Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 319-333
Journal homepage: http://iieta.org/journals/ts

Hydrological Time Series Modeling and Prediction via Time–Frequency Image Representation

Hua Fan[1*], Dongfang Shen[1,2]

[1] School of Electronic Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450011, China
[2] Guizhou Water Conservation Science and Research Institute, Guiyang 550002, China

Corresponding Author Email: fanhua@ncwu.edu.cn

https://doi.org/10.18280/ts.430122

Received: 18 August 2025
Revised: 23 November 2025
Accepted: 16 December 2025
Available online: 28 February 2026

Keywords:
hydrological time series prediction, time–frequency image representation, spatial–frequency–temporal feature coupling, deep convolutional networks, feature fusion

ABSTRACT

Hydrological time series are significantly nonlinear and nonstationary and their accurate prediction is a fundamental challenge in hydrology and water resources. It is difficult for existing deep learning–based time series models to effectively capture dynamic time–frequency characteristics, while the integration of image processing techniques into hydrological forecasting has received insufficient attention, constraining the completeness of temporal feature representation and predictive accuracy. To address these limitations, based on time–frequency image representation, this study proposes an end-to-end deep learning prediction framework by establishing a complete modeling pipeline from temporal signals to high-precision regression prediction through the coordination of three modules: adaptive time–frequency image generation, a deep convolutional backbone network coupling spatial–frequency–temporal features (CFT-Net), and dual-stream fusion prediction. A learnable adaptive time–frequency transform dynamically generates multiscale time–frequency images according to local statistical characteristics of hydrological sequences, encoding one-dimensional temporal dynamics into two-dimensional spatial patterns. Using a parallel spatial–frequency dual-path architecture with a dual attention mechanism, the core backbone network jointly extracts spatial patterns, frequency distributions, and temporal evolution features from time–frequency images. Furthermore, an adaptive gated fusion strategy achieves complementary integration between the time–frequency image stream and the raw time series stream. Experimental results on representative hydrological datasets demonstrate that the proposed framework consistently outperforms conventional time series models and image-based prediction approaches relying on fixed time–frequency transforms in terms of prediction accuracy and robustness. These findings confirm the effectiveness and superiority of image processing techniques for hydrological time series modeling and provide a promising new pathway for high-precision hydrological forecasting.

1. INTRODUCTION

In southwestern China, particularly in Guizhou Province, karst landforms are widely distributed and precipitation exhibits pronounced spatiotemporal heterogeneity. Under the combined influences of complex terrain, climatic variability, and human activities, hydrological time series in this region are characterized by strong nonlinearity, nonstationarity, and multiscale behavior. These characteristics pose substantial challenges to rational water resources regulation, flood disaster early warning in karst mountainous areas, and ecological environmental protection [1, 2]. Accurate hydrological time series prediction therefore constitutes a critical technical foundation for regional water conservancy planning [3, 4], optimal water resource allocation [5], and disaster prevention and mitigation decision-making [6], and is of considerable practical significance and engineering value for safeguarding water security and ecological security in southwestern China. In recent years, the advantages of image processing techniques in time series representation have become increasingly evident [7-9]. By encoding the dynamical characteristics of one-dimensional hydrological sequences into interpretable two-dimensional visual patterns [10, 11], image processing techniques provide a promising avenue for addressing the intrinsic complexity of hydrological time series modeling. However, existing image-based hydrological prediction studies still exhibit notable limitations. Fixed-parameter time–frequency transforms are generally incapable of adapting to the dynamic local characteristics of hydrological time series in southwestern China [12], making it difficult to capture the transient fluctuation features of hydrological signals in Guizhou's karst basins. Moreover, feature extraction is often confined to the spatial dimension of time–frequency images [13, 14], while the frequency distribution patterns and temporal evolution characteristics are insufficiently exploited. As a consequence, deep feature representations remain incomplete. In addition, the fusion between time–frequency domain features and raw value-

domain features is typically implemented in a rigid manner, preventing effective complementary integration and ultimately constraining further improvements in prediction accuracy.

Among existing hydrological time series prediction approaches, deep learning models have been widely adopted owing to their strong feature learning capabilities, including convolutional neural networks, Transformers, and temporal convolutional networks [15-17]. Nevertheless, most of these methods directly model one-dimensional temporal data, without fully exploring the dynamic time–frequency characteristics embedded in hydrological sequences. For hydrological time series in the karst basins of Guizhou, which exhibit strong nonlinearity and pronounced fluctuations, the feature extraction capacity of such models remains limited and insufficient to meet high-precision prediction requirements. In the field of integrating time–frequency analysis with image processing for time series prediction, several studies have attempted to combine time–frequency transforms with deep learning frameworks; however, three critical gaps persist. First, conventional methods based on fixed window functions, such as the short-time Fourier transform or continuous wavelet transform, are unable to dynamically adjust transform parameters according to the local statistical properties of hydrological time series in southwestern China, resulting in limited adaptability [18, 19]. Second, feature extraction primarily focuses on spatial patterns in time–frequency images while neglecting the global distribution of frequency dimensions and the evolutionary patterns along the temporal dimension, thereby failing to comprehensively characterize the complex behavior of hydrological time series in karst basins [20, 21]. Third, effective fusion between visual features in the time–frequency domain and temporal features in the raw value domain has not been adequately achieved. The inherent limitations of relying on a single feature source lead to insufficient model robustness and generalization capability, reducing adaptability to hydrological heterogeneity across different basins in southwestern China [22, 23].

To address the core challenges associated with hydrological time series modeling in southwestern China, particularly in the karst basins of Guizhou Province, as well as the deficiencies of existing prediction approaches in terms of time–frequency adaptability, spatial–frequency–temporal relationship modeling, and feature fusion effectiveness, an end-to-end hydrological time series prediction framework based on time–frequency image representation was established. An integrated design encompassing data preprocessing, feature learning, and prediction output is pursued, with the aim of improving prediction accuracy, robustness, and generalization capability for hydrological time series in southwestern China, thereby providing a reliable technical foundation for regional water resources management and disaster prevention and mitigation. Specifically, a learnable adaptive time–frequency image generation mechanism is introduced to overcome the limitations of traditional fixed window functions. Through dynamic adjustment of transform parameters according to the local statistical characteristics of hydrological time series in southwestern China, highly adaptive, multiscale time–frequency image representations are generated, enabling effective capture of transient fluctuations and multiscale features of hydrological signals in the karst basins of Guizhou Province. In addition, a deep convolutional backbone network coupling spatial–frequency–temporal features (CFT-Net) is designed. By integrating a parallel spatial–frequency dual-path extraction strategy with temporal evolution modeling and a dual attention mechanism, the collaborative learning of spatial patterns, frequency distributions, and temporal evolution characteristics embedded in time–frequency images is achieved, allowing comprehensive exploitation of the complex feature information inherent in hydrological time series from southwestern China. Furthermore, an adaptive gated dual-stream fusion prediction mechanism is constructed, in which visual features from the time–frequency image stream and temporal features from the raw time series stream are jointly integrated. Through dynamic weight allocation, efficient complementary fusion between the two feature sources is realized, significantly enhancing model robustness and generalization capability across different karst basins in southwestern China and alleviating the limitations associated with reliance on a single feature source under complex hydrological conditions.

The remainder of the study is organized below. Section 2 provides a detailed description of the overall architecture of the hydrological time series prediction framework based on time–frequency image representation, along with the technical details of each core module. Section 3 presents experimental validation using representative hydrological datasets from typical karst basins in Guizhou Province, where comparative and ablation experiments are conducted to verify the effectiveness and superiority of the proposed approach. Section 4 offers an in-depth analysis of the experimental results, discussing the advantages, limitations, and potential application prospects of the proposed framework. Section 5 concludes the study by summarizing the main findings, clarifying the research contributions, and outlining directions for future work.

2. METHODOLOGICAL FRAMEWORK

2.1 Overall framework overview

To accurately capture the nonlinear, nonstationary, and multiscale characteristics of hydrological time series in the karst basins of Guizhou Province, southwestern China, an end-to-end hydrological time series prediction framework based on time–frequency image representation was constructed. The framework is composed of three tightly integrated components: a time–frequency image generation module, CFT-Net, and a dual-stream fusion prediction module. These components are hierarchically organized and organically connected, forming a complete deep learning closed loop from data representation and feature learning to prediction output. The core rationale of the framework lies in transforming one-dimensional hydrological time series into two-dimensional time–frequency images. By leveraging image processing techniques, spatial patterns, frequency distributions, and temporal evolution characteristics embedded within time–frequency images are systematically extracted. Through subsequent fusion with temporal features derived from the original time series, complementary enhancement between heterogeneous feature representations is achieved, ultimately enabling high-precision regression prediction of hydrological time series. The principal innovation of the framework resides in the deep coupling between image-based representation and spatial–frequency–temporal feature learning, thereby overcoming the limitations of conventional time series modeling approaches that focus on a single feature dimension.

In this manner, the strong representational capacity of image processing techniques for visual patterns is fully exploited, while the intrinsic temporal dynamics of hydrological sequences are simultaneously preserved, leading to improved accuracy and adaptability in capturing complex hydrological behaviors. From a procedural perspective, time–frequency image generation serves as the front-end preprocessing and feature enhancement stage, CFT-Net functions as the core feature extraction engine, and the dual-stream fusion prediction module acts as the output terminal. Throughout the entire workflow, emphasis is placed on the generation of time–frequency images and the extraction of deep, high-level features, ensuring alignment with modeling paradigms commonly adopted in image processing research while remaining well suited to the complex characteristics of hydrological time series in southwestern China.

