DPTF: Dual-Path Time-Frequency Framework for Long-Term Time Series Forecasting

Pengfei Ding†, Wei Ding†, Zehang Chen, Jingqi Gao, Lvqing Yang*, Fan Lin*
School of Informatics, *Xiamen University*
Xiamen, China
30920241154564@stu.xmu.edu.cn, dingwei@stu.xmu.edu.cn,
ChenZeha@outlook.com, 30920241154565@stu.xmu.edu.cn,
lqyang@xmu.edu.cn, iamafan@xmu.edu.cn

***Abstract*—Long-term time series forecasting plays a vital role in domains such as energy, transportation, and meteorology. However, real-world time series often exhibit fine-grained temporal variations alongside coarse-grained spectral patterns. Constrained by single-domain modeling paradigms, existing approaches fail to simultaneously capture these distinct yet complementary features, leading to sub-optimal performance. Motivated by the need to synergize these complementary perspectives, we propose the Dual-Path Time-Frequency (DPTF) framework, a novel architecture that effectively integrates local and global modeling. Specifically, the temporal branch adopts a hierarchical multi-scale multi-period design to capture the series' trend and seasonal components, thereby explicitly disentangling and fusing local features across different granularities and compositions. In parallel, the frequency branch employs a lightweight spectral transformation to extract global structural patterns. Finally, the two pathways are adaptively aggregated through a learnable weighting mechanism to produce robust forecasts. Extensive experiments on public benchmarks show that DPTF consistently outperforms state-of-the-art baselines, demonstrating superior accuracy and robustness in long-term forecasting tasks.**

***Index Terms*—Long-term time series forecasting, time-frequency analysis, multi-scale modeling, hierarchical fusion**

I. Introduction

Time series forecasting aims to infer future outcomes by analyzing historical data, playing a significant role in key application domains such as intelligent transportation [1], weather prediction [2], and energy consumption [3]. In practice, time series often combine short-term fluctuations with long-term periodic patterns. For example, electricity demand shows hourly variations due to daily activities alongside seasonal and monthly trends [4]. Capturing such long-term structures is crucial for strategic planning, motivating the growing focus on long-term time series forecasting (LTSF) [5]. Due to its unique characteristics, LTSF imposes stricter requirements than short-term forecasting, demanding models that effectively and efficiently capture long-range dependencies [6], [7]. Therefore, achieving a balance between accuracy, efficiency, and generalization has emerged as a critical challenge for both research and industry.

† Equal Contribution.
* Corresponding Author.

LTSF has long been studied using traditional statistical approaches such as ARIMA [8] and exponential smoothing [9]. These methods provide strong interpretability but exhibit substantial limitations when dealing with nonlinear, multivariate, and long-horizon forecasting tasks. The advent of deep learning introduced variants of RNNs [10]–[12] and CNNs [13], [14], which enhanced the ability to capture nonlinear dependencies. However, their limited receptive fields fundamentally restrict their ability to model long-range dependencies in ultra-long sequences. To address this core limitation, transformer-based [6], [15], [16] architectures, with their inherent global attention mechanism, emerged as a powerful alternative and achieved significant progress in LTSF. For instance, Autoformer [17] models periodicity through series decomposition and an autocorrelation mechanism, iTransformer [18] enhances multivariate forecasting with an inverted token representation, and PatchTST [19] expands the receptive field using subsequence patches and a channel-independent design. Despite their powerful performance, the deep architectures and high computational demands of these methods limit their practical deployment.

In pursuit of an optimal balance between efficiency and performance, a growing body of work has proposed lightweight models. For example, in the time domain, DLinear [7] uses a simple linear decomposition framework to achieve results comparable to complex models at a minimal cost. SparseTSF [20] leverages sparse forecasting and periodic priors to perform well under constrained budgets. AMD [21], TimeMixer [22] and TimeMixer++ [23] introduce multi-scale fusion and mixing mechanisms to capture intricate temporal dynamics efficiently. Meanwhile, in the frequency domain, WPMixer [24] and Multi-Order Wavelet Derivative Transform [25] introduce wavelet packet decomposition to achieve efficient multi-resolution mixing within distinct frequency bands. FiLM [26] exploits spectral features to efficiently capture long-term dependencies and global patterns, offering another viable path for LTSF.

However, most existing approaches are limited to a single modeling paradigm. Time-domain models excel at modeling sequential dependencies, but they often lack explicit mechanisms to capture and disentangle fine-grained spectral structures. In contrast, frequency-domain models can effectively extract global periodic components, yet they tend to sacrifice temporal localization by overlooking the evolving dynamics and contextual dependencies present in the original sequence. Consequently, relying on either paradigm alone is insufficient to fully characterize the complex patterns inherent in real-world time series. This limitation motivates the key objective of our work: developing a unified model for LTSF that effectively integrates information from both domains.

