Transactions on Computational and Scientific Methods | Vo. 4, No. 10, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Frequency-Guided Transformer Modeling with Spectral Enhancement for Accurate Time Series Forecasting

Yiyun Lyu[1], Jiejun Jiang[1], Jiahui Hu[1]*

[1]Northeastern University, Seattle, USA

*Corresponding author: lareinahuwork@gmail.com

Abstract: This paper proposes a forecasting framework that integrates frequency-aware spectral enhancement and frequency-guided attention mechanisms to address the challenges of insufficient modeling of periodic structures and interference from redundant information in time series prediction. The method first applies a frequency-aware spectral enhancement module to perform Fourier Transform on the input sequence. It selects and amplifies key frequency components to highlight dominant periodic information and suppress the influence of redundant disturbances. Then, during the encoding phase, a frequency-guided attention mechanism is introduced. It uses spectral response to guide attention distribution, allowing the model to focus more on high-weight structural segments when constructing contextual representations. This improves the ability to model cross-period dependencies and long-term relationships. Both modules are modular and structurally independent. They can be flexibly integrated into various Transformer architectures to enhance structural representation in feature extraction and modeling. Experimental results on multiple real-world time series datasets demonstrate that the proposed method improves prediction accuracy across several evaluation metrics while maintaining computational efficiency, showing strong adaptability to complex periodic structures and dynamic fluctuations.

Keywords: Time series prediction, frequency domain modeling, spectral enhancement, attention mechanism

1. Introduction

In key areas such as intelligent manufacturing, financial risk control, and energy scheduling, accurate time series forecasting plays a vital role in ensuring system stability and improving decision-making efficiency. With the rapid growth of data volume and increasing complexity of sequence structures, traditional statistical methods and shallow models are no longer sufficient for modeling high-dimensional and non-stationary time series. In recent years, deep learning methods, especially Transformer-based models, have attracted extensive attention due to their strong sequence modeling capabilities[1]. These methods have brought significant improvements to time series forecasting. However, effectively capturing periodic structures and long-term dependencies remains a central research challenge.

Although previous studies have attempted to introduce attention mechanisms, multi-scale modeling, or frequency-domain transformations to enhance the representation of complex structures, two main problems remain. First, the perception and utilization of frequency-domain information in existing models are still coarse[2]. These models fail to distinguish key periodic components from noise, leading to redundant feature learning. Second, attention mechanisms often suffer from attention dispersion when handling long sequences. They cannot focus on structural key segments, making it difficult to construct high-quality contextual representations. These limitations result in unstable model performance when dealing with sequences that exhibit frequent periodic changes or structural disturbances[3,4].

To address these issues, this paper proposes a time series forecasting framework that integrates frequency-aware spectral enhancement and frequency-guided attention mechanisms. The framework aims to strengthen the modeling of key periodic features through frequency-domain guidance while suppressing interference from redundant information[5]. It further uses spectral awareness to guide attention distribution, allowing the model to focus more effectively on structurally important segments during context representation generation. This leads to improved prediction accuracy and stability.

The contributions of this work are twofold. First, a frequency-aware spectral enhancement module is designed. It applies the Fourier Transform and selective spectral enhancement to extract key components with periodic structures and remove redundant noise[6]. Second, a frequency-guided attention mechanism is proposed. It uses spectral response information as a guiding signal for attention allocation, making attention focus more directional and structured. This improves the effectiveness of long-term dependency modeling and enhances the discriminability of contextual features. The proposed method offers good scalability and modularity, providing a new solution for modeling complex time series tasks[7].

2. Related work

2.1 Deep Learning for System Load Time Series Forecasting

In backend system load forecasting tasks, deep learning models have gradually replaced traditional statistical approaches due to their powerful feature extraction and nonlinear modeling capabilities. Early studies often employed recurrent neural networks to capture load variation trends, including standard recurrent structures and gated variants. These models modeled temporal dependencies through the transmission of hidden states. They performed well in modeling short-term dynamics but often suffered from vanishing gradients or memory decay when handling long-range dependencies, leading to performance degradation. Moreover, these models usually relied on fixed-length time windows, limiting their ability to represent multi-scale variations. This made them less effective in capturing the sudden fluctuations and structural shifts frequently observed in backend systems[8,9].

To address the limited capability in modeling long-term dependencies, attention-based architectures such as the Transformer have emerged in recent years. These models compute global dependencies between arbitrary time steps using self-attention mechanisms. This significantly improves the modeling of long-range temporal relationships. In backend systems, load variation is often driven by multiple interacting factors with strong nonlinearity and cross-dependencies. Transformer architectures can flexibly focus on key moments during modeling and effectively capture latent dependency structures[10]. Unlike traditional recurrent models, Transformer supports parallel computation and scalable sequence modeling. It is particularly suitable for high-frequency data and multi-source indicators, making it a strong candidate for complex system load forecasting tasks.

Despite the clear advantages of Transformer, its application in backend load forecasting still faces several challenges[11]. On one hand, load sequences typically exhibit strong periodicity, trends, and localized anomalies. Although Transformer can model global dependencies across time steps, it lacks an explicit mechanism to model periodic structures. This may lead to perception bias when capturing long-term trends. On the other hand, various forms of asynchronous behavior exist among backend indicators. Operations such as resource scheduling, delays, and container migration often cause asynchronous responses in load variation. Pure time-domain attention mechanisms struggle to fully interpret such cross-timescale interactions. Moreover, the high computational cost of the Transformer on long, high-frequency sequences raises concerns about modeling efficiency[12,13].

