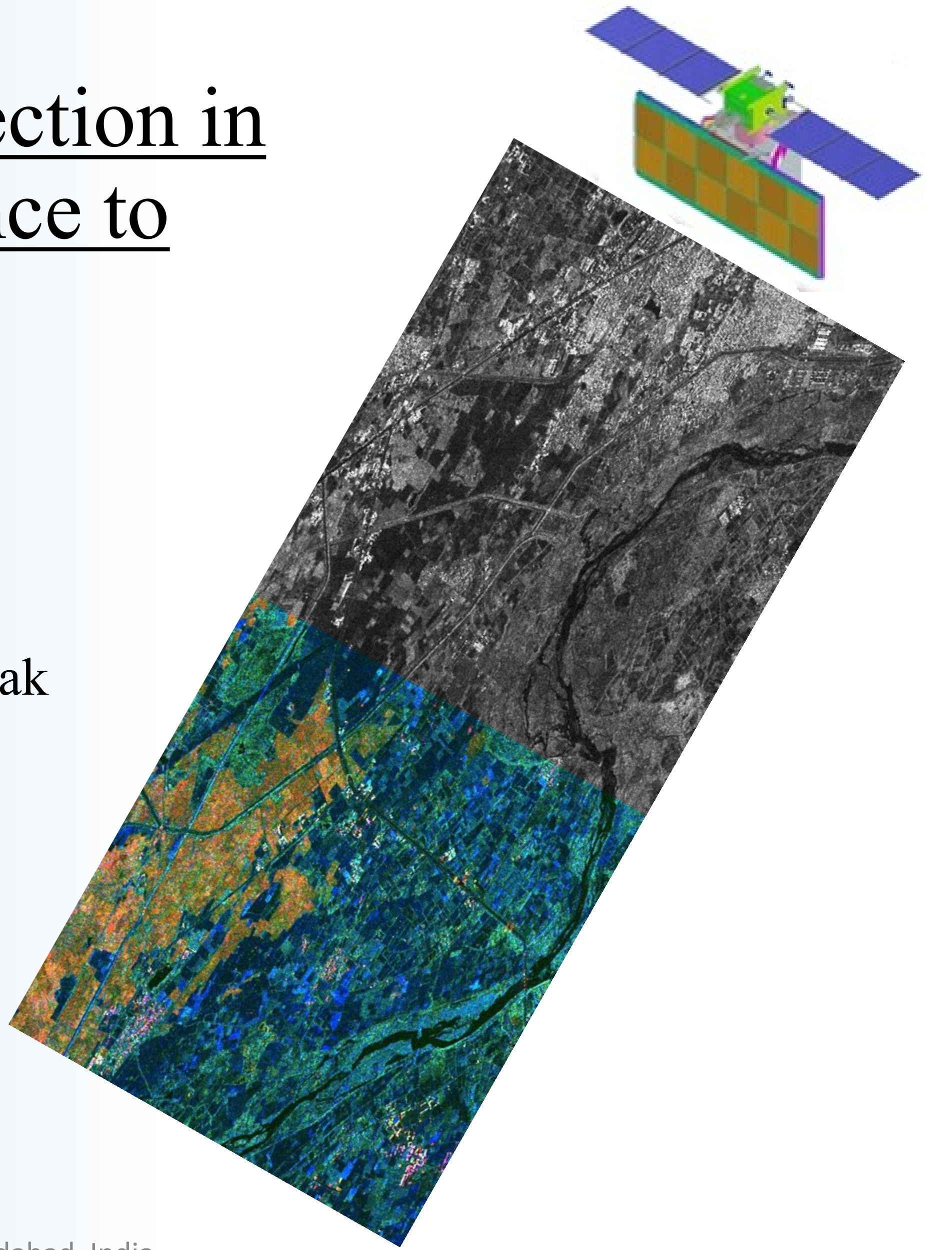


Deep Vision for Seismic Signal Detection in SAR Interferograms with Reference to NISAR Applications

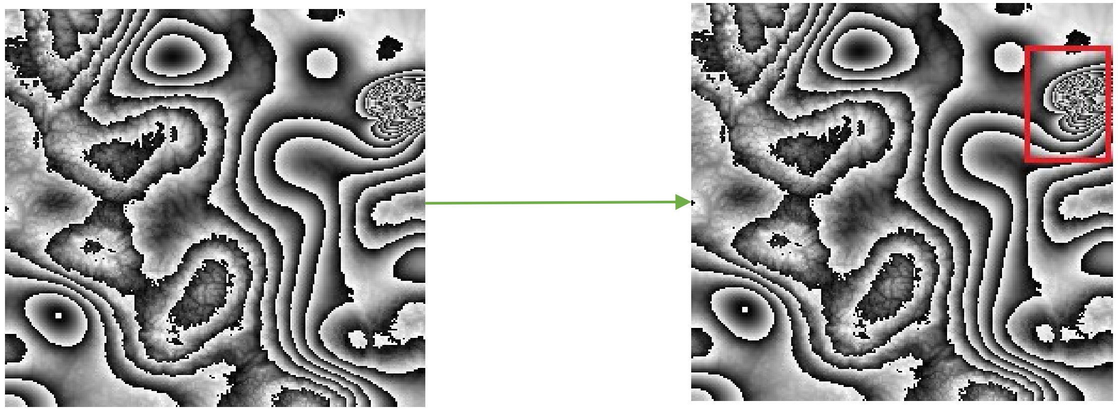
Alka Saini[1*], Sreejith K M[1],Ritesh Agrawal [1], Anuja Sharma[1], Deepak Mishra[2]

