

Multi-Scale Image Representation-Driven Structural Modeling and Non-Stationary Feature Decomposition for Financial Time Series



Xuzhuo Hu^{ID}

School of Statistics and Mathematics, Shandong University of Finance and Economics, Jinan 250014, China

Corresponding Author Email: 18560110501@163.com

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430128>

ABSTRACT

Received: 28 September 2025

Revised: 29 November 2025

Accepted: 27 December 2025

Available online: 28 February 2026

Keywords:

multi-scale image representation, image-domain non-stationary feature decomposition, financial time series, image processing, vision transformer, dynamic feature fusion

Financial time series exhibit pronounced non-stationarity, nonlinearity, and complex temporal dependencies. Traditional time series modeling approaches face inherent limitations in feature extraction and global structural analysis. Meanwhile, existing image-based transformation methods for time series generally suffer from single-representation constraints, insufficient multi-scale fusion, and limited capability in decoupling non-stationary features within the image domain. To address these challenges, this study innovatively reformulates financial time series forecasting as a multi-channel image understanding and analysis problem. We propose an end-to-end multi-scale image-driven framework for structural modeling and non-stationary feature decomposition. Leveraging the strengths of image processing in feature extraction and pattern recognition, the framework employs a dual-path multi-scale image generation mechanism to achieve joint visual encoding of temporal dependencies and frequency dynamics. An image-domain adaptive non-stationary feature decomposition network is introduced to accurately disentangle trend, periodic, and noise components across multiple image scales. Furthermore, a dynamic attention-based gating fusion mechanism is designed to adaptively aggregate multi-source and multi-component image features. The entire system is constructed under a collaborative optimization architecture of “image generation-feature decomposition-dynamic fusion,” enabling precise structural analysis and efficient prediction of complex financial time series. Extensive experiments on multiple publicly available financial time series datasets demonstrate that the proposed framework significantly outperforms existing methods in both forecasting accuracy and structural interpretability. The proposed image-based decomposition and modeling paradigm can be further extended to other non-stationary signal processing tasks, providing valuable insights for cross-domain applications of image processing techniques in non-stationary signal analysis and enriching research at the intersection of image processing and time series modeling.

1. INTRODUCTION

Financial time series generally exhibit non-stationarity, nonlinearity, and cross-scale coupling characteristics [1-3]. Their internal structures are complex and dynamically evolving, making it difficult for traditional time series modeling methods to accurately capture temporal dependency relationships and frequency evolution patterns [4, 5], and resulting in significant limitations in the completeness of feature extraction and the depth of global structural analysis [6]. Image processing techniques possess inherent advantages in complex pattern recognition, multi-scale feature extraction, and global correlation modeling [7, 8]. Transforming financial time series into image forms for analysis has become a research hotspot at the intersection of image processing and time series analysis, providing a new pathway for addressing non-stationary signal modeling problems. However, existing time series imaging transformation methods still suffer from many shortcomings [9-11], which restrict the full utilization of the advantages of image processing techniques and make it

difficult to meet the requirements of accurate modeling of financial time series. Most existing methods adopt a single mapping approach to transform time series into images [12], lacking multi-scale and multi-perspective information fusion, and thus failing to comprehensively characterize the dual properties of financial time series in both the time domain and the frequency domain. Meanwhile, there is a lack of specialized non-stationary feature decomposition techniques designed for the image domain [13, 14]. Traditional signal decomposition methods cannot be directly adapted to image inputs and therefore cannot effectively decouple complex structural components such as trend, periodicity, and noise embedded in images. In addition, most existing feature fusion mechanisms are statically designed [15-17] and cannot adapt to the dynamic changes of financial markets, leading to low utilization efficiency of multi-source image features and further affecting modeling accuracy and robustness.

To address the above problems, this paper relies on core technologies in the field of image processing, including convolution operations, Transformer-based global modeling,

and feature fusion, aiming to solve the key bottleneck that non-stationary features of financial time series are difficult to accurately analyze. At the same time, it extends the application scenarios of image processing techniques in non-stationary signal analysis and fills the gap in existing cross-domain research regarding the collaborative optimization of image generation, feature decomposition, and dynamic fusion. The research objective of this paper is to propose a multi-scale image-based modeling and non-stationary feature decomposition framework adapted to the characteristics of financial time series, so as to achieve precise analysis and efficient utilization of the complex internal structures of financial time series and significantly improve prediction performance under non-stationary scenarios. Meanwhile, it aims to provide a reusable and extensible technical solution for the application of image processing techniques in cross-domain non-stationary signal processing, thereby promoting technological innovation at the intersection of image processing and time series analysis.

The core contributions of this paper are as follows:

(1) A multi-scale dual-path financial time series image generation method is proposed, which integrates Gramian Angular Field (GAF) and Continuous Wavelet Transform (CWT) techniques, combined with an adaptive multi-time-window sampling strategy, to construct multi-scale and multi-perspective image input tensors. This approach achieves dual visual encoding of temporal dependency relationships and frequency dynamic evolution of financial time series, laying a solid data foundation for subsequent image-domain feature analysis and decomposition.

(2) A parameterized, end-to-end adaptive non-stationary feature decomposition network in the image domain is designed. By introducing dedicated convolutional layers and customized constraint loss functions, precise decoupling of trend, periodic, and noise components in multi-scale images is achieved. This approach overcomes the technical bottleneck that traditional signal decomposition methods cannot be directly adapted to image inputs and improves the analytical accuracy of non-stationary features in the image domain.

(3) Vision Transformer is introduced into image-domain component feature modeling. By improving the self-attention mechanism, the global correlation modeling capability of multi-component and multi-scale image features is enhanced. A dynamic attention gating fusion mechanism is proposed to realize adaptive weighted aggregation of multi-source and multi-component image features, effectively adapting to the dynamic changes of financial markets and improving the robustness of feature utilization.

(4) An integrated end-to-end collaborative optimization architecture of image generation-feature decomposition-dynamic fusion-prediction is constructed. A composite loss function is designed to guide joint optimization of parameters throughout the entire process, taking into account prediction accuracy, structural interpretability, and non-stationary adaptability. This forms a new paradigm for deep integration of image processing techniques and financial time series modeling and enriches research achievements in the cross-domain field.

The remaining sections of this paper are organized as follows: The methodology framework section elaborates in detail the design principles and technical details of each module in the proposed end-to-end framework. The experimental validation section verifies the effectiveness and superiority of the proposed method through experiments on

multiple datasets, comparative experiments, and ablation experiments. The discussion section analyzes in depth the academic value, technical advantages, and limitations of the proposed method and proposes future research directions. The conclusion section summarizes the core research findings and academic contributions of this paper and outlines the theoretical and practical significance of the research work.

2. METHODOLOGICAL FRAMEWORK

2.1 Overall framework design

The financial time series structural modeling and non-stationary feature decomposition framework proposed in this paper is an adaptive end-to-end architecture. Its core logic is to reconstruct one-dimensional financial time series into multi-channel image representations, relying on the technical advantages of the image processing field in feature extraction and global modeling. Through a series of specialized image processing operations, it achieves precise decoupling, effective modeling, and efficient fusion of non-stationary features in financial time series, and finally outputs reliable prediction results. The overall processing flow of the framework follows a progressive logic of raw financial time series input, multi-scale image generation, non-stationary feature decomposition and modeling, dynamic feature fusion and prediction. Each module cooperates and links to form a complete technical chain, ensuring full-process optimization from the original sequence to the prediction results. The core advantage of the framework lies in taking images as the core processing object throughout the entire process, deeply integrating image processing techniques with the modeling requirements of financial time series. It can not only fully exploit the inherent advantages of image processing techniques in complex pattern analysis and multi-scale feature capture to achieve precise characterization and global analysis of non-stationary features in financial time series, but also possesses strong architectural interpretability and adaptability to non-stationary scenarios. It can effectively cope with the inherent dynamic changes and complex structural characteristics of financial time series, providing a unified and efficient technical architecture for image-based modeling of non-stationary signals. The specific architecture is shown in Figure 1.

2.2 Multi-scale image generation module

The core innovation of the multi-scale image generation module lies in breaking through the limitations of existing single image transformation methods. It adopts a design strategy combining dual-path parallelism and multi-scale sampling to achieve dual visual encoding of temporal dependencies and frequency dynamics of financial time series. This design abandons the defect of traditional single mapping methods that cannot fully characterize non-stationary structures. Through two parallel image generation paths, it captures the temporal correlation characteristics and frequency evolution patterns of the sequence respectively. Combined with a multi-scale sampling mechanism that covers intraday, inter-day, and long-term time dimensions of market behavior differences, it finally generates a multi-channel image representation with both richness and specificity, providing solid visual information support for subsequent precise

decomposition of non-stationary features in the image domain.