To this end, we propose the Dual-Path Time-Frequency (DPTF) framework, a novel architecture meticulously designed to synergize the complementary modeling paradigms of the temporal and frequency domains. The temporal branch introduces a hierarchical, multi-period architecture that explicitly disentangles and fuses patterns from diverse granularities. By processing decomposed trend and seasonal components through multi-scale analysis and coarse-to-fine mixing, it robustly captures complex local dynamics. The frequency branch leverages a real-valued Fast Fourier Transform and a learnable spectral projection to efficiently extract global structural patterns without incurring the full cost of inverse transforms. Finally, the two pathways are adaptively aggregated through a learnable weighting mechanism, allowing the model to learn the optimal balance between local patterns and global dynamics.

The main contributions of this work are summarized as follows:

1) An Effective Dual-Path Architecture: We propose DPTF, a dual-path time–frequency framework for LTSF. By processing information through parallel temporal and frequency branches, DPTF synergizes fine-grained local patterns and coarse-grained global dynamics within a unified model, addressing the limitations of single-domain paradigms.

2) Key Mechanisms for Robust Forecasting: We design a hierarchical, multi-period temporal module that explicitly disentangles patterns at diverse granularities, complemented by a lightweight frequency module that efficiently captures global dynamics via a learnable spectral projection, bypassing costly inverse transforms. Furthermore, an adaptive fusion mechanism automatically balances the contributions of both paths to enhance robustness.

3) Comprehensive Empirical Validation: Extensive experiments on seven real-world benchmarks demonstrate that DPTF outperforms existing methods. Ablation studies further verify the effectiveness of the dual-path design and the complementary roles of both branches.

II. METHOD

A. Preliminary

Time series forecasting aims to predict future values, denoted as $Y = \{x_{T+1}, x_{T+2}, \ldots, x_{T+L}\} \in \mathbb{R}^{L\times C}$, given a set of historical observations, $X = \{x_1, x_2, \ldots, x_T\} \in \mathbb{R}^{T\times C}$. Here, T is the length of the lookback window, L is the prediction horizon, and C is the number of channels. Building on this general formulation, the goal of long-term time series forecasting (LTSF) is to generate accurate predictions over an extended horizon L.

B. Overall Architecture

Our model, the Dual-Path Time-Frequency (DPTF) framework, is illustrated in Fig. 1. Given an input sequence $X \in \mathbb{R}^{T\times C}$, DPTF first performs instance normalization to stabilize its distribution. This involves subtracting the temporal mean μ_{seq} from the series:

$$\mu_{\text{seq}} = \frac{1}{T}\sum_{t=1}^{T} x_t, \quad \bar{X} = X - \mu_{\text{seq}}, \tag{1}$$

where $\bar{X} \in \mathbb{R}^{T\times C}$ is the normalized sequence and $x_t \in \mathbb{R}^C$ is the vector of variables at time t, $t \in \{1, \ldots, T\}$.

The normalized sequence $\bar{X}$ is then processed in parallel by two specialized branches. The **temporal branch** captures local patterns by decomposing the series and employing multi-scale, multi-period processing with hierarchical mixing to produce a normalized forecast Y_{time}. Meanwhile, the **frequency branch** captures global dynamics via a real-valued Fast Fourier Transform (rFFT) and a lightweight spectrum projection module to generate its normalized forecast Y_{freq}.

Finally, the outputs of the two branches are adaptively fused to produce the final prediction $Y \in \mathbb{R}^{L\times C}$:

$$Y = \alpha \cdot Y_{\text{time}} + (1-\alpha)\cdot Y_{\text{freq}}, \tag{2}$$

where Y_{time} and Y_{freq} have already incorporated the sequence mean μ_{seq} in their computation. The fusion weight $\alpha \in [0, 1]$ is a scalar learnable parameter, allowing the model to automatically balance the contributions from both branches during training.

C. Temporal Branch

The normalized input sequence $\bar{X}$ is first processed by the series decomposition block from Autoformer [17], which separates the sequence into its trend and seasonal components, denoted as $\bar{X}_{\text{trend}}$ and $\bar{X}_{\text{seasonal}}$, respectively:

$$\bar{X}_{\text{trend}}, \bar{X}_{\text{seasonal}} = SeriesDecomp(\bar{X}). \tag{3}$$

Both components are then processed in parallel by the subsequent modules.

