Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

AI Center Projects in Machine Learning Research Spring 2025

ETH AI CENTER

3D Transformer for Earthquake Prediction

Aydin Javadov[1], Anna Kravchenko[1], Fanny Lehmann [2]
[1]ETH Zurich Student, D-INFK [2]ETH AI Center, ETH Zurich

1. Introduction and Objectives

The increasing availability of high-resolution seismic data, coupled with recent advances in machine learning, opens new opportunities for forecasting complex wave dynamics in three-dimensional (3D) domains. While traditional numerical solvers are accurate, they remain computationally expensive for large-scale simulations. Learning-based surrogates, such as Fourier Neural Operators (FNO), have shown promise for wavefield approximation, yet their performance on realistic seismic data is still limited.

Transformer-based architectures, widely successful in natural language processing and computer vision, offer strong capabilities for capturing long-range spatio-temporal dependencies. However, their adaptation to 3D geophysical forecasting has been hindered by computational and memory challenges.

In this work, we explore the effectiveness of transformer-based architectures for seismic forecasting using a recent **large-scale dataset (HEMEW-3D)[1]**. Specifically, we implement and evaluate the **SwinUNETR model**[2], which leverages hierarchical attention and convolutional decoding to learn future wavefield dynamics from historical displacement data.

The main objectives of this study are to:

- Evaluate the SwinUNETR model's capability to predict 3D seismic wave propagation on a realistic 10,000-sample dataset.
- Assess both qualitative and quantitative forecasting performance using slice-wise displacement maps and evaluation metrics (MAE, RMSE, R^2).
- Investigate model behavior through ablation studies and training with varied hyperparameters (e.g., patch size, attention heads, temporal history).

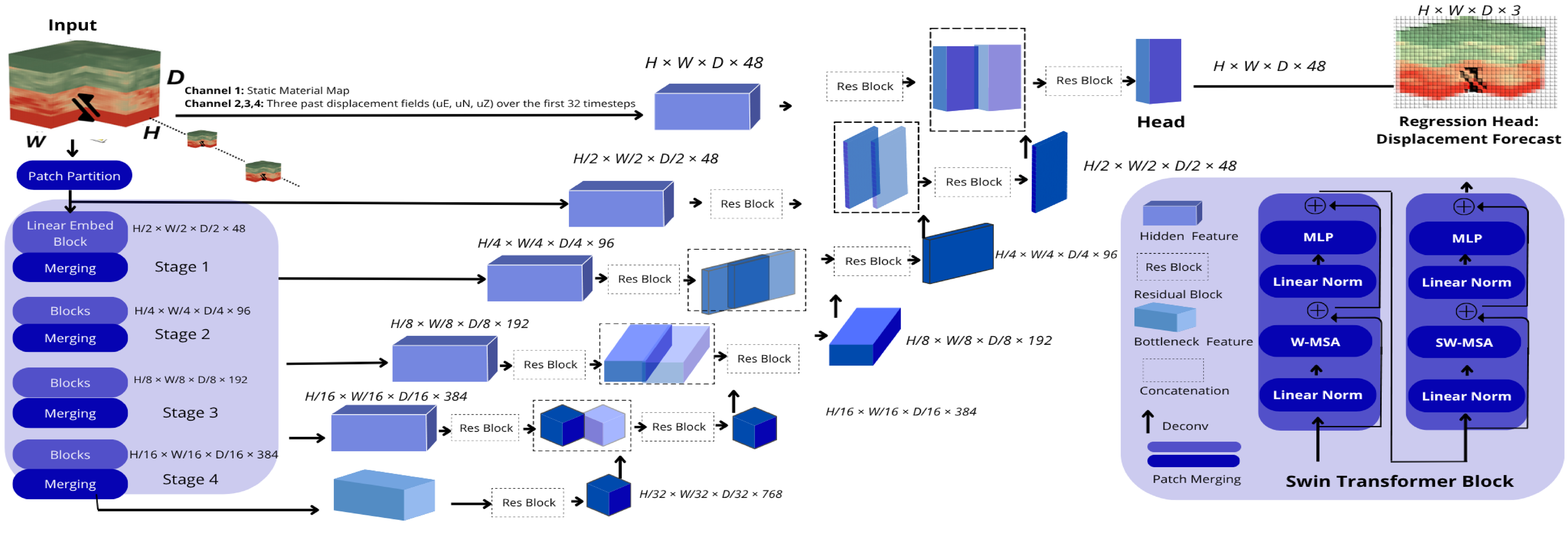

Figure 1: Architecture of the SwinUNETR model for seismic forecasting.

2. Related Work

- **Fourier Neural Operators (FNO)**[3] enable efficient learning of surrogate models for PDEs, including seismic wave propagation, by modeling long-range dependencies in spectral space.
- **Transformers** are widely used across domains for learning spatio-temporal patterns via self-attention, with growing use in physics-informed tasks.
- **Swin Transformers**, designed for vision tasks, apply hierarchical attention with shifted windows and have been adapted to 3D imaging and volumetric learning.
- Seismic forecasting has mostly relied on physics-based numerical simulations.

