GeoFormer: A Vision and Sequence Transformer-based Approach for Greenhouse Gas Monitoring

Madhav Khirwar ♢ Ankur Narang ♠
♢ madhavkhirwar49@gmail.com ♠ankur.narang@fermionai.com

Abstract

Air pollution represents a pivotal environmental challenge globally, playing a major role in climate change via greenhouse gas emissions and negatively affecting the health of billions. However predicting the spatial and temporal patterns of pollutants remains challenging. The scarcity of ground-based monitoring facilities and the dependency of air pollution modeling on comprehensive datasets, often inaccessible for numerous areas, complicate this issue. In this work, we introduce GeoFormer, a compact model that combines a vision transformer module with a highly efficient time-series transformer module to predict surface-level nitrogen dioxide (NO_2) concentrations from Sentinel-5P satellite imagery. We train the proposed model to predict surface-level NO_2 measurements using a dataset we constructed with Sentinel-5P images of ground-level monitoring stations, and their corresponding NO_2 concentration readings. The proposed model attains high accuracy (MAE 5.65), demonstrating the efficacy of combining vision and time-series transformer architectures to harness satellite-derived data for enhanced GHG emission insights, proving instrumental in advancing climate change monitoring and emission regulation efforts globally.

Keywords: geo-spatial imaging, vision transformers, *ProbSparse* attention.

Introduction

- The emission of greenhouse gases (GHGs), primarily from industrial and transportation activities, is a major contributor to the climate change crisis.
- NO_2 is associated with air contaminants like PM2.5 and is released alongside CO_2, making it an effective indicator for CO_2 emissions.
- High-resolution satellite imagery, such as from the Sentinel-5P satellite, provides unprecedented monitoring opportunities for atmospheric pollutants.
- The TROPOMI device on Sentinel-5P enables detailed observation of NO_2 emissions on a global scale.

Contributions

1. Introduced a comprehensive dataset pairing high-resolution Sentinel-5P imagery with surface-level NO_2 concentration measurements.
2. Proposed a novel, compact transformer-based model for efficient spatio-temporal analysis of NO_2 emissions.

Methodology

Vision Transformer Module

- Adapts images as sequences of patches for transformer processing.
- Applies linear transformation to project image patches into embeddings.
- Employs Multi-Head Self-Attention (MHSA) to capture spatial relationships:

$$\text{MHSA}(E) = \text{softmax}\left(\frac{EQE^TK}{\sqrt{d_k}}\right)EV \tag{1}$$

where E denotes patch embeddings, with Q, K, and V as the query, key, and value projections, respectively.

Efficient Time-series Transformer

- Processes sequences with a sparsity-enhanced self-attention mechanism for efficiency.
- Features a reduced time-complexity of $O(N\log(N))$, improving computational tractability:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

focusing on dominant queries for attention calculation, where Q, K, and V are query, key, and value matrices from input data.

Performance Summary

Model	MAE	MSE	Size (MB)
GeoViT*	5.84	58.9	850
CNN Backbone*	6.68	78.4	1964
GeoViT	6.69	72.70	65
CNN Backbone	6.49	67.25	**32**
GeoFormer	**5.65**	**56.95**	70

Table 1: *Model comparison on GHG monitoring. Best Metrics are in* ***bold****. Models trained on the dataset introduced by [1] are marked with an asterisk (*).*

Directions for Future Research:

- **Expansion:** Integrate GeoFormer with optical flow for enhanced sequence modeling from sequential satellite images.
- **Scalability:** Extend the model to monitor additional pollutants and environmental indicators.

GeoFormer sets a new standard for real-time and efficient GHG monitoring, opening pathways for significant advancements in climate change mitigation efforts.

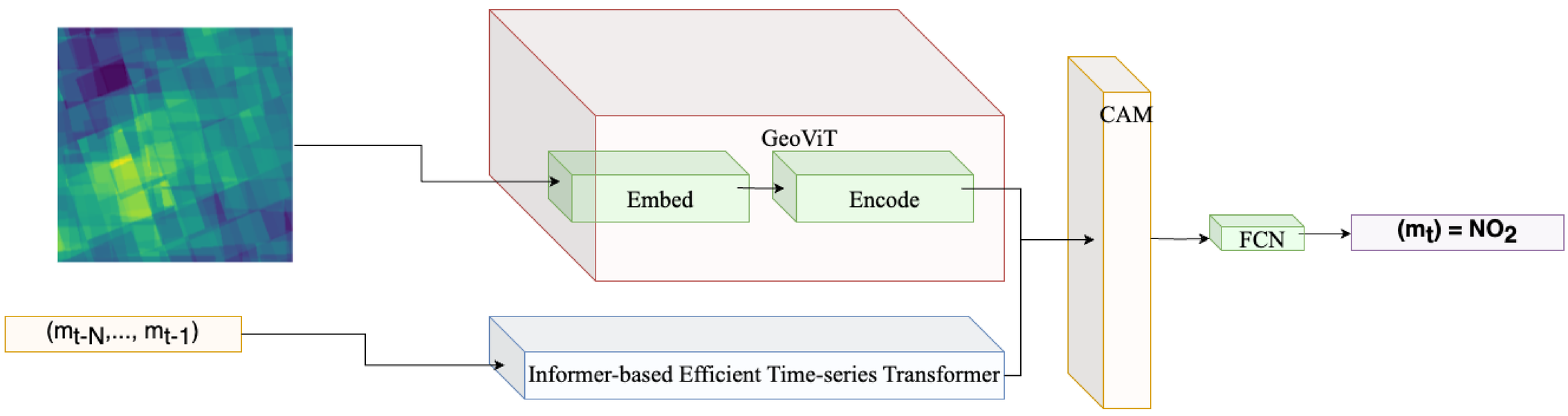

Figure 1: *GeoFormer model architecture. Here, m_t represents an NO_2 prediction at timestamp t, and CAM represents the cross-attention module.*

References

[1] L. Scheibenreif, M. Mommert, and D. Borth. Estimation of air pollution with remote sensing data: Revealing greenhouse gas emissions from space. *arXiv preprint arXiv:2108.13902*, 2021.