Multi-Scale Multi-Period Downsampling. To capture dynamics at different resolutions, we first generate a set of $S+1$ multi-scale representations, denoted as $\bar{X}_{\text{comp}} = \{\bar{X}^{(0)}_{\text{comp}}, \bar{X}^{(1)}_{\text{comp}}, \ldots, \bar{X}^{(S)}_{\text{comp}}\}$. As the operations are identical for both, we describe them here for a generic component, denoted by the subscript $comp$ (where $comp$ can be $trend$ or $seasonal$). The base representation $\bar{X}^{(0)}_{\text{comp}}$ is the original component. Each subsequent representation $\bar{X}^{(s)}_{\text{comp}}$ for the scale index $s \in \{0, 1, \ldots, S\}$ is obtained via an average pooling operation with a stride of 2:

$$\bar{X}^{(s)}_{\text{comp}} = \text{AvgPool}(\bar{X}^{(s-1)}_{\text{comp}}), \tag{4}$$

where $\bar{X}^{(s)}_{\text{comp}} \in \mathbb{R}^{\lfloor \frac{T}{2^s} \rfloor \times C}$.

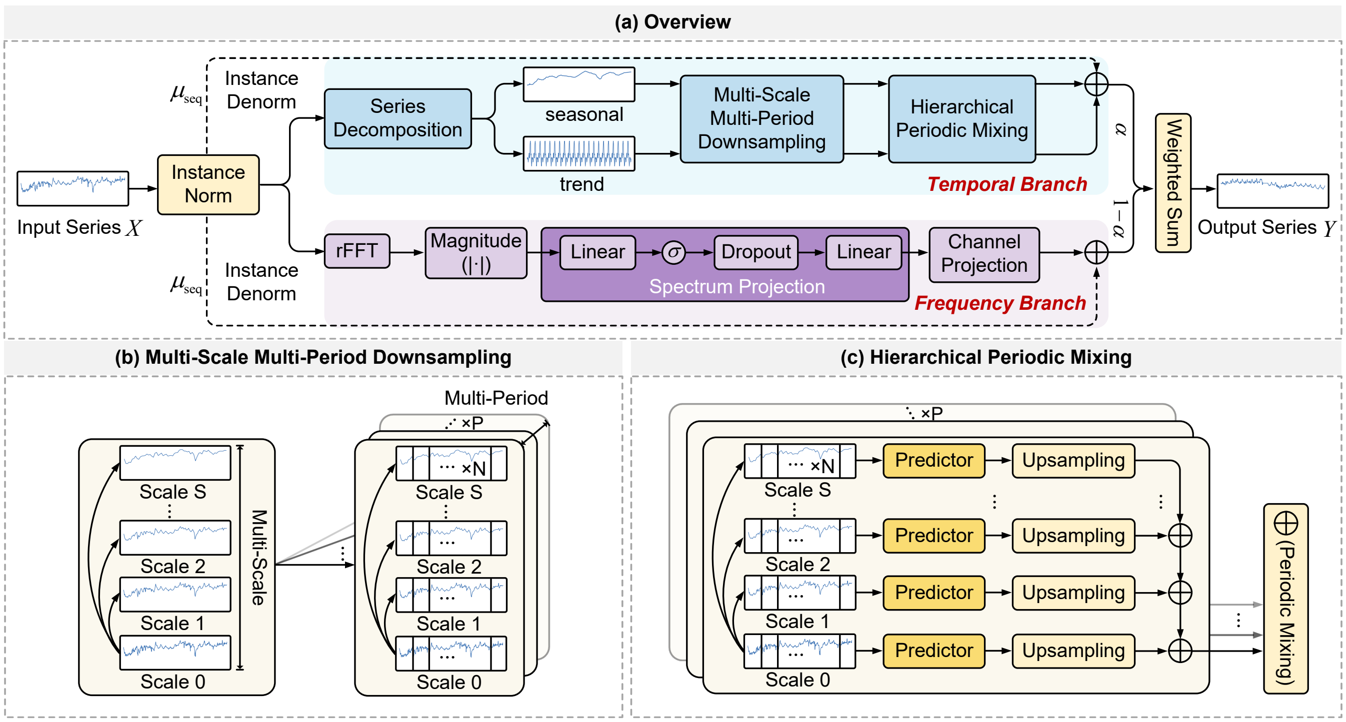

Fig. 1: The overall architecture of the DPTF framework.