[1]Space Applications Centre

[2]Indian Institute of Space Science & Technology

Problem description

- Detection of earthquake-related deformation signals in SAR Interferograms.
- Classifying polarity of Interferogram Whether Earthquake is Present or not.
- By systematically analyzing InSAR data, improved accuracy and reliability can be achieved in identifying earthquake-related deformation signals.
- An effective algorithm and methodology is needed for the automated detection and interpretation of earthquake-related deformation signals in InSAR data.



Objective & Implementation

- Main objective is to develop a Deep learning based model to locate and classify surface deformation patterns related to earthquakes in SAR interferograms automatically.

- A strong and robust attention-based network is needed to detect presence of seismic signal in SAR Interferograms.

- A transformer-based model is employed to classify the observed Interferograms according to the presence or absence of earthquakes.

- By leveraging the capabilities of Vision Transformer, Multi head Self-Attention based Weighing Maps are investigated and extracted for accurate categorization of surface deformation patterns associated with earthquakes .

What is a Vision Transformer(ViT)

- Inspiration from Transformers success in NLP: "Attention is all you need ,2017" .

- Attention allows ViT to integrate information across the entire image.

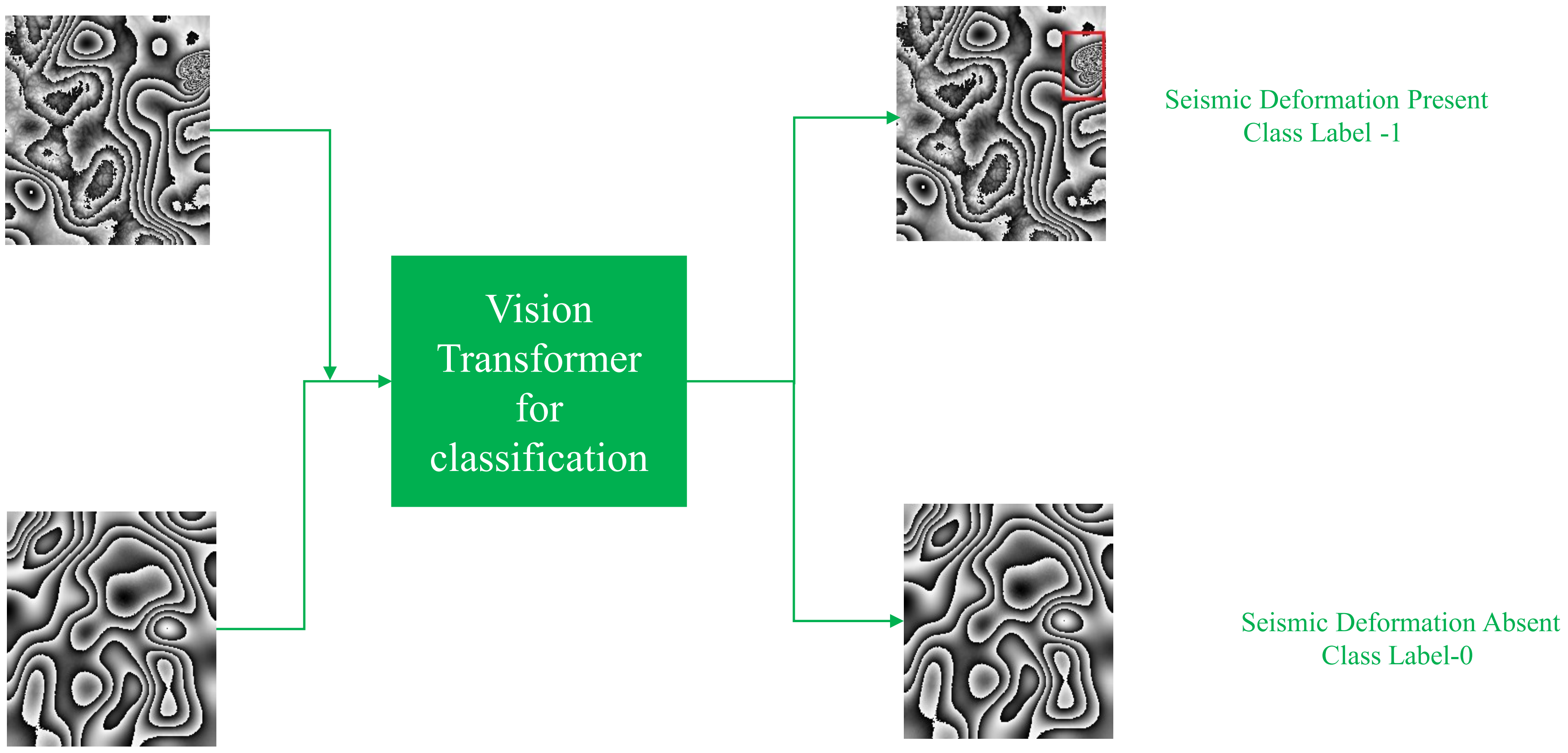
- While transformers take word embedding vectors as input, ViT, is a model for image classification employs a Transformer-like architecture over fixed size patches of the image.

