Journal of
Renewable Energy Research and Applications (RERA)

Vol. 6, No. 2, 2025, 165-178 *DOI: 10.22044/rera.2025.15976.1410*

SP-Transformer: A Medium- and Long-Term Photovoltaic Power Forecasting Model Integrating Multi-Source Spatiotemporal Features

T. Ling*, W. Bin, C. Julong, Z. Yongqing, F. Junqiu, H. Jiang

Power Grid Planning and Research Center Guizhou Power Grid Co, Ltd, 38 Ruijin South Road, Nanming District, Guiyang, Guizhou, China.

Received Date 24 March 2025; Revised Date 11 May 2025; Accepted Date 22 September 2025
**Corresponding author: cillatan0@nuist.edu.cn (T. Ling)*

Abstract
Medium and long-term photovoltaic (PV) power forecasting is crucial for the planning and management of new energy grids. Existing methods often suffer from limited processing capabilities and low prediction efficiency. To address these challenges, this paper proposes a Transformer-based approach called SP-Transformer (Spatiotemporal-ProbSparse Transformer), designed to capture spatiotemporal correlations between meteorological, geographical, and PV power data. The model incorporates geographical location information through spatiotemporal position encoding and employs a spatiotemporal probability sparse self-attention mechanism to enhance correlation capture while reducing complexity. Additionally, a feature pyramid-based self-attention distillation module is introduced to improve the model's ability to generalize complex patterns in medium and long-term forecasting. Experimental results demonstrate that SP-Transformer achieves 93.8% accuracy for forecasting PV power over the next 48 hours and 90.4% for 336 hours, outperforming all comparative algorithms.

Keywords: *PV power forecasting, Medium and Long term forecasting, Transformer, Attention mechanism, Feature pyramid self-attention distillation.*

1. Introduction

Medium and long-term photovoltaic (PV) power forecasting refers to the prediction of electricity generation by photovoltaic power systems over a period ranging from several days to months or even longer. It plays a significant role in energy planning, power system operations, and energy investment [1]. Compared to short-term PV power forecasting, medium and long-term PV power exhibits cyclical variation. Typically, the photovoltaic power curve changes in a similar trend every day, showing a daily cyclicality. Moreover, with the change of seasons, the variation in solar elevation angle and daylight duration also affects the output of photovoltaic power, presenting a seasonal cyclical change. Medium and long-term forecasting requires consideration of a longer time range and a larger spatial scale, which necessitates the model to have more complex spatiotemporal feature capturing capabilities [2]. At the same time, considering the differences in geographical locations, the forecasting model needs to adapt to meteorological differences in different regions.

Current photovoltaic power forecasting methods mainly include statistical methods [3], machine learning methods [4], and deep learning methods [5]. Traditional statistical methods provide intuitive explanations of the relationships between power and various influencing factors, which helps to deeply understand the main factors affecting power fluctuations [6]. These methods perform well on smaller-scale and shorter time-span datasets, suitable for many scenarios in practical applications [7]. However, statistical methods are usually based on linear assumptions, making it difficult to capture complex nonlinear relationships. They also have higher requirements for data quality and sampling frequency, and their response to potential emergencies or uncertain factors that may arise in the future is relatively poor [8]. Therefore, statistical methods have certain limitations when dealing with medium and long-term PV power forecasting tasks.

Machine learning methods, trained on a vast amount of historical data, can automatically adapt to the nonlinear and complex relationships in photovoltaic power forecasting tasks, exhibiting good generalization capabilities [9]. These methods are widely applied in photovoltaic power forecasting tasks. Typically, machine learning

methods require a large volume of data for training and tuning, especially in medium and long-term forecasting where more time series data and related variables need to be considered, involving substantial computational resources [10]. Moreover, machine learning methods carry the risk of overfitting when dealing with large-scale data [11], which is particularly prominent in medium and long-term PV power forecasting.

Deep learning models such as RNN (Recurrent Neural Network) and LSTM [12] (Long Short-Term Memory) possess memory capabilities, enabling them to consider contextual information in time series data, such as seasonal variations and cyclical trends, thereby better predicting the changes in photovoltaic power generation in the medium to long term [13]. Deep learning methods, through their multi-layered neural network structures, can capture complex nonlinear relationships in photovoltaic power and effectively learn spatiotemporal features, better handling the impact of multidimensional factors such as illumination, meteorology on photovoltaic power, enhancing the model's adaptability to the dynamic changes of photovoltaic power systems, and thereby improving forecast accuracy [14]. However, researchers have found in recent years that deep learning methods like RNN and LSTM do not perform ideally when dealing with long-term dependencies in time series data [15]. As a neural network based on recurrent structures, LSTM's computation process is sequential, with each time step depending on the results of the previous time step. This makes effective parallel computation difficult during training, thus limiting the model's training speed. Moreover, when dealing with extremely long sequences, LSTM may face the issue of error accumulation. Additionally, the variation in photovoltaic power is not only influenced by time series factors but also by various meteorological factors such as light exposure, temperature, wind speed, etc. Although LSTM can capture complex relationships in time series, it may not effectively extract and integrate this multi-dimensional feature information.

Existing photovoltaic power forecasting methods are mostly designed for short-term predictions [16], and their performance often falls short in medium and long-term forecasting [17]. The Transformer captures long-term dependencies in sequences through its self-attention mechanism, allowing the model to focus on all other positions when processing each position, unlike RNN and LSTM which need to process in a sequential order by time steps. This makes the Transformer more efficient in training and inference for long sequence data. In recent years, Transformer models have been widely applied to medium and long-term PV power forecasting tasks. For instance, Ran et al [18] proposed a hybrid model that combines adaptive noise, complete ensemble empirical mode decomposition, sample entropy, and Transformer. This model addresses the long memory loss issue by introducing an attention mechanism and combines empirical mode decomposition techniques with the Transformer to verify the final impact of different mode decomposition techniques on forecasting results.

The aforementioned methods only use time series data for photovoltaic power forecasting, without considering the impact of meteorological and geographical factors on photovoltaic power. The lack of support for geographical location and meteorological information can lead to poor performance in perceiving the spatiotemporal characteristics of the data. Zhang et al [19] proposed a new Transformer model for power grid load forecasting, which combines Transformer and graph convolutional networks, using a feedforward neural network to output the predicted load values. The aforementioned methods use global self-attention mechanisms for medium and long-term PV power forecasting. The global self-attention mechanism treats all input sequence information as equally important, which can cause the model to overlook local features and changes, wasting a large amount of computational resources on processing spatiotemporal data with little impact on forecasting, increasing the model's complexity and computational costs. Cao et al [20] proposed an LSTM-Informer model based on an improved Stacking ensemble algorithm. This model utilizes long short-term memory and Informer as the base models and improves the traditional k-fold cross-validation in the Stacking algorithm to time series cross-validation, integrating time series forecasting models. The aforementioned methods use a stacked multi-layer structure of identical encoder-decoder structures to process feature information, which can lead to an excessive number of model parameters. When dealing with large-scale spatiotemporal data, this structure is prone to overfitting, reducing the model's ability to capture potential patterns in complex data, and limiting the model's performance and generalization ability in medium and long-term forecasting tasks.

Compared to traditional statistical methods and machine learning methods, Transformer models have certain advantages in medium and long-term time series forecasting tasks, but they still face

certain challenges when predicting photovoltaic power. The position encoding of Transformer models can only capture temporal information and cannot fully consider the impact of extensive meteorological and geographical factors on photovoltaic power. The self-attention mechanism of Transformer models has excessive computational complexity when processing medium and long-term PV power data. Photovoltaic power data not only has time series characteristics but is also affected by geographical location and meteorological conditions. Therefore, a model is needed that can capture this spatiotemporal correlation. In response to the above issues, this paper proposes a Transformer-based medium and long-term PV power forecasting model called SP-Transformer (Spatiotemporal-ProbSparse Transformer).