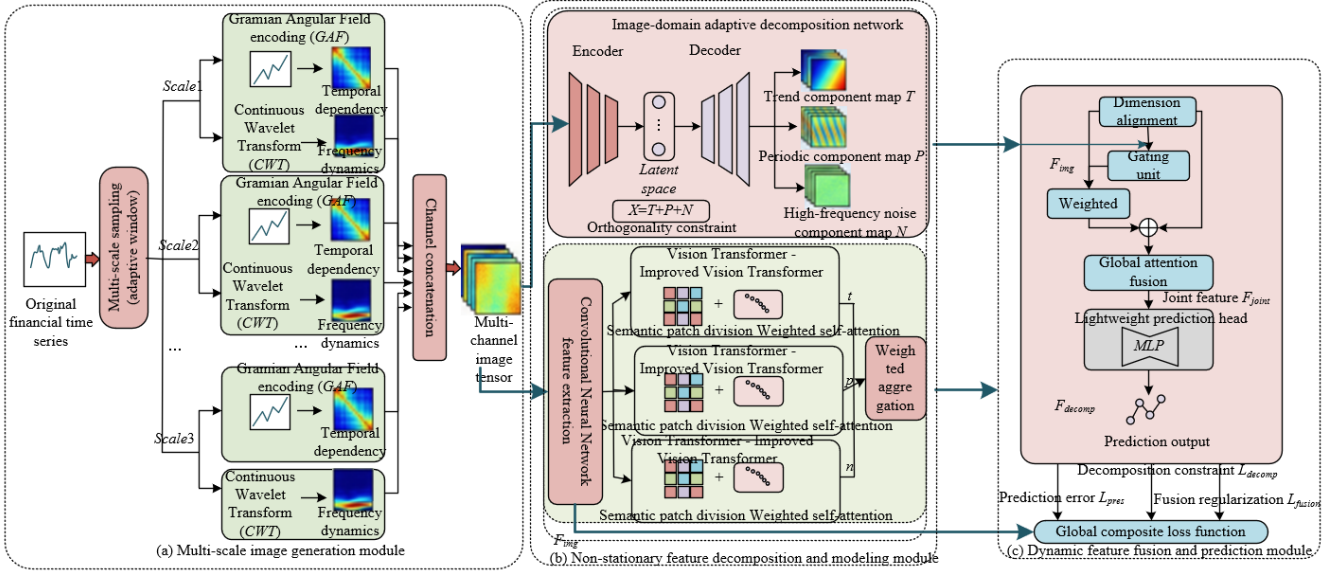


Figure 1. Framework structure of the proposed method

The dual-path image generation strategy is the core technical innovation of the module. It consists of two parallel paths: GAF encoding and CWT time-frequency image generation, which respectively achieve precise visualization of temporal dependencies and frequency dynamics. The GAF encoding path focuses on capturing the temporal dependency relationships of the sequence. First, min-max normalization is adopted to map the original financial time series to the interval $[0, \pi/2]$, adapting to the polar coordinate angle range. The normalization formula is:

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \times \frac{\pi}{2} \quad (1)$$

Subsequently, each data point x'_i in the normalized sequence $x' = [x'_1, x'_2, \dots, x'_n]$ is mapped to polar coordinates (r_i, θ_i) , where r_i is the normalized magnitude reflecting the data value, and $\theta_i = x'_i$ corresponds to the angular feature of the time order. By computing the angle sum of any two time points i and j , $\theta_{i+j} = \theta_i + \theta_j$, and using the cosine function to transform the angular correlation into pixel grayscale values, a GAF feature map of dimension $R^{n \times n}$ is generated. The pixel value calculation formula is:

$$G(i, j) = \cos(\theta_{i+j}) = \cos(\theta_i + \theta_j) \quad (2)$$

A grayscale value closer to 1 indicates a stronger correlation between the two time points. To enhance the recognizability of temporal dependency textures and adapt to subsequent convolutional layer feature extraction, a grayscale enhancement strategy is introduced to optimize the GAF feature map and further strengthen the feature representation capability of the image. The CWT path is used to supplement frequency dynamic information. Considering the transient fluctuation characteristics of financial time series, the complex Morlet wavelet [18] is selected as the basis function, expressed as:

$$\psi(t) = \pi^{-1/4} e^{j\omega_0 t} e^{-t^2/2} \quad (3)$$

where, $\omega_0 = 5$ to balance time-frequency resolution. The

original sequence $x(t)$ is subjected to CWT to obtain the wavelet coefficients:

$$W(a, b) = \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (4)$$

where, a is the scale parameter corresponding to frequency (smaller a indicates higher frequency), b is the translation parameter corresponding to time, and ψ^* denotes the complex conjugate of the wavelet basis function. The modulus value $|W(a, b)|$ of the wavelet coefficients is normalized to $[0, 255]$ and converted into a CWT time-frequency spectrum image of dimension $R^{M \times n}$. The pixel grayscale value corresponds to the intensity of the frequency component, achieving intuitive visualization of frequency dynamics. To ensure compatibility between the dual-path images, bilinear interpolation is adopted to resize the GAF feature map and the CWT time-frequency spectrum image to the same size, laying the foundation for subsequent multi-scale fusion and feature decomposition.

The multi-scale sampling mechanism is another important innovation of the module, aiming to accurately capture behavioral differences of financial markets at different time scales. An adaptive time-window sampling strategy is adopted, where the window size is adaptively adjusted by the sequence volatility σ . The window length is calculated as:

$$L = k \times \sigma \quad (5)$$

where, k is an empirical coefficient ranging from 3 to 5, ensuring that the window size can dynamically adapt to the fluctuation characteristics of the sequence and avoiding the defect that a fixed window cannot balance different fluctuation intensities. Based on this sampling strategy, the original financial time series is sampled with m different scales of windows. Under each scale, GAF feature maps and CWT time-frequency spectrum images are generated in parallel, finally obtaining $m \times 2$ multi-scale and multi-perspective images. To achieve integrated input of multi-source image information, all multi-scale dual-path images are concatenated along the channel dimension to construct a multi-channel image input tensor:

$$X \in \mathbb{R}^{C \times H \times W} \quad (6)$$

where, $C=2m$ is the number of channels corresponding to two types of images under m scales, and H and W denote the height and width of the image, respectively. This tensor integrates

temporal dependency and frequency dynamic information under different scales, forming a multi-dimensional and comprehensive image representation, effectively improving the completeness and precision of subsequent image-domain feature decomposition.

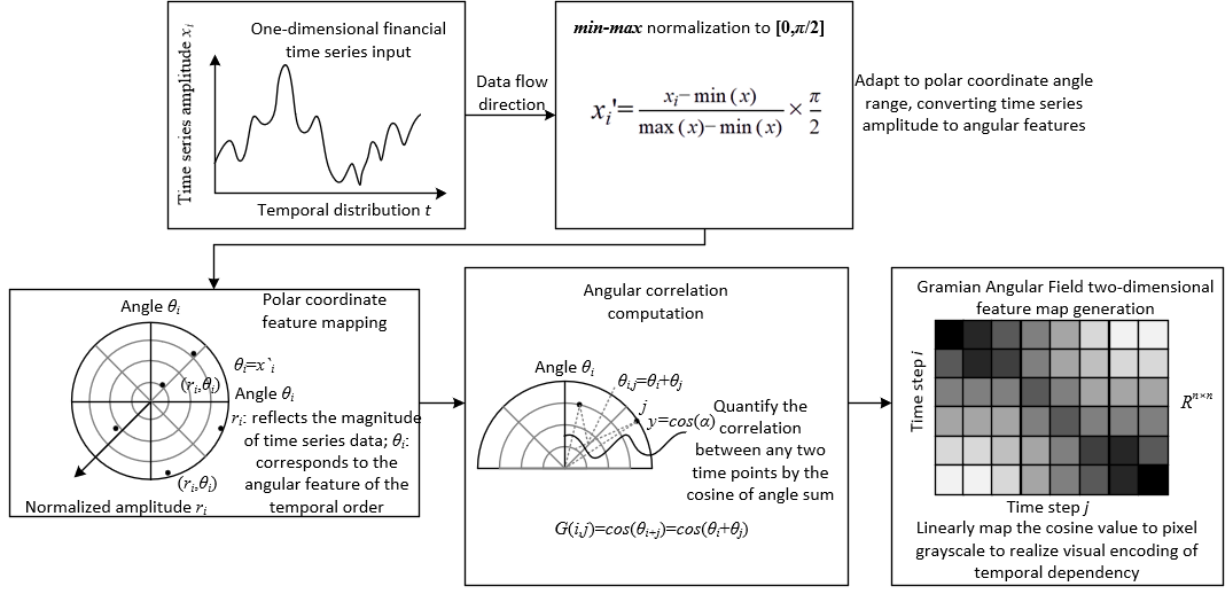


Figure 2. Schematic diagram of Gramian Angular Field (GAF) time series-polar coordinate-image pixel mapping principle