Next, to analyze periodic patterns, each multi-scale sequence $\bar{X}^{(s)}_{\text{comp}}$ is segmented based on a set of predefined periods $P = \{p_1, \dots, p_K\}$. For a given scale s and period $p_k \in P$, the sequence is reshaped into a tensor of subsequences, denoted as: $\bar{X}^{(s,p_k)}_{\text{comp}} = \{Z_0^{(s,p_k)}, Z_1^{(s,p_k)}, \dots, Z_N^{(s,p_k)}\}$, $\bar{X}^{(s,p_k)}_{\text{comp}} \in \mathbb{R}^{M\times N\times C}$. It is composed of $N = \left\lfloor \frac{p_k}{2^s} \right\rfloor$ total segments, where each segment $Z^{(s,p_k)} \in \mathbb{R}^{M\times C}$ represents a periodic subsequence. The length of each subsequence, $M = \left\lfloor \frac{T}{p_k} \right\rfloor$, corresponds to the period's length at that specific scale. This dual strategy of multi-scale and multi-period analysis enables the model to capture both fine-grained local variations and long-range periodic trends synergistically, resulting in more robust and accurate forecasts.

Hierarchical Periodic Mixing. Recent studies have shown the effectiveness of sparse forecasting techniques [20], which capture periodicity in time series with a very low parameter count. Inspired by this, our approach also segments the time series into historical blocks based on predefined periods and performs predictions on these blocks.

We apply a dedicated Multi-Layer Perceptron (MLP)-based predictor to the corresponding tensor of historical segments, $\bar{X}^{(s,p_k)}_{\text{comp}}$. Critically, aligning with the sparse forecasting paradigm, the predictor maps the sequence of historical segments to a sequence of future segments. The process can be formulated as:

$$\hat{X}^{(s,p_k)}_{\text{comp}} = \text{Predictor}^{(s,p_k)}(\bar{X}^{(s,p_k)}_{\text{comp}}), \tag{5}$$

where $\hat{\mathbf{X}}^{(s,p_k)}_{\text{comp}} \in \mathbb{R}^{L_k\times N\times C}$. $L_k = \left\lfloor \frac{L}{p_k} \right\rfloor$ is the number of future segments required to cover the full prediction horizon. The *Predictor* itself consists of two linear layers with a GELU activation and dropout regularization in between.

To integrate the multi-scale predictions, we employ a coarse-to-fine hierarchical fusion strategy. Starting with the prediction from the coarsest scale($s = S$), we progressively upsample it and merge it with the prediction from the next finer scale. First, the tensor of predicted future segments $\hat{\mathbf{X}}^{(s,p_k)}_{\text{comp}}$ is reshaped back into a series of periodic subsequence $Y^{(s,p_k)}_{\text{comp}} = \{\hat{Y}^{(s,p_k)}_{\text{comp},0}, \dots, \hat{Y}^{(s,p_k)}_{\text{comp},L_k}\}$, where $Y^{(s,p_k)}_{\text{comp}} \in \mathbb{R}^{L_k\times N\times C}$, $\hat{Y}^{(s,p_k)}_{\text{comp}} \in \mathbb{R}^{N\times C}$. Then, the iterative fusion process is defined for $s = S, S-1, \dots, 1$ as:

$$Y^{(s-1,p_k)}_{\text{comp}} = \frac{1}{2}\left(Y^{(s-1,p_k)}_{\text{comp}} + \text{Upsampling}(Y^{(s,p_k)}_{\text{comp}})\right), \tag{6}$$

where the Upsampling consists of two linear layers with an activation function and dropout regularization in between.

This recursive refinement mechanism enables the model to progressively embed long-range dependencies from coarser scales into finer-scale predictions. The final output of this process for each period p_k is a pair of fused forecasts, denoted as $Y^{p_k}_{\text{trend}}$ and $Y^{p_k}_{\text{seasonal}}$.

To produce a single, robust forecast, these per-period predictions are aggregated by averaging across all K periods to yield the outputs Y_{trend} and Y_{seasonal}. Finally, these two components are combined with the sequence mean μ_{seq} to produce the prediction of the temporal branch:

$$Y_{\text{time}} = Y_{\text{trend}} + Y_{\text{seasonal}} + \mu_{\text{seq}}. \tag{7}$$