The main contributions of this paper are as follows:

- To address the issue of weak spatiotemporal correlation in medium and long-term PV power data, this paper introduces a spatiotemporal position encoding method. By embedding encoded vectors containing site spatial information into the input time-series meteorological data, the model can more accurately capture the spatiotemporal dependencies between sites. This method uses timestamps to represent different temporal positions and incorporates latitude and longitude as spatial position encodings, revealing the relative spatial relationships between sites. This approach significantly enhances the model's ability to perceive spatiotemporal associations in photovoltaic power data, demonstrating better adaptability in scenarios with significant spatial distribution and climatic condition differences.
- To resolve the challenges of inefficiency and insufficient correlation capture stemming from data redundancy in medium and long-term forecasting, this paper proposes a Spatiotemporal-ProbSparse Self-Attention mechanism. This mechanism not only reduces the computational load but also enhances the utilization of spatiotemporal correlations. By introducing the Haversine distance measure and a probability sparse strategy, the mechanism identifies "active points" in the spatiotemporal dimension, reducing computational expenditure on "nonactive points", thereby improving the accuracy and efficiency of photovoltaic power forecasting. This mechanism provides a more efficient solution for medium and long-term forecasting tasks.
- In response to the inefficiencies encountered in medium and long-term photovoltaic power forecasting, this paper introduces a Feature Pyramid Self-Attention Distillation Module (FPSA) that precisely captures potential patterns within photovoltaic power data. This approach constructs a multi-level feature pyramid structure through deep separable convolutions across various scales, providing an extensive receptive field to aid the model in understanding complex patterns, ensuring efficient feature extraction and complete information transfer. The FPSA achieves efficient feature extraction in complex environments and can be widely applied to medium and long-term forecasting tasks, enabling efficient mining of temporal information.

2. Method

To effectively capture the spatiotemporal correlations between meteorological and geographical elements and photovoltaic power data in medium and long-term forecasting, this paper proposes a Transformer-based medium and long-term PV power forecasting model, the SpatiotemporalProbSparse Transformer (SP-Transformer), aimed at enhancing the accuracy and efficiency of medium and long-term PV power forecasting. The overall architecture of the model is shown in figure 1.

The SP-Transformer model enhances its ability to capture the complex spatiotemporal correlations in photovoltaic power data through the mechanism of spatiotemporal position encoding. This encoding provides the model with crucial information about the geographical locations of sites, thereby improving its capability to capture more intricate spatiotemporal features. With spatiotemporal position encoding, the model can gain deeper insights into the interdependencies between different sites, offering richer contextual information for forecasting tasks. To further reduce the model's time and space complexity while maintaining predictive accuracy and stability, the SP-Transformer model also incorporates a spatiotemporal probability sparse self-attention mechanism. This mechanism selectively focuses on areas that have a key impact on the forecasting task, capturing the spatiotemporal correlations in the input data more effectively while reducing model complexity. This mechanism allows the model to improve predictive accuracy at a more efficient

computational complexity and enhances its adaptability to large-scale photovoltaic power data. The SP-Transformer model utilizes a feature pyramid self-attention distillation mechanism, which reduces information loss and enhances model stability through multi-scale feature extraction and fusion. This enables the model to consider spatiotemporal features at various scales more comprehensively, thus better adapting to different forecasting scenarios. The introduction of self-attention distillation helps ensure that the model maintains accuracy and consistency with input data in long-term forecasting tasks.

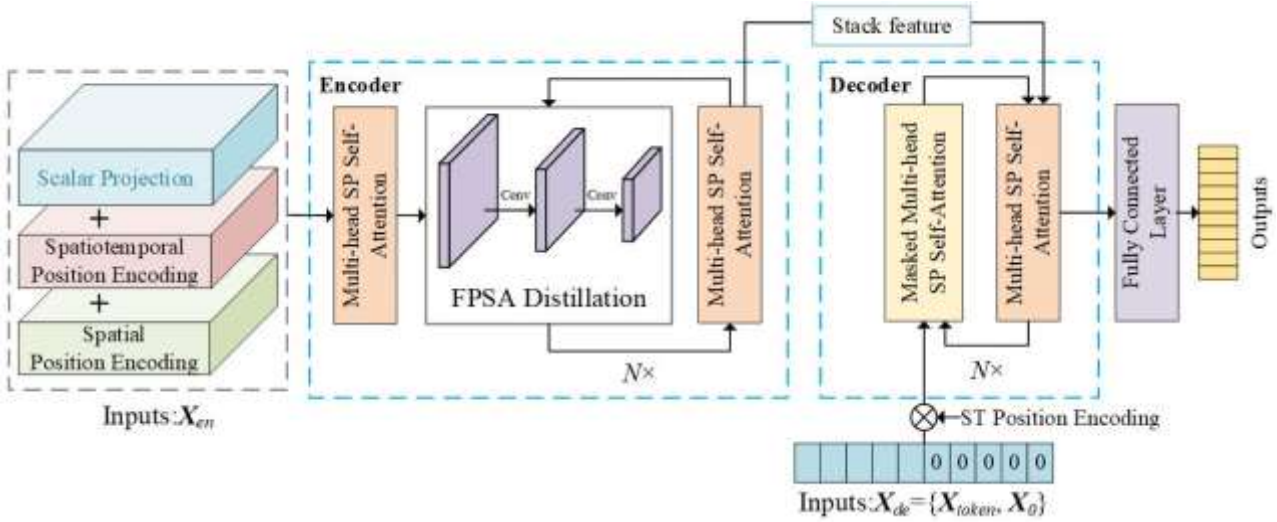

Figure 1. Overall Architecture of SP-Transformer.

2.1. Spatiotemporal position encoding

In the Transformer model, position encoding is used to handle positional information within input sequences, as the Transformer lacks explicit sequencing like Recurrent Neural Networks (RNNs) or Long Short-Term Memory networks (LSTMs). Without position encoding, the Transformer would not be able to distinguish between words or tokens at different positions in the input sequence, as it is an attention-based model that focuses on the content of the input rather than its position. Position encoding allows the Transformer to take into account the relative position of each element in the sequence, but this is limited to one-dimensional sequences. In the medium and long-term photovoltaic power forecasting task, the position encoding in the Transformer can only consider the sequential information along the temporal dimension and cannot account for the relative positions of different nodes in the spatial dimension.

To enable the model to further extract the relative spatial position information of different photovoltaic sites, this paper adds spatial position encoding to the input sequence on this basis. The input vector of the model is obtained by summing the scalar projection, local time stamp, global time stamp, and spatial position encoding, as shown in equation (1):

$$\mathrm{X}_{\mathrm{input}\,[i]}^{t} = \alpha u_i^t + \mathrm{PE}_{(L_x\times(t-1)+i,)} + \sum_p \left[\mathrm{SE}_{(L_x\times(t-1)+i)}\right]_p + \mathrm{SSE}_{(L_x\times(t-1)+i)} \quad (1)$$

where u_i^t represents the scalar projection, and α is a factor that balances the size of the scalar projection with other encodings. If the input sequence has already been normalized, then $\alpha = 1$. The scalars in this paper include historical photovoltaic power, temperature, humidity, horizontal wind speed, vertical wind speed, wind direction, cloud water content, cloud ice content, and solar irradiance. PE stands for the local time stamp, SE for the global time stamp, and SSE for the spatial position encoding. t represents the moment, L_x is the length of the input scalar sequence, i is the current position, and p is the global time stamp. The structure of the spatiotemporal position encoding is shown in figure 2.

