


Article

A Novel Convolutional Neural Net Architecture Based on Incorporating Meteorological Variable Inputs into Ultra-Short-Term Photovoltaic Power Forecasting

Xiaoying Ren ^{1,2}, Fei Zhang ^{1,2}, Junshuai Yan ¹ and Yongqian Liu ^{1,*} 

¹ School of New Energy, North China Electric Power University, Beijing 102206, China; huhurenxy@imust.edu.cn (X.R.); rollrocket@imust.edu.cn (F.Z.)

² School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

* Correspondence: yqliu@ncepu.edu.cn

Abstract: Accurate photovoltaic (PV) power forecasting allows for better integration and management of renewable energy sources, which can help to reduce our dependence on finite fossil fuels, drive energy transitions and climate change mitigation, and thus promote the sustainable development of renewable energy sources. A convolutional neural network (CNN) forecasting method with a two-input, two-scale parallel cascade structure is proposed for ultra-short-term PV power forecasting tasks. The dual-input pattern of the model is constructed by integrating the weather variables and the historical power so as to convey finer information about the interaction between the weather variables and the PV power to the model; the design of the two-branch, two-scale CNN model architecture realizes in-depth fusion of the PV system data with the CNN's feature extraction mechanism. Each branch introduces an attention mechanism (AM) that focuses on the degree of influence between elements within the historical power sequence and the degree of influence of each meteorological variable on the historical power sequence, respectively. Actual operational data from three PV plants under different meteorological conditions are used. Compared with the baseline model, the proposed model shows a better forecasting performance, which provides a new idea for deep-learning-based PV power forecasting techniques, as well as important technical support for a high percentage of PV energy to be connected to the grid, thus promoting the sustainable development of renewable energy.

Keywords: photovoltaic power forecasting; dual input; deep learning; attention mechanism; convolutional neural network



Citation: Ren, X.; Zhang, F.; Yan, J.; Liu, Y. A Novel Convolutional Neural Net Architecture Based on Incorporating Meteorological Variable Inputs into Ultra-Short-Term Photovoltaic Power Forecasting.

Sustainability **2024**, *16*, 2786.
<https://doi.org/10.3390/su16072786>

Academic Editor: Seung-Hoon Yoo

Received: 6 February 2024

Revised: 12 March 2024

Accepted: 25 March 2024

Published: 27 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Motivation

Since the United Nations Intergovernmental Panel on Climate Change put forward the goal of “carbon neutrality” in October 2018, some countries around the world have taken a series of initiatives, among which developing renewable energy sources, achieving energy transitions, and reducing fossil fuel consumption to build a green and low-carbon energy system are some of the important measures for achieving global carbon neutrality [1]. According to the 2022 International Renewable Energy Agency statistics [2], 2021 was a strong year for energy transitions. By the end of 2021, renewables accounted for 38% of the global installed capacity, adding nearly 257 GW of renewable energy. Solar power alone accounted for over half of the renewable additions, with a record 133 GW. As a high percentage of photovoltaic (PV) power continues to be connected to the grid, the efficiency, security, and stability of its access pose great challenges to the robust operation of the power system [3]. Accurate PV power forecasting can improve the efficiency of the PV power system, realize the balance and stability of the power system, reduce energy costs and carbon emissions, and promote the sustainable development of renewable energy sources, which is of great significance to the sustainable development of energy.

1.2. Related Works

In recent years, numerous scholars have studied PV power forecasting techniques. Different forecasting theories have been applied to different forecasting time scales. Based on the different time scales of power system operation, PV power forecasting can be classified into ultra-short-term [4], short-term, and medium- to long-term forecasts. Short-term forecasts are generally used for power system power balance and economic dispatch, day-ahead generation planning, day-ahead trading in the power market, and transient stability assessment. Medium- and long-term forecasts are mainly used for power system maintenance scheduling and generation forecasting. Ultra-short-term forecasts are generally used for real-time dispatching, solving grid frequency regulation problems, and power quality assessment. According to the different theories involved in the forecasting research, they are divided into physical methods, statistical methods, and artificial intelligence methods (machine learning (ML) and deep learning (DL)). The physical methods are based on the principle of PV power generation and combine numerical weather prediction (NWP) data with the PV system's parameters to build a physical model to directly calculate the PV power prediction results. Its modeling relies on detailed system geographical information, system parameters, and accurate meteorological data, and the forecasting accuracy depends heavily on the accuracy of the NWP data, which is still limited [5], and few studies have applied them to PV power forecasting tasks for shorter time scales [6]. Statistical methods are used to establish and optimize the mapping relationship between the historical samples and actual PV power using curve fitting and parameter estimation with the objective of error minimization. Compared with the physical methods, statistical methods do not need to consider many parameters of the PV systems and complex PV conversion models and are relatively simple to model [7]. However, statistical methods require high-quality input data and a large amount of historical data processing work, which increases the difficulty of the data acquisition and processing [8]. In addition, statistical methods cannot extract the nonlinear features in the data, which can produce large forecasting errors for high-dimensional data. Many ML methods have been applied to the domain of PV power forecasting and have shown superior forecasting performance over the physical and statistical methods [9], such as support vector regression (SVR) [10], decision trees (DTs) [11], extreme learning machines (ELMs) [12], and artificial neural networks (ANNs) [13]. Most of the machine learning models have a wide range of applications in solving small sample nonlinear problems, but they are hardly applicable to scenarios with higher-dimensional data inputs and large datasets. Traditional shallow neural network models are still limited in their ability to express complex nonlinear relationships among data, and they are also constrained in their ability to generalize to complex classification problems [14].