To further enhance deep models in backend load forecasting, recent studies have attempted to integrate the Transformer with other architectures. These include multi-scale attention modules, residual decomposition strategies, and graph neural networks. Such extensions partly alleviate the representational limitations by improving the model's ability to capture multi-dimensional dependencies and non-stationary dynamics. However, most of these approaches still rely heavily on time-domain information to construct feature sequences. They often fail to leverage the potential of frequency-domain periodic patterns in load data. In scenarios where load behavior changes frequently and is significantly influenced by system scheduling policies, purely time-domain modeling is insufficient. There is an urgent need to incorporate modeling mechanisms with greater sensitivity to frequency-domain information[14], [15].

2.2 Frequency-based Modeling for Time Series Analysis

Frequency-domain modeling has long played a crucial role in time series analysis across fields such as signal processing, speech recognition, and system diagnostics. From the frequency perspective, a time series is re-expressed as a combination of different frequency components. This enables models to extract more interpretable features from periodic patterns, oscillatory behavior, and hidden structural fluctuations. In load sequences with strong non-stationarity or evident multi-scale variations, frequency-domain modeling helps reduce noise interference and local anomalies in the time domain. By capturing dominant frequencies explicitly, it improves the robustness of sequence modeling. In backend systems, key indicators such as service traffic, container scheduling load, and network bandwidth usage often exhibit complex periodic structures and multi-frequency overlays[16]. Traditional time-domain methods struggle to accurately characterize these internal oscillations and resonant behaviors.

In practical implementation, frequency-domain modeling typically employs techniques such as the Fourier Transform, Wavelet Transform, or Empirical Mode Decomposition to map raw time series into the frequency space. Fourier Transform, as a basic frequency-domain tool, decomposes global signals into sinusoidal functions to extract stable periodic features. Wavelet Transform offers good time-frequency localization, allowing it to detect local mutations in non-stationary signals. This is particularly useful for modeling sudden load disturbances in backend systems. With advancements in computing resources and deep learning frameworks, many studies have begun integrating frequency-domain mechanisms into neural network architectures[17]. These models process both time-domain and frequency-domain signals, enabling a more comprehensive understanding and fusion of features.

In recent years, some deep models for time series forecasting have explicitly incorporated frequency-based designs. Methods such as spectral enhancement, band decomposition, or frequency-guided attention have been used to improve the expression of periodic structures. These models often convert the input sequence to the frequency domain and use the spectral distribution to guide the subsequent modeling. This allows differentiated processing of information from different frequency bands. Compared with traditional convolutional or recurrent models based on fixed windows, this frequency-aware paradigm shows clear advantages in modeling long-term dependencies, stable cycles, and subtle structural changes. In scenarios where load fluctuations follow patterns such as weekday/weekend rhythms, business peak and off-peak cycles, and mid-scale scheduling transitions, frequency-domain modeling offers a complementary approach to time-domain strategies[18,19].

In backend applications, frequency-domain modeling improves the ability to identify periodic load trends. It also helps detect spectral disturbances caused by system interventions, resource contention, or network jitter. When system behavior is affected by high-frequency external shocks or local scheduling, the frequency view enables the model to separate stable trends from abnormal components. This enhances both the accuracy and interpretability of load forecasting. Additionally, frequency-domain modeling supports time-scale compression and feature dimensionality reduction. This allows the model to retain key information while reducing computational cost and improving generalization. The frequency-dominant modeling approach complements the global modeling capacity of Transformer-like architectures and provides a solid structural foundation for building predictive models for complex systems[20].

Recent studies also show that reliable backend forecasting benefits from broader system-level modeling principles. Neural execution verification strengthens the trustworthiness of LLM-agent-oriented data pipeline orchestration, indicating that prediction models should consider operational consistency in addition to point accuracy [21]. Risk-constrained reinforcement learning for cloud-native elastic scaling further emphasizes the connection between forecasting outputs and SLO-aware resource decisions [22]. Cross-period anomaly detection and robust low-resource data mining highlight contrastive representation learning and semi-supervised optimization under temporal distribution shifts [23], [24]. In addition, differentiable graph-based decision modeling provides a structural basis for representing dependency relationships among complex system states [25].

3. Method

This study proposes a load forecasting framework that integrates frequency-domain modeling with a Transformer architecture to address the multi-scale periodic structures and highly dynamic dependencies in the backend system load time series. The method is built around two key innovations. First, a Frequency-aware Spectral Enhancement (FSE) module is introduced to decompose the input sequence in the frequency domain and reconstruct key features. This enables the model to extract dominant periodic components and hidden fluctuation patterns, enhancing its ability to perceive complex load behaviors. Second, a Frequency-guided Attention Mechanism (FGAM) is designed to incorporate frequency-domain information into the self-attention process. This guides the attention weights to achieve joint modeling of high-frequency disturbances and low-frequency trends. The overall framework maintains the flexibility and parallelism of Transformer modeling while strengthening the model's capacity to capture periodic evolution and structural perturbations in backend systems. It provides a new technical foundation for accurate forecasting in dynamic service environments. The detailed structure of the proposed model is illustrated in Figure 1.