Local time stamp refers to the position encoding of Transformer, as shown in equation (2, 3):

$$\mathrm{PE}_{(pos,2j)} = \sin\left(pos / \left(2L_x\right)^{2j/d_{\mathrm{model}}}\right) \quad (2)$$

$$\mathrm{PE}_{(pos,2j+1)} = \cos\left(pos / \left(2L_x\right)^{2j/d_{\mathrm{model}}}\right) \quad (3)$$

where pos represents the position of the data point at the current moment, $2j$ denotes even points, $2j+1$ denotes odd points, and d_{model} refers to the dimensionality of the input sequence features.

The global time stamp selects hierarchical timestamps, which helps to enhance the model's ability to capture long-range dependencies. Considering that photovoltaic power is minimal during the night and early morning, this paper selects data from 8:00 to 18:00 each day as input. Additionally, since photovoltaic power generation is primarily influenced by seasonal changes and the alternation of day and night, the impact of annual, monthly, and daily time features on photovoltaic power is not significant. Therefore, this paper chooses season and hour as the global time stamps.

This paper selects latitude and longitude coordinates as spatial position encoding. The geographical location of photovoltaic stations affects variables such as daily sunlight duration and the angle of solar incidence. By introducing latitude and longitude coordinates, the model can better capture the differences in geographical locations, thereby more accurately reflecting changes in solar radiation and other factors.

Spatial position encoding helps the model capture spatial correlations. Adjacent stations may have similar lighting and weather patterns, and this correlation can be better modeled through spatial position encoding. The model can take into account the spatial relationships between stations, thereby more accurately predicting future power outputs.

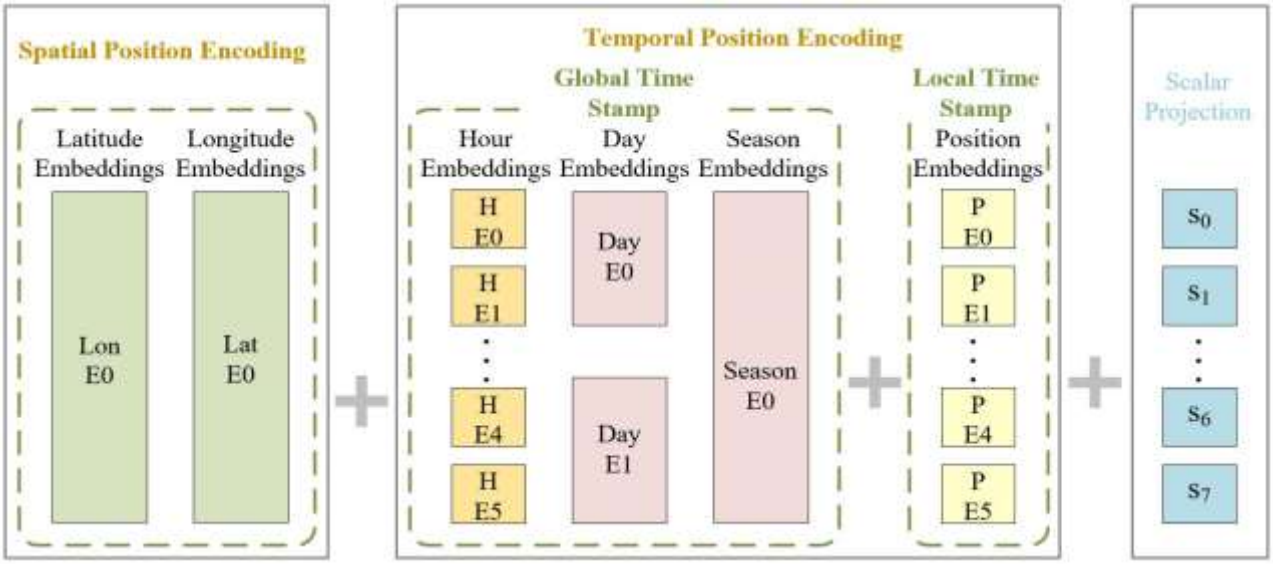

Figure 2. The structure of Spatiotemporal Position Encoding.

2.2. Spatiotemporal probability sparse self-attention mechanism

The self-attention mechanism of the Transformer model leads to high computational complexity when dealing with long sequences, as it requires calculating attention weights with all other positions for each position. The computational complexity increases exponentially when processing long sequence data, reducing the efficiency of handling long sequences. However, not all data points are closely related. To enable the model to better extract key information, this paper proposes a Spatiotemporal-ProbSparse self-attention mechanism, which selects some active data points in the input sequence along the spatiotemporal dimension to calculate attention, thereby reducing the overall computational cost. This allows the model to focus more on the key parts of the sequence without being distracted by unnecessary information.

Adjacent photovoltaic power stations are typically influenced by similar meteorological conditions and environmental factors, such as similar terrain and solar incidence angles. By leveraging the power data from nearby stations, the model can learn these shared pieces of information, thereby better capturing spatial correlations. This paper proposes the Spatiotemporal-ProbSparse self-attention mechanism to select nearby active points of photovoltaic power stations in the spatial dimension. The specific equation is as follows:

$$D_{(i,j)} = \frac{\sum_{j=1}^{L} d_{(i,j)}}{L} - d_{(i,j)} \tag{4}$$

where d represents the distance between two photovoltaic sites, r is the average radius of the Earth, lon_i, lat_i are the longitude and latitude of the target site, and lon_j, lat_j are the longitude and latitude of the selected site. L denotes the total number of photovoltaic power stations. D is defined as the difference between the mean distance of the target site to all other sites and the distance from the target site to the selected site. If D is greater than 0, the selected site is considered an active point that may influence the photovoltaic power forecasting of the target site. Global attention is calculated for the photovoltaic power data of the selected active points and the target site at each moment, resulting in photovoltaic power data that takes into account spatial location information.

The traditional Transformer model utilizes self-attention mechanisms with a time complexity of $O\left(L^2\right)$, which leads to high memory consumption and low computational efficiency when processing long sequence data. The ProbSparse selfattention mechanism addresses this issue by calculating the difference between the target point's attention distribution and a uniform distribution, thereby identifying points that significantly contribute to the attention computation while ignoring others. This approach reduces the time complexity of the Transformer model from $O(L^2)$ to $O(L*\ln(L))$,

$$d_{(i,j)} = 2\mathrm{r}\cdot\arcsin\left(\sqrt{\sin^2\left(\frac{lat_i - lat_j}{2}\right) + \cos\left(lat_i\right)\cdot\cos\left(lat_j\right)\cdot\sin^2\left(\frac{lon_i - lon_j}{2}\right)}\right) \tag{5}$$

substantially enhancing the model's performance in predicting long sequence data. The ProbSparse self-attention mechanism utilizes the Kullback-Leibler (KL) divergence to measure the discrepancy between the attention probability distribution and the uniform distribution. The specific equation is as follows:

$$\bar{M}\left(q_i, K\right) = \max_j\left\{\frac{q_i k_j^{\cdot}}{\sqrt{d}}\right\} - \frac{1}{L_K}\sum_{j=1}^{L_K}\frac{q_i k_j^{\cdot}}{\sqrt{d}} \tag{6}$$

Where q represents the query, K represents the Key, L_K represents the total number of keys, and $\bar{M}\left(q_i, K\right)$ represents the sparsity measure for the i-th query. If the $\bar{M}\left(q_i, K\right)$ for the *i*-th query is large, it is considered to contribute more to the attention. Selecting several points with the largest $\bar{M}\left(q_i, K\right)$ can approximate the attention probability distribution. The final equation for the Spatiotemporal-ProbSparse attention calculation is:

$$A(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{Softmax}\left(\frac{\bar{\mathbf{Q}}\mathbf{K}^{\cdot}}{\sqrt{d}}\right)\mathbf{V} \tag{7}$$

where $\bar{\mathrm{Q}}$ is a sparse matrix that contains only u queries selected by the Spatiotemporal-ProbSparse attention mechanism, with other points filled with zeros, where $u = c*\ln L_Q$ and c is a constant sampling factor. The SpatiotemporalProbSparse self-attention mechanism selects active points on both spatial and temporal dimensions, thereby limiting the model's complexity to $O\left(L_K*\ln L_Q\right)$. This approach not only captures the spatiotemporal correlations within the input data more effectively but also focuses on regions that are critical for the prediction task, thereby enhancing the prediction accuracy with a more efficient computational complexity. This design makes the model more adaptable to large-scale photovoltaic power data, enhancing its feasibility in practical application scenarios.