Figure 2 shows the schematic diagram of the GAF time series-polar coordinate-image pixel mapping principle. Through the innovative design of dual-path parallelism and multi-scale sampling, the multi-scale image generation module successfully transforms one-dimensional financial time series into multi-channel image tensors rich in temporal and frequency information. It not only breaks through the limitations of traditional single image transformation, but also realizes multi-perspective visual encoding of non-stationary features. The image tensor generated by this module fully preserves the complex non-stationary structure of financial time series, accurately characterizes the temporal dependency relationships and frequency evolution patterns of the sequence, and provides high-quality input for the subsequent image-domain adaptive non-stationary feature decomposition network, ensuring that the decomposition process can accurately capture trend, periodic, and noise components in the sequence. This fully reflects the deep integration of image processing techniques and financial time series modeling, laying a data foundation for the performance improvement of the entire framework.

2.3 Non-stationary feature decomposition and modeling module

The non-stationary feature decomposition and modeling module is the innovative hub of the entire framework. Its core breakthrough lies in abandoning the traditional inherent process of “signal-domain decomposition \rightarrow imaging” and constructing a new technical path of “direct decomposition in the image domain \rightarrow feature modeling,” specifically addressing the key problem that traditional signal decomposition methods cannot be directly adapted to image inputs and have difficulty decoupling complex non-stationary components in images. This module realizes precise decoupling of trend, periodic, and high-frequency noise

components in multi-scale images through a parameterized, end-to-end adaptive decomposition network. Combined with an improved Vision Transformer [19, 20] to strengthen global correlation modeling of multi-component image features, it can accurately match the non-stationary characteristics of financial time series and provide high-quality structured feature support for subsequent dynamic feature fusion.

The image-domain adaptive non-stationary feature decomposition network is the core innovation of this module. It adopts an encoder-decoder architecture design, specifically adapted to multi-channel image tensor input, to achieve end-to-end precise decoupling of three types of non-stationary components. The network input is the multi-channel image tensor X output from the multi-scale image generation module, and the outputs are the trend component map T , periodic component map P , and high-frequency noise component map N . The three strictly satisfy the component orthogonality constraint and reconstruction constraint, namely $X=T+P+N$. The encoder part adopts a 3-layer proprietary grouped convolution design. The number of groups is consistent with the input image channel number C . The convolution kernel size is set to 3×3 , step is 1, and padding is 1. While retaining the image spatial structure to the greatest extent, it precisely extracts dedicated features of different channels. After each grouped convolution layer, a batch normalization layer and a LeakyReLU activation function are connected. The slope of LeakyReLU is set to 0.2, effectively alleviating the gradient vanishing problem and enhancing the nonlinear fitting capability of the network. The decoder part corresponds to the encoder output dimensions and adopts 3 layers of transposed convolution to gradually restore the image size. The last layer connects a sigmoid activation function to normalize the three component feature maps to the interval $[0,1]$, ensuring that each component has clear physical meaning and interpretability. To retain multi-scale image detail information and improve decomposition accuracy, U-Net-style skip

connections are introduced, concatenating the outputs of each encoder layer with the corresponding decoder layer inputs to achieve complementary fusion of detail features and global features. To ensure decomposition effectiveness, a composite decomposition constraint loss function is constructed to guide network training. The total decomposition loss is expressed as:

$$L_{decomp} = \alpha \cdot L_{rec} + \beta \cdot L_{orth} + \gamma \cdot L_{period} \quad (7)$$

where, $\alpha=1.0$, $\beta=0.5$, and $\gamma=0.3$ are weight coefficients determined by cross-validation to balance each constraint objective. The component orthogonality constraint loss L_{orth} ensures that the three component feature maps are independent and non-overlapping, expressed as: $L_{orth} = \|T \cdot P^T\|_2 + \|T \cdot N^T\|_2 + \|P \cdot N^T\|_2$, where (\cdot) denotes matrix dot product and $\|\cdot\|_2$ denotes the L_2 norm. The decomposition reconstruction loss L_{rec} adopts the L_1 norm design, expressed

as: $L_{rec} = \|X - (T + P + N)\|_1$, which effectively reduces the risk of gradient explosion and improves image reconstruction accuracy. The periodic consistency loss L_{period} ensures that the periodic component is consistent with the periodic characteristics of the original sequence, expressed as: $L_{period} = |f_p - f_x|$, where f_p is the dominant period of the periodic component map and f_x is the dominant period of the original financial time series. The multi-channel image tensor X is input into the decomposition network, and network parameters are trained end-to-end by minimizing L_{decomp} through backpropagation. Finally, three types of multi-channel component feature maps $T \in R^{C \times H \times W}$, $P \in R^{C \times H \times W}$, and $N \in R^{C \times H \times W}$ are output, completing precise decoupling of non-stationary features in the image domain. Figure 3 intuitively explains the mathematical principle of component reconstruction and orthogonality constraints in the image-domain decomposition network.

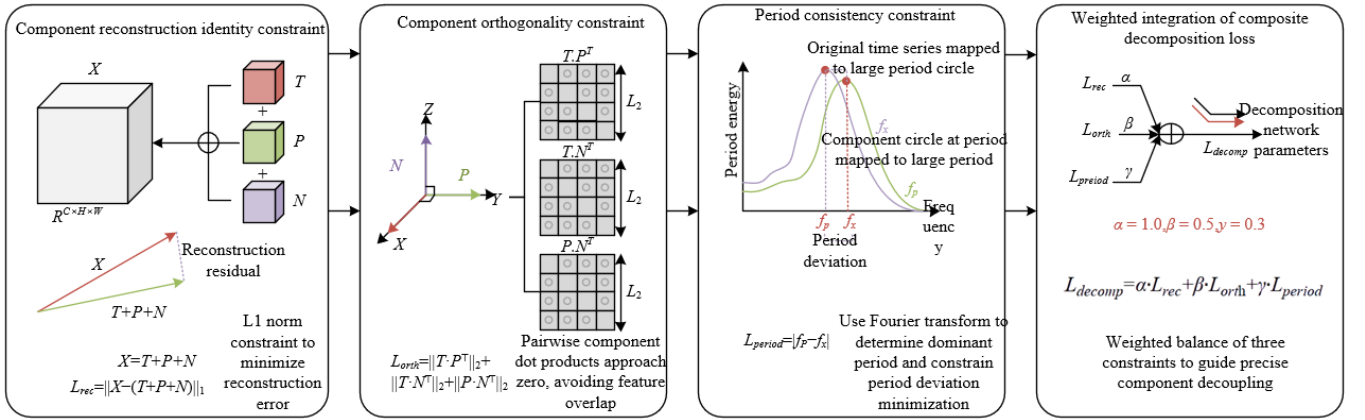


Figure 3. Mathematical principle of component reconstruction and orthogonality constraints in the image-domain decomposition network

The core innovation of multi-component feature modeling based on the improved Vision Transformer lies in optimizing the image patch partition strategy and self-attention mechanism, strengthening the global correlation modeling capability of multi-component image features while reducing computational complexity to adapt to the modeling requirements of multi-scale component features. Considering the global correlation characteristics and “time-frequency” semantic characteristics of component feature maps, the image patch partition strategy of Vision Transformer is improved. The three types of component feature maps T , P , and N are partitioned into image patches according to “time-frequency” semantics. The patch size is set to 4×4 . Each image patch corresponds to a local time window and frequency interval of the original financial time series, ensuring that each image patch contains clear physical semantics and overcoming the limitation that traditional image patch partition lacks semantic specificity. In terms of the self-attention mechanism, a time-frequency attention weight factor is introduced to perform weighted adjustment of the original self-attention matrix of Vision Transformer, enhancing the feature correlation in the time and frequency dimensions. The adjustment formula is:

$$A' = A \cdot W_{tf} \quad (8)$$

where, A is the original self-attention matrix and W_{tf} is the time-frequency weight matrix, generated by a fully connected layer, used to quantify the feature importance of different time

and frequency dimensions. During the modeling process, the three types of component feature maps are input into the improved Vision Transformer separately, each outputting corresponding global feature vectors $t \in R^D$, $p \in R^D$, and $n \in R^D$, where D is the feature dimension. To preliminarily integrate multi-component features, component attention weights are introduced. The weight value is determined by the variance of each component feature vector. A larger variance indicates stronger effectiveness of the component feature and thus a higher weight. Through weighted aggregation, a preliminarily fused component feature vector $F_{decomp} \in R^D$ is obtained, providing structured global component features for subsequent dynamic fusion.