2.2 Learnable parameterized time–frequency image generation module

The time–frequency image generation module serves as the preprocessing and feature enhancement front end of the overall prediction framework. Its input consists of standardized one-dimensional hydrological time series sliding windows, denoted as $X_{1D} \in R^T$, where T represents the sliding window length. In accordance with the temporal scale characteristics of hydrological time series in the karst basins of Guizhou Province, the sliding window length is dynamically determined based on the sampling frequency of the data. Specifically, window lengths of 30–60 are adopted for daily runoff series, while lengths of 24–48 are employed for hourly precipitation series, ensuring that short-term fluctuations and medium-term trends of hydrological processes are adequately captured. During the preprocessing stage, Z-score normalization is applied to eliminate dimensional effects and enhance the stability of the subsequent time–frequency transformation. The normalization is defined as:

$$\hat{x}=\frac{x-\mu}{\sigma} \tag{1}$$

where, μ and σ denote the mean and standard deviation of the hydrological time series within the sliding window, respectively. The normalized sequence is more suitable for the parameter learning process of the adaptive time–frequency transform and provides a stable foundation for high-quality time–frequency image generation.

The core innovation of this module lies in the introduction of a learnable adaptive time–frequency transformation layer. Unlike conventional time–frequency transforms that rely on fixed window functions, the proposed approach enables dynamic learning of window parameters through a lightweight subnetwork, thereby achieving precise adaptation to the local characteristics of hydrological time series in the karst basins of Guizhou Province. The lightweight subnetwork consists of two fully connected layers with rectified linear unit activation functions. The first fully connected layer maps the input dimension T to 64 units, while the second layer outputs three parameters corresponding to the window length, window width, and attenuation coefficient. To guide parameter adjustment, local statistical characteristics of the input sequence are first computed, with particular emphasis placed on instantaneous variance and spectral entropy as key descriptors. The instantaneous variance is calculated as:

$$\sigma_t^2=\frac{1}{K}\sum_{i=t-K/2}^{t+K/2}(x_i-\mu_t)^2 \tag{2}$$

where, K denotes the size of the local window and μ_t represents the mean value within the local window. The spectral entropy is computed as:

$$H=-\sum_{f=1}^{F}p(f)\log p(f) \tag{3}$$

where, $p(f)$ denotes the probability distribution of frequency components. Based on the extracted local statistical characteristics, the lightweight subnetwork dynamically outputs the optimal window function parameters, which are then applied to adaptive short-time Fourier transform or continuous wavelet transform operations. The mapping relationship between the window parameters and the transformation output is expressed as $I_{TF}(f,t)=TF(X_{1D},\theta)$, where $\theta=[len,wid,dec]$ represents the window parameters output by the subnetwork and $TF(\ )$ denotes the time–frequency transformation operator. The resulting output is a high-resolution time–frequency energy spectrum $I_{TF} \in R^{F\times T}$, where F and T correspond to the frequency and time dimensions, respectively. The correspondence between frequency and time dimensions is dynamically adjusted through the window parameters, ensuring that the time–frequency images accurately characterize instantaneous frequency variations and energy distribution patterns of hydrological time series.

To further construct multiscale time–frequency representations and comprehensively capture the dynamic characteristics of hydrological time series across different temporal and frequency scales in the karst basins of Guizhou Province, a pyramid strategy is employed to generate a multiscale set of time–frequency images. By adjusting the resolution-related parameters of the adaptive time–frequency transform—primarily including window length, window width, and overlap ratio—S groups of time–frequency images with distinct resolutions are generated, denoted as $\{I_{TF}^s\}_{s=1}^S$. The value of S is determined through experimental optimization and is set to 3–5, enabling different scales to focus on complementary aspects of the hydrological signal. At higher-frequency scales, shorter window lengths and higher overlap ratios are adopted to emphasize short-term hydrological fluctuations, such as intense precipitation events and abrupt runoff responses commonly observed in karst basins of Guizhou Province. Conversely, at lower-frequency scales, longer window lengths and lower overlap ratios are employed to capture long-term trends, including seasonal variations and interannual cycles in hydrological time series. The resulting multiscale time–frequency image set is jointly used as the input to the subsequent CFT-Net. Compared with single-resolution time–frequency representations, the multiscale strategy provides more comprehensive time–frequency information for deep feature extraction, thereby effectively enhancing the completeness and task relevance of subsequent feature learning. The overall architecture of the learnable parameterized time–frequency image generation module is illustrated in Figure 1.

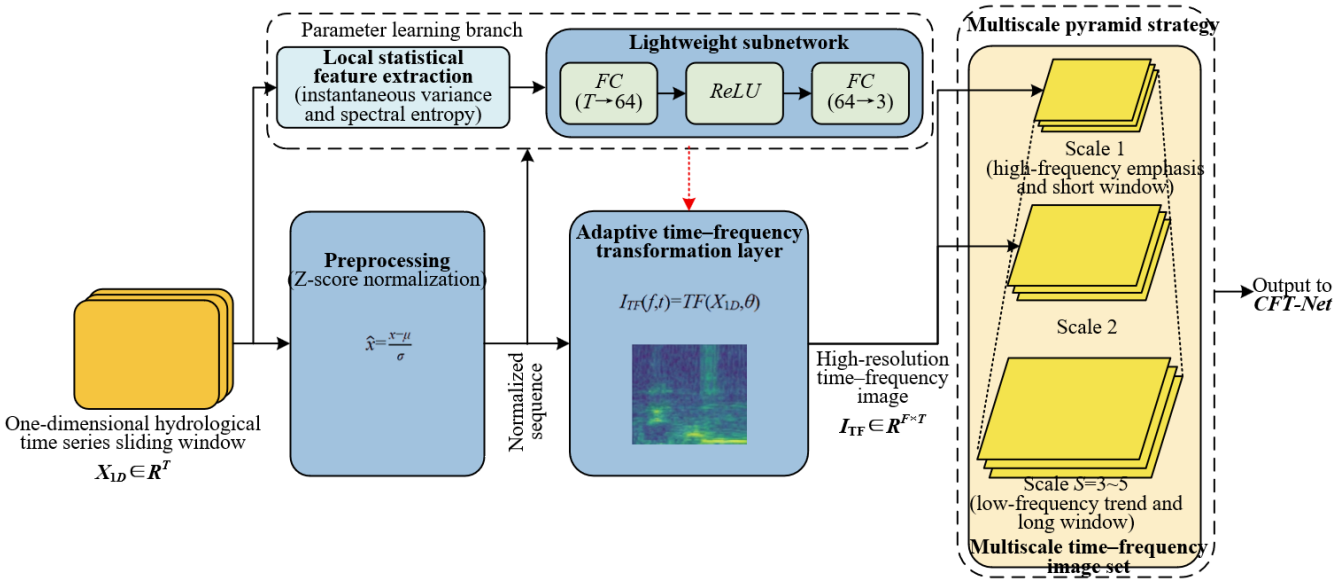


Figure 1. Architecture of the learnable parameterized time–frequency image generation module

The innovative value of this module lies in the realization of adaptive and multiscale time–frequency image generation. Through the design of learnable window functions, the limitations of conventional fixed time–frequency transforms in adapting to the strong nonlinearity and pronounced variability of hydrological time series in the karst basins of Guizhou Province are effectively addressed. The dynamically generated time–frequency images enable precise encoding of the dynamical characteristics of one-dimensional hydrological sequences into spatial patterns within two-dimensional images. Furthermore, the multiscale pyramid strategy compensates for the inherent limitations of single-resolution representations by simultaneously capturing hydrological features across multiple scales. As a result, high-quality visual inputs are provided for the subsequent collaborative learning of spatial–frequency–temporal features, thereby fully exploiting the advantages of image processing techniques for hydrological time series representation.

2.3 Deep convolutional backbone network coupling spatial–frequency–temporal features

CFT-Net serves as the core feature extraction backbone of the overall prediction framework and is responsible for mining deep, fused spatial–frequency–temporal features from multiscale time–frequency image sequences.