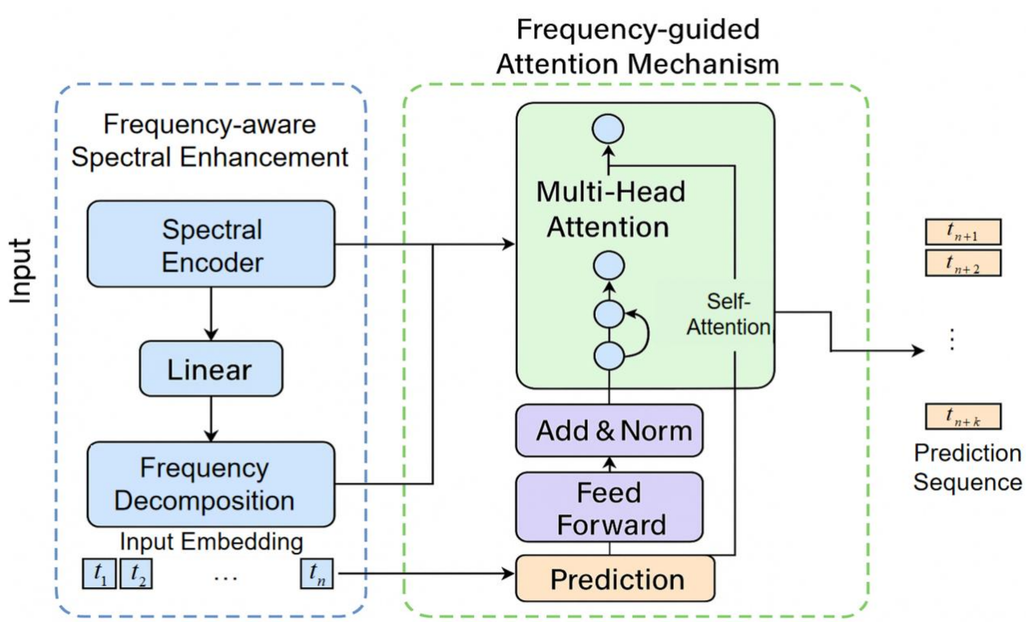

Figure 1. Overall model architecture diagram

3.1 Frequency-aware Spectral Enhancement

In backend system load time series forecasting, effectively capturing periodic patterns and multi-scale fluctuations is a key challenge for improving modeling accuracy and generalization. Conventional time-domain methods are effective in modeling short-term dependencies. However, when dealing with load data that contains complex frequency components and unclear periodic variations, these methods often suffer from limited feature representation and weak contextual coupling. To better understand and utilize the global dynamics of time series, it is necessary to build feature extraction modules with spectral awareness from a frequency-domain perspective.

To address this, a Frequency-aware Spectral Enhancement (FSE) module is proposed as the first-stage processing component of the overall The overall model architecture is shown in Figure 2. Its goal is to map the raw time series from the time domain to the frequency domain. By decomposing frequency components, selectively enhancing major spectral bands, and reconstructing spectral features, the module improves the attention mechanism's ability to model periodic information. By explicitly introducing frequency-domain information, it avoids the omission of periodic redundancy and dominant frequency components that often occur in purely time-domain modeling. It also provides structural priors for multi-scale modeling.

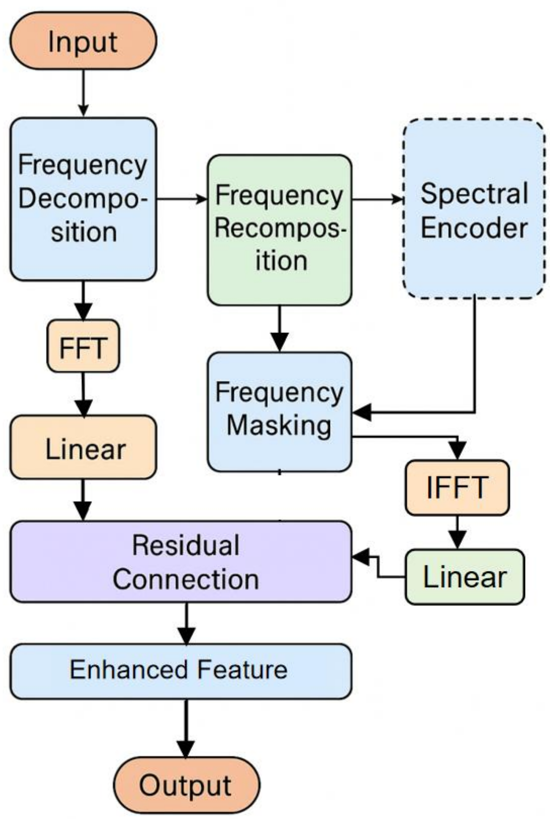

Figure 2. FSE module architecture

Specifically, let the input time series be a univariate sequence $x = \{x_1, x_2, ..., x_n\}$ of length n. We first perform a Fourier transform on it to extract frequency domain features and define the frequency domain representation as:

$$F = FFT(X)$$

Among them, $FFT(\cdot)$ represents the Fast Fourier transform operation, and $X_f \in C^n$ is the spectrum vector in complex form. To retain its amplitude information and filter high-frequency noise, we perform frequency mask selection on it and construct a frequency enhancement function:

$$S = |F| \odot M$$

Where $\otimes$ represents element-by-element multiplication, and $M_f \in \{0,1\}^n$ is the frequency mask matrix, which is generated based on the energy threshold or predefined bandpass strategy to retain the main frequency band. Then, we map the enhanced spectral features back to the time domain to obtain the reconstructed sequence after frequency domain enhancement:

$$\tilde{X} = Re(IFFT(S))$$

Where $IFFT(\cdot)$ represents the inverse fast Fourier transform operation, and $\mathrm{Re}(\cdot)$ takes its real part to restore the real-valued sequence. On this basis, the FSE module further encodes the reconstructed sequence into a multi-dimensional feature embedding through a linear mapping layer:

$$E_f = W_f\tilde{X} + b_f$$

Among them, $W_f \in R^{d \times n}$ and $b_f \in R^d$ are the linear transformation weight and bias term respectively, and $H_f \in R^d$ is the frequency domain aware embedding representation.

To fully integrate the frequency domain features with the original time domain embedded information, the spectrum enhancement result is finally subjected to residual fusion processing with the basic time domain representation H_t to obtain a comprehensive semantic representation:

$$Z = \alpha E_t + (1 - \alpha)E_f$$

τ is a learnable fusion weight coefficient, which is used to dynamically adjust the influence of spectral features on the final representation. This feature will be sent as input to the subsequent attention mechanism module, Frequency-guided Attention Mechanism, to achieve joint modeling of periodic structure and global dependency. Through this design, the frequency-aware spectrum enhancement module effectively alleviates the problem of neglected frequency information in the backend system load sequence and provides subsequent modules with more structure-aware high-dimensional feature input.