Why ViT?

- SAR interferograms can exhibit deformation patterns at different scales, from small localized changes to large-scale deformations covering extensive areas.

- Multi-head attention networks, with their ability to capture both local and global context simultaneously, are inherently more adaptable to varying scales of features compared to CNNs, which may struggle with capturing information at different spatial scales.

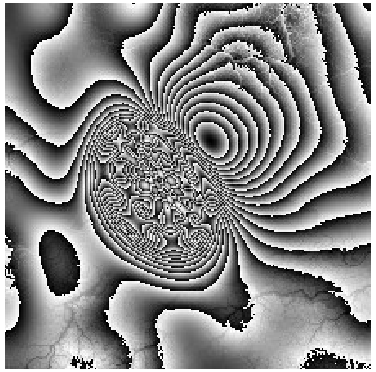
Model for Classification of Interferogram



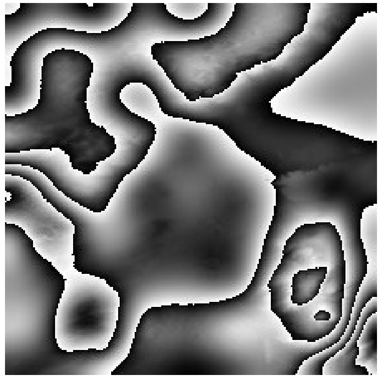
Data Preparation

- A large number of synthetic interferograms were generated to train and feed the model, including both wrapped and unwrapped surface displacements to make the model be applicable on any type of Interferograms, along with labelling (with and without seismic information).
- Splitting Data into Train, Test and Validation parts:
  - ✓ Train set size:805436
  - ✓ Validation set size:100679
  - ✓ Test set size:100681