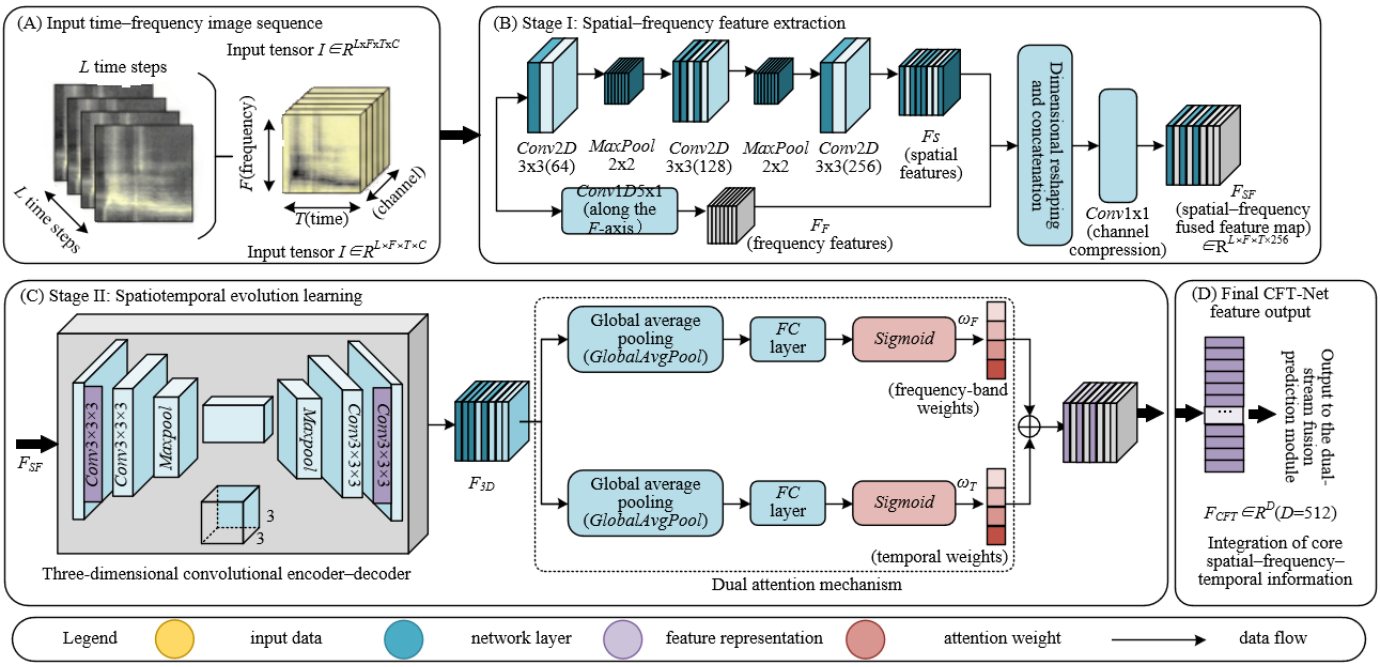


Figure 2. Architecture of the deep convolutional backbone network coupling spatial–frequency–temporal features (CFT-Net)

The input to CFT-Net is a four-dimensional tensor formed by stacking time–frequency images over L consecutive time steps, denoted as $I \in R^{L\times F\times T\times C}$, where L represents the temporal length and is set to 5–8 in accordance with the temporal fluctuation characteristics of hydrological time series in the karst basins of Guizhou Province, ensuring that short-term temporal evolution patterns of time–frequency representations are effectively captured. The dimensions F and T correspond

to the frequency and time axes, respectively, and are consistent with the output dimensions of the preceding time–frequency image generation module. The channel dimension C is set to 1, as the generated time–frequency images are grayscale representations. The construction of the input tensor strictly adheres to the temporal continuity of hydrological time series. Time–frequency images from consecutive time steps are stacked in chronological order to form a video-like temporal sequence of time–frequency images. This structured input representation facilitates the modeling of spatiotemporal evolution features while remaining compatible with deep convolutional feature extraction paradigms commonly adopted in image processing. Through this design, visual patterns embedded in time–frequency images and the underlying temporal dynamics of hydrological sequences are captured in a synchronized manner. The detailed architecture of CFT-Net is illustrated in Figure 2.

CFT-Net adopts a stage-wise encoding design. The first stage corresponds to the spatial–frequency feature extraction module, in which a dual-path parallel architecture is innovatively employed to enable the coordinated capture of features along the spatial and frequency dimensions, thereby overcoming the limitation of conventional deep convolutional networks that primarily focus on spatial information. The spatial path is composed of three two-dimensional convolutional layers interleaved with two max-pooling layers. All convolutional kernels are set to a size of 3 × 3, with a stride of 1 and same padding to preserve feature map resolution. Rectified linear unit activation functions are applied throughout to introduce nonlinear feature transformations. The numbers of output channels for the three convolutional layers are set to 64, 128, and 256, respectively, allowing the representational capacity to be progressively enhanced through channel expansion. Max-pooling layers employ 2 × 2 pooling kernels with a stride of 2, which reduce feature map dimensionality, decrease computational complexity, and preserve salient spatial features. This design facilitates effective extraction of the spatial distribution and texture patterns of energy-concentrated regions in time–frequency images, which are closely associated with localized energy fluctuations in hydrological signals from the karst basins of Guizhou Province. In parallel, the frequency path is specifically designed along the frequency axis (F) of the time–frequency images and consists of a one-dimensional convolutional layer. The convolutional kernel size is set to 5 × 1 with a stride of 1, and rectified linear unit activation is employed. This path is dedicated to capturing the global spectral profile along the frequency dimension while avoiding distortion of frequency information that may arise from spatial convolution operations. To achieve effective fusion of features extracted from the dual paths, the output features of the frequency path are reshaped to match the dimensionality of the spatial path outputs. Channel-wise concatenation is then performed, followed by a 1 × 1 convolution for channel compression and redundancy reduction. The fusion process is formulated as $F_{SF}=Conv_{1\times1}(Concat(F_S,F_F))$, where F_S and F_F denote the output features of the spatial and frequency paths, respectively, and $F_{SF}\in R^{L\times F\times T\times 256}$ represents the resulting spatial–frequency fused feature map. Through this mechanism, local spatial patterns and global frequency distribution characteristics are jointly represented.

The second stage corresponds to the spatiotemporal evolution learning module. In this stage, the spatial–frequency fused feature sequence is treated as a "time–frequency feature video," and the dynamic evolution of time–frequency patterns over time is modeled through the integration of a three-dimensional convolutional encoder–decoder and a dual attention mechanism. This design is well suited to capturing hydrological phenomena in the karst basins of Guizhou Province, such as the migration of energy centers during flood seasons and abrupt changes in time–frequency patterns induced by short-term intense precipitation events. The three-dimensional convolutional encoder–decoder consists of four three-dimensional convolutional blocks organized in a symmetric encoding–decoding structure. During the encoding stage, two three-dimensional convolutional blocks followed by max-pooling layers are used to downsample features, whereas during the decoding stage, two three-dimensional convolutional blocks combined with upsampling layers are employed to restore feature resolution, ensuring consistency between input and output dimensions. All three-dimensional convolutional blocks utilize kernels of size 3 × 3 × 3, a stride of 1, same padding, and rectified linear unit activation functions. The three-dimensional convolution operation can be expressed as:

$$F_{3D}(l,f,t)=\sum_{i=0}^{2}\sum_{j=0}^{2}\sum_{k=0}^{2} w(i,j,k)\cdot F_{SF}(l+i,f+j,t+k)+b \tag{4}$$

where, w denotes the weights of the three-dimensional convolutional kernel, and b represents the bias term. The three-dimensional convolutional kernel is allowed to slide synchronously along the L, F, and T dimensions, enabling direct capture of the dynamic propagation and evolution of time–frequency patterns over time. In this manner, differences in time–frequency characteristics and spatiotemporal evolution between flood and dry seasons in hydrological time series from the karst basins of Guizhou Province are accurately characterized.

The dual attention mechanism constitutes the core innovation of the spatiotemporal evolution learning module and consists of a frequency-band importance attention layer and a temporal attention layer. Through their coordinated operation, precise focusing on task-relevant features is achieved. The frequency-band importance attention layer is designed to emphasize physical frequency bands that are most relevant to hydrological prediction objectives. A lightweight subnetwork composed of a fully connected layer and a Sigmoid activation function is employed to learn a frequency-domain weight vector $\omega_F\in R^F$. The weight generation process is defined as $\omega_F=Sigmoid(FC(GlobalAvgPool(F_{3D})))$, where *GlobalAvgPool* denotes global average pooling and *FC* denotes a fully connected layer. Element-wise multiplication between the weight vector and the three-dimensional convolutional feature map is then performed to enhance informative frequency-band features while suppressing interference from irrelevant bands. This mechanism is particularly well suited to the seasonal variability of dominant frequency bands in hydrological time series from karst basins in Guizhou Province. The temporal attention layer is designed to focus on key historical time steps that are most informative for future prediction. A lightweight subnetwork with the same structural design as the frequency-band attention layer is employed to learn a temporal weight vector $\omega_T\in R^L$, which is computed as $\omega_T=Sigmoid(FC(GlobalAvgPool(F_{3D})))$. Through element-wise multiplication between the temporal weight vector and the feature map, higher importance is

assigned to critical temporal windows, such as flood periods and short-term intense precipitation events, thereby improving the task relevance of feature learning. Through the synergistic action of the two attention layers, key features along both the frequency and temporal dimensions are simultaneously emphasized, allowing the network to adaptively capture the dynamic time–frequency characteristics of hydrological time series in the karst basins of Guizhou Province.

After sequential processing through the spatial–frequency feature extraction stage and the spatiotemporal evolution learning stage, CFT-Net outputs a deep spatial–frequency–temporal fused feature vector $F_{CFT} \in R^D$, where D denotes the feature dimensionality and is set to 512 based on experimental optimization. This feature vector integrates three core sources of information: local spatial patterns, global frequency distributions, and temporal evolution patterns embedded in time–frequency images. As a result, the strong representational capability of image processing techniques for visual features is preserved, while the nonlinear and nonstationary dynamical characteristics of hydrological time series in the karst basins of Guizhou Province are accurately encoded. The limitations associated with single-dimension feature representations are thereby effectively mitigated, and high-quality core features are provided for the subsequent dual-stream fusion prediction module, enabling precise alignment between feature extraction and hydrological prediction requirements.