2.3. FPSA

After the input sequence passes through the SpatiotemporalProbSparse attention layer to obtain sparse features, redundant values are inevitably present in the feature map. To highlight key features and further reduce the computational load of the model, feature distillation is an effective method. Some existing models use max-pooling to reduce the dimensionality of the attention block, and in the stacked Encoder, the input sequence is halved with each additional layer, finally connecting the outputs of all layers to obtain the feature map. However, max-pooling retains only the maximum value and ignores other information; other sub-maximal feature values may also contain important information, and directly downsampling the input sequence can lead to the loss of a significant amount of long-term dependency information. Therefore, this paper proposes a Feature Pyramid Self-Attention Distillation Module (FPSA) to better extract dominant key features, as shown in figure 3.

In this figure, represents the input scalar, denotes the length of the time series, n refers to the number of multi-head attention heads, and indicates the number of encoder layers. For the original input sequence, downsampling is applied starting from the stacked second layer of the Encoder using Depthwise Separable Convolution, with each layer employing convolution kernels of different sizes that increase with the layer number. By utilizing multi-scale convolution along the temporal dimension, the network is enabled to focus on temporal information of varying lengths. In the channel dimension, a 1×1 convolution is employed to extract cross-features among different elements, thereby enriching the network's receptive field and allowing for a more comprehensive understanding of the structure and content of the input sequence. The multi-level convolutional network can learn more abstract and high-level features, which aids the network in establishing an understanding of complex patterns and objects. Furthermore, downsampling through convolution is applied to the attention blocks in each Encoder layer. Compared to max-pooling, convolution can more effectively capture local features of the input data. By stacking multiple convolutional and pooling layers, the model can adapt to feature data of varying scales and learn deeper hierarchical representations of the data. The feature pyramid self-attention distillation process is shown in the equation (8).

$$\mathbf{X}_{j+1}^{i} = \mathrm{ELU}(\mathrm{DSConv}([\mathbf{X}_{j}^{i}]_{\mathrm{AB}})) \tag{8}$$

where represents the number of Encoder layers, j denotes the number of distillation layers within each Encoder, indicates the self-attention block operation, DSConv refers to depthwise separable convolution, and ELU is the activation function. Through the distillation process, the model can extract key features from the stacked multiple Encoders while reducing redundant information. Finally, the outputs of the stacked Encoders are concatenated along the channel dimension to form the output of the Encoder. The feature map produced by the Encoder is then processed through two stacked Decoder layers to obtain the final photovoltaic power prediction value.

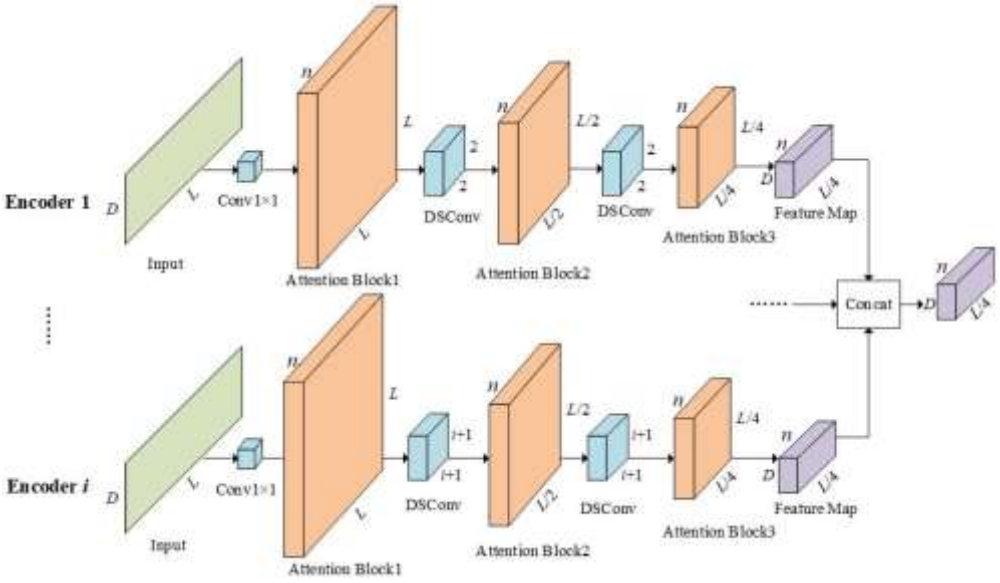

Figure 3. The structure of FPSA.

3. Experiment

3.1. Dataset

The dataset used in this study includes photovoltaic The dataset used in this study includes photovoltaic power data and meteorological data, covering the period from March 2022 to February 2023. In the dataset, 70% of the data is allocated to the training set, 20% to the testing set, and the remaining 10% to the validation set. To eliminate instances of nearly zero photovoltaic power during nighttime, this study selected data only from 8 AM to 6 PM each day. The photovoltaic power data is sourced from the open dataset provided by the Belgian electricity supplier Elia. This paper utilized meteorological data from the WRF model provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), which includes information on temperature, humidity, wind speed, wind direction, cloud water content, cloud ice content, and solar irradiance. The meteorological data has a temporal resolution of 1 hour and a spatial resolution of 1 kilometer.

3.2. Experimental setup and scheme

The experimental setup used in this study features an Intel(R) Core(TM) i9-10900X processor, 32GB of RAM, and an NVIDIA GeForce GTX 2080 Ti GPU, with Ubuntu 18.04 as the operating system. The environment is configured to run PyTorch version 3.6. The training process employs the Adam optimizer, with all models trained for 100 epochs. The batch size is set to 64, and the initial learning rate is set to 0.001.

3.2.1. Comparative experimental setup

To evaluate the performance of the model in the photovoltaic power prediction task, this paper compared it with several common time series forecasting methods, including LSTM, Transformer [21], Log Transformer [22], Informer [23], and Fedformer [24]. To ensure fairness in the experiments, the dataset was split into training and testing sets. After the model training was completed, each model was applied to the testing set, and their performance in the photovoltaic power prediction task was assessed. This study utilized Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to compare the prediction effectiveness of the various models.

3.2.2. Ablation experimental setup

To validate the effectiveness of each module in the model, this study conducted ablation experiments, systematically removing specific components and observing their impact on overall performance. The complete SP-Transformer, which includes all modules, served as the baseline model. In Model 1, the spatiotemporal positional encoding was replaced with sequential positional encoding. In Model 2, the spatiotemporal probabilistic sparse self-attention mechanism was substituted with a global self-attention mechanism. In Model 3, the feature pyramid-based self-attention distillation decoder was replaced with the decoder structure from the Transformer. RMSE and MAE were utilized to evaluate the prediction performance of each model.

3.3. Evaluation metrics

This study employs RMSE and MAE as two metrics to evaluate the performance of the photovoltaic power prediction model. The equations for RMSE and MAE are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_t - y_t\right)^2} \tag{9}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|y_t - y_t\right| \tag{10}$$

where $\hat{y}_t$ represents the predicted photovoltaic power at time t, y_t denotes the actual photovoltaic power at time t, and N is the number of samples.