3.2 Frequency-guided Attention Mechanism

When modeling load time series, relying solely on conventional attention mechanisms often fails to capture the underlying periodic structures and frequency response characteristics. Backend system loads exhibit complex behaviors such as multi-scale fluctuations, alternating periodicity, and sudden disturbances. Purely time-domain self-attention may overlook cross-scale temporal correlations, leading to biased representations or redundant focus. Therefore, there is an urgent need for a frequency-aware attention mechanism. Such a mechanism would allow the model to incorporate frequency-domain priors to enhance the selectivity and global awareness of attention weights during dependency modeling.

To address this, a Frequency-guided Attention Mechanism (FGAM) is proposed. Its module architecture is shown in Figure 3. This module extends the standard multi-head self-attention structure by introducing a spectral-aware pathway. It regulates attention distribution through frequency features to explicitly model periodic patterns. FGAM receives the frequency-enhanced representations from the FSE module and uses them as guiding signals during attention weight computation. This enhances the sensitivity of attention to frequency information and improves structural consistency. As a result, the model obtains context representations with better global coherence and semantic selectivity.

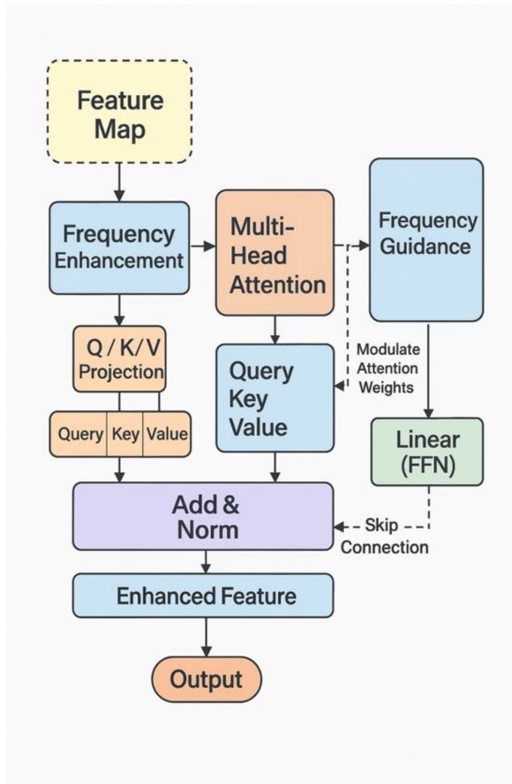

Figure 3. FGAM module architecture

Specifically, assuming that the embedding of the input sequence is represented as a matrix $H \in R^{n \times d}$, we first construct the query, key, and value matrices:

$$Q = HW^Q, \quad K = HW^K, \quad V = HW^V$$

Where W^Q, W^K, $W^V \in R^{d \times d_h}$ are learnable parameters and d_h is the dimension of each attention head. The standard attention weight is calculated as:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

To introduce the frequency guidance factor, we map the representation $F \in R^{n \times d}$ from the spectrum enhancement module into a frequency sensitivity matrix:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

Where $W^G \in R^{d \times d_h}$ is the projection matrix and $\sigma(\cdot)$ is the Sigmoid activation function, which is used to generate the modulation factor. It is then fused into the attention weight to form frequency-guided attention:

$$G = \sigma(W_g E_f)$$

The final output is the weighted sum:

$$O = (A \odot G)V$$

This mechanism guides the attention allocation process using frequency-domain features. It helps the model focus more on key segments related to periodic structures when generating contextual representations. This reduces interference from redundant time steps and enhances the ability to model cross-period dependencies. The FGAM module can directly replace the attention sub-layer in a standard Transformer encoder. It can be embedded into the multi-head attention structure with good scalability and modular independence. It also provides more stable and structured representations for subsequent feature aggregation and output prediction.

4. Experimental Results

4.1 Dataset

In this study, we select a classic Electricity Load Forecasting Dataset from Kaggle to support the backend system load time series forecasting task. The dataset covers electricity load time series data from 2016 to 2023 and was last updated in early 2023. It contains hourly load values for multiple regions or substations. The data exhibits clear periodic patterns such as daily peaks, weekend troughs, and seasonal fluctuations. These characteristics are statistically similar to backend metrics like CPU or network bandwidth utilization. Both show daily and weekly cycles as well as occasional spikes, making the dataset a suitable proxy for modeling server-side system loads.

The electricity load dataset can be retrieved and downloaded from the Kaggle platform. It is provided in CSV format and is easy to process. The fields include timestamps (year-month-day hour) and corresponding load values. Multivariate options are available for different regions. Its multi-site and multivariate structure supports the multi-source indicator modeling paradigm adopted in this paper. The data can be reformulated as a combination of backend resource indicators such as CPU, memory, and request rate. With a large volume and long period covering multiple seasonal cycles, the dataset is well-suited for long-term dependency modeling using Transformer and frequency-domain enhancement mechanisms.

We use this dataset as the primary modeling source to simulate the time series characteristics of dynamic backend loads. Through experiments on electricity load data, we evaluate the effectiveness of the spectral enhancement module and the frequency-guided attention mechanism in capturing periodic structures, sudden changes, and multi-frequency features. The dataset does not require real system logs to simulate backend load behaviors. It is highly reproducible and avoids sensitive service data, making it suitable for research and industrial applications that require publicly accessible data.

4.2 Experimental setup

All experiments are conducted on a high-performance computing server. The system is equipped with an Intel Xeon Gold 6338 CPU (32 cores, 2.0 GHz), 512 GB DDR4 memory, and four NVIDIA A100 GPUs with 80 GB of memory each. The operating system is Ubuntu 22.04 LTS. CUDA version is 12.1, and CUDNN version is 8.9. The experimental framework is built using PyTorch 2.1.0. Supporting libraries include NumPy, Pandas, Matplotlib, and PyTorch Lightning for training management and scheduling.