2.4 Dual-stream fusion prediction module

The core innovation of the dual-stream fusion prediction module lies in the construction of a dual-source complementary architecture, in which visual features from the time–frequency image stream and temporal features from the raw sequence stream are jointly integrated. Through this design, the inherent limitations of relying on a single feature source for characterizing the complex hydrological behavior of karst basins in Guizhou Province are effectively addressed, leading to substantial improvements in model robustness and generalization capability and enabling deep cross-domain integration between image processing and time series modeling. The overall architecture is illustrated in Figure 3. In the image-stream branch, CFT-Net is employed as a dedicated feature extractor. The input consists of the multiscale time–frequency image pyramid generated by the preceding modules. Through coupled spatial–frequency–temporal feature extraction within CFT-Net, a deep fused feature vector $F_{img} \in R^D$ is produced, where the feature dimensionality D is set to 512 based on experimental optimization and representational requirements, ensuring sufficient expressive capacity while avoiding unnecessary model redundancy. In parallel, the sequence stream branch directly processes the original one-dimensional hydrological time series X_{1D}. A one-dimensional dilated causal convolutional network is adopted to capture long-term temporal dependencies in the raw sequence. The dilation rates are configured using an exponentially increasing strategy with values of 1, 2, 4, and 8, which effectively expands the receptive field to model long-period hydrological trends while maintaining moderate network complexity. The sequence stream branch outputs a temporal feature vector $F_{seq} \in R^D$, whose dimensionality is deliberately aligned with that of the image stream features to facilitate subsequent fusion.

The adaptive gated fusion unit constitutes the key innovative component of this module. Unlike conventional fusion strategies based on fixed weighting, a learnable gating network is introduced to dynamically generate fusion weights according to the characteristics of the current hydrological input, thereby enabling adaptive complementary integration of dual stream features. This mechanism is particularly well suited to the pronounced variability and heterogeneity of hydrological time series in the karst basins of Guizhou Province.

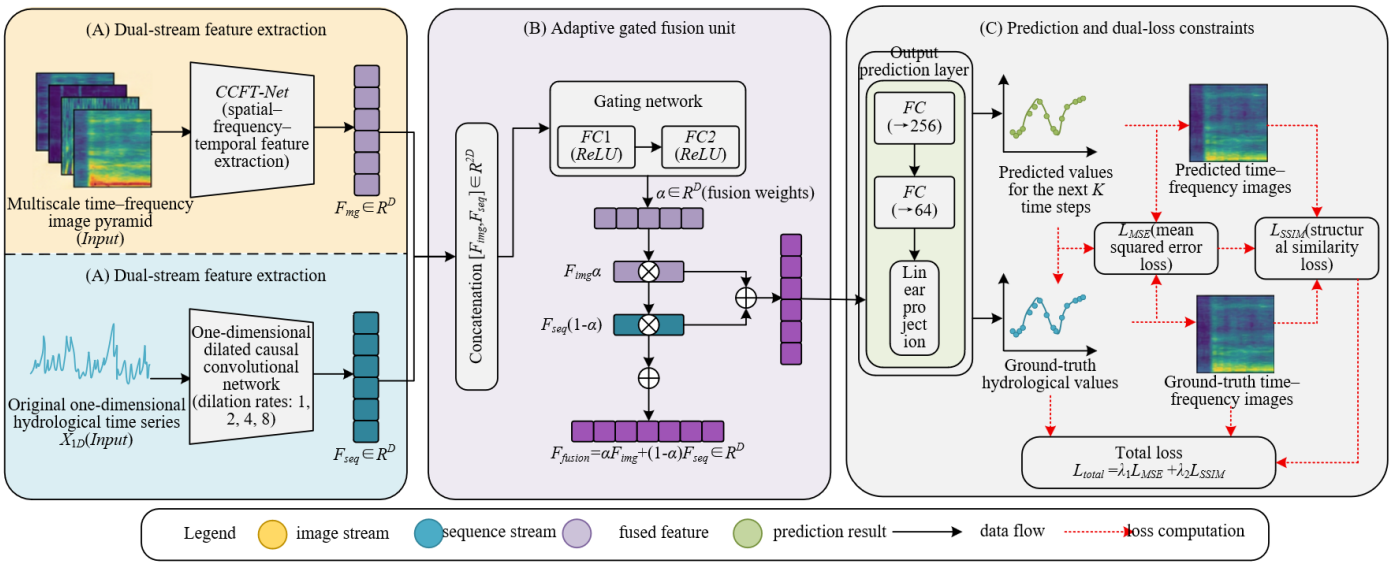


Figure 3. Architecture of the dual-stream fusion prediction module

The gating network consists of two fully connected layers followed by a Sigmoid activation function. The input to the gating network is the concatenated feature vector $[F_{img}, F_{seq}] \in R^{2D}$. The first fully connected layer maps the input from $2D$ to D dimensions and employs a rectified linear unit activation function to introduce nonlinearity, while the second fully connected layer preserves the dimensionality and applies a Sigmoid activation function to produce a fusion weight vector $\alpha \in R^D$, with values constrained to the interval [0, 1]. The weight generation process is expressed as

$\alpha=Sigmoid(FC_2(ReLU(FC_1([F_{img},F_{seq}]))))$, where FC_1 and FC_2 denote the two fully connected layers. Adaptive fusion is implemented through element-wise weighted summation, defined as $F_{fusion}=\alpha\cdot F_{img}+(1-\alpha)\cdot F_{seq}$. When hydrological fluctuations are pronounced, higher weights are automatically assigned to the sequence stream features to emphasize temporal dynamics; conversely, when the sequence exhibits relatively stable behavior, the contribution of image-stream features is increased to highlight time–frequency structural information. Through this dynamic complementary mechanism, the fused feature vector $F_{fusion}\in R^D$ integrates the strengths of both feature sources, providing comprehensive feature support for high-precision prediction.

The output prediction head and the dual-loss constraint design together constitute the core of prediction and training in this module, ensuring high numerical accuracy while preserving time–frequency structural fidelity. This design aligns with the characteristics of image-processing–oriented journals and the practical requirements of hydrological forecasting. The output prediction head is composed of two fully connected layers followed by a linear projection layer. The first fully connected layer maps the fused feature dimension from D to 256, the second maps it to 64, and the final linear projection layer transforms the resulting feature vector into predicted values for the next K time steps, denoted as $\hat{Y}\in R^K$. In accordance with practical hydrological forecasting demands in the karst basins of Guizhou Province, K is set to 3–7 depending on the prediction scenario. A linear activation function is adopted in the output layer to avoid prediction bias introduced by nonlinear transformations and to ensure numerical prediction accuracy. To jointly constrain numerical accuracy and time–frequency structural fidelity, the loss function is formulated as a weighted combination of mean squared error and structural similarity loss. The mean squared error loss is used to quantify numerical discrepancies between predicted and ground-truth values and is defined as:

$$L_{MSE}=\frac{1}{K}\sum_{k=1}^{K}(y_k-\hat{y}_k)^2 \tag{5}$$

where, y_k denotes the ground-truth hydrological value and $\hat{y}_k$ denotes the predicted value. The structural similarity loss is computed based on time–frequency images generated from the ground-truth and predicted sequences, serving to constrain the learning of time–frequency dynamic structures. It is defined as $L_{SSIM}=1-SSIM(I_{TF}^{true},I_{TF}^{pred})$. The total loss function is expressed as $L_{total}=\lambda_1\cdot L_{MSE}+\lambda_2\cdot L_{SSIM}$, where the weighting coefficients λ_1 and λ_2 are determined via five-fold cross-validation and are optimally set to 0.7 and 0.3, respectively. This configuration effectively prevents the model from focusing exclusively on numerical accuracy while neglecting time–frequency structural learning, thereby ensuring robust learning of the dynamic time–frequency characteristics inherent in hydrological time series from the karst basins of Guizhou Province.

The innovative value of this module lies in the organic integration of adaptive dual-source feature fusion and time–frequency structural constraints. Through the dual-stream architecture, the representational limitations associated with a single feature source are effectively mitigated. The adaptive gated unit enables feature fusion to dynamically adjust to variations in hydrological sequence characteristics, while the dual-loss constraint simultaneously enforces numerical precision and time–frequency structural fidelity. In this manner, the strong representational capacity of image processing techniques for time–frequency features is fully exploited, while the intrinsic temporal dynamics of the original hydrological sequences are preserved. As a result, model robustness and generalization capability across diverse hydrological scenarios in the karst basins of Guizhou Province are substantially enhanced, providing a critical foundation for the high-precision output of the overall prediction framework.

2.5 Model training strategy

The proposed hydrological time series prediction framework is trained in an end-to-end manner, enabling coordinated optimization of all parameters across the entire pipeline, from time–frequency image generation and deep feature extraction to prediction output. All training hyperparameters are determined through five-fold cross-validation in conjunction with the characteristics of hydrological datasets from the karst basins of Guizhou Province, thereby ensuring training stability and effectiveness. During training, the AdamW optimizer is employed, as it effectively alleviates issues related to gradient vanishing and weight decay and is well suited for deep network optimization. The initial learning rate is set to $1e$-4, and a cosine annealing strategy is adopted for learning rate decay to improve convergence speed and overall model performance. The batch size is set to 32 based on hardware constraints and data scale considerations, while the number of training epochs is fixed at 150 to achieve a balance between training efficiency and convergence quality.