3. Data

The HEMEW-3D dataset provides 30,000 high-fidelity 3D simulations of elastic wave propagation in a 9.6 km³ domain with heterogeneous geological media. Each sample includes a stochastic shear-wave velocity (V_S) model.

Seismic motion is recorded as 3-component velocity wavefields over a 16×16 spatial grid for 20 seconds (2000 time steps at 0.01s). Geological inputs and wavefields are discretized into voxelized 3D grids of size 32×32×32 and provided in tabular format.

The model was trained on a **set of 10,000 samples** at 10 Hz resolution, using **9000 for training** and **1000 for validation**. Each input tensor comprises 4 channels: 3 past velocity components (u_E, u_N, u_Z) and 1 velocity model — all encoded as volumetric data.

4. Methods

We train a 3D SwinUNETR model to forecast future ground motion velocity using past wavefield snapshots and geological inputs from the HEMEW-3D dataset. Preprocessing steps include temporal slicing of input sequences, z-score normalization of each feature channel, and spatial alignment via cropping and padding to obtain uniform input volumes. The model is trained using a **pure L1 loss function**:

$$\mathcal{L}(a, b) = \frac{1}{N} \sum_{c,d,h,w} \frac{|a_{c,d,h,w} - b_{c,d,h,w}|}{|b_{c,d,h,w}| + \varepsilon}, \quad \varepsilon = 0.01$$

We monitor performance using MAE, RMSE, and R^2 metrics across three displacement components: Easting, Northing, and Vertical.

The SwinUNETR architecture combines shifted window self-attention with a U-Net–like decoder for efficient modeling of spatial dependencies and hierarchical feature fusion. Its ability to capture long-range interactions while preserving fine-scale structure makes it well suited for volumetric learning tasks such as 3D seismic forecasting.

5. Results

The model was trained for 100 epochs using pure L1 loss. Figure 2 shows heatmaps of predicted vs. true East–West velocity (u_E) at fixed x, over height (y) and time (t), capturing key structural patterns. Figure 3 plots velocity magnitude over the width axis for fixed depths (z). The model follows true trends but underestimates peaks in high-gradient zones.

Visualization of $u_E(x = X_{pos}, y, t)$ *at fixed* x *position.*

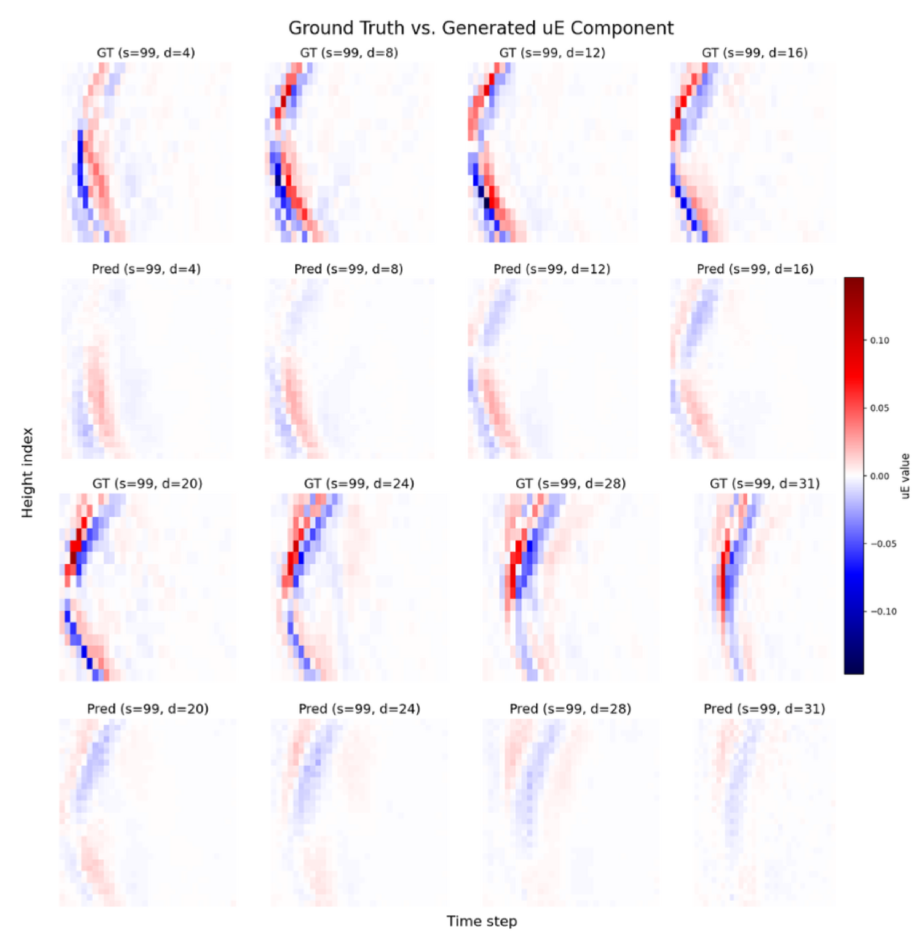

Figure 2: Ground truth vs. predicted East–West velocity (u_E) at fixed x position, shown as heatmaps over North–South position (Y) and time (T).