3.4. Experimental results

3.4.1. Comparative experimental results

To validate the advantages of the proposed SP-Transformer in the photovoltaic power prediction task, this study compared the prediction results of SP-Transformer with several common time series forecasting methods, including LSTM, Transformer, LogTransformer, Informer, and FEDformer. The prediction time steps were gradually extended to assess the performance of each model in long sequence prediction tasks.

The RMSE values of the photovoltaic power prediction results from the aforementioned methods on the testing set are visually represented in figure 4. From the figure, it can be observed that as the prediction time steps are extended, most models begin to exhibit significant deviations in their predictions. However, the SP-Transformer, utilizing the spatiotemporal probabilistic sparse self-attention mechanism, is better able to capture the spatiotemporal relationships between photovoltaic power and geographical as well as meteorological factors. By selectively focusing on regions that have a critical impact on the prediction task, the model enhances prediction accuracy and stability while maintaining a more efficient computational complexity. Consequently, compared to other models, the SP-Transformer demonstrates stable predictive performance across all time steps.

Figure 5 displays a comparison of the prediction results for the next 40 time steps from each model at the 48-hour and 336-hour nodes. Figure 6 further illustrates the declining trend in prediction accuracy of these models from the 48-hour to the 336-hour forecast horizon.

From the figures, it can be observed that while some models exhibit performance close to that of the SP-Transformer at the 48-hour prediction mark, there is a noticeable decline in prediction accuracy for all models, except for the SP-Transformer, during the 336-hour long-term forecasting task. The average prediction accuracy of the SP-Transformer at 336 hours decreases by approximately 10% compared to 48 hours, representing the smallest decline among all methods. In contrast, the second-best performing model, FEDformer, experiences a 24% drop in

accuracy, while the least effective model, LSTM, sees a 32% decline. This indicates that the SP-Transformer is effective in capturing the spatiotemporal relationships in time series data when dealing with long sequences, thereby demonstrating a clear advantage in stability and accuracy. By employing spatiotemporal positional encoding, the model embeds geographical location information of relevant sites, enhancing its ability to capture the spatiotemporal dependencies of photovoltaic power data.

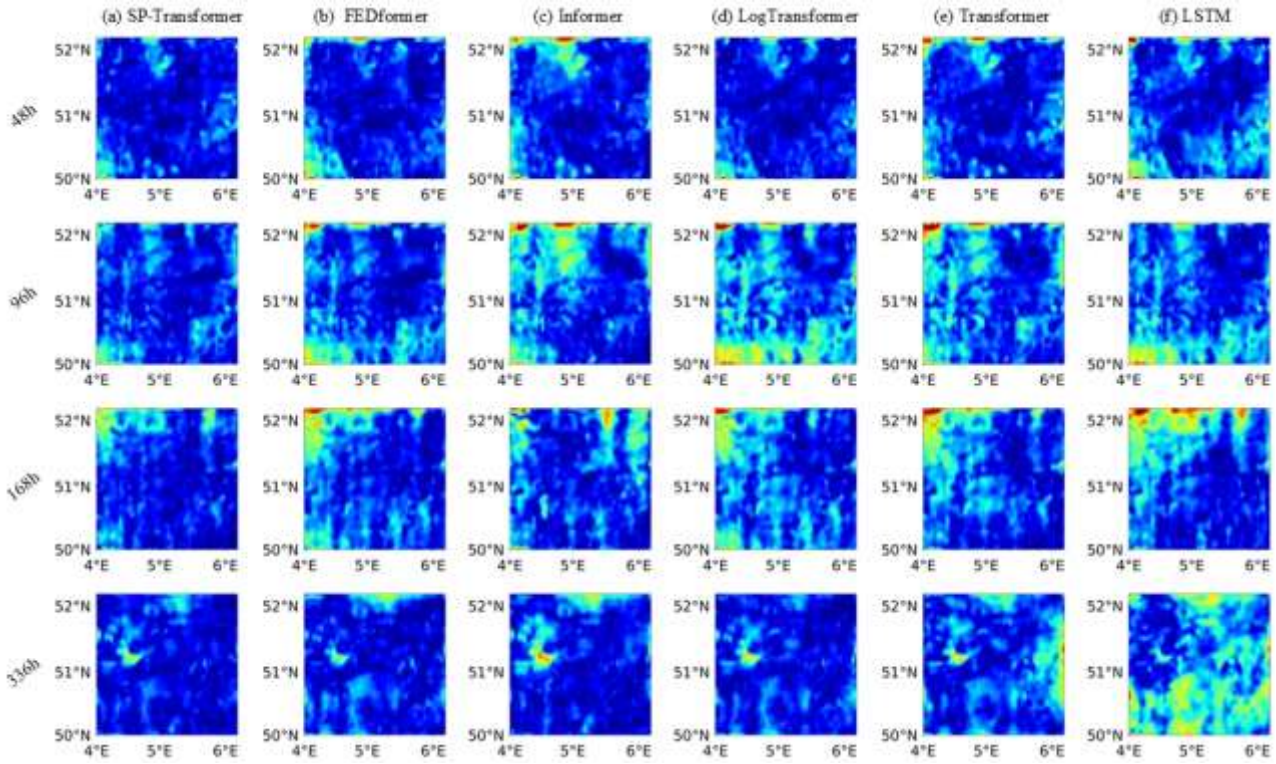

Figure 4. Visualization of RMSE in comparative experiments.

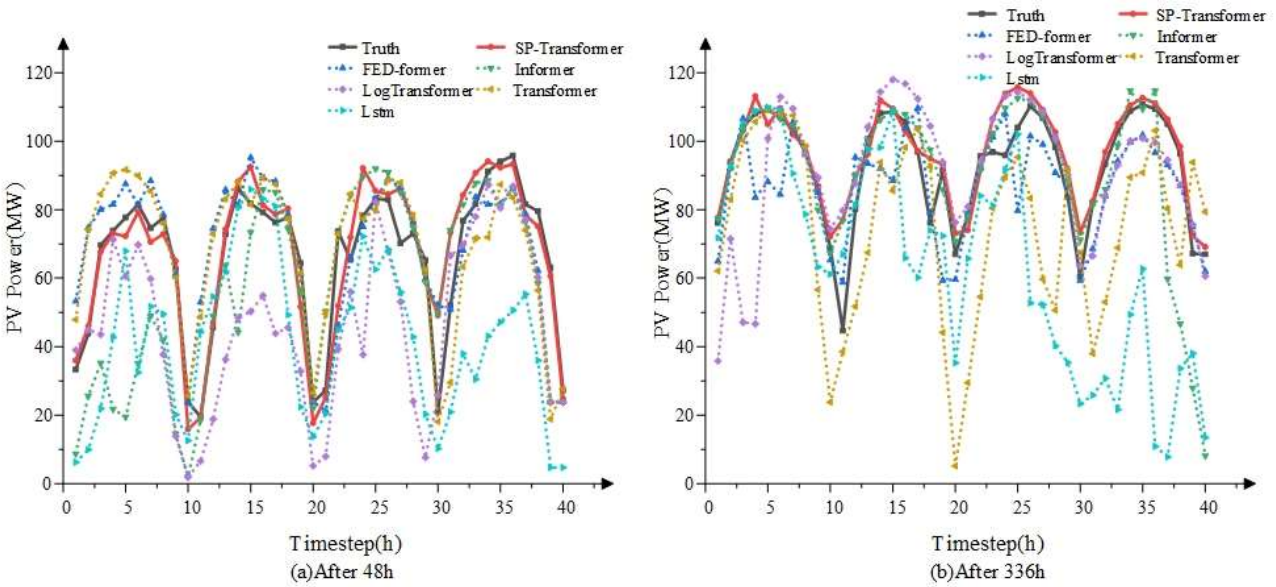

Figure 5. Comparison of prediction results for future 40 time steps at 48-hour and 336-hour nodes across models.