WRAPPED INTERFEROGRAM

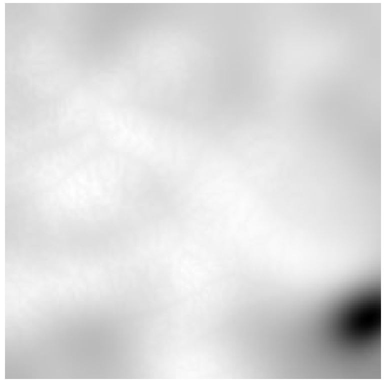


Seismic Deformation Present
Class Label -1

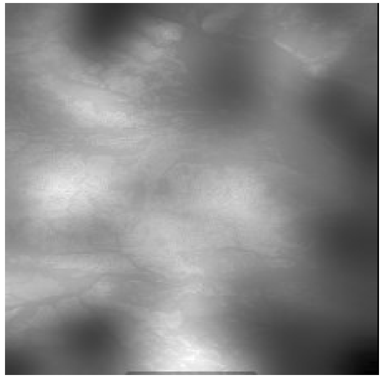


Seismic Deformation Absent
Class Label -0

UNWRAPPED INTERFEROGRAM



Seismic Deformation Present
Class Label -1



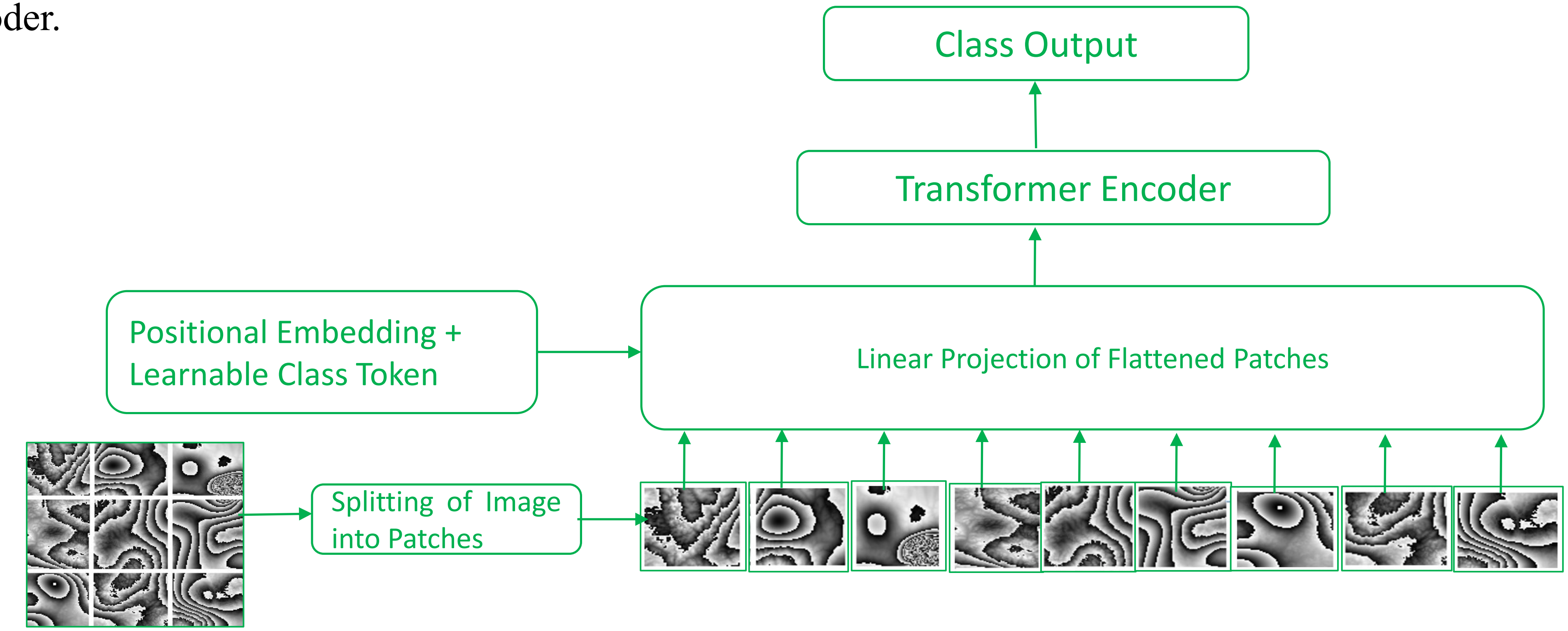
Seismic Deformation Absent
Class Label -0

Model Overview

- To feed images to the Transformer encoder, each image is split into a sequence of fixed-size non-overlapping patches, which are then linearly embedded.