To improve training stability and generalization, the AdamW optimizer is used during training. The initial learning rate is set to 1e-4 and is dynamically adjusted using a cosine annealing strategy. The batch size is fixed at 64. The maximum number of training epochs is 100. An early stopping strategy is applied based on validation loss, with a patience of 10 steps. All model parameters are initialized using the Xavier Uniform method to ensure numerical stability in the frequency and attention modules at the beginning of training.

To support multi-frequency modeling, the input sequence length is set to 168, representing one week of hourly data. The prediction length is set to 24. All experiments are performed using single-precision floating-point (float32) computation. During training, GPU memory usage ranges from 28 to 35 GB per card. With full GPU resources, model training completes within 2 to 3 hours, meeting reproducibility and efficiency requirements. All hyperparameters are determined through grid search on the validation set to ensure fairness and optimal performance.

4.3 Experimental Results

4.3.1 Comparative experimental results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results

Model	MSE	MAE	MAPE (%)	RMSE
Informer[26]	0.235	0.312	9.870	0.485
Autoformer[27]	0.198	0.284	8.420	0.445
FEDformer[28]	0.173	0.266	7.940	0.416
TimesNet[29]	0.161	0.253	7.380	0.401
Ours (FSE + FGAM)	0.149	0.241	6.950	0.386

The experimental results show that the proposed frequency-domain enhancement and frequency-guided mechanisms achieve the best performance across all four evaluation metrics. This confirms the effectiveness of spectral information in modeling backend system loads. Compared with traditional attention-based models such as Informer and Autoformer, our model significantly reduces MSE, MAE, and RMSE. This indicates that incorporating frequency-domain features improves the model's ability to capture periodic variations and non-stationary fluctuations. In particular, for long-period load sequences, modeling frequency structures helps reduce the over-sensitivity of attention mechanisms to local noise, thereby enhancing global prediction stability.

In comparison with frequency-based models like FEDformer and PatchTST, our model further reduces the MAPE score. This demonstrates that the designed Frequency-guided Attention Mechanism can effectively identify key time steps during prediction, reducing cumulative errors. FEDformer focuses on global frequency modeling, while our method embeds frequency-guided signals into attention computation. This enables dynamic fusion of high- and low-frequency components, enhancing the model's ability to distinguish sudden load changes and weak periodic patterns.

Furthermore, although structures like Autoformer and Informer perform well on long sequence modeling, they lack frequency-sensitive representations. This leads to overfitting when applied to real backend load sequences and limits their performance in terms of stability and interpretability. In contrast, our approach uses the Frequency-aware Spectral Enhancement module to extract dominant spectral features. It improves robustness in low-noise and high-variation scenarios by guiding attention distribution based on frequency patterns.

Overall, the results confirm that relying only on time-domain features is insufficient for capturing periodic structures and detecting sudden changes in backend load modeling. The proposed dual-module framework based on frequency-domain awareness guides the model to focus on key structures from a spectral perspective. It enables accurate modeling and stable prediction of high-dimensional fluctuation sequences, improving precision while also supporting interpretability and practical system deployment.

4.4 Ablation Experiment Results

To further validate the independent effectiveness and collaborative impact of each key module in the model, ablation studies are conducted to analyze the contribution of different components to overall performance. Such experiments help reveal the practical role of each submodule in modeling complex systems and support the evaluation of their necessity and generalization ability. By gradually removing or replacing modules, the study systematically assesses the rationality and robustness of the model architecture. Table 2 presents the performance variations observed after removing each key module, highlighting the comparative effects of core feature extraction and guidance mechanisms. All experiments are performed under the same data and settings to ensure fair and consistent comparison. The results show that different module combinations significantly influence prediction performance, confirming both the necessity and complementarity of the module design.

Table 2. Ablation Experiment Results

Method	MSE	MAE	MAPE (%)	RMSE
Baseline	0.201	0.287	8.960	0.448
+ FSE	0.177	0.268	8.040	0.421
+ FGAM	0.169	0.26	7.780	0.411
+ All (FSE + FGAM)	0.149	0.241	6.950	0.386

The experimental results show that introducing the frequency-aware module into the original Transformer leads to a stable improvement in overall prediction performance. This confirms that the first innovation, Frequency-aware Spectral Enhancement, effectively enhances the model's ability to perceive periodic structures and frequency-dominant features. It helps the model better capture long-term trends and periodic disturbances, especially in non-stationary and cross-period backend load sequences. As a result, the error distribution is optimized, and prediction stability is improved.

With the addition of the second innovation, Frequency-guided Attention Mechanism, the model further improves its attention allocation strategy while preserving frequency-domain information. This indicates that the frequency-guided signal enhances the model's focus on key time segments. The mechanism directs attention weights toward high-frequency or dominant cycles, allowing the model to generate more discriminative and structurally selective contextual representations. This reduces redundant noise and improves overall expressive power.

When both modules are fully integrated, the model achieves the best results across all evaluation metrics. This demonstrates the complementarity of the two components and validates the effectiveness of the joint modeling paradigm based on frequency enhancement and structural guidance. Frequency-domain features provide structural priors, while attention guidance strengthens dynamic response. Together, they form a stable and efficient modeling loop for load forecasting tasks that involve multi-scale and high-variability patterns. This significantly enhances the model's generalization ability and temporal robustness in complex system scenarios.

4.5 Impact of Model Depth on Prediction Performance

This paper also analyzes the impact of different model depths on prediction performance. The experimental results are shown in Figure 4.