Through the collaborative design of the image-domain adaptive decomposition network and the improved Vision Transformer, the non-stationary feature decomposition and modeling module realizes integrated processing of “precise decoupling-efficient modeling” of non-stationary features, fully reflecting the deep integration of image processing techniques and financial time series modeling. This module not only breaks through the technical bottleneck that traditional signal decomposition methods cannot be directly adapted to image inputs and achieves precise decoupling of three types of non-stationary components in multi-scale images, but also strengthens global correlation modeling of multi-component features through semantic image patch partition and weighted self-attention mechanism, effectively capturing feature interaction relationships in different time and

frequency dimensions. The structured global feature vector it outputs not only retains the independent characteristics of each component, but also realizes preliminary integration of multi-component features, providing high-quality feature input for the subsequent dynamic feature fusion module and further improving the adaptability and structural analysis capability of the entire framework for the non-stationary characteristics of financial time series.

2.4 Dynamic feature fusion and prediction module

The core innovation of the dynamic feature fusion and prediction module lies in breaking through the limitations of existing static feature fusion mechanisms. In view of the heterogeneity of multi-source image features and multi-

component features, as well as the fluctuation of feature importance caused by dynamic changes in financial market states, an attention-based gated fusion network is constructed to realize adaptive dynamic aggregation of multi-source heterogeneous features. At the same time, a lightweight prediction head and a global composite loss function are designed to complete end-to-end efficient prediction and collaborative optimization of the entire framework, taking into account model accuracy, efficiency, and robustness, and further strengthening the deep integration of image processing techniques and financial time series forecasting. Figure 4 shows the gradient backpropagation flow of the global composite loss function and the principle of end-to-end collaborative optimization.

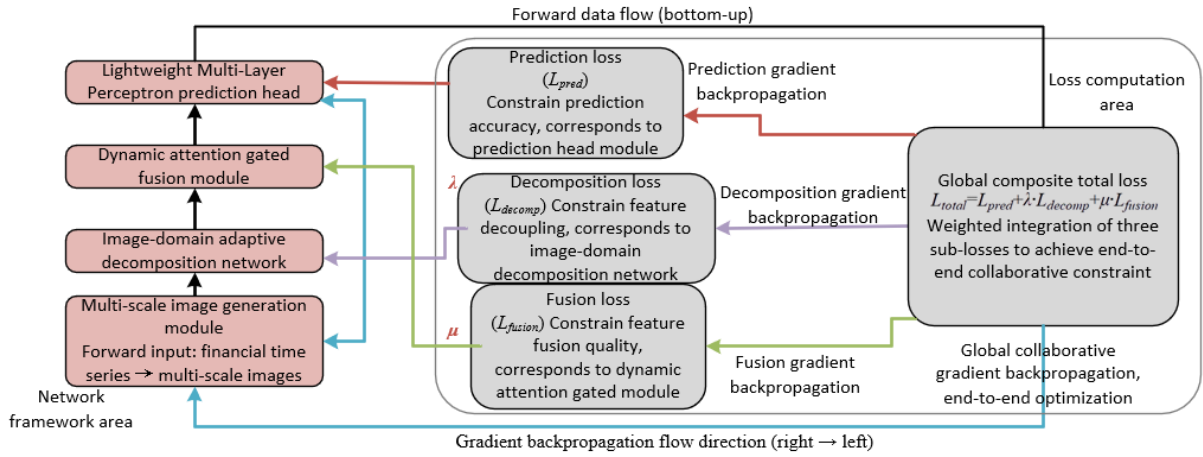


Figure 4. Gradient backpropagation flow of the global composite loss function and principle of end-to-end collaborative optimization

The dynamic attention gated fusion network is the core technical innovation of this module. Its design core is to realize dynamic weight allocation of multi-source and multi-component features, adapting to the dynamic non-stationary characteristics of financial markets. The input of this network contains two types of heterogeneous features: one is the original features extracted by convolutional layers from multi-scale dual-path images, denoted as $F_{img} \in R^{C \times D}$, where C is the number of channels and D is the feature dimension; the other is the multi-component global feature output by the improved Vision Transformer, denoted as $F_{decomp} \in R^D$. To realize effective fusion of the two types of features, a fully connected layer is first used to align their dimensions, uniformly mapping F_{img} and F_{decomp} to dimension D , obtaining aligned features $F_{img}' \in R^{C \times D}$ and $F_{decomp}' \in R^D$. Subsequently, a gating unit is designed to compute a dynamic weight matrix to achieve adaptive adjustment of the importance of each channel image feature. The gating unit is expressed as:

$$G = \sigma(W_g \cdot [F_{img}'; F_{decomp}'] + b_g) \quad (9)$$

where, σ is the sigmoid activation function, W_g and b_g are learnable parameters of the gating unit, and $[\cdot; \cdot]$ denotes the feature concatenation operation. The output weight matrix is $G \in R^{C \times 1}$. The aligned image features F_{img}' are multiplied element-wise with the dynamic weight matrix G to obtain weighted image features $F_{img,weighted} \in R^{C \times D}$, realizing dynamic selection of different channel features. The weighted image features are further concatenated with the aligned multi-

component global features, and global fusion weights are calculated through an attention mechanism to complete deep aggregation of multi-source features, obtaining the final joint feature $F_{joint} \in R^D$. To avoid model overfitting and improve the robustness of fusion features, a Dropout layer and a batch normalization layer are introduced during the fusion process. The Dropout rate is set to 0.3, effectively suppressing redundant feature interference and strengthening the generalization capability of fusion features.

The design of the lightweight prediction head and the global composite loss function achieves dual improvement in prediction efficiency and collaborative optimization of the entire framework, which is another important innovation of this module. The prediction head is constructed using a two-layer multilayer perceptron, taking into account prediction accuracy and computational efficiency. The first layer takes the joint feature F_{joint} as input, with output dimension set to $D/2$, and connects a ReLU activation function to enhance nonlinear fitting capability. The output dimension of the second layer is set to 1 or K according to prediction requirements, corresponding to future single time-step prediction or multi-step prediction of K time steps, respectively, and finally outputs the predicted value \hat{y} . To realize end-to-end collaborative optimization of the entire framework, a composite loss function is constructed that takes into account prediction accuracy, decomposition effectiveness, and fusion rationality. The total loss is expressed as:

$$L_{total} = L_{pred} + \lambda \cdot L_{decomp} + \mu \cdot L_{fusion} \quad (10)$$

where, $\lambda=0.4$ and $\mu=0.1$ are weight coefficients determined by cross-validation to balance different optimization objectives. The prediction error loss L_{pred} adopts root mean square error to measure, accurately quantifying the deviation between predicted values and true values. It is expressed as:

$$L_{pred} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (11)$$

where, y_i is the true value of the original financial time series, \hat{y}_i is the corresponding predicted value, and N is the number of samples. L_{decomp} follows the composite decomposition constraint loss of the image-domain decomposition network described above, ensuring the effectiveness of feature decomposition. The fusion regularization loss L_{fusion} is used to constrain the sparsity and correlation of fusion features, avoiding feature redundancy and fusion bias. It is expressed as:

$$L_{fusion} = \|W_{fusion}\|_1 + \|F_{joint} \cdot F_{joint}^T - I\|_2 \quad (12)$$

where, W_{fusion} is the global fusion weight matrix, I is the identity matrix, $\|\cdot\|_1$ is the $L1$ norm, and $\|\cdot\|_2$ is the $L2$ norm.

