Remote Sensing Image Change Detection with Transformers

Internship Report

Ayesha Siddiqa	407198
Navaal Iqbal	409977
Amna Ahmed	408099

https://github.com/amnaahmed567/Remote-Sensing-Image-Change-Detection-with-Transformers

Contents

Remote Sensing Image Change Detection with Transformers:

This project focuses on detecting changes in remote sensing images using transformer-based models. Remote sensing involves the acquisition of information about an area or object without direct contact, typically via satellite or aerial imagery. Change detection in such imagery is crucial for various applications such as urban expansion monitoring, deforestation tracking, disaster management, and environmental change analysis.

In this project, deep learning techniques, specifically transformer models, are employed to analyze and detect changes between pairs of satellite images taken at different times. Transformer architectures, which have proven effective in natural language processing, are adapted for this task to capture spatial and temporal changes in the images.

The project is implemented using Python and leverages libraries such as TensorFlow, PyTorch, and the Earth Engine API to preprocess satellite imagery and train transformer models for accurate change detection.

PART I

Change Detection in Amazon Rainforest Using Machine Learning and Deep Learning Techniques

Introduction:

Deforestation is a critical environmental concern that impacts biodiversity, climate, and the overall health of our planet. Monitoring deforestation effectively requires advanced techniques to detect changes in land cover over time. This project focused on detecting deforestation using satellite imagery and applying machine learning (ML) and deep learning (DL) techniques to identify changes in the Amazon rainforest. The Amazon, being the largest tropical rainforest, plays a crucial role in global carbon cycles and biodiversity preservation. The primary objective of this project was to monitor regions within the Amazon where forests were converted to other land cover types over time, specifically using a pixel-wise classification approach to detect changes from forest to vegetation, urban areas, and water bodies.

Data Collection and Preparation:

The study utilized satellite imagery from Google Earth Engine, covering the Amazon region from 2015 to 2021. The images were accompanied by corresponding land cover data, which allowed us to create a comprehensive dataset for training and testing our models.

We generated our training dataset through a pixel-by-pixel comparison of satellite images (2015-2020) and their corresponding land covers. Each entry in the dataset represented a pixel with its associated 5-band values and a final class label, which could be 'forest,' 'vegetation,' 'urban,' or 'water bodies.' This method resulted in a dataset containing approximately 21.6 million entries. The primary goal was to detect regions where forests were converted into other land cover classes.

Given the significant size of the dataset, one of the challenges faced was dealing with imbalanced class representation. Classes such as 'urban' and 'water bodies' were underrepresented, which impacted the performance of our models.

Model Training and Challenges:

Initially, the models did not achieve satisfactory accuracy levels due to the class imbalance in the Amazon dataset. To mitigate this issue, we implemented various techniques, such as:

- **SMOTE Analysis (Synthetic Minority Over-sampling Technique): To** generate synthetic samples and balance the class distribution.
- **Stratified Shuffling:** To ensure that each batch of data used in training maintained the original class distribution.
- **Random Sampling**: To reduce the dataset size while maintaining representative samples.
- **Hyperparameter Tuning:** To find the optimal parameters for each model, enhancing performance.

The models utilized for training included:

1. Random Forest Classifier (achieved the best accuracy)
2. XGBoost
3. Support Vector Classifier (SVC)
4. Gradient Descent-based models
5. Decision Tree Classifier
6. Convolutional Neural Networks (CNNs)

Experimental Results:

We conducted experiments on two smaller Regions of Interest (ROI) to evaluate the performance of our models. The results for each ROI at different resolutions and dataset sizes are summarized below:

1. **ROI 1 (Predominantly Vegetation Area):**

- **1000m Resolution:**
  - Trained on 200,000 entries: 41.9% accuracy
  - Trained on 20,000 entries: 84.79% accuracy

- **100m Resolution:**
  - Trained on 200,000 entries: 80.9% accuracy

2. **ROI 2 (Mixed Vegetation and Forest Area):**

- **1000m Resolution:**
  - Trained on 200,000 entries: 87.6% accuracy
  - Trained on 20,000 entries: 81.63% accuracy

- **100m Resolution**
  - Trained on 200,000 entries: 81.5% accuracy

The experiments indicated that training the models on smaller subsets of the data (e.g., 20,000 entries) often yielded better results compared to larger datasets (200,000 entries or the entire 21.6 million entries). This outcome is likely due to the more focused data representation and improved balance in smaller samples.