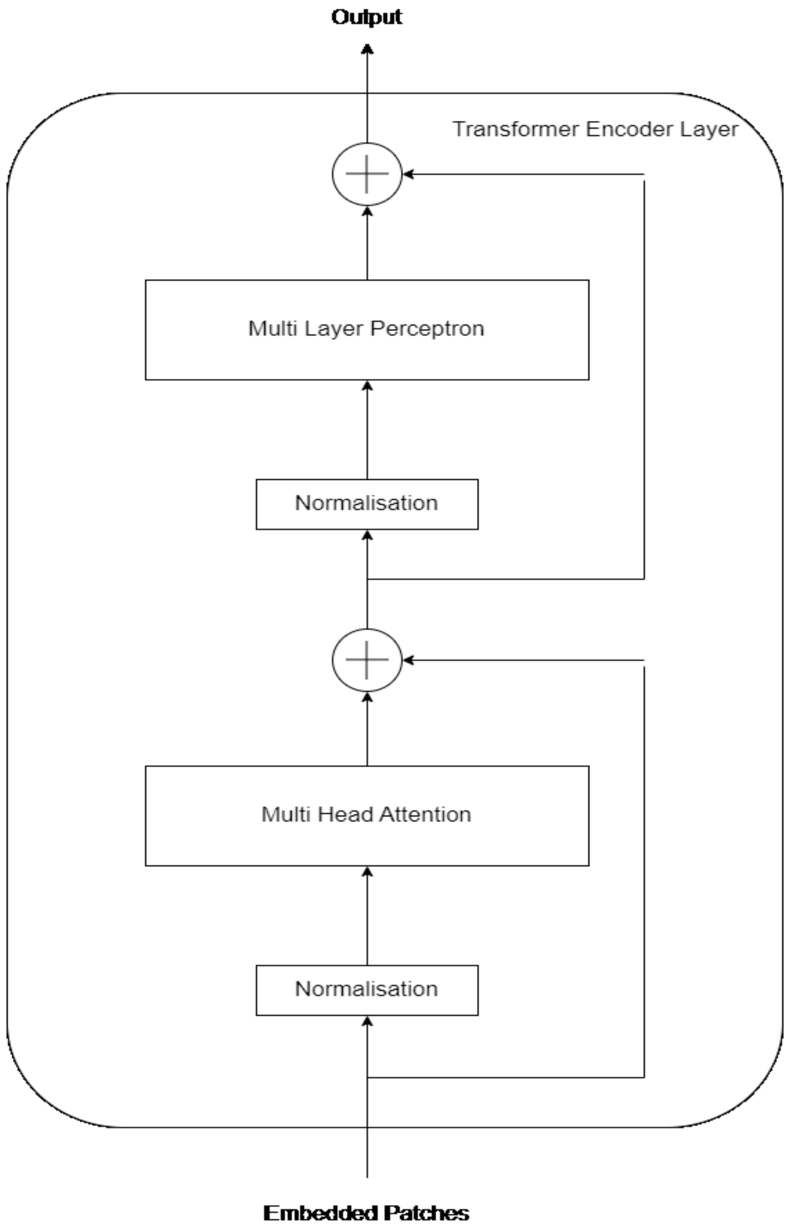
  224x224 → 16x16 :196 Patches

- A [CLS] token is added to serve as representation of an entire image, which can be used for classification. Position embeddings are also added along with tokens , and feed the resulting sequence of vectors to a standard Transformer encoder.



Transformer Encoder

- The encoder component is a stack of identical encoders.
- Each one is further broken down into two sub-layers, a multi attention head layer and a MLP layer.
- The input tokens are first embedded using a patch embedding layer, and then passed through the transformer encoder.
- The Encoder's inputs first flow through an attention layer and then it is passed through a MLP. Finally classification labels are obtained.
- The output of the transformer encoder is a sequence of contextualized token embeddings, which capture the meaning of the input tokens in the context of the entire input sequence.
- Normalization is done make training of model stable and to reduce effect of any shift in outputs.



Model working

An image is split into fixed-size patches, each of them are then linearly embedded, position embedding are added

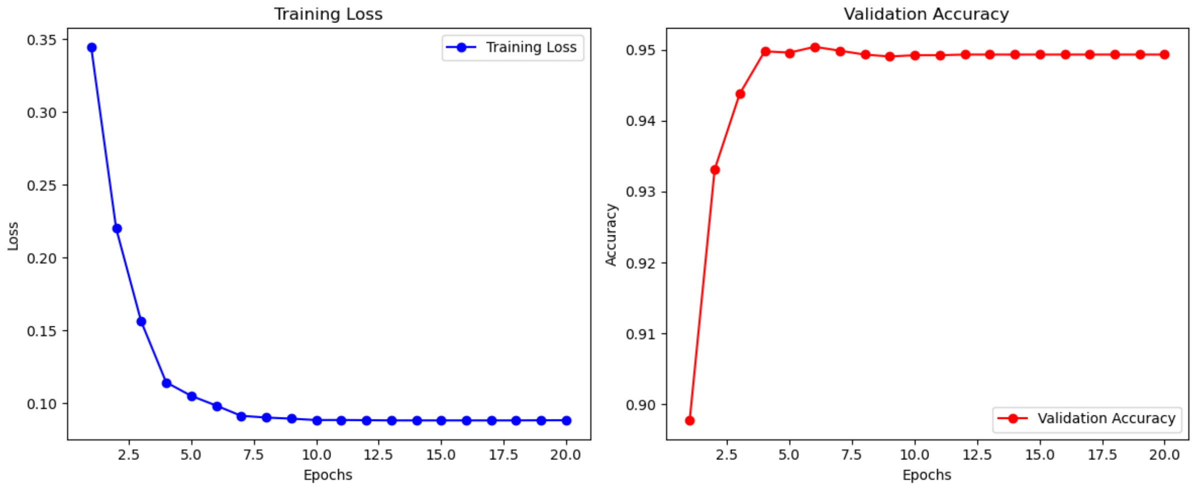
Provide sequence of linear embedding of these patches as an input to transformer.

Add a learnable embedding class along with patch embedding

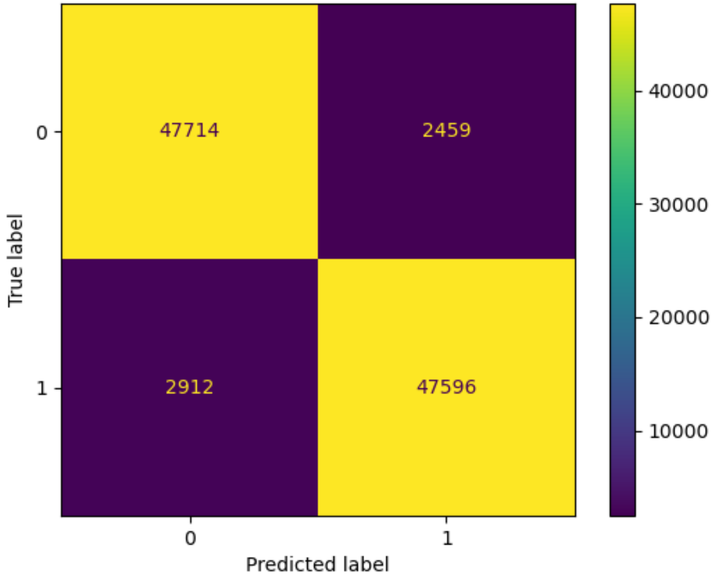
Pass these embedding through Transformer encoder and get the output values for each class tokens.