To mitigate overfitting and enhance generalization capability across diverse hydrological scenarios in the karst basins of Guizhou Province, a dual regularization strategy combining Dropout and L2 regularization is applied. The Dropout probability is set to 0.3, enabling random deactivation of network units to reduce feature redundancy, while the L2 regularization weight decay coefficient is set to $1e$-5 to constrain model parameter magnitude. In addition, an early stopping strategy is introduced, with validation loss used as the monitoring criterion. Training is automatically terminated and the current optimal model parameters are saved when no decrease in validation loss is observed over 20 consecutive epochs, effectively preventing degradation in generalization performance caused by overtraining. The entire training process is driven by minimization of the total loss function. All learnable network parameters are updated via backpropagation, ensuring that the model fully captures the dynamic time–frequency characteristics and temporal evolution patterns inherent in hydrological time series from the karst basins of Guizhou Province. Through this training strategy, high-accuracy and high-robustness hydrological time series prediction is ultimately achieved.

3. EXPERIMENTAL VALIDATION

3.1 Experimental datasets and preprocessing

Experimental validation was conducted using two representative hydrological datasets from karst basins in Guizhou Province to ensure alignment with the intended application scenarios of the proposed framework and to rigorously assess its generalization capability. Specifically, a

daily runoff time series from the Wujiang River Basin and an hourly precipitation time series from a tributary of the Chishui River were employed. The daily runoff dataset from the Wujiang River Basin spans the period from January 2010 to December 2020, with a sampling frequency of one day, yielding a total of 3,948 observations. The mean runoff is 128.6 m^3/s, with a variance of 3,862.3, a maximum value of 1,892.5 m^3/s, and a minimum value of 15.8 m^3/s. The hourly precipitation dataset from the Chishui River tributary covers the period from January 2015 to December 2020, with a sampling frequency of one hour and a total of 52,584 observations. This dataset exhibits a mean precipitation of 2.3 mm, a variance of 42.8, a maximum value of 89.7 mm, and a minimum value of 0 mm. Data preprocessing strictly followed standardized procedures. Missing values were filled using linear interpolation, while outliers were identified and removed according to the three-sigma (3σ) criterion. Subsequently, Z-score normalization was applied to eliminate dimensional effects and improve numerical stability. A sliding window strategy was then adopted to construct samples, with the dataset split into training, validation, and testing subsets at a ratio of 7:2:1. The sliding window length was kept consistent with the settings used in the preceding methodological modules. Parameters related to time–frequency image generation were fixed at S=3 and L=5 to ensure consistency between the experimental setup and the proposed framework configuration. Through this preprocessing pipeline, a reliable data foundation was established for subsequent experimental evaluation.

3.2 Experimental setup

The experimental setup was designed with the primary objective of rigorously validating the effectiveness of the proposed framework's methodological innovations. To ensure fairness and rigor in performance comparison, all baseline models and the proposed framework were trained using identical data preprocessing procedures, training configurations, and evaluation metrics. The comparison models were categorized into three groups to comprehensively assess the performance advantages of the proposed approach. First, traditional time series prediction models, including the Autoregressive Integrated Moving Average and Support Vector Machine, were selected to compare the performance differences between conventional statistical or machine learning methods and deep learning–based approaches. Second, deep learning–based time series prediction models, including the Convolutional Neural Network–Long Short-Term Memory, Transformer, and Temporal Convolutional Network, were employed to evaluate the effectiveness of one-dimensional temporal modeling relative to the proposed spatial–frequency–temporal coupled modeling strategy. Third, time–frequency image–based prediction models were constructed by combining fixed Short-Time Fourier Transform representations with Residual Neural Network and fixed Continuous Wavelet Transform representations with Visual Geometry Group Network, respectively. These models were used to compare fixed versus adaptive time–frequency transformations, as well as the conventional Convolutional Neural Network versus the proposed CFT-Net architecture. To further examine the necessity and contribution of each key innovation, three ablation models were designed. Ablation Model 1 removed the adaptive time–frequency transformation and replaced it with a fixed Short-Time Fourier Transform. Ablation Model 2 removed the dual attention mechanism within CFT-Net. Ablation Model 3 removed the adaptive gated fusion mechanism and instead employed simple channel concatenation. By comparing the performance of these ablation models with that of the complete framework, the individual contributions of each proposed module were clearly identified.

3.3 Experimental results and analysis

Experimental results were analyzed through a combination of quantitative and qualitative evaluations, with particular emphasis placed on validating the overall performance of the proposed framework, the necessity of each methodological innovation, as well as robustness and generalization capability. All experiments were conducted on two hydrological datasets from karst basins in Guizhou Province, namely the daily runoff series of the Wujiang River Basin and the hourly precipitation series of a tributary of the Chishui River. Evaluation metrics included root mean square error, mean absolute error, coefficient of determination (R^2), structural similarity, and peak signal-to-noise ratio, enabling comprehensive assessment of both numerical prediction accuracy and time–frequency structural learning capability. Detailed quantitative results for all comparison models across the two datasets are summarized in Table 1.

Table 1. Overall performance comparison of different models on datasets from karst basins in Guizhou Province

Model		Autoregressive Integrated Moving Average	Support Vector Machine	Convolutional Neural Network–Long Short-Term Memory	Transformer
Wujiang River Basin (daily runoff)	Root mean square error (m^3/s)	48.62	42.35	35.78	33.42
	Mean absolute error (m^3/s)	32.57	28.91	24.16	22.58
	Coefficient of determination (R^2)	0.78	0.82	0.86	0.88
	Structural similarity	0.61	0.65	0.70	0.73
	Peak signal-to-noise ratio (dB)	22.35	23.18	24.52	25.16
Chishui River tributary (hourly precipitation)	Root mean square error (mm)	3.89	3.42	2.87	2.63

	Mean absolute error (mm)	2.17	1.92	1.65	1.51
	R^2	0.72	0.76	0.81	0.83
	Structural similarity	0.58	0.62	0.67	0.70
	Peak signal-to-noise ratio (dB)	21.87	22.69	23.85	24.42

Model		Temporal Convolutional Network	Short-Time Fourier Transform+Residual Neural Network	Continuous Wavelet Transform+Visual Geometry Group Network	Proposed Framework
Wujiang River Basin (daily runoff)	Root mean square error (m^3/s)	34.61	29.85	28.61	22.35
	Mean absolute error (m^3/s)	23.35	19.87	18.92	14.68
	Coefficient of determination (R^2)	0.87	0.90	0.91	0.95
	Structural similarity	0.71	0.78	0.80	0.91
	Peak signal-to-noise ratio (dB)	24.83	26.35	26.89	29.42
Chishui River tributary (hourly precipitation)	Root mean square error (mm)	2.75	2.31	2.18	1.56
	Mean absolute error (mm)	1.58	1.32	1.25	0.89
	R^2	0.82	0.86	0.87	0.93
	Structural similarity	0.68	0.75	0.77	0.89
	Peak signal-to-noise ratio (dB)	24.11	25.68	26.12	28.75

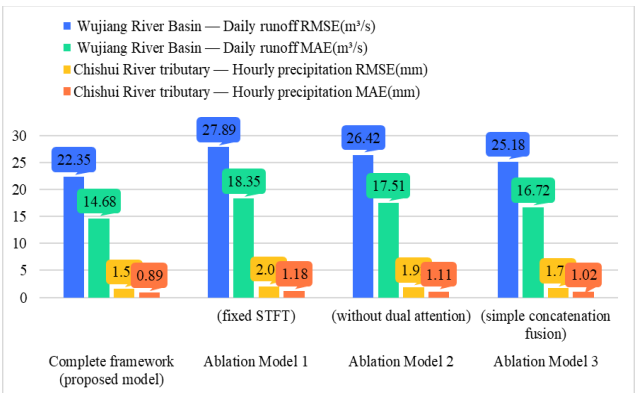
The overall performance comparison experiments provide compelling evidence of the superiority of the proposed framework. Quantitative analysis based on the results reported in Table 1 indicates that all evaluation metrics achieved by the proposed model on both hydrological datasets from karst basins in Guizhou Province are markedly superior to those of all comparison models, with statistically meaningful advantages. For daily runoff prediction in the Wujiang River Basin, the root mean square error and mean absolute error are reduced to 22.35 m^3/s and 14.68 m^3/s, respectively, while R^2 is increased to 0.95. Compared with the traditional time series model Support Vector Machine, root mean square error and mean absolute error are reduced by 47.18% and 49.22%, respectively, and R^2 is improved by 15.85%. Relative to the best-performing deep learning–based time series model, namely the Transformer, root mean square error and mean absolute error are reduced by 33.12% and 35.00%, respectively, with a corresponding increase of 7.95% in R^2. These results clearly demonstrate the advantage of spatial–frequency–temporal coupled modeling over single-domain temporal modeling. With respect to time–frequency structural metrics, the proposed framework achieves structural similarity and peak signal-to-noise ratio values of 0.91 and 29.42 dB, respectively, which are substantially higher than those obtained by the fixed time–frequency transformation–based Short-Time Fourier Transform+Residual Neural Network model. Specifically, structural similarity and peak signal-to-noise ratio are improved by 16.67% and 11.65%, respectively, indicating that the adaptive time–frequency image generation module in conjunction with CFT-Net is capable of capturing time–frequency structural characteristics of hydrological sequences with substantially higher fidelity. These findings are well aligned with the emphasis placed on representational completeness in image-processing–oriented journals.