Compared with shallow models, DL emphasizes the depth of the model feature structure and automatically learns the abstract features in the samples through layer-by-layer feature transformation, which is more capable of expressing the complex nonlinear relationships between the data and maximizing the completeness and diversity of the feature information. Deep learning methods mainly include generative adversarial networks (GANs) [15], recurrent neural networks (RNNs) [16], CNNs [17], stacked autoencoders (SAEs) [18], deep belief networks (DBNs) [19], etc.

In recent years, more and more scholars have applied DL methods to PV power forecasting tasks. The most typical of them are CNNs and hybrids of CNN and RNN variants, which exhibit excellent prediction performance. The significant advantage of CNNs over other deep learning models is the use of local connectivity and a shared weight structure, which greatly reduces the number of parameters to be trained and increases the training speed. When large numbers of complex data are involved, the CNN multilayer structure can extract deeper nonlinear features of the data and better explain the abstract features and deep invariant structures inherent in the data [20]. In addition, CNNs do not require a pre-training process, so they are more suitable for real-time applications [21]. These advantages make CNNs an attractive choice among the PV power forecasting methods. Research work [22] proposes a hybrid model based on

wavelet transform (WT) [23] with a CNN applied to a short-term PV power forecasting task. The input PV power sequence is decomposed into several different frequency sequences using WT, and then a separate CNN is used for each frequency sequence to generate the predictions. The results show that a higher forecasting accuracy is achieved than when using the model without WT. Research work [24] proposes an integrated multivariate PV power forecasting model based on VMD-CNN-BiGRU. The PV power is decomposed into several sub-models using variable mode decomposition (VMD) [25] and then combined with other meteorological variables to form several new multivariate sub-models. Then, the CNN_BiGRU network is used for the forecasting of each sub-model. The results show that the model considering the meteorological variables has better performance. Research work [26] applies CNNs to daily wind and solar radiation forecasts, respectively. The inputs to the CNN are taken from a multivariate numerical weather prediction system, and the study shows that the forecasts outperform a support vector machine (SVM). Research work [27] applies two novel CNNs, namely ResNet and DenseNet, to day-ahead PV power forecasting. The method uses historical PV power sequences of different frequencies jointly with other meteorological elements' data as the input to the model. The study performed point forecasting and probabilistic forecasting tasks, and the accuracy and reliability of the proposed forecasting method are demonstrated. Research work [28] compares the day-ahead PV power forecasting results of three models, a CNN, Long Short-Term Memory (LSTM) [29], and CNN_LSTM, on multivariate PV datasets of different lengths. The results show that both the CNN and CNN_LSTM achieved better forecasting results compared to LSTM. In research work [30], two hybrid models (CNN_LSTM and ConvLSTM) are proposed to forecast the power of a PV plant from one day to one week in advance. The forecasts are performed on univariate (containing only historical power) and multivariate (containing multiple weather variables) datasets, respectively. The forecasting results are compared with the LSTM model, and the proposed method obtains more accurate forecasts.

Considering that neural networks are limited by their own computational power and their information processing capabilities [31], attention mechanisms (AMs) have been introduced into neural network structures. In the field of artificial intelligence, AMs have become an important component of neural network structures. AMs allow neural networks to focus on a subset of their input features, thus improving their ability to process information [32]. They have been applied to PV power forecasting by individual scholars in the latest research. Research work [33] combines an AM with ConvLSTM to utilize the AM to adjust the weights of physical a priori features and historical PV data in the input data. The experimental results show that the proposed method significantly improves the accuracy of annual PV generation forecasting. In research work [34], an attention-based long- and short-term spatio-temporal neural network prediction model (ALSM) is proposed using multivariate sequences as the input, which combines a CNN and LSTM and applies an AM to the output part of the whole network, aiming to capture both the short- and long-term temporal patterns, and finally achieves hourly PV power generation prediction during the daytime.

From the above recent studies, it can be found that deep learning models with CNN structures are increasingly used for PV power forecasting tasks. These studies mainly focus on investigating the methods of combining CNNs with other models, with the aim of combining the respective advantages of multiple models to obtain better prediction performance than using a single model. However, this approach of combining other models exploits the advantages of each model while mixing in their respective disadvantages (e.g., LSTM is limited in its ability to extract spatial features [33]). How to design architectures that maximize the advantages of the models based on the actual scenario's requirements can be further investigated. In addition, these studies basically use a single-input model, with a single target sequence as the only input to the model or a multivariate sequence (a combination of target and weather variable sequences) as the only input to the model. The single target series input mode only provides the model with the dependencies between the elements of the target series and does not take into account the effects of other weather

variables on the trend in the target series. The multivariate sequence input model takes into account the effects of the weather variables but lacks targeted extraction of the dependencies between the target sequence elements. How to provide the model with a more comprehensive and fine-grained input mode for the constraint relationships between the target series elements can be further investigated.