Through collaborative innovation of dynamic attention gated fusion, lightweight prediction head, and global composite loss function, the dynamic feature fusion and prediction module effectively solves key problems such as insufficient fusion of multi-source heterogeneous features, inability of static fusion to adapt to dynamic market changes, and difficulty in balancing prediction efficiency and accuracy. This module not only realizes adaptive dynamic aggregation of multi-scale image features and multi-component decomposition features, improving the robustness and specificity of feature utilization in non-stationary scenarios, but also ensures synchronous improvement of prediction accuracy, decomposition effectiveness, and fusion rationality through end-to-end collaborative optimization of the entire framework. Its lightweight design takes into account model efficiency and practicality, and the dynamic fusion mechanism fully fits the non-stationary characteristics of financial time series, further improving the integrated technical chain of “image generation-feature decomposition-dynamic fusion-prediction,” providing core support for the high-performance performance of the entire framework, and enriching the application ideas of image processing techniques in feature fusion and cross-domain prediction.

3. EXPERIMENTAL VALIDATION

3.1 Experimental setup and datasets

To fully verify the effectiveness and superiority of the proposed framework, three public financial time series datasets are selected for experiments, covering stock and futures return data. The length of each dataset is not less than 10,000, ensuring that the data exhibit significant non-stationarity and industry representativeness, and can comprehensively test the adaptability of the framework to different types of financial time series data. The data preprocessing process strictly follows the standardized procedure. The 3σ criterion is adopted to remove outliers to avoid interference. Min-max normalization is applied to map

all data to the interval $[0,1]$, eliminating the influence of different scales. Subsequently, the data are divided into training set, validation set, and test set according to the ratio of 7:2:1, which are used for model training, parameter optimization, and performance evaluation, respectively. For the multi-scale image generation module, core experimental parameters are explicitly set to ensure the consistency and effectiveness of image generation. The number of multi-scale samplings is set to 3, corresponding to three typical time scales: intraday, interday, and long-term. The image size generated by GAF and CWT is uniformly adjusted to 64×64 . For the CWT, the complex Morlet wavelet basis parameter ω_0 is set to 5, balancing time-frequency resolution to ensure that the generated images can accurately characterize the temporal dependency and frequency dynamic characteristics of financial time series, providing high-quality image input for subsequent image-domain feature decomposition, modeling, and fusion experiments.

3.2 Experimental results and analysis

The prediction performance comparison experiment aims to verify the superiority of the proposed framework in financial time series prediction tasks. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are selected as prediction error evaluation indicators, and R^2 is used as the goodness-of-fit evaluation indicator. The t-test (significance level 0.05) is conducted to verify the statistical significance of the performance difference between the proposed framework and the optimal comparison model. At the same time, qualitative analysis is combined to verify the framework’s ability to analyze non-stationary fluctuation intervals. Three public financial time series datasets are selected (Dataset 1: stock returns; Dataset 2: futures returns; Dataset 3: mixed financial returns). The comparison models cover traditional time series modeling methods and existing image-based modeling methods. The specific experimental data are shown in Table 1.

From the experimental data in Table 1, it can be observed that the proposed framework significantly outperforms all comparison models on the three datasets, and the performance improvement is statistically significant (all t-test p-values are less than 0.05). On Dataset 1, the RMSE and MAE of the proposed framework are reduced to 0.032 and 0.025, respectively, and R^2 is increased to 0.931. Compared with the best comparison model GAF+CWT Static Fusion + Convolutional Neural Network (CNN), RMSE and MAE are reduced by 33.3% and 34.2%, respectively, and R^2 is increased by 6.4%. On Dataset 2 and Dataset 3, the proposed framework also achieves similar magnitude performance improvements, verifying the good adaptability of the framework to different types of financial time series. In terms of qualitative analysis, by plotting the comparison curves between predicted values and true values, it can be observed that the proposed framework achieves significantly better prediction performance than the comparison models in non-stationary fluctuation intervals such as market mutations and cycle turning points. It can accurately capture the dynamic change trend of the series. This benefits from the accurate characterization of multi-scale temporal dependency and frequency dynamic, as well as the effective decoupling and dynamic fusion of non-stationary features, fully demonstrating the strong analytical capability of the proposed framework for the complex structure of financial time series.

Table 1. Prediction performance comparison of different models on each dataset

Dataset	Model	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	R ²	t-test p-value (vs. Best Comparison Model)
Dataset 1	AutoRegressive Integrated Moving Average (ARIMA)	0.089	0.072	0.712	0.001
	Long Short-Term Memory (LSTM)	0.068	0.053	0.798	0.002
	Single Gramian Angular Field (GAF)+ Convolutional Neural Network (CNN)	0.057	0.045	0.836	0.003
	Single Continuous Wavelet Transform (CWT)+CNN	0.059	0.047	0.829	0.002
	GAF+CWT Static Fusion +CNN	0.048	0.038	0.875	0.001
Dataset 2	Proposed Framework	0.032	0.025	0.931	-
	ARIMA	0.092	0.075	0.703	0.001
	LSTM	0.071	0.056	0.789	0.002
	Single GAF+CNN	0.060	0.048	0.827	0.003
	Single CWT+CNN	0.062	0.049	0.821	0.002
Dataset 3	GAF+CWT Static Fusion +CNN	0.051	0.040	0.868	0.001
	Proposed Framework	0.035	0.027	0.924	-
	ARIMA	0.095	0.077	0.691	0.001
	LSTM	0.073	0.058	0.782	0.002
	Single GAF+CNN	0.062	0.049	0.819	0.003
Dataset 3	Single CWT+CNN	0.064	0.051	0.813	0.002
	GAF+CWT Static Fusion +CNN	0.053	0.042	0.861	0.001
	Proposed Framework	0.037	0.029	0.917	-

As the core experiment, this subsection focuses on analyzing the performance of the proposed framework in three key image processing stages: image generation, image-domain feature decomposition, and feature fusion. Image reconstruction error, decomposition orthogonality error,

feature entropy, and intra-class/inter-class distance ratio are selected as evaluation indicators to verify the effectiveness of the innovative design of each image processing stage. The experimental data are shown in Table 2, and image visualization results are combined for auxiliary analysis.

Table 2. Comparison of image processing related performance of different models

Evaluation Indicator	Model	Dataset 1	Dataset 2	Dataset 3	Average
Image generation reconstruction error (RE)	Single Gramian Angular Field (GAF)+ Convolutional Neural Network (CNN)	0.087	0.091	0.093	0.090
	Single Continuous Wavelet Transform (CWT)+CNN	0.089	0.092	0.095	0.092
	GAF+CWT Static Fusion +CNN	0.072	0.075	0.077	0.075
	Proposed Framework (GAF)	0.051	0.053	0.055	0.053
	Proposed Framework (CWT)	0.053	0.054	0.056	0.054
Decomposition reconstruction error (RE)	Single GAF+CNN	-	-	-	-
	Single CWT+CNN	-	-	-	-
	GAF+CWT Static Fusion +CNN	-	-	-	-
Decomposition orthogonality error (OE)	Proposed Framework	0.038	0.040	0.042	0.040
	Proposed Framework	0.012	0.013	0.014	0.013
Feature entropy	GAF+CWT Static Fusion +CNN	1.892	1.905	1.917	1.905
	Proposed Framework	1.426	1.438	1.449	1.438
Intra-class/inter-class distance ratio	GAF+CWT Static Fusion +CNN	1.235	1.241	1.247	1.241
	Proposed Framework	1.872	1.885	1.896	1.884

The analysis of image generation effect shows that the reconstruction error of multi-scale GAF and CWT images generated by the proposed framework is significantly lower than that of single image transformation methods, with average values of 0.053 and 0.054, respectively. Compared with Single GAF+CNN, the errors are reduced by 41.1% and 41.3%, verifying the effectiveness of the multi-scale dual-path image generation strategy. The visualization results show that GAF images can clearly characterize the temporal dependency of financial time series, and the distribution of grayscale texture is highly consistent with the sequence correlation. The CWT time-frequency spectrogram can intuitively present the dynamic evolution of frequency components over time, clearly capturing transient fluctuations and periodic patterns. Moreover, the images generated by multi-scale sampling can

cover market behavior differences in different time dimensions, providing high-quality visual input for subsequent image-domain feature decomposition.