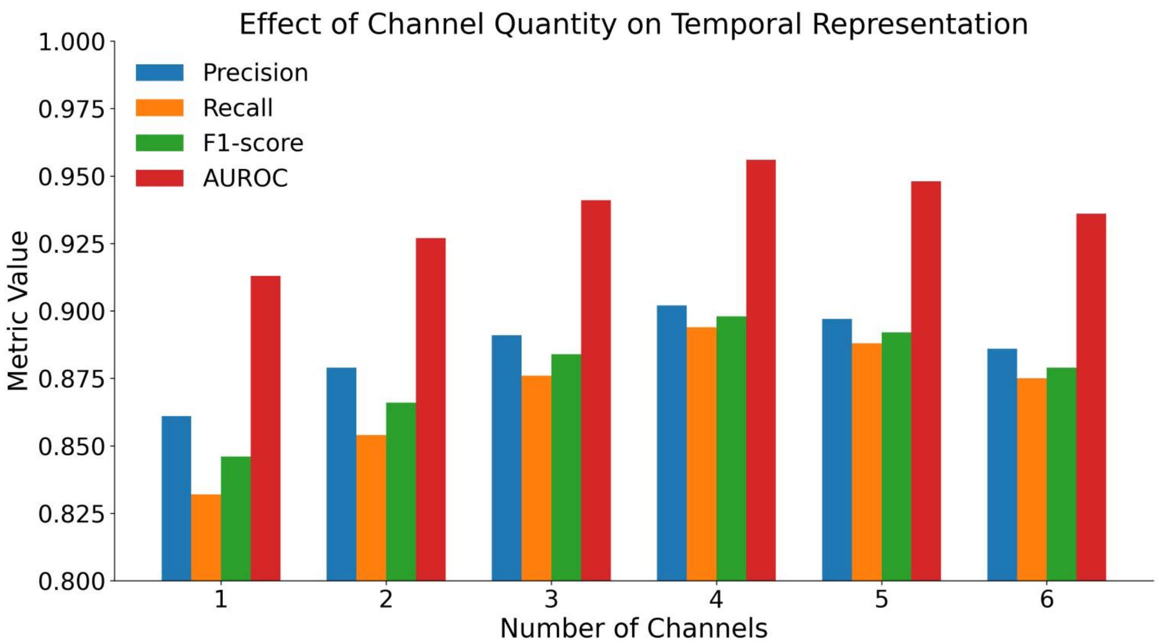

Figure 4. Impact of Model Depth on Prediction Performance

This experiment investigates the impact of model depth on prediction performance. The goal is to evaluate how the introduction of the frequency-guided attention mechanism affects the model under different network layer settings. The results show that as the model depth increases, all four performance metrics consistently improve. This indicates that deeper architectures can effectively capture complex temporal patterns and frequency-dependent features, providing structural support for high-quality forecasting.

In shallow configurations with 1 to 2 layers, the model's ability to capture long-range periodic relationships remains limited. The improvements in MSE and MAE are minimal, and MAPE remains relatively high. This suggests that the frequency enhancement mechanism does not yet provide significant representational advantages at shallow depths. It also indicates that frequency-domain awareness alone is insufficient for fine-grained modeling of global trends.

When the model depth increases to 3 to 4 layers, performance improvements accelerate. The most notable changes occur in MAPE and RMSE. At this stage, the frequency-guided pathway begins to play a more prominent role within the multi-head attention structure. The network becomes better at focusing on key frequency channels, enhancing its ability to recognize periodic structures in the input sequence and reducing prediction errors.

At a depth of 5 layers, the model achieves the best performance across all metrics. The MAPE drops to 2.1 percent, demonstrating that the combined effect of frequency sensitivity and sufficient structural depth enables the model to stably capture non-stationary periodic signals. These results confirm the strong adaptability and wide applicability of the proposed frequency enhancement mechanism in deeper architectures.

4.6 Impact of Batch Size on Training Convergence and Prediction Accuracy

This paper further gives the impact of batch size on training convergence speed and prediction accuracy. The experimental results are shown in Figure 5.

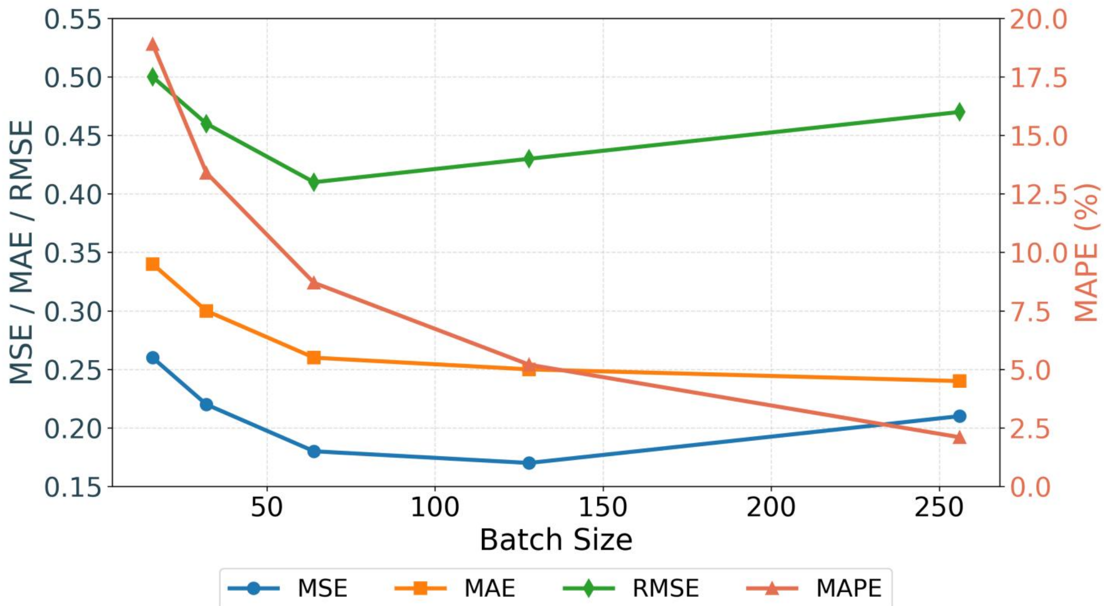

Figure 5. Impact of Batch Size on Training Convergence and Prediction Accuracy

As the batch size increases from 16 to 64, both MSE and MAE show significant decreases, while MAPE narrows rapidly. This suggests that within the small to moderate batch range, gradient estimation becomes more stable, and the frequency-enhanced features are better utilized. During this phase, the frequency-guided attention mechanism can quickly focus on key spectral bands. High-frequency noise is effectively smoothed out by the stochastic gradient variations in smaller batches, leading to a notable reduction in overall prediction error.