1.3. The Research Work in This Paper

Inspired by the above studies, this paper proposes a dual-input CNN model based on a two-head AM incorporating multiple meteorological variables, DI_ACNNs, for ultra-short-term PV power forecasting tasks. Different from previous studies, this method makes full use of the PV system data, innovatively splits the input data into two combinations, and provides the model with input features containing two different levels of information; the dual-branch architecture is used to extract and focus on important feature constraints for different time scales and to improve the upper limit of the model's own learning ability. The overall approach focuses on the deep integration of PV system data with the deep learning model to maximize the diversity and completeness of the extracted features, so as to obtain accurate and comprehensible PV power forecasting results.

The main research work can be summarized as follows:

- The DI_ACNN model is proposed. A deep learning network is designed to improve the forecasting accuracy by innovatively considering three aspects simultaneously: the data source, algorithm operation mechanism, and information processing capability.
- The PV input data innovatively are split into two sets (a historical power dataset and a mixed historical power and meteorological variable dataset), which are deeply fused with the subsequent two sets of dual-scale convolution to provide finer input features for the model.
- A dual-scale convolutional cascade double-head AM structure is innovatively proposed. It fully utilizes the CNN's spatio-temporal cross-feature extraction capability and focuses on the important features while extracting the long-time and short-time dependencies in the sequences, thus enhancing the feature learning capability of the CNN.
- The proposed method is validated using a case study based on three real PV generation datasets with different meteorological conditions in China.

The paper is organized as follows: Section 2 describes the data used and the methodology employed in this study and the predictive performance evaluation index. Section 3 details the multiple experiments conducted in this study and the corresponding rated conclusions. Section 4 summarizes the entire study.

2. Materials and Methods

2.1. Data Collection and Preprocessing

The experimental data in this paper are actual operational data from three PV plants with meteorological differences in China, with a resolution of 15 min and a data length of one year (35,040 sample points). The rated capacities of plants 1, 2, and 3 are 50 MW, 20 MW, and 35 MW. The data contain seven variables, namely active power (kW), total irradiance (W/m^2), normal direct irradiance (W/m^2), horizontal scattered irradiance (W/m^2), temperature ($^{\circ}C$), humidity (%), and air pressure (hPa). To illustrate the meteorological differences between the three PV plants, Figure 1 gives a box plot of all the meteorological variables provided in the data, with the top and bottom part of the box representing the 75% quantile and 25% quantile, which means that the box contains 50% of the data. Thus, the width of the box also represents the degree of volatility of the data over a one-year time horizon. The symbol "×" in Figure 1 represents the annual mean of the corresponding variable. In addition, the maximum and minimum values of the variables, the values of the top and bottom edges of the boxes, and the median line (short horizontal line on the boxes) are indicated in Figure 1, where the median lines of the three irradiances almost coincide with the 25% quantile of the boxes. As can be seen from Figure 1, the total irradiance

and horizontal scattered irradiance of both plant 1 and plant 3 are relatively close to each other and are both higher than those of plant 2. The normal direct irradiance of plant 1 is the lowest, while that of plant 3 is the highest. Plant 3 has the highest temperature and a relatively concentrated temperature range throughout the year. Plant 2 has the lowest temperature. The temperature of plant 1 is between that of plant 2 and plant 3. Plant 2 has the highest air pressure. Plant 3 has the lowest air pressure. Plant 1 has the lowest relative humidity. Plant 2 has the highest relative humidity, much higher than that of plant 1. Plant 3's relative humidity is between plant 1's and plant 2's.