Model Adaptation and Change in Region of Interest:

Based on the insights gained from the Amazon region experiments, we decided to change our Region of Interest to an area where there was a more balanced representation of all classes. We selected the Islamabad region for this purpose, where the presence of different land cover types, such as forests, urban areas, vegetation, and water bodies, was more comparable.

Upon training the model on the new dataset for the Islamabad region, we achieved an accuracy of 76%.

PART II

Introduction to Transformer Architectures in Change Detection

Initially, our machine learning models performed change detection manually by predicting a specific class value for each pixel in the test images. To enhance this process, we transitioned to using transformer architectures that could automatically generate change maps, providing a more comprehensive understanding of changes over time.

Following our initial experiments with traditional machine learning models, we explored transformer architectures, which have recently gained prominence for their efficacy in various computer vision tasks, including change detection.

We began our exploration with the Vision Transformer (ViT), which applies the self-attention mechanism to image patches, enabling the model to effectively capture global patterns and dependencies. However, we discovered that several specialized transformer models are specifically designed and optimized for the task of change detection, offering more tailored performance for this application.

Survey and Implementation of Tailored Transformer Models

We studied survey papers on state-of-the-art transformer models tailored for change detection, including:

- **SwinSUNet:** A transformer-based model that integrates a U-Net-like architecture with Swin Transformers for enhanced multi-scale feature extraction.
- **STANet (Spatial-Temporal Attention Network):** Designed to capture spatial and temporal changes between bitemporal images effectively.
- **BIT (Bi-temporal Image Transformer):** A model specifically focused on capturing differences between two temporal images for change detection tasks.
- **ChangeFormer:** A model that leverages transformers to model change patterns across different temporal datasets.
- **ScratchFormer:** A transformer model designed for robust change detection with scratch learning.

After understanding the architecture and strengths of each model, we chose to experiment with BIT and ScratchFormer on publicly available datasets, namely LEVIR-CD (Large-scale Earth Vision Intelligence Research) and WHU-CD (Wuhan University Change Detection).

Implementation and Results

Both the BIT and ScratchFormer models were implemented for the task of change detection using the LEVIR-CD and WHU-CD datasets. These datasets are widely recognized in the remote sensing community for evaluating change detection algorithms due to their large-scale and diverse scenarios, including urban expansion, construction changes, and vegetation shifts.

Our implementation process involved:

1. **Data Preprocessing:** Preparing the datasets for input into the transformer models by aligning and normalizing bitemporal images.
2. **Model Fine-Tuning:** Adjusting hyperparameters such as learning rate, batch size, and the number of training epochs to achieve optimal performance.
3. **Training**: Running the models on google colab.
4. **Evaluation:** Assessing model performance using standard metrics such as overall accuracy, precision, recall, and F1-score.

Both models, BIT and ScratchFormer, demonstrated excellent performance, achieving an accuracy of over 90% on both datasets. This high accuracy indicates the effectiveness of transformer-based models in detecting changes over time in remote sensing images.

Overview of BIT_CD Transformer

BIT_CD stands for **Bitemporal Image Transformer for Change Detection**, and its architecture leverages the strengths of transformers to process pairs of images taken at different times and detect meaningful changes between them.

Key Features of BIT_CD Transformer:

- **Bitemporal Analysis**: The model processes two images taken at different times (bitemporal images) to detect changes. It uses attention mechanisms to focus on regions where changes are most likely to occur, such as deforestation, urban expansion, or environmental shifts.