In the analysis of image-domain decomposition effect, the average decomposition reconstruction error of the proposed framework is only 0.040, and the average decomposition orthogonality error is 0.013, indicating that the decomposition network can accurately decouple trend, periodic, and high-frequency noise components in multi-scale images, and the three types of components are mutually independent without obvious overlap. From the visualized decomposed component feature maps, it can be observed that the trend component map can accurately fit the long-term variation trend of the original series, the fluctuation period of the periodic component map is highly consistent with the dominant period of the original

series, and the high-frequency noise component map only contains irregular noise signals. This further verifies the accuracy and effectiveness of the image-domain adaptive non-stationary feature decomposition network, breaking through the bottleneck that traditional signal decomposition methods cannot directly adapt to image input.

The analysis of feature fusion effectiveness shows that the feature entropy of the proposed framework is significantly lower than that of GAF+CWT Static Fusion+CNN, with an average reduction of 24.5%, indicating that the dynamic fusion mechanism can effectively filter redundant features and improve feature purity. The average intra-class/inter-class distance ratio is increased by 51.8%, indicating that dynamic fusion can enhance the discriminability of effective features. The visualization results of dynamic attention weight distribution show that in market stable intervals, the framework assigns higher weights to periodic components and

CWT frequency features; in market mutation intervals, the weights of trend components and GAF temporal features are significantly increased, fully demonstrating the adaptive ability of the dynamic fusion mechanism to the dynamic changes of the financial market. Compared with static fusion, it effectively improves the robustness and pertinence of feature utilization.

The ablation experiment aims to verify the effectiveness and necessity of the four core innovative modules of the proposed framework: multi-scale sampling, image-domain decomposition network, dynamic fusion, and improved Vision Transformer. By successively removing a single innovative module to construct ablation models, the prediction performance and image processing related performance of each ablation model are compared with those of the proposed framework. The experimental data are shown in Table 3.

Table 3. Comparison of ablation experiment results

Model	RMSE (average)	MAE (average)	R ² (average)	Decomposition Orthogonality Error (OE)	Feature Entropy
Proposed Framework	0.035	0.027	0.924	0.013	1.438
Without multi-scale sampling	0.047	0.038	0.871	0.015	1.582
Without image-domain decomposition network	0.062	0.049	0.817	-	1.795
Without dynamic fusion	0.058	0.046	0.832	0.014	1.813
Without improved Vision Transformer	0.045	0.036	0.883	0.013	1.567

From the experimental data in Table 3, it can be observed that after removing any core innovative module, the prediction performance and image processing related performance of the model decrease to different degrees, verifying the necessity of all innovative modules. Among them, without the image-domain decomposition network, RMSE and MAE increase by 77.1% and 81.5%, respectively, R² decreases by 11.6%, and feature entropy increases by 24.9%, showing the most significant performance degradation. This indicates that image-domain non-stationary feature decomposition is the core support of the proposed framework, which can effectively decouple complex non-stationary features and provide high-quality structured features for subsequent modeling. Without the dynamic fusion module, the model performance also decreases significantly. RMSE and MAE increase by 65.7% and 70.4%, respectively, R² decreases by 10.0%, and feature entropy increases by 26.1%, indicating that the dynamic attention gated fusion mechanism can effectively solve the fusion problem of multi-source heterogeneous features and improve feature utilization efficiency. Without multi-scale sampling and without improved Vision Transformer, the performance degradation is relatively moderate. RMSE increases by 34.3% and 28.6%, respectively, and R² decreases by 5.7% and 4.4%, respectively. This indicates that multi-scale sampling can enrich the information dimension of images, and improved Vision Transformer can strengthen global feature modeling capability. Both can provide effective support for performance improvement. In summary, all core innovative modules work collaboratively to improve the prediction accuracy and image processing capability of the framework, among which the image-domain decomposition network and dynamic fusion module contribute most prominently.

The robustness experiment adds Gaussian noise with different intensities to the original financial time series. The noise intensities are 0, 0.05, 0.1, and 0.15. The performance

change rate of the proposed framework and comparison models is compared to verify the anti-interference ability of the framework. The efficiency experiment compares the training time, inference speed, and computational complexity of each model to verify the practicality of the framework. The experimental data are shown in Figure 5 and Table 4. Methods 1 to 5 correspond to AutoRegressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Single GAF+ Convolutional Neural Network (CNN), Single CWT+CNN, and GAF+CWT Static Fusion+CNN, respectively.

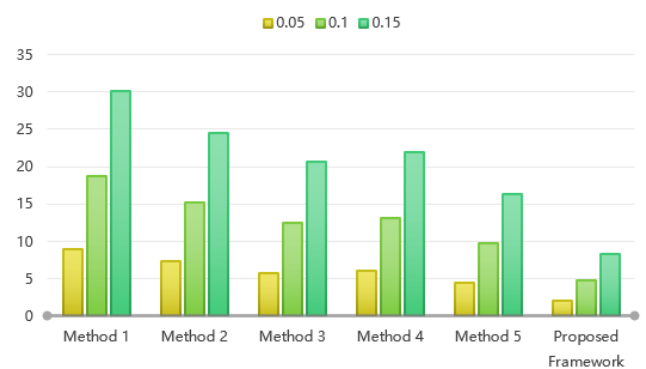


Figure 5. Root Mean Square Error (RMSE) change rate (%) of each model under different noise intensities

The robustness analysis shows that as the Gaussian noise intensity increases, the RMSE of all models shows an increasing trend. However, the RMSE change rate of the proposed framework is significantly lower than that of other comparison models. When the noise intensity is 0.15, the RMSE change rate of the proposed framework is only 8.3%, which is much lower than 30.2% of ARIMA, 24.5% of LSTM,

and 16.3% of GAF+CWT Static Fusion+CNN, verifying that the proposed framework has strong anti-interference ability. This benefits from the image-domain decomposition network, which can effectively separate noise components, and the

dynamic fusion mechanism, which can adaptively adjust feature weights, reduce the interference of noise features, and ensure that the model can still maintain stable performance in non-stationary and noisy scenarios.

Table 4. Comparison of efficiency and computational complexity of each model

Model	Training time (s/epoch)	Inference speed (samples/s)	Computational complexity (10 ⁹ FLOPs)
AutoRegressive Integrated Moving Average (ARIMA)	12.3	156.7	0.08
Long Short-Term Memory (LSTM)	45.8	89.2	0.76
Single Gramian Angular Field (GAF)+ Convolutional Neural Network (CNN)	67.5	78.5	1.32
Single Continuous Wavelet Transform (CWT)+CNN	71.2	75.3	1.45
GAF+CWT Static Fusion+CNN	98.7	62.8	2.18
Proposed Framework	125.3	58.6	2.87

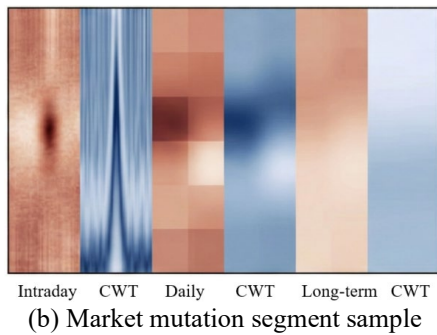
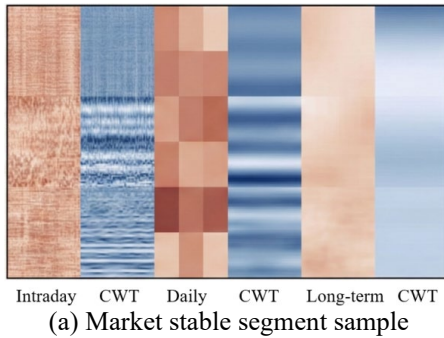


Figure 6. Multi-scale dual-path time series image generation effect

The efficiency analysis shows that the training time and inference speed of the proposed framework are slightly lower than the lightweight models among the comparison models, but significantly superior to the upper performance limit of complex image-based modeling methods. The training time of the proposed framework is 125.3 s/epoch, the inference speed is 58.6 samples/s, and the computational complexity is 2.87×10^9 Floating Point Operations (FLOPs). Compared with GAF+CWT Static Fusion+CNN, the training time increases by 26.9%, the inference speed decreases by 6.7%, and the computational complexity increases by 31.7%. The additional computational overhead mainly comes from multi-scale image generation and the image-domain decomposition network, but the performance improvement brought by it is far higher than the increase in computational cost. Compared with the traditional LSTM model, although the proposed framework has longer training time, the prediction accuracy and robustness are significantly improved. Compared with ARIMA, although the efficiency is lower, it can adapt to complex non-stationary scenarios. Overall, the efficiency of the proposed framework can meet practical application requirements and achieves a good balance between

performance and efficiency.