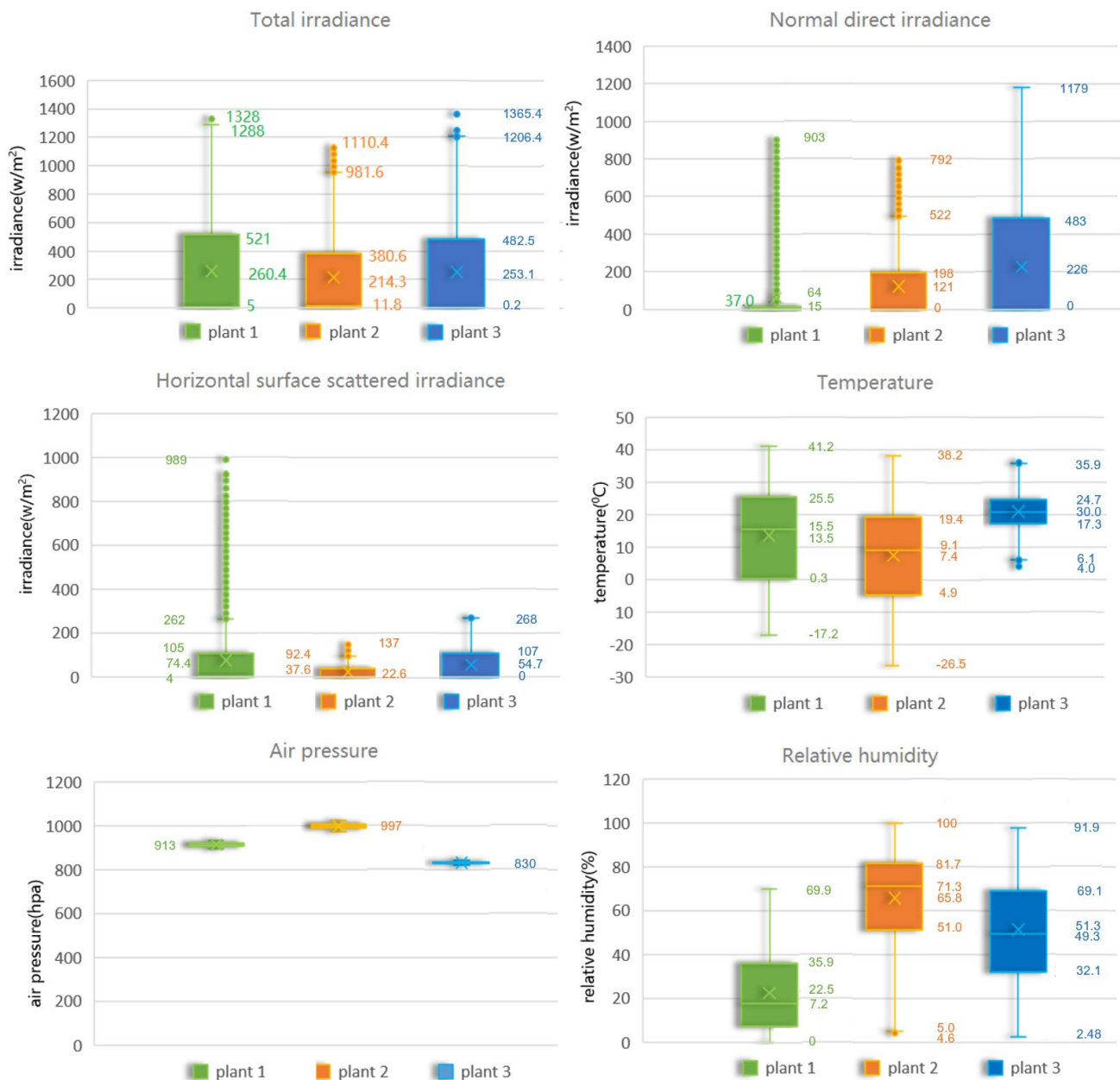


Figure 1. Box diagrams of the meteorological variables.

In this study, we considered all meteorological variables and finally selected five meteorological variables, namely, total irradiance, normal direct irradiance, horizontal scattered irradiance, temperature, and relative humidity. Figure 2 shows the trends of the five meteorological variables and power for five consecutive days on the three plants (All three irradiance levels have been reduced by a factor of 10 for clarity), and it can be seen from this that there is an intuitive connection between the five meteorological variables

selected and the power. These five meteorological variable sequences will be used together with the power sequences as inputs to the multivariate input branch of the proposed model and to the baseline models with multivariate inputs.