A consistent pattern is observed for hourly precipitation prediction in the Chishui River tributary. The proposed framework achieves root mean square error and mean absolute error values of 1.56 mm and 0.89 mm, respectively, with an R^2 of 0.93, while structural similarity and peak signal-to-noise ratio reach 0.89 and 28.75 dB. All metrics attain their optimal levels among the evaluated models, demonstrating strong capability in capturing the instantaneous fluctuations associated with short-term intense precipitation events. From a qualitative perspective, the predicted hydrological curves exhibit a high degree of agreement with the corresponding ground-truth sequences, enabling accurate reconstruction of abrupt changes in runoff and precipitation characteristic of karst basins in Guizhou Province. Furthermore, visual comparison between ground-truth and predicted time–frequency images reveals that the proposed framework is able to faithfully reproduce the energy distribution and evolutionary patterns observed in true time–frequency representations. In particular, transitions in time–frequency characteristics between flood and dry seasons are reconstructed with notably higher fidelity than those produced by fixed time–frequency transformation–based models. These qualitative observations further corroborate the core advantage of the proposed framework in learning complex time–frequency structures.

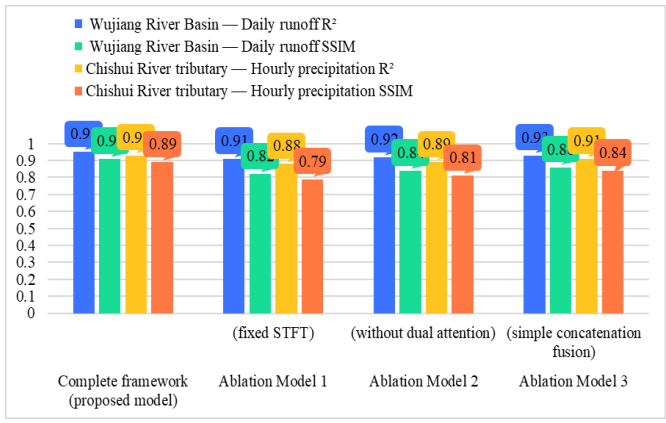
The ablation experiments, in conjunction with the results shown in Figure 4, clearly demonstrate the necessity and functional significance of each proposed innovation module. When any individual module is removed, model performance exhibits varying degrees of degradation, and the magnitude of these declines can be interpreted in a physically meaningful manner. In Ablation Model 1, the adaptive time–frequency

transform is replaced with a fixed Short-Time Fourier Transform. Under this setting, root mean square error and mean absolute error in the Wujiang River Basin increase by 24.8% and 25.0%, respectively, while structural similarity and peak signal-to-noise ratio decrease by 9.89% and 7.71%. A similar degradation trend is observed for the Chishui River dataset. These results indicate that the adaptive time–frequency transform, through dynamic adjustment of window parameters, is able to better accommodate the strong nonlinearity and pronounced variability of hydrological time series in karst basins of Guizhou Province. As a result, higher-quality time–frequency images and improved time–frequency feature adaptability are achieved, effectively overcoming the limited adaptability of fixed time–frequency transformations. In Ablation Model 2, the dual attention mechanism within CFT-Net is removed. In this case, root mean square error and mean absolute error increase by 18.19% and 19.28%, respectively, while structural similarity decreases by 7.69%. These results demonstrate that the dual attention mechanism plays a critical role in selectively emphasizing key frequency bands and temporal stages that are most relevant to the prediction task, while suppressing interference from irrelevant features. This capability is particularly important for distinguishing between hydrological characteristics associated with flood and dry seasons in karst basins of Guizhou Province, thereby enhancing the task specificity of deep feature extraction.

In Ablation Model 3, adaptive gated fusion is replaced with simple feature concatenation. As a consequence, root mean square error and mean absolute error increase by 12.66% and 13.89%, respectively, and structural similarity decreases by 5.49%. These results confirm that adaptive gated fusion enables efficient complementary integration of dual-source features through dynamic weight allocation. Compared with fixed-weight concatenation, this mechanism is more capable of adapting to variations in hydrological sequence characteristics, thereby improving model adaptability under complex hydrological conditions. Qualitative analysis further supports these quantitative findings. Feature heat maps clearly indicate that the dual attention mechanism significantly amplifies feature responses in key frequency bands and critical temporal windows during flood periods. In addition, visual comparison of time–frequency images shows that adaptive time–frequency transformation generates richer and more detailed texture patterns, capturing localized fluctuations that are not identifiable by fixed time–frequency transformations. These observations further substantiate the technical value of each proposed module.



(a) Root mean square error/mean absolute error



(b) Coefficient of determination (R^2)/structural similarity

Figure 4. Performance comparison of ablation models on datasets from karst basins in Guizhou Province

The robustness and generalization evaluations, interpreted in conjunction with the results in Table 2, indicate that the proposed framework exhibits strong stability and practical applicability, and is capable of accommodating the randomness and complexity of hydrological data from karst basins in Guizhou Province. In the robustness tests, Gaussian noise and outlier disturbances of varying intensities were injected into the input data. Under all disturbance settings, the performance degradation observed for the proposed framework was markedly smaller than that of the Transformer and Short-Time Fourier Transform+Residual Neural Network baselines, with the performance gap widening as disturbance intensity increased. When Gaussian noise with an intensity of 0.10 was introduced, the root mean square error of the proposed framework in the Wujiang River Basin increased by only 8.32%, while R^2 decreased by merely 0.02. In contrast, root mean square error increases of 16.52% and 16.61% were observed for Short-Time Fourier Transform+Residual Neural Network and Transformer, respectively, accompanied by R^2 reductions of 0.04 and 0.05. When 10% outliers were added, the R^2 of the proposed framework decreased by only 0.01 and the mean absolute error increased by 7.49%, outperforming all comparison models. These results demonstrate effective resistance to random noise and anomalous values commonly present in hydrological data from karst regions, confirming strong adaptability to complex data conditions.

For generalization assessment, a cross-dataset evaluation was conducted using a daily runoff time series from the Beipan River Basin in Guizhou Province. In this setting, the proposed framework achieved root mean square error, mean absolute error, and R^2 values of 24.78 m³/s, 16.23 m³/s, and 0.93, respectively. Relative to performance on the Wujiang River Basin dataset, root mean square error increased by only 10.87% and R^2 decreased by just 0.02, indicating minimal performance degradation. By comparison, the R^2 values of the Transformer and Short-Time Fourier Transform+Residual Neural Network models dropped to 0.79 and 0.82, respectively, reflecting substantial loss of predictive capability under cross-basin conditions. These findings confirm that the proposed framework is capable of adapting to hydrological characteristics across different karst basins in Guizhou Province, thereby demonstrating strong generalization capability and providing reliable technical support for hydrological time series prediction in southwestern China.

Table 2. Robustness test results of different models under varying disturbance intensities (Wujiang River Basin, daily runoff)

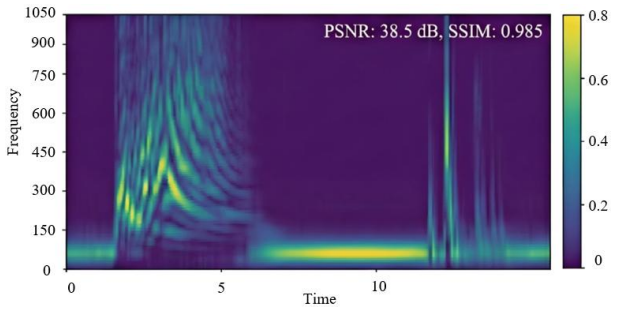
Disturbance Type	Intensity	Model	Root mean Square Error (m^3/s)	Mean Absolute Error (m^3/s)	Coefficient of Determination (R^2)	Structural Similarity
None	——	Proposed framework	22.35	14.68	0.95	0.91
		Transformer	33.42	22.58	0.88	0.73
		Short-Time Fourier Transform+Residual Neural Network	29.85	19.87	0.90	0.78
Gaussian noise	0.05	Proposed framework	23.78	15.72	0.94	0.88
		Transformer	36.89	24.92	0.85	0.69
		Short-Time Fourier Transform+Residual Neural Network	32.94	21.75	0.88	0.74
	0.10	Proposed framework	24.21	16.45	0.93	0.85
		Transformer	38.97	26.35	0.83	0.66
		Short-Time Fourier Transform+Residual Neural Network	34.78	23.12	0.86	0.71
	0.15	Proposed framework	25.87	17.89	0.92	0.82
		Transformer	41.23	28.17	0.81	0.63
		Short-Time Fourier Transform+Residual Neural Network	37.21	24.89	0.84	0.68
Outlier	5%	Proposed framework	22.98	15.12	0.94	0.89
		Transformer	34.76	23.89	0.87	0.71
		Short-Time Fourier Transform+Residual Neural Network	30.56	20.45	0.89	0.76
	10%	Proposed framework	23.56	15.78	0.94	0.87
		Transformer	36.52	25.17	0.85	0.68
		Short-Time Fourier Transform+Residual Neural Network	31.89	21.67	0.87	0.73
	15%	Proposed framework	24.89	16.92	0.93	0.84
		Transformer	38.75	26.98	0.83	0.65
		Short-Time Fourier Transform+Residual Neural Network	33.45	22.98	0.86	0.70

Table 3. Sensitivity analysis results of key innovation parameters (Wujiang River Basin, daily runoff)