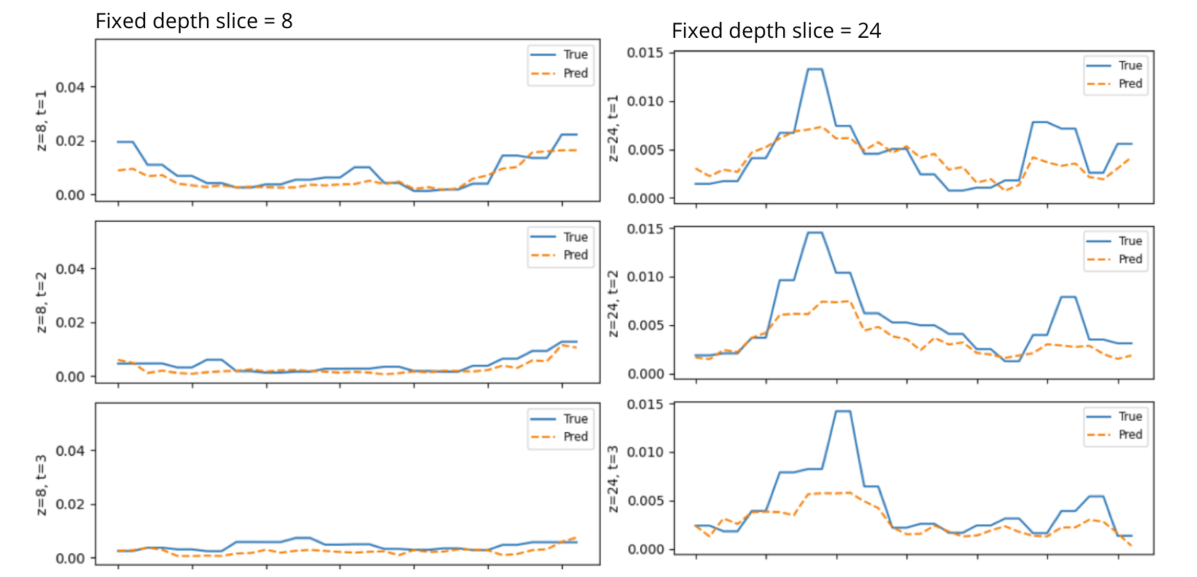

Figure 3: Line plots of predicted vs. true velocity magnitudes along the width axis for fixed slices $z = 8$ and $z = 24$ over the first 3 timesteps. The vertical axis represents $\sqrt{u_E^2 + u_N^2 + u_Z^2}$ at location (x, z, t).

6. Conclusions

The SwinUNETR model demonstrates robust potential for forecasting 3D seismic displacement fields from historical data. Trained on a large-scale subset of the HEMEW-3D dataset, the model achieves low prediction error under a pure L1 loss formulation. Qualitative results confirm the model's ability to capture structural patterns across depth slices, demonstrating its suitability for learning complex 3D spatio-temporal dynamics.

7. Future Work

- Perform ablation studies on architecture (e.g., patch size, attention depth) and input settings (e.g., history length, normalization).
- Explore model formulations that predict the full future wavefield directly, rather than relying on displacement history as input.
- Test formulations predicting full wavefields without displacement history input.
- Investigate multi-scale supervision and enhanced encoding.

References

1. F. Lehmann, F. Gatti, M. Bertin, D. Clouteau. *Synthetic ground motions in heterogeneous geologies: the HEMEW-3D dataset for scientific machine learning*. Earth System Science Data, 2024. https://doi.org/10.5194/essd-16-3949-2024
2. A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, D. Xu. *Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images*. arXiv preprint arXiv:2201.01266, 2022. https://arxiv.org/abs/2201.01266

F. Lehmann, F. Gatti, D. Clouteau. *Multiple-Input Fourier Neural Operator (MIFNO) for source-dependent 3D elastodynamics*. Journal of Computational Physics, 2025. https://doi.org/10.1016/j.jcp.2025.113813

Contributions

Aydin Javadov: Model building and implementation (SwinUNETR), conceptualization, experiments, generation of poster visualizations & quantitative outputs. Anna Kravchenko: Initial CNN-based model development and experimentation, poster concept and layout structuring, theoretical content writing, and integration of methodology and figures.