D. Frequency Branch

To capture global periodic patterns, the frequency-domain branch processes the normalized input sequence $\bar{X}$. First, real-valued Fast Fourier Transform (rFFT) is applied, and the magnitude spectrum is extracted to serve as the frequency-domain features:

$$\bar{X}_f = \left|\text{rFFT}(\bar{X})\right|, \tag{8}$$

where $\bar{X}_f \in \mathbb{R}^{\lfloor \frac{T}{2}+1 \rfloor \times C}$. The extracted magnitude spectrum $\bar{X}_f$, is then enhanced and projected back to the temporal resolution through a lightweight spectrum projection network consisting of two linear layers with a ReLU activation and dropout:

$$\hat{X}_f = \text{Linear}_2\Big(\text{Dropout}\big(ReLU(\text{Linear}_1(\bar{X}_f))\big)\Big). \tag{9}$$

Finally, a channel projection layer, implemented as a 1-D pointwise convolution, maps the enhanced spectral representation to the output channel dimension, and the sequence mean μ_{seq} is added to yield the frequency-domain forecast:

$$Y_{\text{freq}} = \text{ChannelProj}(\hat{X}_f) + \mu_{\text{seq}}. \tag{10}$$

III. Experiments

A. Experimental Setup

1) Datasets: To evaluate the proposed DPTF, we conduct extensive experiments on seven real-world benchmark datasets [6], [17], including ETT(ETTh1, ETTh2, ETTm1, ETTm2), Weather, Electricity, Traffic for long-term forecasting. The details of datasets are given in table I.

TABLE I: Detailed dataset descriptions.

Dataset	Dimension	Dataset Size	Frequency
ETTh1	7	(8445, 2785, 2785)	Hourly
ETTh2	7	(8545, 2881, 2881)	Hourly
ETTm1	7	(34369, 11425, 11425)	15 min
ETTm2	7	(34369, 11425, 11425)	15 min
Weather	21	(36696, 5175, 10440)	10 min
Electricity	321	(18221, 2537, 5165)	Hourly
Traffic	862	(12089, 1661, 3413)	Hourly

2) Baselines: We compare DPTF with 7 baselines, which comprise the state-of-the-art long-term forecasting models including AMD [21], TimeMixer [22], SparseTSF [20], iTransformer [18], PatchTST [19], DLinear [7], FiLM [26].

3) Implementation details: All experiments are conducted using PyTorch on a single NVIDIA RTX A6000 GPU with 48GB of memory. We use mean squared error (MSE) as the loss function and Adam as the optimizer, with an initial learning rate η sampled log-uniformly from $[10^{-5}, 10^{-2}]$. The batch size is tuned over $\{32, 64, 128, 256, 512, 1024\}$ to explore different training scales, and early stopping is applied if the validation performance does not improve within 30 epochs. For the proposed DPTF, the structural hyperparameters are configured to match the dataset properties. Specifically, we employ a hierarchical structure where the downsampling factors are set to $\{2, 4, 8\}$, and the period set P is configured to match the dominant granularities of each dataset(e.g., $\{24, 168\}$ for hourly data). Hyperparameters for baselines are set according to their original papers to ensure a fair comparison. All the experiments are conducted 5 times to produce the average results.

4) Evaluation Metrics: Following standard practice, we use Mean Squared Error (MSE) and Mean Absolute Error (MAE) as our evaluation metrics.

B. Main Results

As reported in Table II, the proposed DPTF model is comprehensively evaluated against competitive baselines on long-term time series forecasting tasks. In terms of best counts, DPTF achieves 38 occurrences, achieving the highest number of best results. More importantly, DPTF provides substantial improvements over TimeMixer, AMD and SparseTSF. Specifically, on the ETTh1 dataset, DPTF yields average MSE reductions of **21.60%**, **11.18%** and **10.03%** compared with TimeMixer, AMD, and SparseTSF, respectively. On the ETTh2 dataset, the reductions reach **20.41%**, **16.75%** and **11.49%**. Furthermore, DPTF also demonstrates consistently strong performance across other datasets, except for the Weather dataset, where its performance is less competitive. This phenomenon is attributable to the inherent characteristics of the Weather dataset, which exhibits ultra-long-term seasonality (e.g., annual cycles) relative to its high sampling rate, significantly exceeding the receptive field of the lookback window. Consequently, DPTF's mechanism, which prioritizes the explicit disentanglement of granular periodic patterns, may exhibit diminished efficacy when the dominant periodicity is unobservable. In contrast, simpler baselines like DLinear [7] or decomposition-based methods like AMD [21] may perform better in such scenarios by focusing primarily on trend fitting or multi-scale fusion without enforcing strict periodic matching. This suggests that while DPTF excels in capturing complex, entangled temporal dynamics (as seen in Traffic and ETT datasets), simpler paradigms may offer greater robustness on datasets dominated by smooth, ultra-long-range trends. Overall, these results clearly validate the modeling capability of DPTF.