Compared to other models, the SP-Transformer delves deeper into the relationships between meteorological and geographical factors and photovoltaic power, improving prediction accuracy. Additionally, the SP-Transformer utilizes a spatiotemporal probabilistic sparse self-attention mechanism, which selectively focuses on key areas that significantly influence the prediction task. This approach reduces computational complexity when handling large-

scale data while enhancing the model's adaptability and predictive performance. The introduction of a feature pyramid-based self-attention distillation module, along with multi-scale depthwise separable convolution for feature extraction, effectively minimizes information loss and improves model stability, thus enhancing its ability to capture and generalize complex patterns in photovoltaic power data. In summary, the SP-Transformer addresses issues of weak data correlation and low prediction efficiency in medium and long-term PV power forecasting through its spatiotemporal positional encoding, spatiotemporal probabilistic sparse self-attention mechanism, and feature pyramid self-attention distillation module, resulting in superior performance in these tasks.

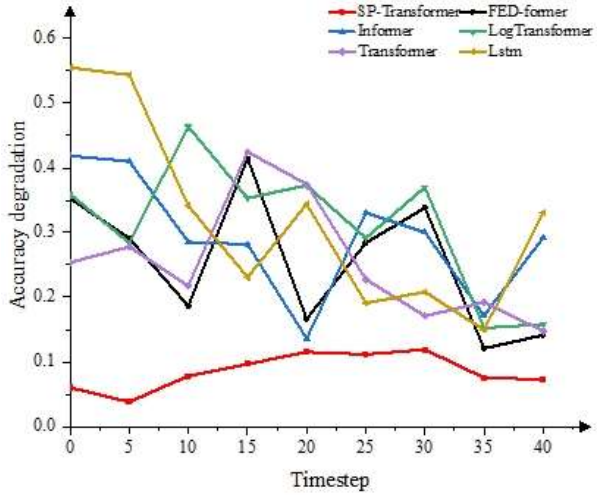

Figure 6. Prediction accuracy degradation in comparative experiments.

Table 1 presents the results of the comparative experiments for each model. This paper utilized RMSE and MAE as evaluation metrics, with the best result for each metric highlighted in bold.

Ten random data sets were selected from the testing set to conduct experiments on each model, and the RMSE and MAE metrics were averaged over ten trials. From the data in table 1, it is evident that the SP-Transformer outperforms the other methods. The RMSE for the SP-Transformer at 48 hours is 0.761, which is 4.2% lower than that of FEDformer [24], 10.8% lower than Informer [23], 20.3% lower than LogTransformer [22], and 33.9% lower than Transformer [21]. The SP-Transformer demonstrates high prediction accuracy across all four prediction time steps, maintaining an RMSE of 1.061 at 336 hours. Although this represents an increase in error compared to the 48-hour prediction, the growth trend is steady and gradual, remaining significantly lower than that of the other methods. Overall, the SP-Transformer achieves the highest accuracy in predicting photovoltaic power and exhibits the most stable performance in long sequence predictions.

Transformer [21]. The SPTransformer demonstrates high prediction accuracy across all four prediction time steps, maintaining an RMSE of 1.061 at 336 hours. Although this represents an increase in error compared to the 48-hour prediction, the growth trend is steady and gradual, remaining significantly lower than that of the other methods. Overall, the SP-Transformer achieves the highest accuracy in predicting photovoltaic power and exhibits the most stable performance in long sequence predictions.

Table 1. Comparative experimental results.

method	48 h		96 h		168 h		336 h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Ours	0.761	0.558	0.784	0.562	0.905	0.655	1.061	0.768
[24]	0.795	0.659	0.885	0.725	0.987	0.799	1.123	0.924
[23]	0.854	0.663	0.913	0.706	1.053	0.811	1.096	0.844
[22]	0.956	0.714	1.002	0.737	1.139	0.874	1.209	0.952
[21]	1.151	0.932	1.162	0.959	1.190	0.952	1.232	0.995
[12]	2.360	1.934	2.363	1.946	2.418	1.934	4.384	3.494

3.4.2. Ablation experimental eesults

To investigate the impact of each component of the SP-Transformer model on the accuracy of photovoltaic power prediction, ablation experiments were conducted. Specific components of the model were systematically removed, and their effects on overall performance were observed. The Transformer was selected as the baseline model for comparison. In Model 1, the spatiotemporal positional encoding was eliminated, and traditional sequential positional encoding was used instead. Model 2 replaced the

spatiotemporal probabilistic sparse self-attention mechanism with a global self-attention mechanism. In Model 3, the decoder structure based on feature pyramid self-attention distillation was substituted with the decoder structure from the Transformer.

Figure 7 displays the prediction results of each model from the ablation experiments at the 48-hour and 336-hour marks, while figure 8 illustrates the decline in prediction accuracy for these models as the forecast horizon extends from 48 hours to 336 hours.

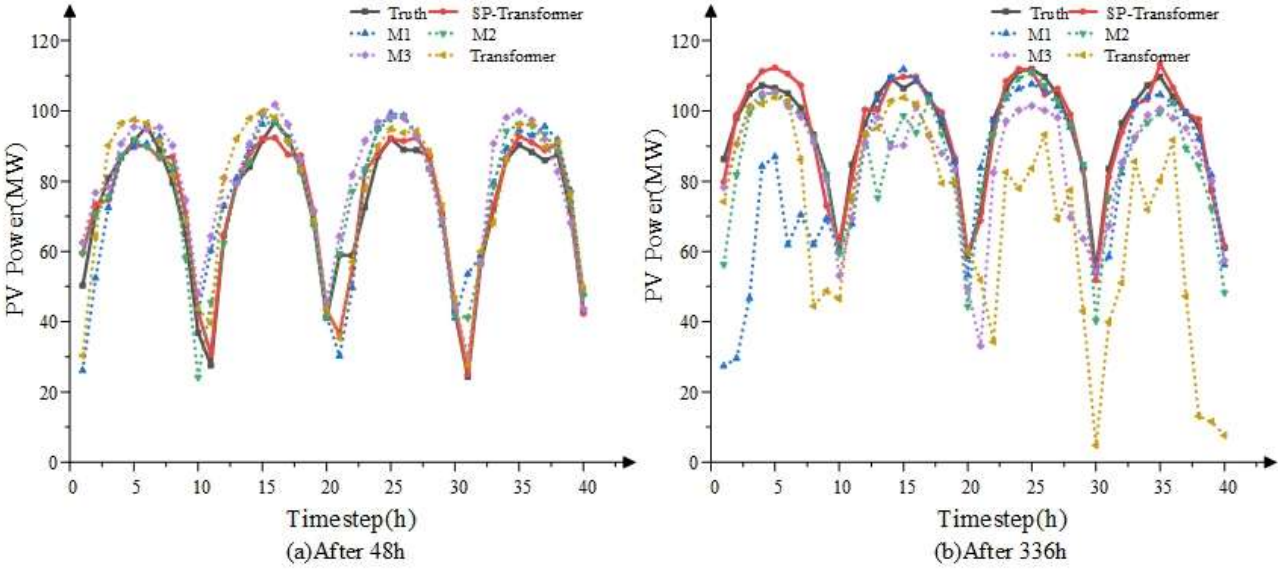

Figure 7. Prediction results of ablation experiments for each model at 48-hour and 336-hour nodes.