Pass the representation of class through MLP head and get the final class predictions.

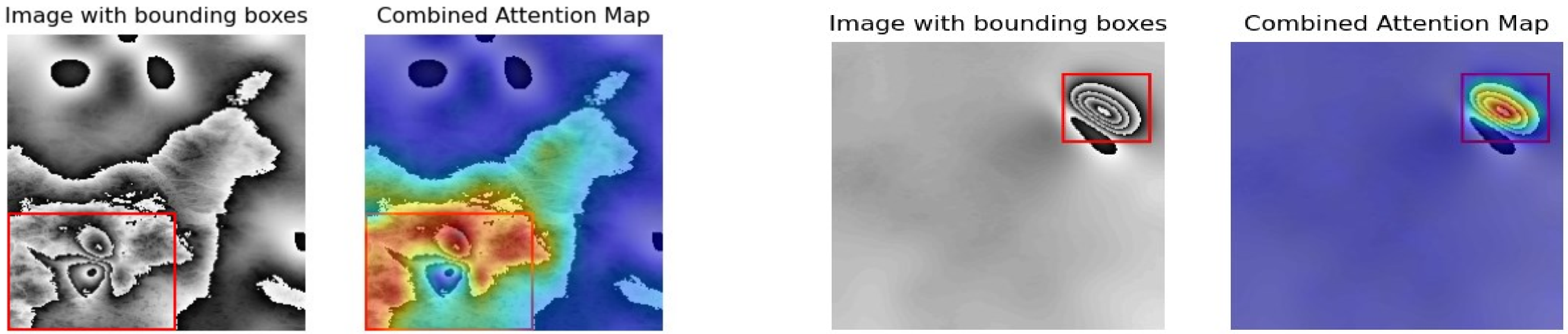
Results



Loss and Accuracy

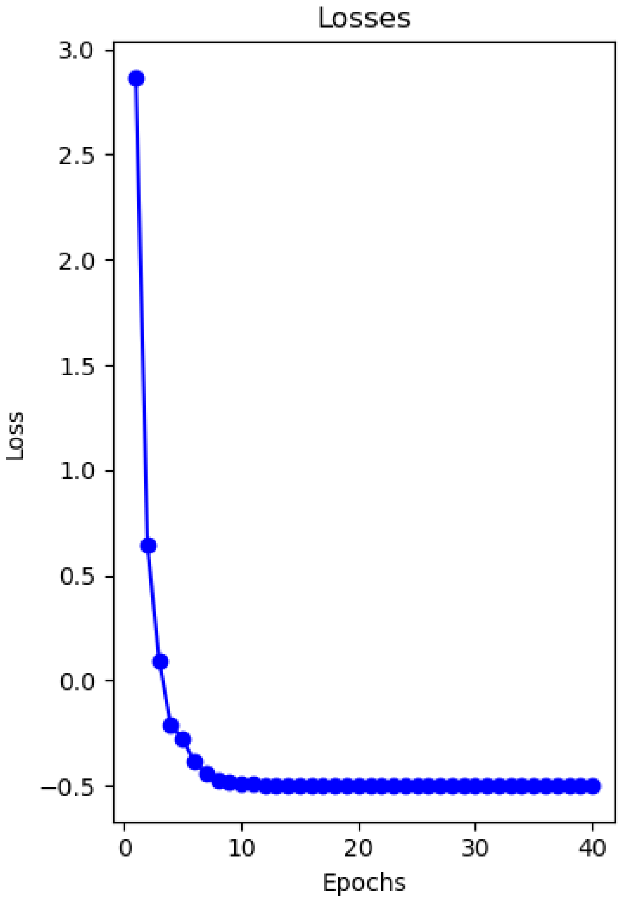


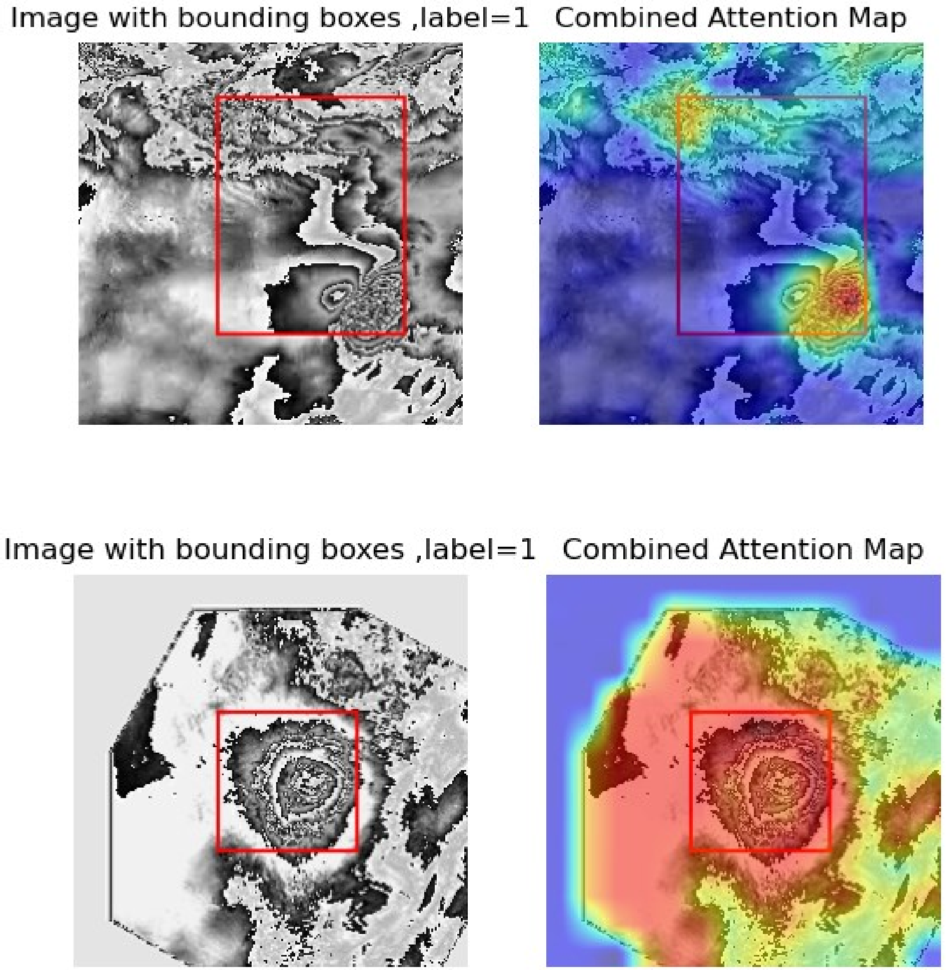
Confusion Matrix



Visualisations of Interferogram and Attention Maps

Results on Real Data





After Fine tuning over real earthquake datasets
Precision: 75 %
Recall: 73 %
F1 Score: 73 %
Test Accuracy: 75 %

Conclusion & Future Scope

❖ Through this systematic model development process, we aimed to create a robust and reliable deep learning model capable of accurately identifying and categorizing surface deformation patterns associated with earthquakes in SAR Interferograms..

❖ This research highlighted how incorporating a multi-head attention network can enhance the model's ability to capture complex patterns and spatial dependencies within SAR Interferograms data, ultimately improving its performance in earthquake detection and analysis.

Future Work :

❖ Minimization of false positives and false negatives.

❖ Optimization of model for Improved Accuracy & Precision.

❖ Estimation of seismic characteristics via Deep Learning Techniques.

References

1. [Brengman and Barnhart, 2021] Brengman, C. M. and Barnhart, W. D. (2021). Identification of surface deformation in insar using machine learning. *Geochemistry, Geophysics, Geosystems*, 22(3):e2020GC009204.

2. [Dosovitskiy et al., 2020] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