C. Ablation Study

To evaluate the effectiveness of our dual-branch architecture, we conducted three ablation experiments: 1) the original DPTF model with both temporal and frequency branches; 2) removing the frequency branch while keeping the temporal branch; 3) removing the temporal branch while keeping the frequency branch. As shown in Table III, the original dual-branch DPTF consistently achieves the best performance across all settings. We observe that the stand-alone Temporal branch significantly outperforms the stand-alone Frequency branch, indicating that explicit multi-scale disentanglement of local trend and seasonality serves as the foundational backbone for accurate forecasting. However, the Frequency branch, despite its weaker individual performance due to the loss of temporal context, plays a critical role as a global

Models		DPTF (Ours)		AMD (2025)		TimeMixer (2024)		SparseTSF (2024)		iTransformer (2024)		PatchTST (2023)		DLinear (2023)		FiLM (2022)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.334**	**0.374**	0.370	0.398	0.378	0.399	0.362	0.388	0.386	0.404	0.378	0.398	0.378	0.397	0.422	0.431
	192	**0.349**	**0.384**	0.403	0.420	0.440	0.430	0.403	0.411	0.441	0.436	0.424	0.431	0.420	0.426	0.462	0.458
	336	**0.360**	**0.398**	0.422	0.431	0.499	0.459	0.434	0.428	0.490	0.462	0.470	0.457	0.446	0.442	0.501	0.483
	720	**0.419**	**0.442**	0.451	0.461	0.548	0.510	0.426	0.447	0.509	0.493	0.524	0.507	0.486	0.486	0.545	0.526
ETTh2	96	**0.246**	**0.320**	0.278	0.343	0.288	0.342	0.294	0.346	0.299	0.350	0.291	0.345	0.292	0.358	0.326	0.371
	192	**0.303**	**0.360**	0.363	0.397	0.378	0.396	0.339	0.377	0.381	0.398	0.378	0.403	0.389	0.420	0.391	0.415
	336	**0.295**	**0.367**	0.380	0.418	0.428	0.434	0.359	0.397	0.420	0.430	0.424	0.440	0.454	0.463	0.415	0.440
	720	**0.373**	**0.422**	0.441	0.466	0.435	0.446	0.383	0.424	0.446	0.454	0.435	0.453	0.587	0.552	0.435	0.455
ETTm1	96	0.306	0.355	0.288	0.343	0.318	0.358	0.312	0.362	0.341	0.376	0.323	0.364	**0.287**	**0.336**	0.312	0.357
	192	0.331	0.368	**0.328**	0.365	0.367	0.388	0.347	0.381	0.382	0.395	0.370	0.391	0.330	**0.361**	0.342	0.369
	336	**0.355**	0.385	0.365	**0.381**	0.390	0.403	0.373	0.398	0.418	0.418	0.399	0.407	0.368	0.382	0.369	0.385
	720	**0.395**	**0.409**	0.422	0.417	0.453	0.441	0.430	0.426	0.487	0.456	0.458	0.444	0.427	0.418	0.418	0.413
ETTm2	96	**0.161**	**0.254**	0.167	0.259	0.175	0.262	0.178	0.270	0.185	0.272	0.185	0.267	0.193	0.281	0.180	0.269
	192	**0.217**	0.298	0.221	**0.294**	0.237	0.298	0.232	0.306	0.253	0.313	0.250	0.310	0.234	0.312	0.223	0.296
	336	**0.266**	0.336	0.270	**0.327**	0.294	0.335	0.284	0.342	0.315	0.350	0.311	0.349	0.281	0.347	0.276	0.331
	720	**0.334**	**0.379**	0.355	0.381	0.412	0.411	0.360	0.389	0.413	0.406	0.423	0.415	0.387	0.418	0.362	0.391
Weather	96	0.168	0.224	0.148	**0.202**	0.162	0.209	0.169	0.223	0.175	0.226	0.175	0.218	**0.147**	0.213	0.203	0.282
	192	0.216	0.266	0.192	**0.243**	0.207	0.251	0.214	0.262	0.228	0.269	0.221	0.256	**0.191**	0.258	0.234	0.296
	336	0.255	0.295	**0.242**	**0.281**	0.262	0.292	0.257	0.293	0.280	0.298	0.279	0.298	0.243	0.299	0.267	0.326
	720	**0.310**	0.342	0.315	**0.332**	0.344	0.345	0.321	0.343	0.357	0.349	0.356	0.348	0.320	0.361	0.323	0.362
Electricity	96	**0.130**	**0.225**	**0.130**	0.227	0.156	0.247	0.134	0.226	0.147	0.250	0.180	0.273	0.134	0.232	0.165	0.268
	192	**0.139**	0.241	0.150	0.244	0.169	0.260	0.148	**0.240**	0.172	0.268	0.187	0.279	0.149	0.247	0.173	0.259
	336	0.169	0.269	0.166	0.262	0.186	0.277	**0.164**	**0.258**	0.195	0.302	0.204	0.295	0.166	0.265	0.192	0.281
	720	**0.199**	0.293	0.200	0.292	0.227	0.312	0.201	**0.290**	0.259	0.331	0.245	0.328	0.202	0.299	0.238	0.337
Traffic	96	**0.371**	**0.264**	0.386	0.277	0.468	0.293	0.383	0.269	0.393	0.268	0.458	0.298	0.436	0.307	0.425	0.298
	192	**0.387**	**0.267**	0.402	0.282	0.509	0.301	0.395	0.281	0.413	0.279	0.468	0.300	0.451	0.314	0.437	0.311
	336	**0.402**	**0.277**	0.412	0.288	0.524	0.313	0.410	0.294	0.427	0.290	0.483	0.307	0.464	0.322	0.452	0.338
	720	**0.447**	**0.306**	0.449	0.307	0.559	0.323	0.452	0.318	0.458	0.312	0.517	0.326	0.495	0.340	0.517	0.359
Best Count		**38**		10		0		4		0		0		5		0	