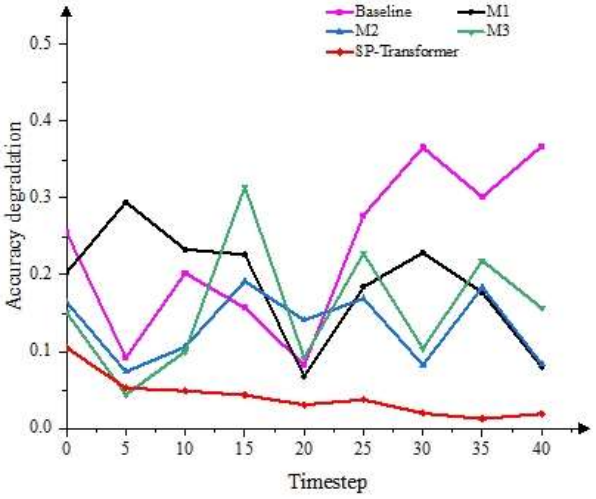

Figure 8. Prediction accuracy degradation in ablation experiments.

In the ablation experiments, Model 1 eliminated the spatiotemporal positional encoding in favor of traditional sequential positional encoding. The results indicated a decline in prediction accuracy compared to the complete SP-Transformer, highlighting the significant role of spatiotemporal positional encoding in capturing the spatiotemporal relationships in photovoltaic power data. The embedding of geographical location information enables the model to better understand the variations in photovoltaic power across different sites, thereby improving the accuracy of medium and long-term predictions.

Model 2 replaced the spatiotemporal probabilistic sparse self-attention mechanism with a global self-attention mechanism, resulting in decreased prediction accuracy. This suggests that the spatiotemporal probabilistic sparse self-attention mechanism is effective in addressing data redundancy issues. By selectively focusing on regions that have a critical impact on the prediction task, this mechanism effectively reduces spatiotemporal complexity and enhances the model's adaptability and predictive performance. Model 3 substituted the feature pyramid-based self-attention distillation decoder with the decoder structure from the Transformer.

The experimental results showed a negative impact on model performance, indicating that the feature pyramid-based self-attention distillation decoder, which employs multi-scale depthwise separable convolution for feature extraction, helps reduce information loss and improve model stability. The self-attention distillation method enhances the model's ability to capture and generalize complex patterns in photovoltaic power data, allowing it to maintain high accuracy and robustness in long-term prediction tasks.

The experimental results indicate that the complete SP-Transformer achieves an accuracy of 93.8% in 48-hour predictions and 90.4% in 336-hour predictions. In contrast, the prediction accuracy of Models 1 to 3, each with a module

removed, declines but remains higher than that of the traditional Transformer model. This demonstrates that each module in the SP-Transformer plays a crucial role in enhancing the accuracy and efficiency of medium and long-term PV power predictions. The collaborative contribution of these modules improves the model's performance in long-term forecasting tasks, resulting in superior stability and accuracy across different time scales compared to other models.

The RMSE values of the photovoltaic power prediction results from each model on the testing set are visually represented in figure 9.

Table 2 presents the results of the ablation experiments for each model. The data indicates that all models outperform the traditional Transformer, with the complete SP-Transformer exhibiting the best performance across all time steps. Model 1 shows a reduction in RMSE and MAE of 25.8% and 28.8% respectively, at the 48-hour prediction step compared to the Transformer, and reductions of 1.8% and 4.3% at the 336-hour step. The spatiotemporal positional encoding provides the model with the ability to perceive the spatiotemporal relationships between photovoltaic sites, enhancing its capability to capture complex spatiotemporal features. Model 2 achieves a reduction in RMSE and MAE of 16.9% and 23.3% respectively, at the 48-hour prediction step, and reductions of 11.3% and 16.1% at the 336-hour step. The spatiotemporal probabilistic sparse self-attention mechanism identifies active points that significantly impact the current prediction. This approach to computational efficiency, making it more suitable for large-scale datasets and real-time prediction tasks. Model 3 demonstrates a reduction in RMSE and MAE of 30.4% and 29.2% respectively, at the 48-hour prediction step, and reductions of 13.8% and 14.5% at the 336-hour step. The feature pyramid-based self-attention distillation, through multi-scale feature extraction, reduces information loss and enhances model stability. This improvement strengthens the model's ability to capture and generalize complex patterns in photovoltaic power data, helping it maintain accuracy and coherence in long-term predictions, thus addressing the issue of low efficiency in medium and long-term PV power forecasting.

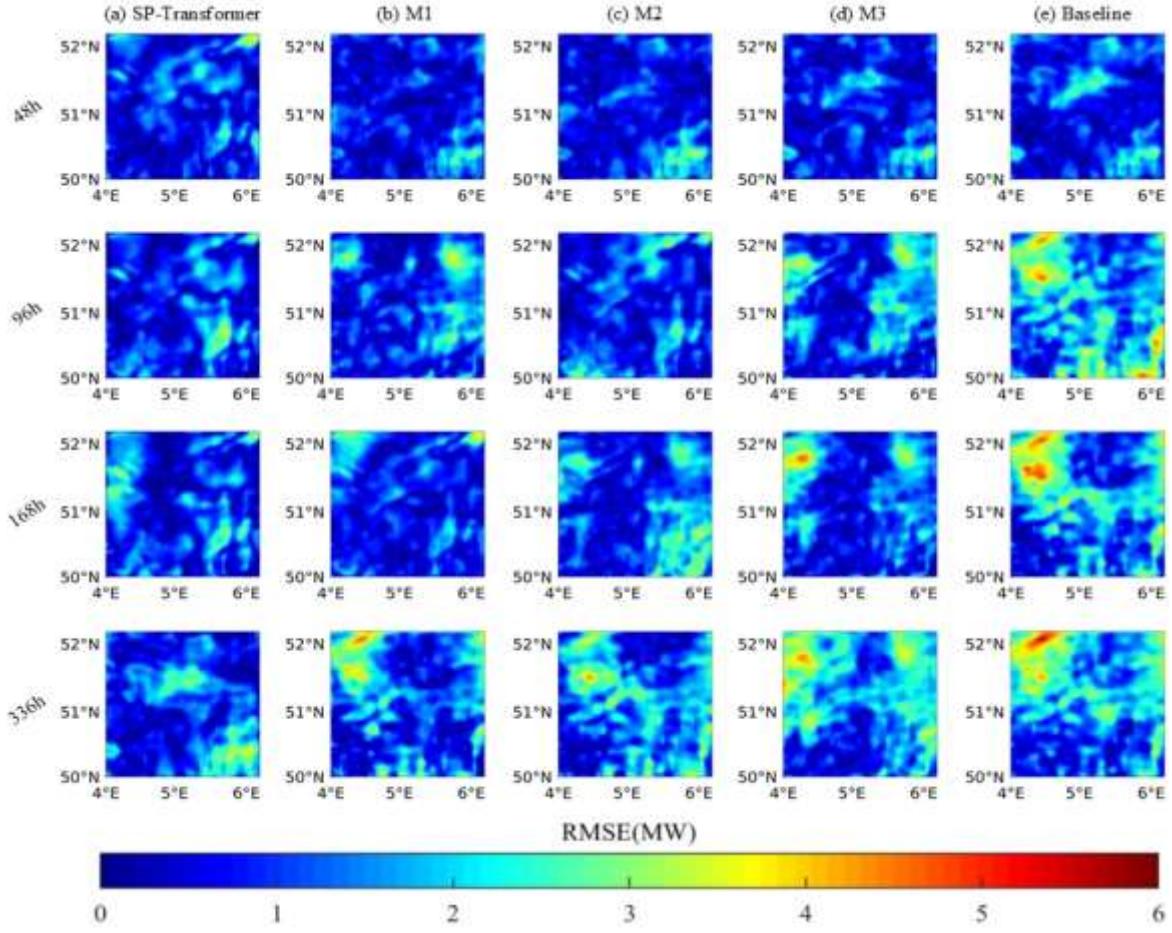

Figure 9. Visualization of RMSE in ablation experiments.