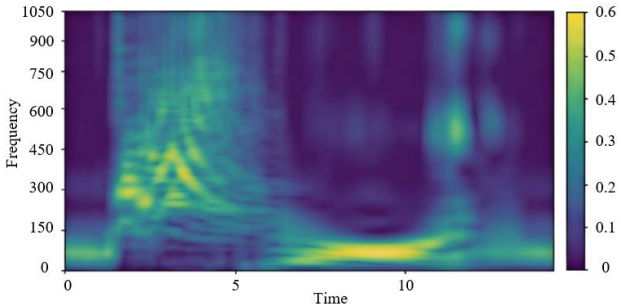
Parameter Type	Parameter Value	Root mean Square Error (m^3/s)	Mean Absolute Error (m^3/s)	Coefficient of Determination (R^2)	Structural Similarity
Time–frequency image scale S	1	28.97	19.23	0.90	0.81
	2	25.42	17.15	0.92	0.85
	3	22.35	14.68	0.95	0.91
	4	22.89	15.03	0.94	0.89
	5	23.76	15.87	0.93	0.87
Three-dimensional convolution kernel size of the deep convolutional backbone network coupling spatial–frequency–temporal features (CFT-Net)	1 × 1 × 1	27.51	18.02	0.91	0.83
	3×3×3	22.35	14.68	0.95	0.91
	5 × 5 × 5	24.12	16.05	0.93	0.86
Fusion weighting coefficient λ_1 (λ_2=1-λ_1)	0.5	24.87	16.32	0.93	0.88
	0.6	23.51	15.47	0.94	0.89
	0.7	22.35	14.68	0.95	0.91
	0.8	23.18	15.21	0.94	0.87
	0.9	24.56	16.09	0.93	0.84

Sensitivity analysis was conducted with respect to three key innovation parameters—namely the time–frequency image scale S, the three-dimensional convolution kernel size in CFT-Net, and the fusion weighting coefficients λ_1 and λ_2—to examine the impact of parameter variations on model performance and to verify the rationality and stability of the model design. The analysis was performed using the results summarized in Table 3. The analysis of the time–frequency image scale S indicates that optimal performance across all evaluation metrics is achieved when S is 3. When S<3, multiscale feature extraction is insufficient, preventing comprehensive capture of both high-frequency details and low-frequency trends in hydrological time series from karst basins in Guizhou Province, which results in higher root mean

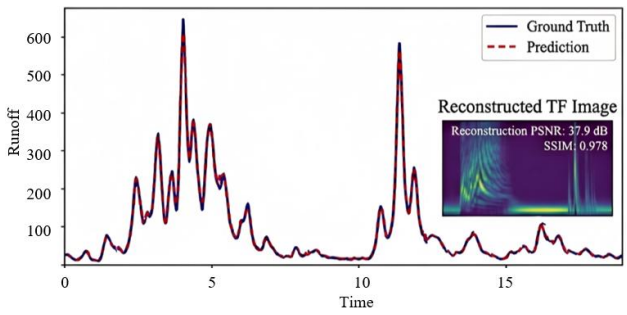
square error and lower structural similarity. When $S>3$, model redundancy increases and computational efficiency decreases, while feature redundancy yields no further performance improvement and may even lead to slight degradation. Accordingly, the optimal range of S is identified as 3–4, which is consistent with the experimentally optimized configuration adopted in the framework. Analysis of the three-dimensional convolution kernel size in CFT-Net demonstrates that the 3 × 3 × 3 kernel yields the best overall performance. When the kernel size is too small (1 × 1 × 1), spatiotemporal evolution features cannot be adequately captured, limiting the model's ability to characterize the dynamic propagation of hydrological time–frequency patterns. Conversely, excessively large kernels (5 × 5 × 5) introduce feature redundancy and substantially increase computational cost, while potentially obscuring local features. These observations confirm the rationality of the selected kernel size.



(a) Adaptive time–frequency image representation produced by the proposed approach



(b) Comparison result using conventional Short-Time Fourier Transform with a fixed window



(c) Hydrological prediction results with corresponding time–frequency reconstruction

Figure 5. Visualization of image-processing effects produced by the proposed method

The analysis of the fusion weighting coefficients λ_1 and λ_2 shows that the cross-validated optimal values of 0.7 and 0.3 achieve the best balance between numerical prediction accuracy and time–frequency structural fidelity. When λ_1 is excessively large, numerical accuracy is overemphasized at the expense of time–frequency structural preservation, leading to reduced structural similarity. Conversely, when λ_2 is overly dominant, numerical prediction accuracy deteriorates and root mean square error increases. These findings further substantiate the scientific soundness of the proposed dual-loss constraint design. Overall, the key parameter configurations of the proposed framework exhibit clear physical interpretability and well-founded optimization rationale. The model architecture is shown to be stable and reliable, enabling precise adaptation to the prediction requirements of hydrological time series in karst basins of Guizhou Province, while maintaining an effective balance between computational efficiency and predictive performance.

To provide an intuitive verification of the effectiveness and superiority of the proposed hydrological time series modeling approach based on time–frequency image representation—in terms of feature representation, image processing quality, and predictive performance—a four-panel visualization experiment was designed, incorporating raw hydrological time series, adaptive time–frequency processing results, conventional method comparisons, and prediction–reconstruction outcomes. The corresponding results are presented in Figure 5. In the adaptive time–frequency image subfigure, time–frequency details are rendered with high clarity and without spectral leakage. Concentrated high-frequency energy during flood periods, stable low-frequency distributions during dry periods, and instantaneous time–frequency characteristics associated with extreme peaks are accurately captured. The resulting peak signal-to-noise ratio and structural similarity values are consistently high, demonstrating that the adaptive time–frequency transform effectively accommodates local dynamic characteristics of hydrological sequences and generates high-quality time–frequency representations. In contrast, the subfigure produced using the conventional Short-Time Fourier Transform with a fixed window exhibits pronounced spectral smearing and blurred details, accompanied by distorted energy distributions. These deficiencies prevent accurate characterization of the dynamic time–frequency behavior of hydrological sequences, thereby highlighting the technical advantages of the proposed adaptive time–frequency module.

In the prediction and reconstruction subfigure, strong agreement is observed between predicted and ground-truth curves, with high accuracy achieved for extreme peaks and fluctuation turning points. Moreover, the reconstructed time–frequency images attain peak signal-to-noise ratio and structural similarity values close to those of the ground-truth time–frequency representations. These results jointly confirm not only the numerical prediction accuracy of the model but also its capability to preserve time–frequency structural fidelity. The findings further indicate that the dual-stream fusion strategy combined with the dual-loss constraint design enables effective joint optimization of numerical accuracy and time–frequency structural consistency, thereby providing comprehensive support for the core advantages of the proposed approach in hydrological time series modeling and prediction.

4. DISCUSSION

By integrating the experimental results with the design logic of the proposed framework, the intrinsic mechanisms

underlying the performance improvements associated with each innovation can be systematically elucidated. The adaptive time–frequency transform dynamically learns window function parameters through a lightweight subnetwork, enabling real-time adaptation to the strong nonlinearity and pronounced variability of hydrological time series in karst basins of Guizhou Province. By adjusting time–frequency resolution according to local statistical characteristics of the sequence, finer-grained local fluctuations can be captured compared with fixed time–frequency transforms. As a result, the generated time–frequency images exhibit enhanced representational capacity, providing a high-quality foundation for subsequent image processing and feature extraction. The parallel spatial–frequency dual-path architecture of CFT-Net enables coordinated extraction of local spatial patterns and global frequency distributions. The dual attention mechanism further concentrates model capacity on key frequency bands and critical temporal stages, effectively amplifying responses to core hydrological features such as flood periods and short-term intense precipitation events, while suppressing interference from irrelevant features. Through this mechanism, the specificity and completeness of deep feature extraction are substantially enhanced.

The adaptive gated fusion mechanism achieves efficient complementarity between features from the time–frequency image stream and the raw sequence stream via dynamic weight allocation. By adaptively adjusting the relative contributions of the two feature sources in response to changes in hydrological sequence characteristics, the limitations of single-source feature representation are effectively mitigated. The synergistic interaction among these components collectively drives significant improvements in both prediction accuracy and robustness. From both academic and practical perspectives, hydrological time series modeling is reformulated as a problem of spatiotemporal feature understanding in dynamic time–frequency images. Image processing and feature learning strategies tailored to hydrological data from karst basins are thereby introduced, enriching research at the intersection of time series prediction and image processing. Moreover, the end-to-end framework design facilitates engineering implementation and can be directly applied to real-world scenarios such as flood early warning and water resources regulation in karst regions of Guizhou Province. Consequently, strong practical applicability and considerable potential for broader deployment are demonstrated.

Although substantial performance gains are achieved by the proposed approach, several objective limitations remain, which is consistent with the rigor expected in academic research. During multi-scale time–frequency image generation, time–frequency transformations and feature extraction across multiple scales inevitably increase computational complexity. Compared with single-scale models, inference speed is therefore reduced. In addition, the three-dimensional convolution operations employed in CFT-Net impose nontrivial requirements on hardware resources, which may hinder deployment on resource-constrained embedded devices. Furthermore, in the prediction of extreme hydrological events in karst basins of Guizhou Province—such as catastrophic floods and extreme rainstorms—the associated time–frequency characteristics are highly abrupt and event-specific. Under such conditions, the ability to capture these anomalous patterns remains improvable, and a slight degradation in prediction accuracy is observed relative to routine hydrological scenarios. Consequently, the stringent accuracy requirements of extreme-event forecasting are not yet fully satisfied. Moreover, the current framework relies primarily on a single hydrological time series and does not fully integrate multi-source correlated information, such as meteorological, topographic, and anthropogenic factors, leaving room for further enhancement of representational completeness.