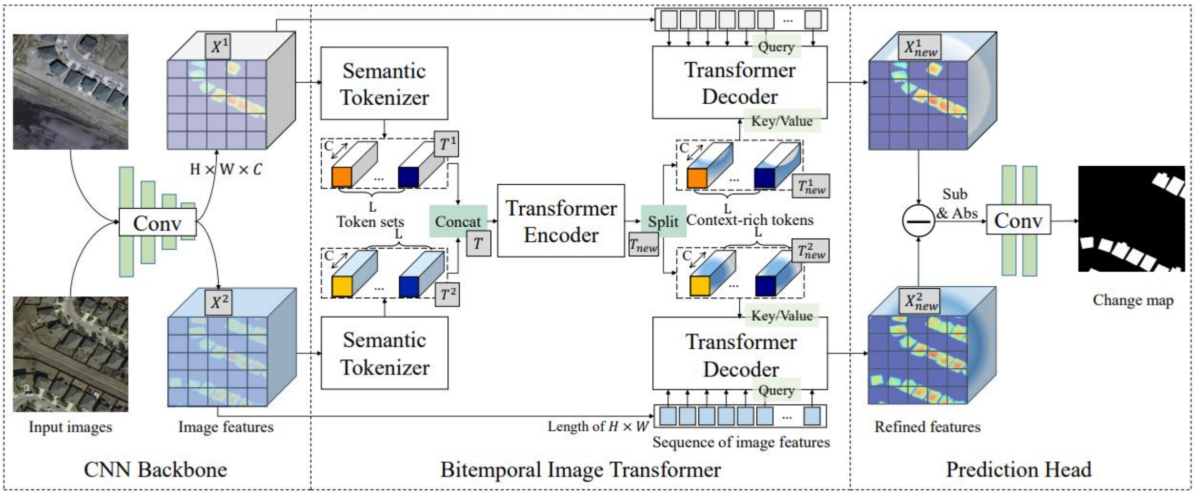
- **Self-Attention Mechanism**: One of the core strengths of the BIT_CD Transformer is its ability to use the self-attention mechanism to learn dependencies and relationships between pixels in the images, both spatially and temporally. This allows the model to capture long-range context and subtle differences between the two images.

- **Adaptability to Remote Sensing Data**: The BIT_CD Transformer is tailored to handle the unique characteristics of remote sensing data, such as multispectral bands, varying resolutions, and large geographic coverage.

- **Improved Change Detection Performance**: By leveraging the attention mechanisms of transformers, BIT_CD achieves higher accuracy in identifying changes over traditional machine learning and CNN-based models. This is because it can better capture both global and local changes in land cover.

Overall, the BIT_CD Transformer offers a modern, efficient, and highly accurate approach to remote sensing change detection, pushing the boundaries beyond what traditional models can achieve.

Architecture of BIT_CD Transformer:



Overview of ScratchFormer

The **ScratchFormer** is a transformer-based model designed from scratch for remote sensing change detection tasks, similar to other transformer architectures but built specifically with a focus on simplicity and adaptability. The aim of ScratchFormer is to leverage the power of transformers while reducing model complexity and training it from the ground up, without relying on pre-trained models or sophisticated initialization techniques.

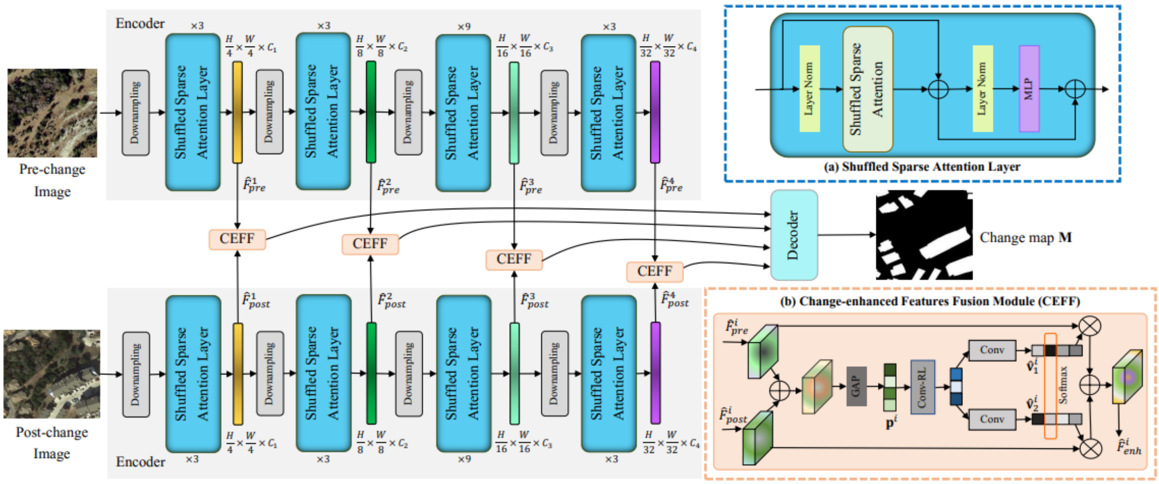
Key Features of ScratchFormer:

