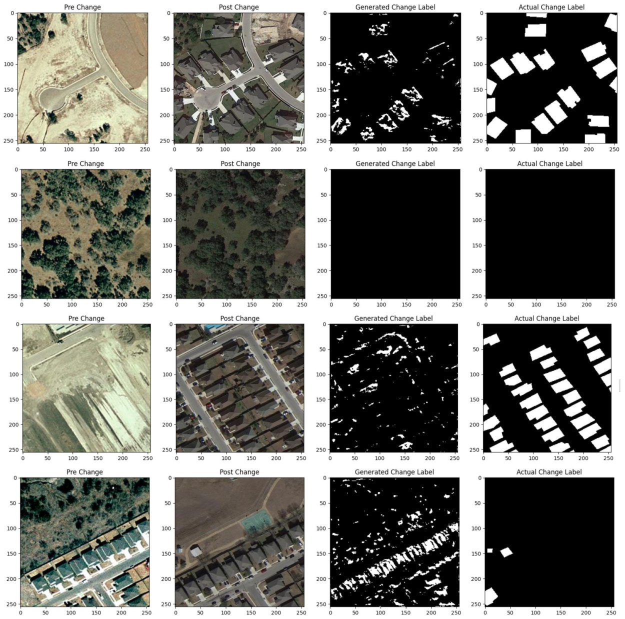

Fig. 4. SiameseSegmentationNet with regularization Outputs

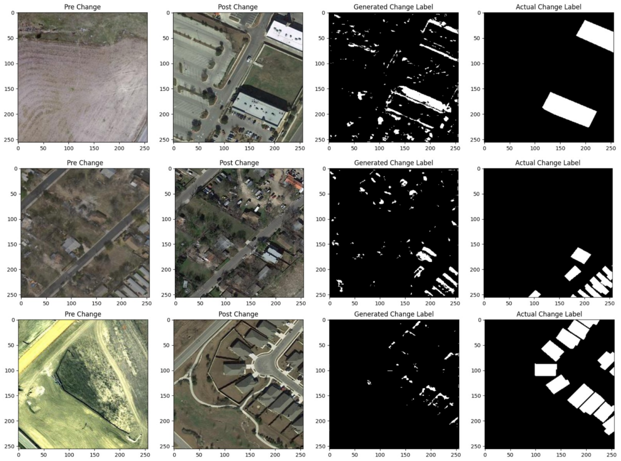

Fig. 5. SiameseSegmentationNet with regularization Outputs

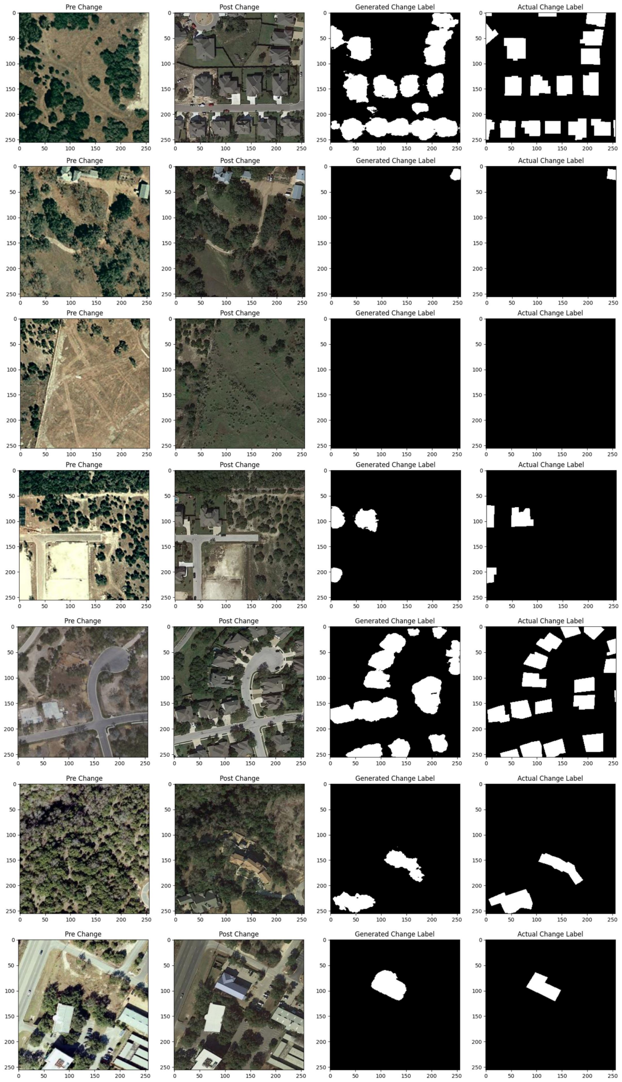

Fig. 6. SiameseSegmentationNet without regularization Output

The following are some of the metrics calculated on our model with regularization:

- **Precision:** 0.3129
- **Recall:** 0.2189
- **F1 score:** 0.2361
- **IoU or Jaccard score:** 0.1386

The following are some of the metrics calculated on our model without regularization:

- **Precision:** 0.6102
- **Recall:** 0.8611
- **F1 score:** 0.7114
- **IoU or Jaccard score:** 0.5579

V. Conclusion

In our study, the SiameseSegmentationNet model showcases potential as a foundational model for pixel-level change detection in geospatial imagery. While current results are moderate, they highlight the model's nuanced capabilities in identifying subtle changes. The regularized model's limitations, primarily in consistency and precision, suggest a significant scope for enhancement. The model without regularization has showcased much better results but is confined to structural changes whereas the model with regularization was able to detect minute pixel-level changes that are not even present in the change map. We advocate for further development, particularly through integrating Convolutional LSTM layers and data augmentation techniques, to refine its performance. The regularized model's inherent ability to detect changes beyond structural differences, though imperfect, is a promising aspect. With strategic fine-tuning, we believe this model could evolve into a robust tool, offering deeper insights into geospatial data analysis. This research paves the way for future explorations, aiming to harness and elevate the model's capabilities for more accurate and comprehensive change detection.

References

[1] A. Tahraoui, R. Kheddam, A. Bouakache, and A. Belhadj-Aissa, "Land change detection using multivariate alteration detection and chi squared test thresholding," in *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, 2018.

[2] J. Qian, M. Xia, Y. Zhang, J. Liu, and Y. Xu, "Tcdnet: Trilateral change detection network for google earth image," *Remote Sensing*, vol. 12, p. 2669, 08 2020.

[3] 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, 2020.

[4] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Advances in neural information processing systems*, vol. 6, 1993.

[5] A. Nandy, S. Haldar, S. Banerjee, and S. Mitra, "A survey on applications of siamese neural networks in computer vision," in *2020 International Conference for Emerging Technology (INCET)*, pp. 1–5, 2020.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[9] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.