AdaptFNO: Adaptive Fourier Neural Operator with Dynamic Spectral Modes and Multiscale Learning for Climate Modeling

Hiep Vo Dang
School of Computer Science
Yeshiva University
New York, United States
hiep.dang@mail.yu.edu

Bach D.G. Nguyen
Department of Computer Science Engineering
Michigan State University
East Lansing, Michigan
nguy1104@.msu.edu

Phong C. H. Nguyen
Faculty of Mechanical Engineering and Mechatronics
Phenikaa University
Hanoi, Vietnam
phong.nguyenconghong@phenikaa-uni.edu.vn

Truong-Son Hy*
Department of Computer Science
The University of Alabama at Birmingham
Birmingham, Alabama, United States
thy@uab.edu

Abstract

Fourier Neural Operators (FNOs) are effective for modeling spatio-temporal dynamics but often favor low-frequency patterns, overlooking fine-scale details critical in climate forecasting. We propose AdaptFNO, an adaptive variant that dynamically adjusts spectral modes based on input frequency content and partitions domains into frequency-specific patches. A cross-attention mechanism aligns global and local features, enabling efficient multiscale learning. Evaluated on ERA5 reanalysis data, AdaptFNO captures both large-scale circulation and fine-grained events, such as typhoon trajectories, while maintaining long-range stability. Preliminary results on Typhoon Yagi highlight its ability to preserve details of cyclone formation, showing promise for high-resolution climate forecasting. The source code is available at: `https://github.com/HySonLab/AdaptFNO`.

1 Introduction

Accurate weather forecasting is essential for daily planning, agriculture, and preparation for severe storms such as tropical cyclones and hurricanes. Forecasting these events requires models that capture both global atmospheric circulation and local dynamics. Traditional numerical weather prediction (NWP) systems, such as the ECMWF's Integrated Forecasting System (IFS), and NOAA's Global Forecast System (GFS), remain the gold standard. While accurate, they demand massive computational resources and long runtimes on supercomputers, limiting their ability to deliver

*Correspondence to Phong C. H. Nguyen and Truong-Son Hy.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Machine Learning and the Physical Sciences.

frequent high-resolution updates, particularly for localized forecasts. Recent advances in machine learning offer a more efficient alternative. Fourier Neural Operators (FNOs) [4] learn mappings in frequency space, allowing fast modeling of spatio-temporal dynamics. FourCastNet [5] demonstrated strong medium-range forecasting using ERA5 reanalysis data [2]. However, standard FNOs suffer from low-frequency bias as a result of fixed spectral configurations, restricting their ability to represent fine-scale, high-frequency structures critical for local predictions.

We introduce AdaptFNO, an adaptive FNO that harmonizes global and local details through multiscale learning. AdaptFNO employs (i) a global operator trained in low-resolution ERA5 data to capture large-scale circulation patterns, (ii) a local operator trained on high-resolution inputs for fine-grained regional forecasts, and (iii) a cross-attention mechanism to selectively transfer global information to local domains. This design improves efficiency while preserving local variability, such as atmospheric rivers and storm tracks. We evaluated AdaptFNO on the prediction of the ERA5 wind field, showing its ability to maintain fine-scale structures while achieving competitive global skill. Our contributions are as follows.

- **Dynamic spectral modes:** A mechanism for adjusting the number of modes based on the input frequency content, allowing the capture of low- and high-frequency features.
- **Multiscale learning:** A global operator for large-scale patterns and a lightweight local operator for fine-grained details.
- **Cross-attention mechanism:** An integral operator cross-attention that transfers relevant global features to the local forecast, enhancing accuracy without costly global high-resolution modeling.

2 Methodology

Our proposed method, AdaptFNO, extends Fourier Neural Operators with adaptive spectral resolution to capture multi-scale climate patterns, such as temperature, precipitation, and wind fields, more effectively. Figure 1 provides an overview of the architecture.