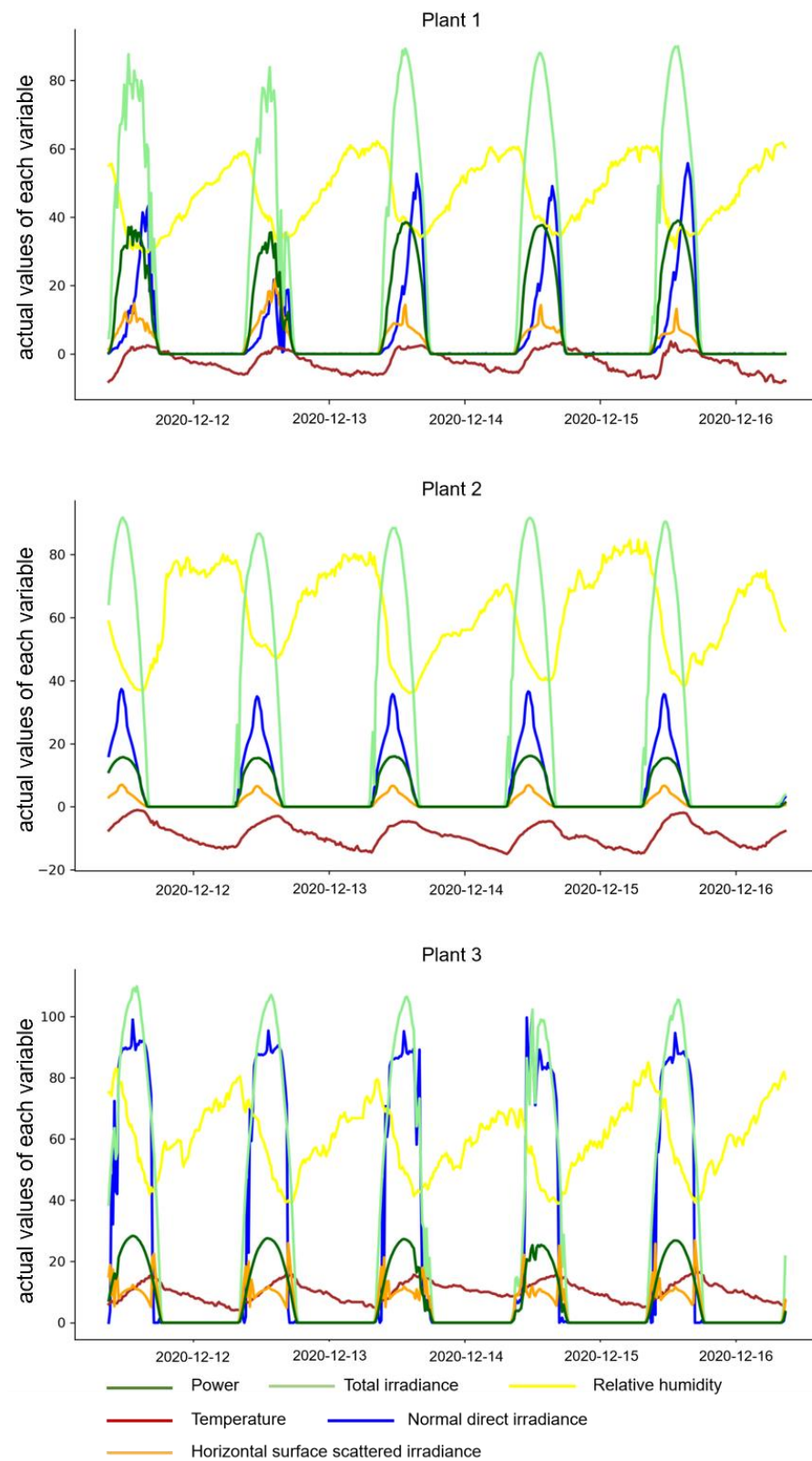


Figure 2. Five meteorological variables and power variations for five consecutive days for the three plants.

Before all PV historical data (PV power and selected meteorological variables) are fed into the forecasting model, basic pre-processing such as outlier identification, vacant value filling and normalization operations need to be performed to reduce the impact on model performance due to data quality. The annual datasets for the three PV plants are divided as shown in Figure 3 [35].

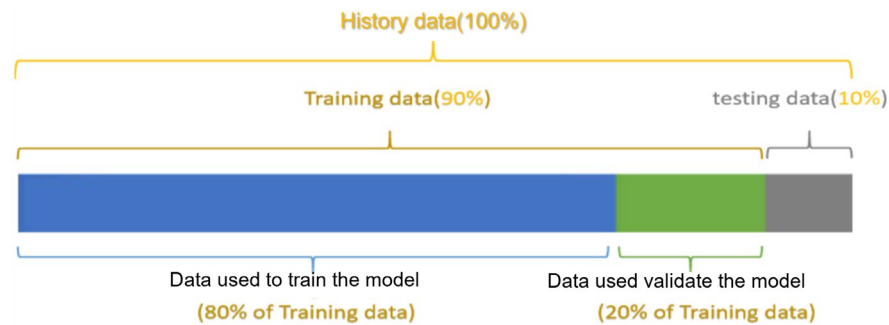


Figure 3. The division of the input history data.

2.2. The One-Dimensional CNN Structure and Feature Extraction Mechanism

A CNN is a multilayer supervised learning neural network with a convolutional structure designed to greatly reduce memory usage, and its three key operations are local perceptual field, weight sharing, and pooling operations [36]. A CNN can be summarized as a three-layer structure, with an input layer, a hidden layer, and an output layer. The convolutional and down-sampling layers of the low hidden layer are important units for realizing the feature extraction function, and the fully connected (FC) layers of the high hidden layer correspond to the hidden layer of a traditional multilayer perceptron and the classifier of logistic regression.

A 1D CNN has 2D input and output data and is mainly used for time series. So, this paper uses a 1D CNN to perform the PV power forecasting task. Figure 4 depicts the feature extraction process of a 1D CNN with one convolutional kernel and one convolutional layer. Multiple kernels and multiple convolutional layers can be obtained by extending on this basis. The input layer is an $n \times m$ matrix. The size of the convolution kernel in the convolution operation part is 2 (2 rows), and the kernel slides along the 2D matrix of the input layer from top to bottom with steps of 1 (stride = 1). Each step is a convolution operation for a local data window, and the symbol “ \otimes ” represents the convolution operation. The above operation is the so-called local perception mechanism. The data window slides, resulting in changing input, but the weights W of the convolution kernel are fixed, i.e., the weight sharing mechanism. After the convolution kernel convolves the input feature map X , a bias term b is added, and then an activation function (sigmoid, tanh, or ReLU, etc.) is passed to obtain the convolution layer feature map Y . The mapping relationship between the input and output of the convolution operation part of the figure is shown in Equation (1).

$$y_i = \sigma(W \otimes X_i + b) \quad (1)$$

where i represents the number of the top-down sliding window in the input matrix, y represents the element of the convolutional layer feature map Y , and σ represents the activation function. The activation function introduces nonlinear factors to enhance the ability of neural networks to fit complex nonlinear functions. The pooling operation part is performed by a pooling kernel (kernel size = 2, stride = 2) to carry out simple operations on the neurons of Y in the window range, such as averaging, maximizing, etc., to obtain the pooled layer feature map Z . The pooling operation reduces the dimensionality of the convolutional layer and extracts more important features, while also effectively controlling overfitting. The FC layer in Figure 4 is used to integrate the distributed features extracted from the previous pooling layer and pass them to the output layer by matching the scale of the model output.