TABLE II: Long-term time series forecasting results. All the results are averaged from 4 different prediction lengths $L \in \{96, 192, 336, 720\}$, the input sequence length T is searched among {96, 336, 512, 720, 864, 1440, 1680}. The best results are highlighted in red, while the second-best results are underlined. A lower MSE or MAE indicates a better prediction.

Dataset			ETTh1				ETTh2			
T	F	Metrics	96	192	336	720	96	192	336	720
✓	✓	MSE	**0.334**	**0.349**	**0.360**	**0.419**	**0.246**	**0.303**	**0.295**	**0.373**
		MAE	**0.374**	**0.384**	**0.398**	**0.442**	**0.320**	**0.360**	**0.367**	**0.422**
✓	×	MSE	0.350	0.369	0.383	0.443	0.263	0.314	0.323	0.384
		MAE	0.383	0.405	0.409	0.454	0.327	0.367	0.380	0.445
×	✓	MSE	0.493	0.498	0.490	0.533	0.336	0.366	0.342	0.413
		MAE	0.428	0.440	0.446	0.477	0.378	0.382	0.399	0.458

TABLE III: Ablation of Temporal branch (T) and Frequency branch (F). A check mark ✓ and a cross mark × indicate the presence and absence of the branch, respectively. The table shows the MSE and MAE results on the ETTh dataset for different prediction lengths $L \in \{96, 192, 336, 720\}$. The best results are shown in **bold**.

Dataset		ETTm1				ETTm2			
Models	Metrics	960	1200	1440	1680	960	1200	1440	1680
DPTF	MSE	**0.407**	**0.416**	**0.420**	**0.427**	**0.352**	**0.360**	**0.346**	**0.344**
	MAE	**0.417**	**0.419**	**0.424**	**0.435**	0.397	**0.406**	**0.401**	**0.406**
TimeMixer	MSE	0.484	0.502	0.520	0.527	0.442	0.467	0.477	0.470
	MAE	0.458	0.471	0.483	0.488	0.424	0.446	0.460	0.455
AMD	MSE	0.453	0.471	0.485	0.497	0.382	0.402	0.407	0.398
	MAE	0.433	0.443	0.450	0.457	0.401	0.416	0.421	0.419
SparseTSF	MSE	0.432	0.444	0.450	0.450	0.374	0.392	0.386	0.377
	MAE	0.421	0.426	0.429	0.436	**0.396**	0.409	0.410	0.408

TABLE IV: The table shows the MSE and MAE results on the ETTm dataset for different ultra-long prediction lengths $L \in \{960, 1200, 1440, 1680\}$. The best results are shown in **bold**.

regularizer. This demonstrates that while the temporal branch excels at capturing short-term dynamics, the frequency branch becomes increasingly vital for preserving long-range global structures where local temporal correlations weaken. Thus, the integration of both branches is not merely additive but complementary, ensuring robustness across varying horizons.