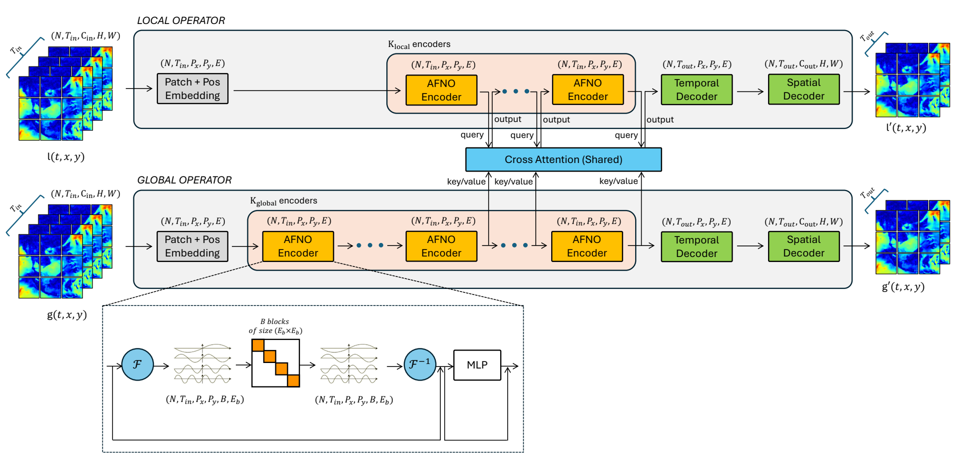

Figure 1: Overview of the AdaptFNO architecture. The model adaptively adjusts spectral resolution to capture both large-scale and fine-grained climate patterns in spatial-temporal data.

Patch Embedding and Positional Encoding. Each input climate field (e.g., a spatial grid of temperature or wind) is divided into $P = P_h \times P_w$ patches of size $p_h \times p_w$. Each patch is flattened and projected into a higher-dimensional latent space:

$$\mathbf{z}_0 = \left[\mathbf{x}_p^1\mathbf{E}_{\text{patch}}; \mathbf{x}_p^2\mathbf{E}_{\text{patch}}; \cdots ; \mathbf{x}_p^N\mathbf{E}_{\text{patch}}\right] + \mathbf{E}_{\text{pos}}, \tag{1}$$

where $\mathbf{E}_{\text{patch}} \in \mathbb{R}^{(P^2 \cdot C)\times D}$ is the learnable projection, $\mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N+1)\times D}$ is the learnable positional embedding, and N is the number of patches.

Adaptive Fourier Neural Operator. We extend the Adaptive FNO [1] for multi-frame climate forecasting. Given an input tensor $u(t, x, y)$ with T_{in} time frames, the operator predicts T_{out} future frames. The operator applies Fourier transforms in the spatial domain:

$$\text{AFNO}(u(t, x, y)) = \mathcal{F}^{-1}\big(\mathcal{F}(\kappa) \cdot \mathcal{F}(u)\big)(t, x, y), \tag{2}$$

where κ is the kernel function, and $\mathcal{F}/\mathcal{F}^{-1}$ are the Fourier and inverse Fourier transforms. For discrete grids:

$$\text{DFT}(u)(t, k_x, k_y) = \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} u(t, n_x\Delta x, n_y\Delta y) e^{-2\pi i \left(\frac{k_x n_x}{N_x} + \frac{k_y n_y}{N_y}\right)}, \tag{3}$$

$$\text{IDFT}(u)(t, x, y) = \frac{1}{N_x N_y} \sum_{k_x=0}^{N_x-1} \sum_{k_y=0}^{N_y-1} \text{DFT}(u)(t, k_x, k_y) e^{2\pi i \left(\frac{k_x n_x}{N_x} + \frac{k_y n_y}{N_y}\right)}. \tag{4}$$

The kernel DFT(κ) is parameterized as a block-diagonal matrix $W \in \mathbb{C}^{N_B \times B \times B}$ to operate on subsets of frequency modes:

$$W = \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_{N_B} \end{bmatrix},$$

with each $W_i \in \mathbb{C}^{B \times B}$.

Global & Local Operators. The global operator models large-scale atmospheric dynamics using downsampled low-resolution data for efficiency. It stacks 12 AdaptFNO encoders to capture temporal dependencies and spatial interactions, and serves as a context provider for the local operator. Pre-training with historical ERA5 data allows it to share global patterns while keeping computational cost low. The local operator focuses on high-resolution forecasts for a target region. Lightweight with only two AdaptFNO encoders, it can be easily adapted to different domains. Its performance is enhanced through information transferred from the global operator. To couple the two operators, we design a frequency-domain cross-attention mechanism that efficiently selects relevant global modes for local forecasts. This enables high-resolution predictions with minimal overhead.