Table 2. Ablation experimental results.

method	48 h		96 h		168 h		336 h	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Ours	0.761	0.558	0.784	0.562	0.905	0.655	1.061	0.768
M1	0.854	0.663	0.913	0.706	1.053	0.811	1.209	0.952
M2	0.956	0.714	1.002	0.737	1.139	0.874	1.096	0.835
M3	0.800	0.659	0.885	0.706	0.987	0.800	1.061	0.835
Transformer	1.151	1.932	1.162	0.959	1.190	0.952	1.232	0.995

4. Conclusion

This paper presents the SP-Transformer model, aimed at addressing the issues of insufficient consideration of spatiotemporal relationships between sites and the low predictive efficiency in medium and long-term PV power forecasting. The main contributions are as follows: First, the introduction of spatiotemporal positional encoding enriches the photovoltaic site data with temporal and spatial information, enabling the model to better capture the relative positional relationships between sites. This encoding provides insights into the spatiotemporal associations among different sites, enhancing the model's ability to capture more complex spatiotemporal features. It increases sensitivity to geographic differences between sites and offers deeper contextual understanding. Second, this paper proposes the SpatiotemporalProbSparse self-attention mechanism, which selects active points in the spatiotemporal dimension to reduce model complexity and better capture the spatiotemporal correlations in the input data. This mechanism selectively focuses on areas that significantly impact the prediction task, improving the model's prediction accuracy while maintaining efficient computational complexity. This design enhances the model's adaptability when processing large-scale photovoltaic power data, providing greater feasibility for practical applications. This paper introduced a feature pyramid-based self-attention distillation approach, which reduces information loss through multi-scale feature extraction and improves model stability. This allows the model to consider spatiotemporal features across different scales more comprehensively, better adapting to various prediction scenarios. The inclusion of self-attention distillation helps ensure accuracy and consistency in long-term prediction tasks. Finally, this paper demonstrates through experiments that the SP-Transformer exhibits superior performance in medium and long-term PV power forecasting tasks.

5. Acknowledgment

This work was supported by the National Key Research and Development Plan of China Scientific and Technological Innovation 2030 - "New Generation Artificial Intelligence" Major Project (2021ZD0102100).

6. References

[1] T. Qiu, L. Wang, Y. Lu, M. Zhang, W. Qin, S. Wang, and L. Wang, ‘‘Potential assessment of photovoltaic power generation in china,’’ Renewable and Sustainable Energy Reviews, vol. 154, p. 111900, 2022.

[2] Y. Yang, J. Che, C. Deng, and L. Li, ‘‘Sequential grid approach based support vector regression for short-term electric load forecasting,’’ Applied energy, vol. 238, pp. 1010–1021, 2019.

[3] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu, M. Seyedmahmoudian, T. K. Soon, B. Horan, S. Mekhilef, and A. Stojcevski, ‘‘Shortterm pv power forecasting using hybrid gasvm technique,’’ Renewable energy, vol. 140, pp. 367–379, 2019.

[4] C. Zhu, M. Wang, M. Guo, J. Deng, Q. Du, W. Wei, and Y. Zhang, ‘‘Innovative approaches to solar energy forecasting: unveiling the power of hybrid models and machine learning algorithms for photovoltaic power optimization,’’ The Journal of Supercomputing, vol. 81, no. 1, p. 20, 2025.

[5] R. Nguyen, Y. Yang, A. Tohmeh, and H.-G. Yeh, ‘‘Predicting pv power generation using svm regression,’’ in 2021 IEEE Green Energy and Smart Systems Conference (IGESSC). IEEE, 2021, pp. 1–5.

[6] Z. Zhou, L. Liu, and N. Y. Dai, ‘‘Day-ahead power forecasting model for a photovoltaic plant in macao based on weather classification using svm/pcc/lm-ann,’’ in 2021 IEEE Sustainable Power and Energy Conference (iSPEC). IEEE, 2021, pp. 775–780.

[7] D. Niu, K. Wang, L. Sun, J. Wu, and X. Xu, ‘‘Short-term photovoltaic power generation forecasting based on random forest feature selection and ceemd: A case study,’’ Applied soft computing, vol. 93, p. 106389, 2020.

[8] M. Ali, R. Prasad, Y. Xiang, M. Khan, A. A. Farooque, T. Zong, and Z. M.Yaseen, ‘‘Variational

mode decomposition based random forest model for solar radiation forecasting: new emerging machine learning technology,'' Energy Reports, vol. 7, pp. 6700–6717, 2021.

[9] Y. Gao, J. Wang, L. Guo, and H. Peng, ''Short-term photovoltaic power prediction using nonlinear spiking neural p systems,'' Sustainability, vol. 16, no. 4, p. 1709, 2024.

[10] X. Huang, Q. Li, Y. Tai, Z. Chen, J. Liu, J. Shi, and W. Liu, ''Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and bi-lstm,'' Energy, vol. 246, p. 123403, 2022.

[11] L. Wang, M. Mao, J. Xie, Z. Liao, H. Zhang, and H. Li, ''Accurate solar pv power prediction interval method based on frequency-domain decomposition and lstm model,'' Energy, vol. 262, p. 125592, 2023.

[12] S. Hochreiter, ''Long short-term memory,'' Neural Computation MITPress, 1997.

[13] J. Qu, Z. Qian, and Y. Pei, ''Day-ahead hourly photovoltaic power forecasting using attention-based cnn-lstm neural network embedded with multiple relevant and target variables prediction pattern,'' Energy, vol. 232, p. 120996, 2021.

[14] T. Limouni, R. Yaagoubi, K. Bouziane, K. Guissi, and E. H. Baali, ''Accurate one step and multistep forecasting of very short-term pv power using lstm-tcn model,'' Renewable Energy, vol. 205, pp. 1010–1024, 2023.

[15] S. Huang, Y. Liu, F. Zhang, Y. Li, J. Li, and C. Zhang, ''Crosswavenet: A dual-channel network with deep cross-decomposition for long-term time series forecasting,'' Expert Systems with Applications, vol. 238, p. 121642, 2024.

[16] S. Wang and J. Ma, ''A novel gbdt-bilstm hybrid model on improving dayahead photovoltaic prediction,'' Scientific Reports, vol. 13, no. 1, p. 15113, 2023.

[17] V. Kushwaha and N. M. Pindoriya, ''A sarima-rvfl hybrid model assisted by wavelet decomposition for very short-term solar pv power generation forecast,'' Renewable Energy, vol. 140, pp. 124–139, 2019.

[18] P. Ran, K. Dong, X. Liu, and J. Wang, ''Short-term load forecasting based on ceemdan and transformer,'' Electric Power Systems Research, vol. 214, p. 108885, 2023.

[19] Q. Zhang, J. Chen, G. Xiao, S. He, and K. Deng, ''Transformgraph: A novel short-term electricity net load forecasting model,'' Energy Reports, vol. 9, pp. 2705–2717, 2023.

[20] Y. Cao, G. Liu, D. Luo, D. P. Bavirisetti, and G. Xiao, ''Multi-timescale photovoltaic power forecasting using an improved stacking ensemble algorithm based lstm-informer model,'' Energy, vol. 283, p. 128669, 2023.

[21] A. Vaswani, ''Attention is all you need,'' Advances in Neural Information Processing Systems, 2017.

[22] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, ''Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting,'' Advances in neural information processing systems, vol. 32, 2019.

[23] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, ''Informer: Beyond efficient transformer for long sequence time-series forecasting,'' in Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[24] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, ''Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting,'' in International conference on machine learning. PMLR, 2022, pp. 27 268–27 286.