D. Ultra-long-term Time Series Forecasting

To assess the robustness of DPTF on ultra-long prediction horizons, we selected the strongest baselines from the main experiments—TimeMixer, AMD, and SparseTSF—for comparison. As shown in Table IV, DPTF achieves average MSE reductions of **21.15%**, **12.07%** and **7.14%** on the ETTm dataset compared to these methods, demonstrating its superior predictive performance. This robustness on ultra-long horizons is largely attributed to the frequency branch, which extracts global spectral invariants that remain stable over extended timeframes. Unlike purely time-domain baselines that may suffer from error accumulation or information loss across long distances, DPTF effectively preserves the macroscopic structure of the series via spectral projection, ensuring consistent accuracy even as the prediction horizon expands.

IV. Conclusion

We propose DPTF, a dual-branch framework for long-term time series forecasting that jointly models temporal and frequency domains. By bridging the gap between fine-grained temporal variations and coarse-grained spectral patterns, DPTF addresses the limitations of single-paradigm approaches. Its temporal branch captures local dependencies by explicitly modeling multi-scale and multi-period dynamics, while the frequency branch extracts complementary global dynamics to prevent long-term information loss. An adaptive fusion mechanism balances both representations to enhance robustness. Extensive experiments show DPTF consistently outperforms strong baselines, demonstrating exceptional stability particularly in ultra-long forecasting scenarios. Acknowledging the sensitivity to predefined periods observed in specific datasets (e.g., Weather), future work will explore adaptive period learning mechanisms and extensions to irregular sampling to further enhance generalization.

References

[1] Ling Cai, Krzysztof Janowicz, Gengchen Mai, Bo Yan, and Rui Zhu, "Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting," *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, 2020. I

[2] Rafal A Angryk, Petrus C Martens, Berkay Aydin, Dustin Kempton, Sushant S Mahajan, Sunitha Basodi, Azim Ahmadzadeh, Xumin Cai, Soukaina Filali Boubrahimi, Shah Muhammad Hamdi, et al., "Multivariate time series dataset for space weather data analytics," *Scientific data*, vol. 7, no. 1, pp. 227, 2020. I

[3] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017. I

[4] Tao Hong and Shu Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016. I

[5] Bryan Lim and Stefan Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, pp. 20200209, 2021. I

[6] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 11106–11115. I, III-A

[7] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu, "Are transformers effective for time series forecasting?," in *Proceedings of the AAAI conference on artificial intelligence*, 2023, vol. 37, pp. 11121–11128. I, III-A, III-B

[8] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015. I

[9] Peter R Winters, "Forecasting sales by exponentially weighted moving averages," *Management science*, vol. 6, no. 3, pp. 324–342, 1960. I

[10] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. I

[11] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014. I

[12] Xueli Zhang, Cankun Zhong, Jianjun Zhang, Ting Wang, and Wing WY Ng, "Robust recurrent neural networks for time series forecasting," *Neurocomputing*, vol. 526, pp. 143–157, 2023. I

[13] Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, et al., "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, vol. 12, pp. 1, 2016. I

[14] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager, "Temporal convolutional networks for action segmentation and detection," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 156–165. I

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017. I

[16] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International conference on machine learning*. PMLR, 2022, pp. 27268–27286. I

[17] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in neural information processing systems*, vol. 34, pp. 22419–22430, 2021. I, II-C, III-A

[18] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long, "itransformer: Inverted transformers are effective for time series forecasting," *arXiv preprint arXiv:2310.06625*, 2023. I, III-A

[19] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022. I, III-A

[20] Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang, "Sparsetsf: Modeling long-term time series forecasting with 1k parameters," *arXiv preprint arXiv:2405.00946*, 2024. I, II-C, III-A

[21] Yifan Hu, Peiyuan Liu, Peng Zhu, Dawei Cheng, and Tao Dai, "Adaptive multi-scale decomposition framework for time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, vol. 39, pp. 17359–17367. I, III-A, III-B

[22] Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and Jun Zhou, "Timemixer: Decomposable multiscale mixing for time series forecasting," *arXiv preprint arXiv:2405.14616*, 2024. I, III-A

[23] Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin, "Timemixer++: A general time series pattern machine for universal predictive analysis," *arXiv preprint arXiv:2410.16032*, 2024. I

[24] Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz, "Wpmixer: Efficient multi-resolution mixing for long-term time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, vol. 39, pp. 19581–19588. I

[25] Ziyu Zhou, Jiaxi Hu, Qingsong Wen, James T Kwok, and Yuxuan Liang, "Multi-order wavelet derivative transform for deep time series forecasting," *arXiv preprint arXiv:2505.11781*, 2025. I

[26] Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al., "Film: Frequency improved legendre memory model for long-term time series forecasting," *Advances in neural information processing systems*, vol. 35, pp. 12677–12690, 2022. I, III-A