3. [Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*,

4. [Silva et al., 2022] Silva, B., Sousa, J. J., Lazecky, M., and Cunha, A. (2022). Deformation fringes detection in sar interferograms using deep learning. *Procedia Computer Science*, 196:151–158.

5. Anantrasirichai et al. (2018, 2019a, 2019b) and Valade et al. (2019)

Thank you for your kind attention!

☺

Model Components

Patch embedding layer: Patches in a sequence are embedded. This layer assigns each token's meaning to a high-dimensional embedding vector.

Position embedding : Along with patch embedding, position embedding is also passed as an input to the transformer encoder to retain positional information.

Transformer encoder: A multi-layer transformer encodes the input patch embeddings. Multiple self-attention layers allow the transformer encoder to focus on different regions of the input sequence and capture long-range relationships.

Output layer: The transformer encoder outputs contextualized token embeddings that represent the input tokens in the context of the complete input sequence. The output layer predicts classes of seismic deformation

Model Parameters

Optimizer =Adam

Learning Rate =1e-5

Epochs=20

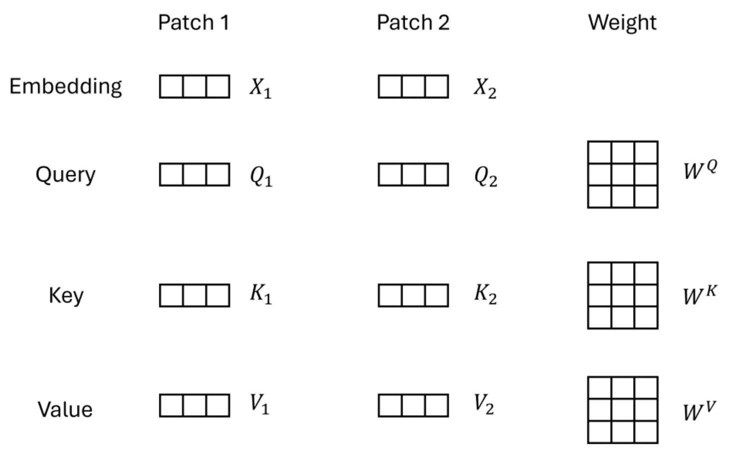
Loss=Cross Entropy

MHA & MLP

- Self-attention is computed not once but multiple times in the Transformer's architecture, in parallel and independently, allowing the model to process different aspects of the input sequence concurrently.