To address these limitations, several future research directions are identified in alignment with practical hydrological forecasting needs in karst regions and current academic trends. First, lightweight designs of the time–frequency image generation module and the CFT-Net architecture can be explored. Through techniques such as feature compression and lightweight convolution, computational complexity can be reduced and inference efficiency improved, thereby facilitating deployment on embedded platforms and enhancing engineering practicality. Second, more efficient gated attention mechanisms can be introduced to refine key feature extraction strategies, with particular emphasis placed on strengthening the representation of extreme hydrological events, so that predictive accuracy under catastrophic flood and extreme rainfall scenarios can be further improved. Third, the integration of multi-source data—including meteorological variables, terrain information, and indicators of human activity—can be pursued to enrich feature representation dimensions and further enhance adaptability and predictive accuracy for complex hydrological dynamics in karst basins of Guizhou Province. Finally, the proposed time–frequency image representation and spatial–frequency–temporal coupled modeling strategy can be extended to other time series prediction tasks, such as soil moisture forecasting and water quality prediction, thereby fully leveraging the generality of the approach and broadening its academic and application impact.

5. CONCLUSION

To address the pronounced nonlinearity, nonstationarity, and multiscale characteristics of hydrological time series in karst basins of Guizhou Province, as well as the limitations of existing prediction methods in time–frequency feature adaptability, deep feature extraction, and feature fusion effectiveness, an end-to-end hydrological time series prediction framework based on time–frequency image representation is presented. The core of the framework lies in the coordinated design of three innovation modules. An adaptive time–frequency image generation module dynamically learns window function parameters through a lightweight subnetwork, enabling precise alignment between time–frequency transformation and local characteristics of hydrological sequences and producing multiscale time–frequency images with strong representational capacity. CFT-Net adopts a parallel spatial–frequency dual-path architecture in combination with a dual attention mechanism, allowing coordinated extraction of spatial patterns, frequency distributions, and temporal evolution, thereby enhancing the completeness and task relevance of feature representation. An adaptive gated dual-stream fusion module achieves efficient complementarity between features from the time–frequency image stream and the raw sequence stream through dynamic weight allocation, effectively overcoming the limitations of

single-source feature representation. Through organic integration of these components, a complete closed loop for modeling and prediction is formed.

The present study enriches research at the intersection of time series prediction and image processing, and provides an effective new paradigm for hydrological time series modeling with significant academic value and engineering potential. From a theoretical perspective, hydrological sequence modeling is reformulated as a problem of spatiotemporal feature understanding in dynamic time–frequency images, and image processing and feature learning strategies tailored to karst hydrological data are introduced, thereby extending the application boundaries of image processing techniques in the field of hydrology and water resources. From a practical perspective, the end-to-end framework design facilitates real-world deployment and can be directly applied to scenarios such as flood early warning and optimal water resources allocation in karst basins of Guizhou Province, providing reliable technical support for hydrological prediction in southwestern China. Future research will focus on further optimization of model performance in response to limitations such as relatively high computational complexity and reduced accuracy under extreme hydrological events. Improvements will be pursued through lightweight network design, the introduction of more efficient attention mechanisms, and the integration of multi-source data. In addition, the proposed method will be extended to other time series prediction tasks, including soil moisture and water quality forecasting, thereby providing broader methodological references for precise modeling and prediction in hydrology and water resources research.

REFERENCES

[1] Zhang, Z.C., Chen, X., Shi, P., Ou, G.X. (2013). Study of canopy transpiration based on a distributed hydrology model in a small karst watershed of southwest China. Carbonates and Evaporites, 28(1): 111-117. https://doi.org/10.1007/s13146-013-0146-5

[2] Hu, K., Chen, G., Gregory-Eaves, I., Huang, L., Chen, X., Liu, Y., Leavitt, P. R. (2019). Hydrological fluctuations modulate phototrophic responses to nutrient fertilization in a large and shallow lake of Southwest China. Aquatic Sciences, 81(2): 37. https://doi.org/10.1007/s00027-019-0633-4

[3] Schaake, J.C., Hamill, T.M., Buizza, R., Clark, M. (2007). HEPEX: The hydrological ensemble prediction experiment. Bulletin of the American Meteorological Society, 88(10): 1541-1548. https://doi.org/10.1175/BAMS-88-10-1541

[4] Papacharalampous, G., Tyralis, H., Koutsoyiannis, D., Montanari, A. (2020). Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale. Advances in Water Resources, 136: 103470. https://doi.org/10.1016/j.advwatres.2019.103470

[5] Wu, X., Wang, Z.C. (2022). Multi-objective optimal allocation of regional water resources based on slime mould algorithm. Journal of Supercomputing, 78(16): 18288-18317. https://doi.org/10.1007/s11227-022-04599-w

[6] Seddiky, M.A., Giggins, H., Gajendran, T. (2020). International principles of disaster risk reduction informing NGOs strategies for community based DRR mainstreaming: The Bangladesh context. International Journal of Disaster Risk Reduction, 48: 101580. https://doi.org/10.1016/j.ijdrr.2020.101580

[7] Xu, J.F., Zhang, Y., Tang, X., Chen, X. (2006). Research on real-time signal processing technology of large view field infrared image detecting system. Journal of Infrared and Millimeter Waves, 25(6): 421-425.

[8] Lai, S., Xiong, Z., Chen, L., Tan, X., Zhang, M. (2014). Real-time distortion correction of fish-eye lens based on Bayer image signal. Optical Review, 21(2): 162-173. https://doi.org/10.1007/s10043-014-0025-x

[9] Anisimov, I.Y. (2005). Using signal processors for automatic image processing in real time. Journal of Optical Technology, 72(4): 327-329. https://doi.org/10.1364/JOT.72.000327

[10] Kadiyala, L.A., Mermer, O., Samuel, D.J., Sermet, Y., Demir, I. (2024). The implementation of multimodal large language models for hydrological applications: A comparative study of GPT-4 vision, gemini, LLaVa, and multimodal-GPT. Hydrology, 11(9): 148. https://doi.org/10.3390/hydrology11090148

[11] Fuhrmann, S. (2000). Designing a visualization system for hydrological data. Computers & Geosciences, 26(1): 11-19. https://doi.org/10.1016/S0098-3004(99)00040-0

[12] Stanković, S., Orović, I., Žarić, N., Ioana, C. (2010). Two dimensional time-frequency analysis based eigenvalue decomposition applied to image watermarking. Multimedia Tools and Applications, 49(3): 529-543. https://doi.org/10.1007/s11042-009-0446-x

[13] Han, B., Bao, B.K. (2021). River channel extraction in SAR images using level sets driven by symmetric Kullback–Leibler distance. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-16. https://doi.org/10.1109/TGRS.2021.3130684

[14] Sghaier, M.O., Foucher, S., Lepage, R. (2016). River extraction from high-resolution SAR images combining a structural feature set and mathematical morphology. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(3): 1025-1038. https://doi.org/10.1109/JSTARS.2016.2609804

[15] Tian, W., Wu, J., Cui, H., Hu, T. (2021). Drought prediction based on feature-based transfer learning and time series imaging. IEEE Access, 9: 101454-101468. https://doi.org/10.1109/ACCESS.2021.3097353

[16] Sadler, J.M., Koenig, L.E., Gorski, G., Carter, A.M., Hall Jr, R.O. (2024). Evaluating a process-guided deep learning approach for predicting dissolved oxygen in streams. Hydrological Processes, 38(9): e15270. https://doi.org/10.1002/hyp.15270

[17] Bevainis, L., Bielinis, M., Cesnulevicius, A., Bautrénas, A. (2023). Lithuanian river ice detection and automated classification using machine-learning methods. Baltica, 36(1): 1-12. https://doi.org/10.5200/baltica.2023.1.1

[18] Dai, K., Ma, C., Wang, Z., Long, Y., Li, X., Feng, S., Ye, Y. (2023). Exploiting spatial–temporal dynamics for satellite image sequence prediction. IEEE Geoscience and Remote Sensing Letters, 20: 1-5. https://doi.org/10.1109/LGRS.2023.3261317

[19] Potocnik, B., Zazula, D. (2002). Automated analysis of a sequence of ovarian ultrasound images. Part II: prediction-based object recognition from a sequence of images. Image and Vision Computing, 20(3): 227-235.

https://doi.org/10.1016/S0262-8856(01)00097-X
[20] Boashash, B., Boubchir, L., Azemi, G. (2012). A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals. EURASIP Journal on Advances in Signal Processing, 2012(1): 117. https://doi.org/10.1186/1687-6180-2012-117
[21] Ma, Y., Wang, C., Yang, D., Wang, C. (2021). Adaptive extraction method based on time-frequency images for fault diagnosis in rolling bearings of motor. Mathematical Problems in Engineering, 2021(1): 6687195. https://doi.org/10.1155/2021/6687195
[22] Jin, P., Yang, S., Xu, X., Li, C., et al. (2025). Multiview state-of-health estimation for lithium-ion batteries using time–frequency image fusion and attention-based deep learning. Plos One, 20(11): e0335351. https://doi.org/10.1371/journal.pone.0335351
[23] Li, K., Sun, Z., Jin, H., Xu, Y., et al. (2022). Proposal and experimental study on a diagnosis method for hermetic refrigeration compressor using dual time-frequency image fusion. Applied Sciences, 12(6): 3033. https://doi.org/10.3390/app12063033