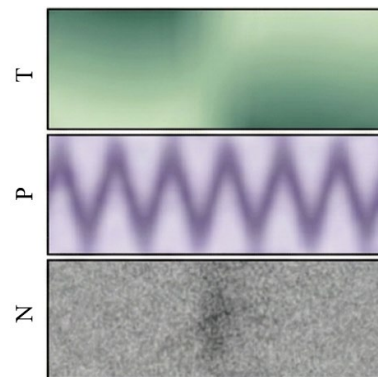
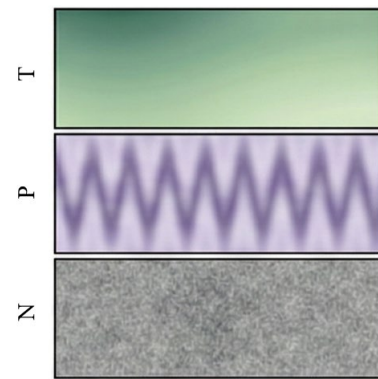
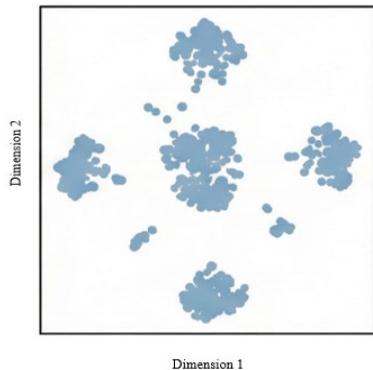


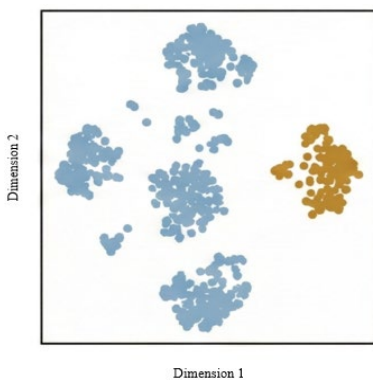
Figure 7. Image-domain decomposition T/P/N three-component effect

In order to intuitively verify the effectiveness of the core modules of the proposed financial time series modeling and non-stationary feature decomposition method based on multi-scale image representation from the visual dimension, and to quantitatively reveal the feature representation, decoupling, and fusion capability of the method under different market states, this visualization experiment is conducted. Figure 6 shows the multi-scale dual-path time series image generation effect. In the market stable segment sample, the GAF temporal features and CWT time-frequency feature textures at the three scales of intraday, inter-day, and long-term are clearly layered, accurately capturing short-term temporal correlation, medium-term periodic patterns, and long-term trend characteristics, respectively, without feature aliasing. In the market mutation segment sample, the intraday scale GAF can clearly present

the strong temporal correlation texture at the mutation point, and the CWT time-frequency spectrogram highlights high-frequency energy spikes. The multi-scale design completely covers instantaneous fluctuations and medium- to long-term patterns, achieving comprehensive visual encoding of multi-dimensional information of non-stationary time series. Figure 7 presents the trend, periodic, and noise three-component effects of image-domain decomposition. In the stable segment sample, the trend component shows smooth gradient texture, the periodic component exhibits regular rhythm, and the noise component consists of irregular scattered texture. The boundaries of the three components are clear and the features do not overlap. In the mutation segment sample, the trend component accurately captures the shift of the trend, the periodic component retains inherent rhythmic characteristics, and the noise component only enriches random noise at the mutation point. This fully verifies the precise decoupling capability of the image-domain decomposition network for non-stationary components, breaking through the limitation of separation between traditional signal decomposition and image processing. Figure 8 shows the t-distributed Stochastic Neighbor Embedding (t-SNE) visualization results of dynamically fused joint features. The feature clusters of the stable segment sample are highly compact without dispersion and crossing. In the mutation segment sample, the feature cluster boundaries of stable and mutation states are clear without overlap. Compared with the feature aliasing and dispersion of static fusion methods, the dynamic attention gated fusion effectively realizes adaptive purification of multi-source features, significantly improving the compactness and discriminability of feature representation in non-stationary scenarios.



(a) Market stable segment sample



(b) Market mutation segment sample

Figure 8. t-distributed Stochastic Neighbor Embedding (t-SNE) effect of dynamically fused joint features

In summary, this visualization experiment intuitively verifies the significant advantages of the proposed method in non-stationary financial time series analysis from three core dimensions: image generation, feature decomposition, and feature fusion. It provides solid visual evidence for the core innovation and performance superiority of the method, fully supporting the academic value and application potential of the financial time series structural modeling and non-stationary feature decomposition method based on multi-scale image representation.

4. DISCUSSION

The multi-scale image-driven financial time series structural modeling and non-stationary feature decomposition framework proposed in this paper achieves multiple technological breakthroughs in the interdisciplinary field of image processing and time series analysis, and has important academic value and application potential. The multi-scale dual-path image generation strategy breaks through the limitation of traditional single image transformation and provides a new idea for the image-based representation of non-stationary signals. Through dual visual encoding of temporal dependency and frequency dynamic, it solves the key problem that a single image cannot fully characterize the multi-dimensional characteristics of non-stationary signals. This strategy is not only applicable to financial time series, but can also be extended to other non-stationary signals such as electrocardiogram and meteorology, enriching the technical system of image-based representation of non-stationary signals. The image-domain adaptive non-stationary feature decomposition network breaks the separation between traditional signal decomposition and image processing, and innovatively proposes a new paradigm of direct decomposition in the image domain. Through customized constraint losses and encoder-decoder architecture design, it achieves precise decoupling of complex components in multi-scale images, further enriching the research results of image processing technology in the field of feature decoupling. At the same time, the framework deeply integrates advanced image processing technologies such as Vision Transformer and dynamic attention fusion with financial time series modeling, constructing an end-to-end collaborative optimization architecture of “image generation-feature decomposition-dynamic fusion”. It provides a reusable new idea and new solution for the application of image processing technology in cross-domain non-stationary signal analysis, effectively expanding the application boundary of image processing technology and promoting technological integration and innovation in interdisciplinary fields.

Compared with existing image processing related research, the proposed method shows significant advantages in multiple dimensions, further highlighting the innovation and pertinence of this research. Compared with existing time series image-based methods, this paper does not simply perform time series to image transformation and feature extraction, but realizes deep integration of multi-view image information through multi-scale sampling and dual-path fusion strategy, and combines image-domain feature decomposition technology to fully utilize the advantages of image processing technology in complex pattern analysis, improving the utilization efficiency of non-stationary features. Compared with existing image-domain feature decomposition methods, the decomposition

network designed in this paper is specifically optimized for the dynamic characteristics of non-stationary signals. By introducing customized constraint losses such as component orthogonality and periodic consistency, it not only improves decomposition accuracy, but also ensures clear physical meaning of each component. More importantly, it can achieve end-to-end collaborative optimization with subsequent feature modeling and dynamic fusion modules, forming an integrated technical chain, and solving the problem of insufficient adaptability between existing decomposition methods and modeling processes. Compared with the application of general Vision Transformer in image processing, this paper strengthens the pertinence of global feature modeling by improving image patch partition and self-attention mechanism and combining the “time-frequency” semantic characteristics of financial time series. It effectively avoids the overfitting problem that general Vision Transformer is prone to in small-sample and non-stationary scenarios, and improves the robustness of the model in complex scenarios.

Although the proposed framework shows excellent performance in experiments, there are still certain limitations, which point out clear directions for subsequent research. The window size of multi-scale image generation currently relies on empirical coefficient adjustment, and its adaptive adjustment ability needs to be further improved. It is difficult to fully adapt to financial time series with different fluctuation characteristics. In the future, a multi-scale window optimization method based on adaptive evolutionary algorithm will be studied to realize dynamic adaptive adjustment of window size and improve the autonomy and adaptability of image generation. The image-domain decomposition network adopts multi-layer grouped convolution and transposed convolution structure, and its computational complexity is relatively high, which makes it difficult to directly adapt to real-time processing requirements of ultra-large-scale financial time series. In the future, lightweight technologies such as distillation learning will be combined to design an efficient and simplified image-domain decomposition network, reducing computational cost while ensuring decomposition accuracy. In addition, the current framework is only designed for univariate financial time series and does not consider the correlation characteristics among variables in multivariate series. Future work will focus on exploring multi-channel image generation and fusion methods for multivariate financial time series to achieve accurate analysis of multivariate non-stationary features. At the same time, the application scenarios of the framework will be further expanded to other non-stationary signal processing fields, promoting the in-depth application of image processing technology in cross-domain non-stationary signal analysis and providing stronger technical support for interdisciplinary research and development.