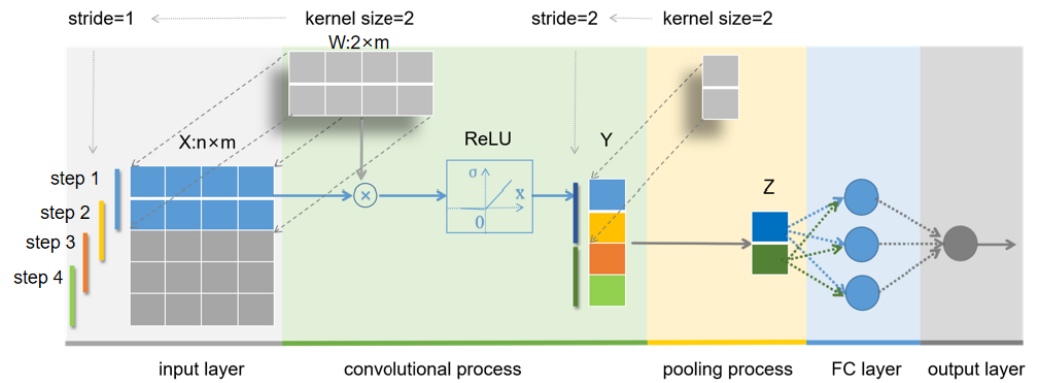


Figure 4. One-dimensional CNN data processing flow.

2.3. The Attention Mechanism

An AM is essentially a resource allocation mechanism that changes the way resources are allocated according to the importance level of the target of attention, so that the resources are tilted more toward the object of attention. In this study, we apply it in the convolutional network layer. In CNNs, the resources to be allocated by the AM are the weight parameters, and more weight parameters are assigned to the objects of attention during the training process of the model so as to filter out useful information from a large amount of information and achieve the purpose of improving the feature extraction ability for the objects of attention.

In this paper, we construct a two-headed attention network, as shown in Figure 5. The two input layer feature vectors are $x = [x_1, x_2, \dots, x_n]$ and $x' = [x'_1, x'_2, \dots, x'_n]$. First, two parallel dense layers (dense layer 1 and dense layer 2, i.e., two heads) are used to assign weights to the corresponding input layer feature vectors x and x' , respectively, to obtain weight vectors $a = [a_1, a_2, \dots, a_n]$ and $a' = [a'_1, a'_2, \dots, a'_n]$, whose elemental expressions are shown in Equations (2) and (3) as:

$$a_i = \sigma(z_i) \quad (2)$$

$$a'_i = \sigma(z'_i) \quad (3)$$

where σ is the activation function ReLU, and z_i and z'_i have the expressions shown in Equations (4) and (5):

$$z_i = \text{Dense}(x_i) \quad (4)$$

$$z'_i = \text{Dense}(x'_i) \quad (5)$$

Then, vectors a and a' are concatenated to obtain a new attention weight vector A , and the original feature map vectors x and x' are concatenated to obtain a new vector X . Finally, vectors X and A are multiplied by elements to obtain the output feature vector C weighted by the AM, whose elemental expression is shown in Equation (6):

$$C_m = A_m \bullet X_m \quad (m = 1, 2, \dots, 2n). \quad (6)$$

So far, adaptive adjustment of the importance of the input features is achieved.

2.4. The Framework and Functions of the Proposed Model

Neural networks mimic the human brain, and the more information is obtained from the outside world, the more this helps the network to make more accurate judgments about things. Following this idea, this study designs the model DI_ACNNs for ultra-short-term PV power forecasting, and the proposed model framework is shown in Figure 6.