- **Designed from Scratch**: Unlike many transformer models that build upon pre-trained networks (often from natural language processing tasks), ScratchFormer is trained from scratch for change detection, allowing it to learn directly from remote sensing data.
- **Basic Transformer Architecture**: It follows the standard transformer design, consisting of encoder layers with self-attention mechanisms, but it is optimized for satellite imagery. This simplicity in design ensures that it is highly adaptable to various remote sensing applications and can be customized as needed.
- **Bitemporal Image Input**: Similar to other change detection models, ScratchFormer takes bitemporal satellite images (two images from different times) as input. It then processes these images to detect meaningful changes in land cover or environmental conditions over time.
- **Self-Attention for Spatial and Temporal Changes**: ScratchFormer uses the self-attention mechanism to capture dependencies between pixels in both spatial (within an image) and

temporal (across images from different times) dimensions. This helps in identifying subtle changes and maintaining a global context over the scene.

In summary, ScratchFormer offers a simple yet powerful approach to change detection in remote sensing, providing a balance between model effectiveness and flexibility, especially for custom applications or novel datasets.

Architecture of Scratch Former:



Initial Attempts and Challenges

After achieving promising results with publicly available datasets, we moved on to create our own dataset tailored for detecting urban expansion. Initially, we aimed to replicate the high-resolution satellite imagery used in datasets like LEVIR-CD (0.5m resolution), but due to limited access to such high-resolution images, we focused on regions that experienced the most significant urban expansion between 2020 and 2021. We selected diverse areas across multiple countries, including India, China, Japan, Korea, Bangladesh, the Netherlands, and the UK.
When we tested the transformer models (BIT and ScratchFormer) on this dataset, we encountered significant challenges. The accuracy was low, around **45%**, and the models failed to detect any changes between the pre-change and post-change images. After further analysis, we realized that the primary issue was the short time frame between the images, resulting in minimal visible differences. This lack of detectable changes rendered the models ineffective.

Creation of an Enhanced Dataset

To overcome these challenges, we developed a new dataset with a longer time frame and a more targeted selection of regions. We increased the timeline from **2000 to 2020** to capture more substantial changes and carefully identified regions with visible changes in satellite imagery. This process required extensive effort and involved manually scanning satellite images to pinpoint areas with significant urban development.