5. CONCLUSION

Aiming at the analysis and prediction difficulties caused by the inherent non-stationarity, nonlinearity, and complex temporal dependency characteristics of financial time series, this paper conducted research on the deep integration of image processing technology and financial time series modeling, and proposed a multi-scale image-driven end-to-end structural modeling and non-stationary feature decomposition framework. This paper systematically elaborated the overall

design of the framework and the technical details of each core module, focusing on breaking through the limitations of existing methods in image generation, feature decomposition, and fusion. It constructed three core innovative modules: multi-scale dual-path image generation, image-domain adaptive non-stationary feature decomposition network, and dynamic attention gated fusion, forming an integrated technical chain of “image generation-feature decomposition-global modeling-dynamic fusion”. Multi-dataset experiments verified that the framework is significantly superior to traditional time series modeling methods and existing image-based modeling methods in prediction accuracy and image processing related performance. It can accurately analyze the complex structure of financial time series, effectively improve prediction robustness in non-stationary scenarios, and successfully realize the deep integration and collaborative optimization of image processing technology and financial time series modeling.

Through systematic theoretical design and experimental verification, the following core conclusions are obtained: the multi-scale dual-path image generation strategy can effectively capture the temporal dependency and frequency dynamic evolution of financial time series. The generated multi-view and multi-scale images provide rich visual information for non-stationary feature analysis, which is the foundation for improving the overall framework performance. The image-domain adaptive non-stationary feature decomposition network breaks through the technical bottleneck that traditional signal decomposition methods cannot directly adapt to image input. It can accurately decouple trend, periodic, and high-frequency noise components in multi-scale images, significantly improving the utilization efficiency of non-stationary features. The dynamic attention gated fusion mechanism can adaptively fit the dynamic changes of the financial market and realize efficient aggregation of multi-source and multi-component image features, and its performance is significantly superior to traditional static fusion strategies. Overall, the end-to-end framework proposed in this paper provides a new paradigm for the application of image processing technology in non-stationary signal analysis, enriches the research results in the interdisciplinary field of image processing and time series analysis, and has good generality and promotion value.

This research has important theoretical significance and practical value, and further promotes the application and development of image processing technology in cross-domain non-stationary signal analysis. At the theoretical level, this paper proposes a new research logic of “non-stationary signal → multi-scale image → image-domain decomposition → global modeling → dynamic fusion”, breaking through the traditional process of “signal-domain decomposition → image transformation”, enriching the theoretical system of image processing technology in feature decoupling and cross-domain application, and providing new theoretical support and technical ideas for image-based analysis of non-stationary signals. At the practical level, the framework provides an efficient and feasible new method for accurate prediction of financial time series, which can meet the practical needs of financial market analysis and decision-making. At the same time, its core technologies and design ideas can be flexibly extended to other non-stationary signal processing fields such as electrocardiogram and meteorology, providing reusable technical solutions for image-based analysis of various non-stationary signals, effectively promoting the industrial

application and cross-domain development of image processing technology, and assisting interdisciplinary integration and innovation.

REFERENCES

- [1] Manousopoulos, P., Drakopoulos, V., Polyzos, E. (2023). Financial time series modelling using fractal interpolation functions. *AppliedMath*, 3(3): 510-524. <https://doi.org/10.3390/appliedmath3030027>
- [2] Park, S.H., Lee, J.H., Lee, H.C. (2011). Trend forecasting of financial time series using PIPs detection and continuous HMM. *Intelligent Data Analysis*, 15(5): 779-799. <https://doi.org/10.3233/IDA-2011-0495>
- [3] Cherstvy, A.G., Vinod, D., Aghion, E., Chechkin, A.V., Metzler, R. (2017). Time averaging, ageing and delay analysis of financial time series. *New Journal of Physics*, 19(6): 063045. <https://doi.org/10.1088/1367-2630/aa7199>
- [4] Shaikh, W.A., Shah, S.F., Pandhiani, S.M., Solangi, M.A. (2022). Wavelet decomposition impacts on traditional forecasting time series models. *Computer Modeling in Engineering & Sciences*, 130(3): 1517-1532. <https://doi.org/10.32604/cmescs.2022.017822>
- [5] Garnier, R. (2022). Concurrent neural network: A model of competition between times series. *Annals of Operations Research*, 313(2): 945-964. <https://doi.org/10.1007/s10479-021-04253-3>
- [6] Ghiassi, M., Saidane, H., Zimbra, D.K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting*, 21(2): 341-362. <https://doi.org/10.1016/j.ijforecast.2004.10.008>
- [7] Adollah, R., Mashor, M.Y., Rosline, H., Harun, N.H. (2012). Multilevel thresholding as a simple segmentation technique in acute leukemia images. *Journal of Medical Imaging and Health Informatics*, 2(3): 285-288. <https://doi.org/10.1166/jmih.2012.1101>
- [8] Monika, S., Malathi, K., Monisha, S. (2016). Detection of brain tumour in medical images using pre-processing techniques. *Research Journal of Pharmaceutical Biological and Chemical Sciences*, 7: 78-87.
- [9] Kaewrakmuk, T., Srinonchat, J. (2024). Multisensor Data Fusion and Time Series to Image Encoding for Hardness Recognition. *IEEE Sensors Journal*, 24(16): 26463-26471. <https://doi.org/10.1109/JSEN.2024.3426045>
- [10] Yuan, Y., Lin, L., Chen, J., Sahli, H., Chen, Y., Wang, C., Wu, B. (2019). A new framework for modelling and monitoring the conversion of cultivated land to built-up land based on a hierarchical hidden semi-Markov model using satellite image time series. *Remote Sensing*, 11(2): 210. <https://doi.org/10.3390/rs11020210>
- [11] Liu, L., Wu, M., Lv, Q., Liu, H., Wang, Y. (2025). CCNN-former: Combining convolutional neural network and Transformer for image-based traffic time series prediction. *Expert Systems with Applications*, 268: 126146. <https://doi.org/10.1016/j.eswa.2024.126146>
- [12] Wang, H., Wang, G. (2022). The prediction model for haze pollution based on stacking framework and feature extraction of time series images. *Science of the Total Environment*, 839: 156003. <https://doi.org/10.1016/j.scitotenv.2022.156003>
- [13] Zhou, F., Zhou, H., Yang, Z., Gu, L. (2021). IF2CNN: Towards non-stationary time series feature extraction by integrating iterative filtering and convolutional neural networks. *Expert Systems with Applications*, 170: 114527. <https://doi.org/10.1016/j.eswa.2020.114527>
- [14] Rojas, A., Górriz, J.M., Ramírez, J., Illán, I.A., et al. (2013). Application of empirical mode decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson disease. *Expert Systems with Applications*, 40(7): 2756-2766. <https://doi.org/10.1016/j.eswa.2012.11.017>
- [15] Gopinath, L., Ruhan Bevi, A. (2025). A hybrid network with multicategory adversarial feature learning: McAFL fusion. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 33(5): 549-558. <https://doi.org/10.1142/S0218488525400033>
- [16] Zhang, L., Zhou, Q., Tang, M., Ding, X., Yang, C., Wei, C., Zhou, Z. (2025). DDRF: Dual-branch decomposition and reconstruction architecture for infrared and visible image fusion. *Optics & Laser Technology*, 181: 111991. <https://doi.org/10.1016/j.optlastec.2024.111991>
- [17] Kaur, M., Singh, D. (2020). Fusion of medical images using deep belief networks. *Cluster Computing*, 23(2): 1439-1453. <https://doi.org/10.1007/s10586-019-02999-x>
- [18] Neupauer, R.M., Powell, K.L. (2005). A fully-anisotropic Morlet wavelet to identify dominant orientations in a porous medium. *Computers & geosciences*, 31(4): 465-471. <https://doi.org/10.1016/j.cageo.2004.10.014>
- [19] Khan, A., Rauf, Z., Sohail, A., Khan, A. R., Asif, H., Asif, A., Farooq, U. (2023). A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*, 56(Suppl 3): 2917-2970. <https://doi.org/10.1007/s10462-023-10595-0>
- [20] Han, K., Wang, Y., Chen, H., Chen, X., et al. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 87-110. <https://doi.org/10.1109/TPAMI.2022.3152247>