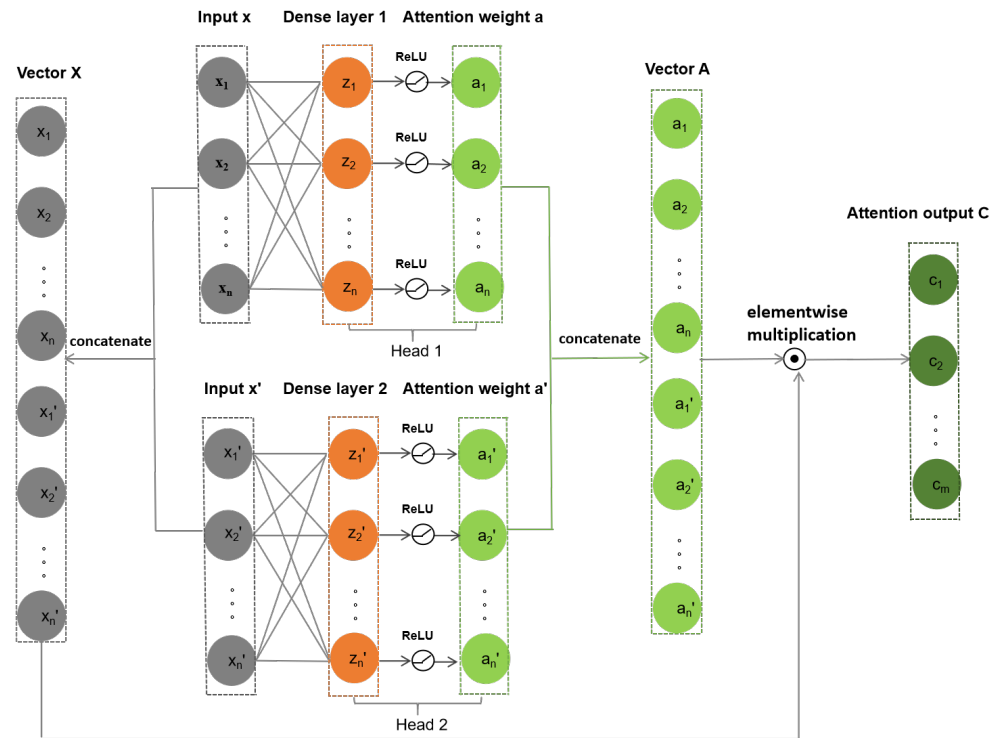


Figure 5. The structure of the attention network.

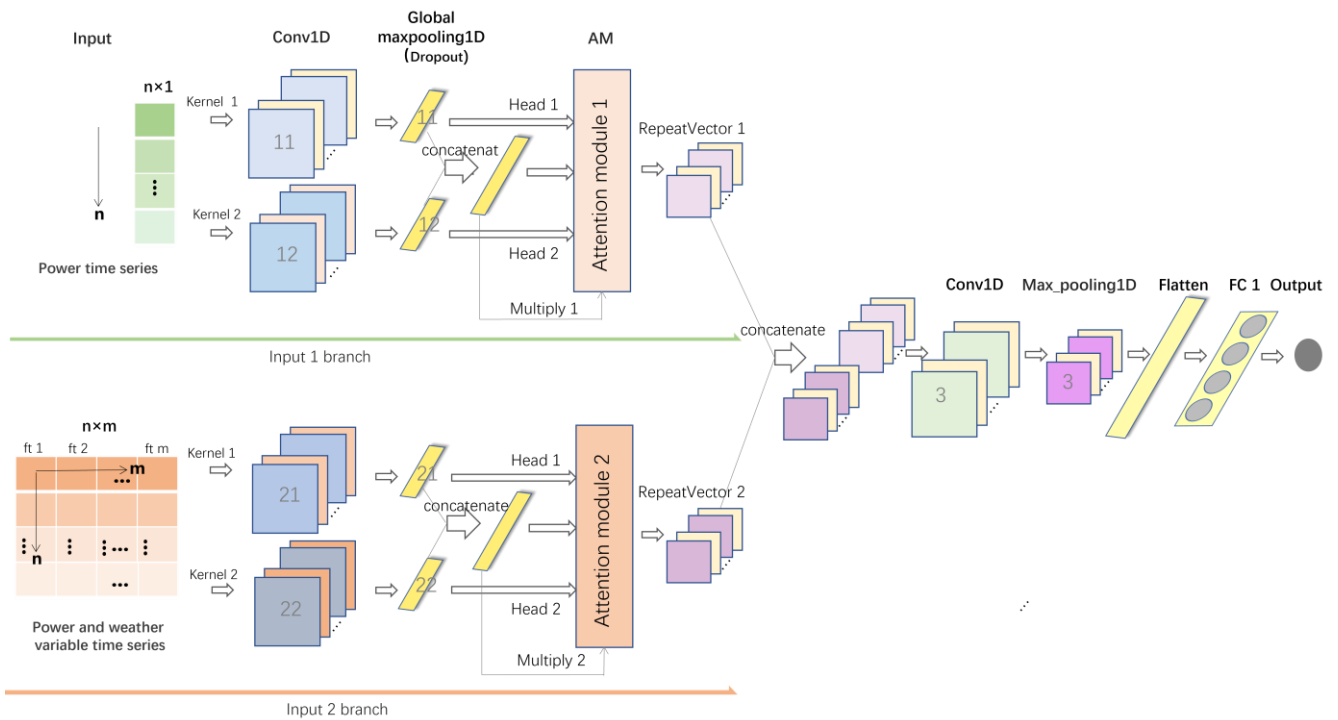


Figure 6. The framework of the proposed model.

The pre-processed data will be used as the input for the proposed model. The model uses a dual-input mode with univariate data as one type of input (an $n \times 1$ -dimensional matrix of historical power series) and multivariate data as the other type of input (an $n \times m$ -dimensional feature matrix containing historical power series), as shown in the matrix on the left side of Figure 6, where n represents the length of the input series and m represents the number of features. Each small rectangle in the two input matrixes represents

a sequence element, and $ft_1, ft_2 \dots, ft_m$ represent the input feature variables, such as power, irradiance, temperature, etc. The multivariate input branch mainly provides the model with information on the temporal and spatial constraints between each feature sequence element. Although the constraint information between the target sequence elements is also included in this feature information, it is not intuitive, which will make the learning process of the subsequent model difficult. Therefore, in this study, a single target sequence input branch is added to provide the model with the constraint information between the target sequence elements intuitively. The dual-input mode deeply fuses the PV data with the model structure to provide the model with more refined input feature information and improve the model's learning efficiency.