Decoders & Loss Function. Finally, temporal and spatial decoders generate multi-step forecasts, rolling out sequences up to 3 days and producing high-resolution weather maps. We used a temporal Mean Square Error (TemporalMSE) which accounts for multistep prediction to optimize AdaptFNO. We utilized the Mean Squared Error and added temporal weights for the function so that earlier predictions will matter more, penalizing the difference between predicted fields $\hat{y}_t$ and ground truth y_t across all forecast horizons. The per-timestep squared error is defined as $\ell_t = \|\hat{y}_t - y_t\|_2^2$. The overall loss is a weighted sum across timesteps:

$$\mathcal{L}_{\text{TemporalMSE}} = \frac{1}{Z} \sum_{t=1}^{T} w_t\, \ell_t. \tag{5}$$

We will have the first 72-hour forecast to have the largest weight, and the weight will decay over time. We will be using an exponential decay weighting scheme:

$$w_t = \frac{e^{-\alpha(t-1)}}{\sum_{k=1}^{T} e^{-\alpha(k-1)}}\, T, \tag{6}$$

with a decay rate $\alpha > 0$. This formulation ensures that $\sum_{t=1}^{T} w_t = T$, so that the weighted loss remains on the same scale as the standard mean squared error.

3 Experiments

In Figure 2, we have tracked the formation of Typhoon Yagi from August 31st 2024 at 18:00 with a time gap of 6 hours. We have rolled out prediction until September 1st 2024 at 18:00. Our results showed that the prediction when compared with the ground truth has a promising result, as the mean absolute error for each of the prediction time frames is approximately 0.5. In Figure 3, we also created a visualization of the trajectory of Typhoon Yagi throughout the time, our prediction also indicates promising results, as the prediction of the typhoon trajectory compared to the grouth truth has also showcased the likelihood of latitude throughout the time. There is a point in time where the trajectory error over time increased, which is when Typhoon Yagi reached Taiwan. When a typhoon reaches the mainland, the wind power will decrease and also the trajectory of the typhoon will have a slight change in trajectory due to obstacles when reaching the mainland.

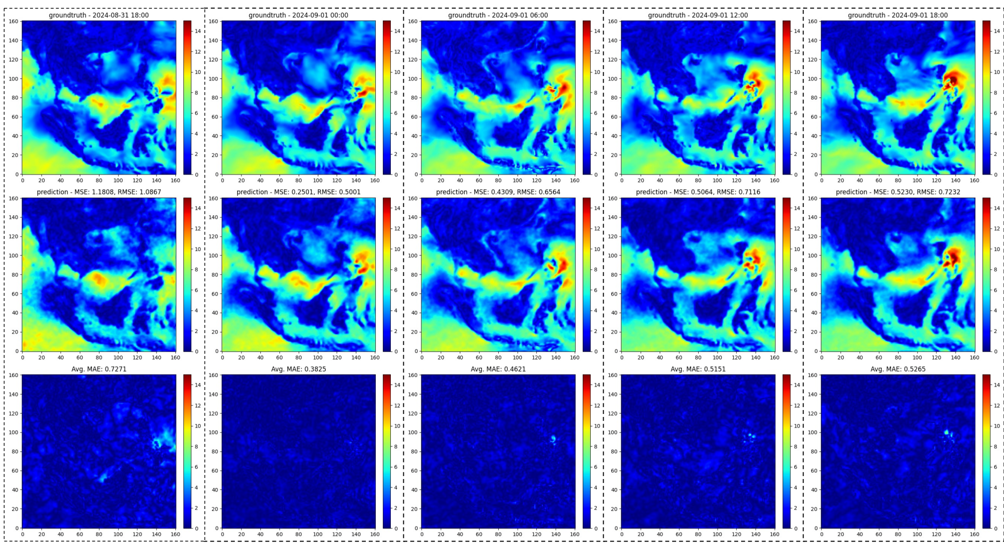

Figure 2: Performance of AdaptFNO tracking the formation and evolution of Typhoon Yagi from August 31st 2024 to September 1st 2024.