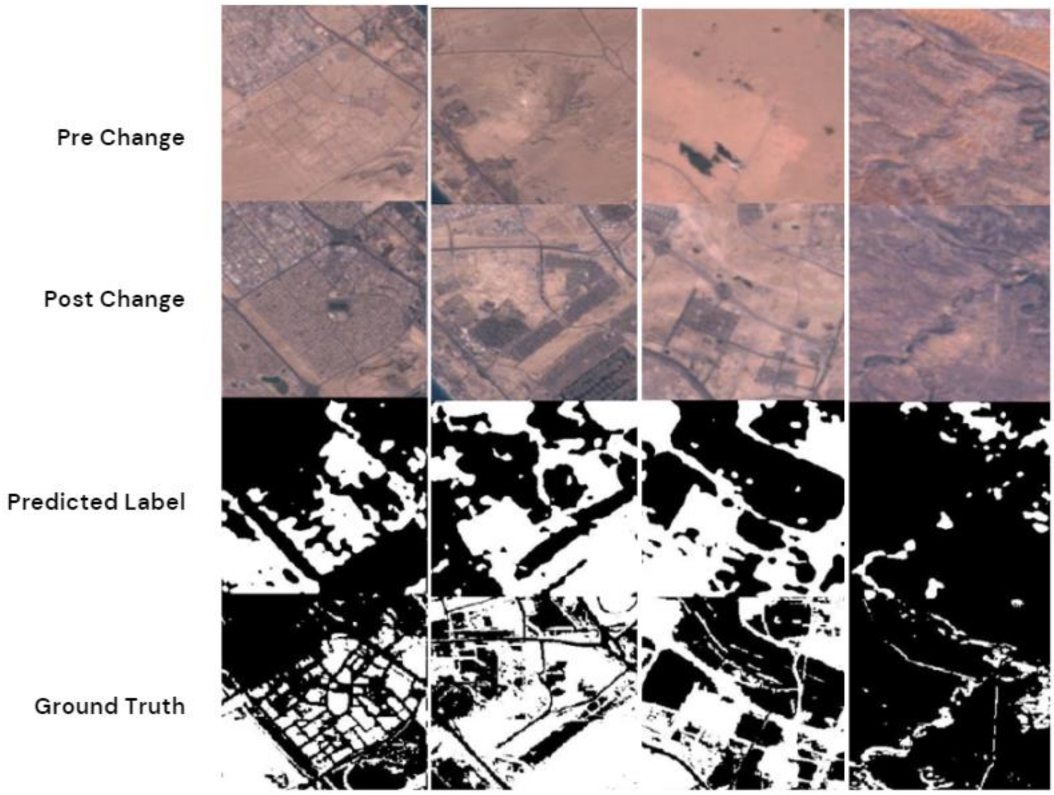
The final dataset comprised 20 different regions with noticeable urban expansion, including:

- **Middle East and Asia:** Dubai, Bahrain, Qatar, KSA, Abu Dhabi, Karachi, Lahore, Faisalabad, Quetta, Islamabad, KPK, Gujranwala, and Myanmar.
- **China:** Zhenzhou, Tianjin, and other regions experiencing rapid development.
- **Other Countries:** South Korea and selected urban areas in other nations.

This dataset was built with a strong focus on urban expansion and was specifically designed to enhance the model's ability to detect meaningful changes over time.

Model Training and Results

BIT_CD: Achieved an accuracy of **75%** after 200 epochs us.



ScratchFormer : Achieved an accuracy of **66%** at 60 out of 300 epochs.

These results were a substantial improvement over our previous attempts, as the models were now able to detect changes in urban areas rather than predicting blank images. The better performance was attributed to the extended timeline and the careful selection of regions with visible changes.

Conclusion:

This project explored both traditional machine learning techniques and modern transformer-based models to detect changes in remote sensing images, focusing on the Amazon Rainforest and urban expansion regions. Initially, machine learning models such as Random Forest, XGBoost, and Decision Tree were applied to satellite imagery, but the performance varied due to class imbalance and dataset size challenges. The best results were achieved using Random Forest Classifier on a smaller dataset.

Building on this, transformer architectures like BIT and ScratchFormer were implemented for more sophisticated change detection. These models leveraged the self-attention mechanism to capture spatial and temporal patterns, yielding significantly improved results. Despite initial challenges with minimal visible changes in the dataset, an enhanced dataset with a longer timeframe and targeted regions led to substantial performance gains, particularly in urban expansion detection.

Overall, the project demonstrated the powerful potential of transformer models in remote sensing change detection tasks, especially when applied to well-prepared datasets. The insights gained through this work highlight the transformative impact of using advanced architectures for monitoring environmental changes and urbanization over time.

References:

https://github.com/justchenhao/BIT_CD

https://github.com/mustansarfiaz/ScratchFormer

https://paperswithcode.com/paper/efficient-transformer-based-method-for-remote

https://ieeexplore.ieee.org/abstract/document/10489990