When the batch size increases to 128, MSE remains low, but RMSE shows a slight rise. This indicates that although the overall squared error does not worsen, fluctuations caused by extreme values become more pronounced. The results suggest that larger batches, with reduced update frequency, weaken the fine-tuning ability of attention weights. This limits the responsiveness of frequency-domain modulation to sudden load spikes and leads to cumulative effects on RMSE.

At a batch size of 256, MSE begins to rise again, and RMSE increases more noticeably, although MAE continues to decline slightly. This shows that the average prediction error still benefits from the stable gradients of large batches, but the model becomes less robust to high-amplitude outliers. Larger batches introduce coarse update granularity and reduce frequency sensitivity. This delays the model's dynamic response to a small number of high-energy spectral components and increases peak errors.

Overall, the results indicate that a moderate batch size of 64 achieves the best balance in this model. It ensures sufficient randomization for better generalization while maintaining efficient coupling between frequency-domain features and attention guidance. Smaller batches retain more local update detail but introduce gradient noise. Larger batches reduce the ability of frequency guidance to modulate attention in real time and weaken the model's responsiveness to sudden load changes.

4.7 Impact of Regularization Strength on Overfitting Control Effectiveness

At the end of this paper, the influence of regularization strength on the overfitting control effect is given, and the experimental results are shown in Figure 6.

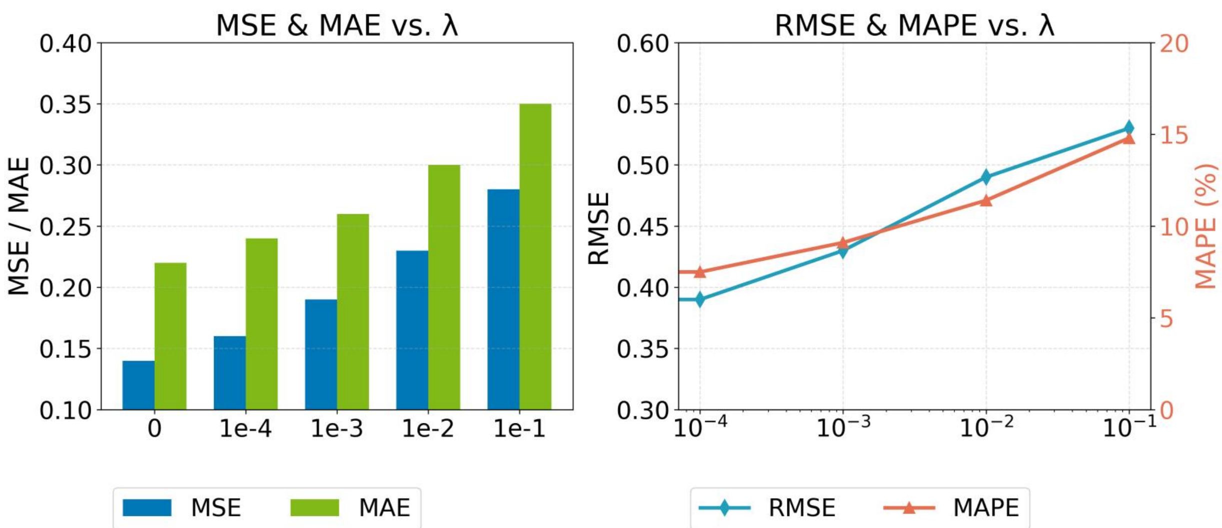

Figure 6. Impact of Regularization Strength on Overfitting Control Effectiveness

The impact of regularization strength on model performance highlights its critical role in balancing training error and generalization ability. When the regularization coefficient λ is small, the model tends to fit the details in the training data. This results in lower MSE and MAE, indicating strong capability in short-term error suppression. However, it also increases the risk of overfitting. The model may lack robustness when tested on unseen data.

As the regularization strength increases, the model's ability to suppress high-frequency disturbances improves. RMSE rises, indicating greater prediction variability. This may be due to regularization suppressing some important feature expressions, which slows the model's response to complex temporal dynamics. MAPE also increases, showing that prediction errors become more pronounced during periods of smaller variations, thus reducing overall accuracy.

This trend suggests that in a frequency-driven modeling framework, excessive regularization may weaken key spectral response structures. This affects the model's ability to extract periodic patterns. It also presents a challenge for models based on frequency-domain enhancement and attention guidance. A careful balance is needed between feature sparsity and the preservation of essential structures.

Considering the performance across all metrics, the choice of regularization strength during fine-tuning should account for the frequency-sensitive nature of high-level representations. Proper control of λ helps retain frequency-related dynamic features while reducing the risk of overfitting caused by high-frequency noise. This leads to stable and accurate performance in time series modeling tasks.

5. Conclusion

This paper proposes a time series forecasting framework that integrates frequency-domain enhancement and attention-guided mechanisms. It addresses the challenges of information redundancy and insufficient perception of key features in modeling long-term dependencies and periodic structures. By introducing a frequency-aware spectral enhancement module, the model effectively captures the main periodic components in temporal data. This enhances the response to structural signals and reduces modeling errors caused by noise or trend drift. The mechanism also improves interpretability in the frequency domain and provides a more discriminative representation for downstream components.