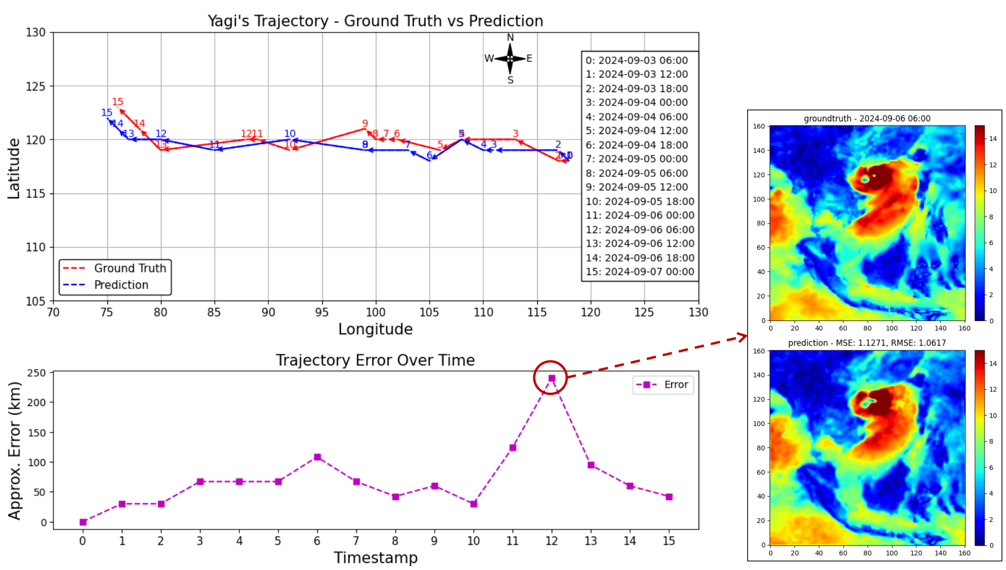

Figure 3: Trajectory comparison demonstrating the improved accuracy of AdaptFNO in capturing high-frequency phenomena.

Our experiments demonstrated that AdaptFNO can effectively capture large-scale and fine-grained climate patterns thanks to our architecture. Our results have been tested against the ground-truth, and the results showed promising results. However, in order to prove and test our architecture, future benchmarks and comparisons will be made against well-known baseline such as Convolution Neural Networks (CNNs) and Fourier Neural Operators (FNOs). Overall, the results support our hypothesis that integrating frequency-aware mechanisms and multi-scale attention improves the model's ability to predict high-resolution climate fields. Experimental details are included in the Appendix.

4 Conclusion & Future Work

Our proposed method AdaptFNO, an adaptive Fourier Neural Operator, is created to improve the current issue of traditional Fourier Neural Operators by dynamically choosing the modes based on the input content so that we can have the most important information that is relevant to our problem. The model demonstrates state-of-the-art accuracy in capturing both local variability and large-scale climate structures while maintaining computational efficiency. In future research, we plan to further develop AdaptFNO in several directions. We will be conducting a comprehensive comparison against standard FNOs and CNN baselines to evaluate the results of other methods and climate models. We also plan to extend AdaptFNO to predict other extreme events such as hurricanes, heat waves, and precipitation extremes. Finally, we will explore the integration of AdaptFNO into forecasting pipelines, where its efficiency and accuracy can provide rapid high-resolution updates for extreme weather events.

References

[1] John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv preprint arXiv:2111.13587*, 2021.

[2] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[4] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

[5] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

[6] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

A Related Work

Traditional Climate Modeling. Traditional climate models relied on numerical weather prediction (NWP) systems, and it has been the backbone of weather forecasting. It solves governing physical equations on global grids, providing consistent forecasts across multiple spatial and temporal scales. These models often utilize massive computational resources, and hours will be spent on supercomputers to be able to roll out a forecast. This limits the ability of the models to generate frequent forecasts in real time, and in critical scenarios such as cyclones or hurricanes, the latency of several hours can reduce the value of the forecast to make decisions. AdaptFNO runs faster at inference time while retaining fine spatial resolution, allowing rapid updates without the need for full-scale numerical integration.