- A single attention head may struggle to capture all of these relationships simultaneously. It has a limited capacity to attend to different aspects of the input sequence.

- By using multiple attention heads, the transformer can attend to various parts of the input sequence in parallel and capture different dependencies.



Attention formulation

$$attention(XW_q, XW_k, XW_v) = attention(Q, K, V) = softmax\frac{QK^T}{\sqrt{d_k}}V$$

- The final MLP block also called the MLP head, is used as an output of the transformer. An application of softmax on this output can provide classification labels.

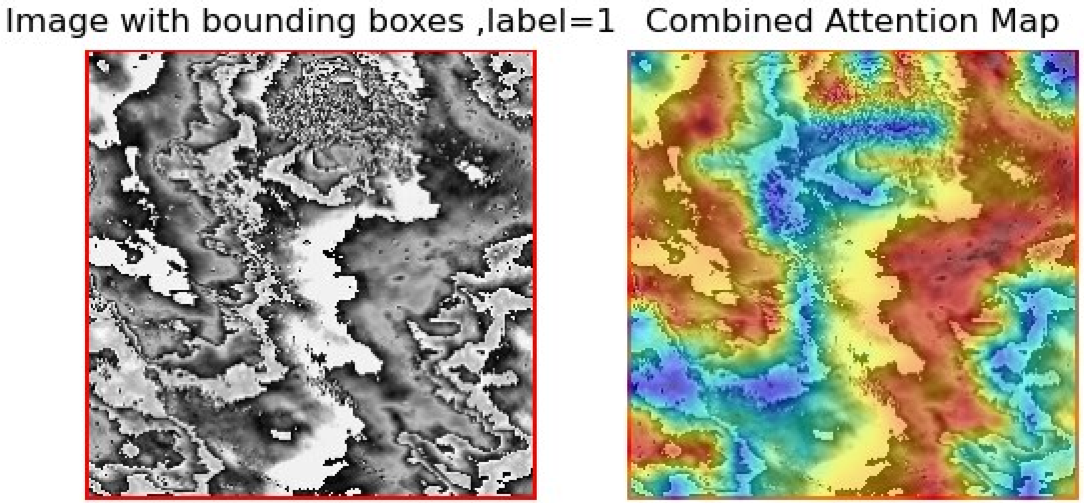
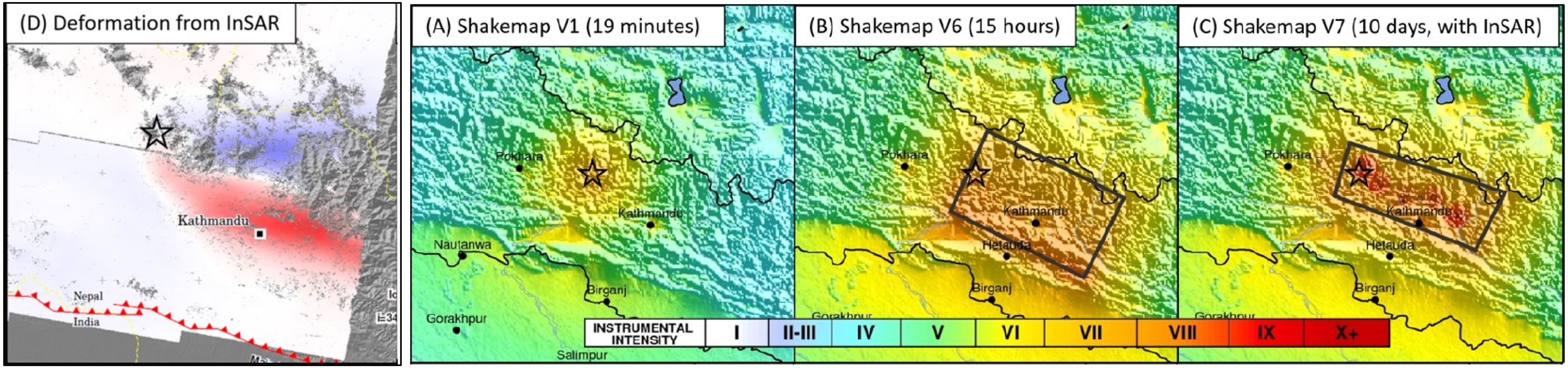


Image with bounding boxes ,label=1
Combined Attention Map

2015 Nepal Earthquake Hazard Analysis



Seismic Only Analysis

Seismic + InSAR Analysis