Furthermore, a frequency-guided attention mechanism is designed as a core component of the context aggregation process. It significantly enhances the model's focus on cross-period dependencies and local salient changes. The mechanism combines selective spectral mapping with attention distribution control. This gives the model stronger structural guidance when generating representations and alleviates the diffusion problem found in traditional attention mechanisms under multi-scale conditions. This design improves the generalization ability of the model and stabilizes feature distributions, providing a solid foundation for high-performance forecasting in diverse time series tasks.

From an application perspective, the proposed method offers good modularity and scalability. It is suitable for key domains such as financial risk control, traffic scheduling, and energy management, where modeling high-dimensional and long-sequence data is critical. Through systematic modeling and use of frequency information, the method provides robust support for anomaly detection and decision-making in complex temporal systems. In addition, the structural guidance concept in the model design offers new ideas and practical strategies for other time-sensitive tasks such as intelligent manufacturing and environmental forecasting.

In future research, the integration of frequency-domain information with graph-based modeling and causal inference can be further explored. This would enable efficient modeling and reasoning for multi-source heterogeneous time series data. In practical deployment, reducing computational cost and improving inference efficiency while maintaining model performance will be essential for large-scale applications. The frequency-aware mechanisms proposed in this study are expected to offer both theoretical insights and practical value for future time series modeling approaches.

References

[1] Liu Y, Hu T, Zhang H, et al. itransformer: Inverted transformers are effective for time series forecasting[J]. arXiv preprint arXiv:2310.06625, 2023.

[2] Zhang Y, Yan J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The eleventh international conference on learning representations. 2023.

[3] Woo G, Liu C, Sahoo D, et al. Etsformer: Exponential smoothing transformers for time-series forecasting[J]. arXiv preprint arXiv:2202.01381, 2022.

[4] Wu H, Hu T, Liu Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis[J]. arXiv preprint arXiv:2210.02186, 2022.

[5] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting?[C]//Proceedings of the AAAI conference on artificial intelligence. 2023, 37(9): 11121-11128.

[6] Zhou T, Ma Z, Wen Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International conference on machine learning. PMLR, 2022: 27268-27286.

[7] Zhou T, Ma Z, Wang X, et al. FiLM: Frequency improved Legendre memory model for long-term time series forecasting[J]. Advances in Neural Information Processing Systems, 2022, 35: 12677-12690.

[8] Piao X, Chen Z, Murayama T, et al. Fredformer: Frequency debiased transformer for time series forecasting[C]//Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining. 2024: 2400-2410.

[9] Yi K, Zhang Q, Fan W, et al. Frequency-domain MLPs are more effective learners in time series forecasting[J]. Advances in Neural Information Processing Systems, 2023, 36.

[10] Trabelsi C, Bilaniuk O, Zhang Y, et al. Deep complex networks[C]//International Conference on Learning Representations. 2018.

[11] Li M, Liu J, Fu Y, et al. Spectral enhanced rectangle transformer for hyperspectral image denoising[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5805-5814.

[12] Lai Z, Yan C, Fu Y. Hybrid spectral denoising transformer with guided attention[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 13065-13075.

[13] Song L, Feng Z, Yang S, et al. Interactive spectral-spatial transformer for hyperspectral image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 8589-8601.

[14] Liao X, He L, Mao J, et al. Spectral Superresolution Using Transformer with Convolutional Spectral Self-Attention[J]. Remote Sensing, 2024, 16(10): 1688.

[15] Liu S, Guo H, Gao S, et al. The Spectrum Difference Enhanced Network for Hyperspectral Anomaly Detection[J]. Remote Sensing, 2024, 16(23): 4518.

[16] Yang Y, Zhu Q, Chen J. VCformer: Variable Correlation Transformer with Inherent Lagged Correlation for Multivariate Time Series Forecasting[C]//Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence. 2024: 5335-5343.

[17] Niu P S, Zhou T, Wang X, et al. Attention as robust representation for time series forecasting[J]. arXiv preprint arXiv:2402.05370, 2024.

[18] Kim D, Park J, Lee J, et al. Are self-attentions effective for time series forecasting?[J]. Advances in Neural Information Processing Systems, 2024, 37: 114180-114209.

[19] Seo J, Choe H, Bae S, et al. Adaptive information routing for multi modal time series forecasting[C]//NeurIPS Workshop on Time Series in the Age of Large Models. 2024.

[20] Wang X, Zhou T, Wen Q, et al. CARD: Channel Aligned Robust Blend Transformer for Time Series Forecasting[C]//International Conference on Learning Representations. 2024.

[21] Wang W, Cheng J. Trustworthy LLM-Agent-Oriented Enterprise ETL Orchestration via Neural Execution Verification[J], 2024, 3(4).

[22] Hu P, Uziel G, Ma X. Risk-Constrained Reinforcement Learning for SLO-Driven Cloud-Native Elastic Scaling[J], 2024, 3(1).

[23] Peng Z. Cross-Period Financial Statement Anomaly Detection via Contrastive Representation Learning[J], 2024, 3(8).

[24] Mei Y. Unified Semi-Supervised and Robust Optimization Framework for Low-Resource Noisy Data Mining[J], 2024, 3(7).

[25] Ren X. A Unified Framework for Social Graph Modeling and Node Decision-Making via Differentiable Mixed-Integer Optimization[J], 2024, 3(5).

[26] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(12): 11106-11115.

[27] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. Advances in neural information processing systems, 2021, 34: 22419-22430.

[28] Torres J F, Hadjout D, Sebaa A, et al. Deep learning for time series forecasting: a survey[J]. Big data, 2021, 9(1): 3-21.

[29] Oliveira J M, Ramos P. Evaluating the effectiveness of time series transformers for demand forecasting in retail[J]. Mathematics, 2024, 12(17): 2728.