Physics-Informed Neural Networks (PINNs). Physics-informed neural networks (PINNs) [6] apply physics laws to the loss function, improving the consistency even with limited training data. It has been widely applied in physics equations and is used specifically in geophysical fluid dynamics, allowing us to achieve stable simulations without labeled datasets. However, when it comes to global climate problems, PINNs tend to scale poorly because Partial Differential Equation complexity, domain size, and multiscale variability make convergence difficult and time consuming. Their dependence on PDE also poses a restriction when only reanalysis data, such as ERA5 [2], is available. The global scale reanalysis dataset contains rich dynamics, but it often has tractable PDE forms for PINN. AdaptFNO learns directly from reanalysis data without requiring PDE supervision, while preserving multiscale dynamics through the architecture.

Fourier Neural Operators (FNOs). Unlike standard neural networks that map input vectors to outputs, the Fourier Neural Operator (FNOs) [4] learns the mapping between function spaces, mapping entire functions to functions. This is a great tool to use for spatio-temporal dynamics, in this case, climate predictions. Once we have trained the FNO, the inference will be extremely fast, faster than numerical solutions by a large margin. However, with the configuration of the architecture, the FNO typically selects a certain fixed range of spectral modes, which introduces a low-frequency bias. This will result in a loss of high-frequency structures and fine-scale details that only high frequency possesses. Fine-scale weather features are crucial for weather forecasting and early warning as they contain features of phenomena such as cyclones and storms, and losing them would reduce the credibility of the forecast. AdaptFNO fixes this problem by dynamically allocating spectral modes based on input frequency content, ensuring efficiency while maintaining fine-grained details.

Multiscale Learning for Climate Forecasting. Multiscale modeling can enable us to learn both global circulation patterns and fine-scale features. It excels at capturing rare but impactful phenomena thanks to its flexible design, where it can combine local convolutions for fine-grained details and global attention for context. However, many multiscale methods process global and local branches independently; therefore, there will be weak coupling between the scales. Additionally, most of these models do not implement frequency domain adaptivity. Climate forecasts require that fine-scale events be placed in the right large-scale context to be able to be accurate, as seen in phenomena such as atmospheric rivers. AdaptFNO merges multiscale features through a cross-attention mechanism that aligns global and local features, while adaptive Fourier layers provide the frequency resolution control.

B Experimental Details

B.1 Dataset

We evaluated AdaptFNO using the ERA5 hourly data on pressure levels from 1940 to present, and chose the data from 1940 to 2024. The dataset combines model data with observations from all over the world into a globally complete and consistent dataset using the laws of physics, and it will be downloaded from the main website of the climate data store. The dataset includes hourly observations for multiple variables, including the U and V components of wind, vertical velocity, temperature, relative humidity, and geopotential. These variables will be selected across 3 pressure levels at 250, 500 and 850 hPA, making our dataset containing 18 variables across 3 pressure levels. The data have been regridded to a regular grid of 0.25 degrees for the reanalysis, which means that for every 0.25

degrees of latitude and longitude, there will be a data sensor and, depending on the global region, there will be hundreds and hundreds of data points. The global operator for the global domain would contain a ginormous amount of data; we would then need to down-sample the global data to a 64 x 64 resolution, keeping the data in row resolution which means that the data point will be more scarce, allowing us to stack more layers of AFNO encoder. The local domain will retain high resolution, allowing us to have more fine-grained and detailed data, allowing us to capture phenomena and roll out accurate predictions.

B.2 Baselines and Metrics

We will compare AdaptFNO against baseline models such as Convolutional neural networks (CNNs) and Fourier Neural Operators (FNOs). The performance will be evaluated using a TemporalMSE with decaying weights, on the first day, the weights will be the highest in order to have the most accurate prediction and we will optimize the weight in the days later by decreasing it accordingly.

B.3 Training Setup

We downloaded the raw data from the climate data store and then pre-processed and downscaled the sample to later use for training. We want to forecast the future state of the weather variables based on historical data using an input window of 10 days of historical data, and we will predict 3 days into the future. The data is chronologically split into a training set and a validation set with the training set using the data from 1980 to 2022 and the validation set will be using data from 2023 to 2024. We will set the hyper parameter for the training with a learning rate of 5×10^{-4} and train for a maximum of 1000 epochs, an early stopping mechanism with a patience of 50 epochs which will stop training if the validation loss does not improve. We will be using a batch size of 4 for training and 2 for validating together with the optimizer Adam [3].