Then, a cascade structure of *dual-scale convolution + a two-headed AM* is designed for each of these two inputs, as shown in the area inside the orange and green arrows in Figure 6. The dual-scale convolution layer is set with two convolution kernels of a small and large size (kernel 1 and kernel 2 in Figure 6), the smaller one for extracting the short-time dependencies between the sequence elements and the larger one for extracting long-time dependencies. Then, the convolutional layer feature map is downscaled (2D tensor) using the global maximum pooling method to match the input dimensionality required by the attention module. The two-head attention module is used to focus on the important features of the dual-scale convolutional structure feature map; for example, in attention module 1 in Figure 6, *head 1* focuses on feature map *Global maxpooling1D_11* and *head 2* focuses on feature map *Global maxpooling1D_12*, and the whole module finally implements the adaptive adjustment of the feature importance of the combined feature *concatenate 11*. The proposed cascade structure deeply fuses the input feature data with the feature extraction mechanism of the CNN to extract finer features while putting focus on the important features, thus improving the learning ability of the model.

Then, *RepeatVector* is used as a network adapter to convert the two sets of feature maps after the feature focus into a 3D tensor and concatenate the two to match the data format required by the high hidden layer of the CNN.

Finally, the above 3D feature map is extracted further using the high hidden layer of the CNN to obtain more abstract feature information, and then the FC layer is used to classify and output it to complete the forecasting. This design of a multi-CNN and an AM cooperatively extracting features gives full play to the layer-by-layer feature extraction ability of the CNN and maximizes the diversity and completeness of the feature information extracted by the model.

More specific parameter selection will be described in detail in Section 3.

2.5. Evaluation Indexes

In order to be able to objectively assess the forecasting performance of the proposed model from multiple perspectives, five commonly used evaluation indexes are selected in this study: *MAE* (Mean Absolute Error), *NMAE* (Normalized Mean Absolute Error), *RMSE* (Root Mean Square Error), *NRMSE* (Normalized Root Mean Square Error), and R^2 (Coefficient of Determination).

MAE measures the mean of the forecast residuals and is insensitive to outliers in the data, while *RMSE* measures the variance in the residuals and is sensitive to outliers and penalizes larger errors more. They have the same units as the actual values and can visually express the accuracy of the forecasting results, with smaller *MAE* and *RMSE* values implying the higher accuracy of the forecasting model.

NMAE and *NRMSE* are two evaluation metrics normalized to the *MAE* and *RMSE*, respectively, which are used to evaluate the performance of the model on different data. A smaller value means a better performance of the model on the data.

The above evaluation indexes can assess the accuracy of the forecasting results but cannot judge the degree of fit between the forecasting curve and the actual curve. R^2 quantifies the degree of the fit between the forecasted values and the actual measurements and represents the proportion of the actual measurements that is explained by the forecasting

model, which is a unit-less fraction with a value from 0 to 1. The closer the value is to 1, the better the forecasting curve is fitted to the curve of the actual measurements, and the more explainable the model is. The formula for each evaluation index is as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |p_t - \hat{p}_t| \quad (7)$$

$$NMAE = \frac{MAE}{\bar{p}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (p_t - \hat{p}_t)^2} \quad (9)$$

$$NRMSE = \frac{RMSE}{\bar{p}} \quad (10)$$

$$R^2 = 1 - \frac{\sum_t (p_t - \hat{p}_t)^2}{\sum_t (p_t - \bar{p}_{test})^2} \quad (11)$$

where p_t denotes the measured values of power, \hat{p}_t denotes the predicted values of power, \bar{p} denotes the mean value of the entire historical power data, \bar{p}_{test} is the mean value of the test set of power data, and n represents the number of predicted power sample points.

3. Experiments and Results

3.1. Experimental Settings

The following five experiments are conducted in this paper, each on the three sets of PV data presented above: (1) Input branch ablation experiment: To demonstrate the superiority of the dual-input mode of the proposed model, we first let the historical power input branch and the multivariate input branch perform the forecasting task separately as the input of the whole model and then compare the two sets of results with the forecasting results of the proposed model. (2) Dual-scale convolution ablation experiment: To verify the feature extraction ability of dual-scale convolution, in the proposed model, we keep one of the two scale convolution structures to perform the forecasting task separately and compare their forecasting results with those of the proposed model. (3) AM ablation experiments: To verify the role of the two-headed attention module in improving the model performance, the experimental results of the dual-input multi-CNN model with the attention module removed are compared with the proposed model. (4) To demonstrate the forecasting performance of the proposed model, the forecasting results of the proposed model are compared with the baseline model with traditional input modes. (5) To investigate the effect of different meteorological conditions on the model performance, the forecasting results of the proposed model on three sets of PV data with different meteorological conditions are compared.

The parameters involved in the proposed model are shown in Table 1. The labeling of each layer corresponds to the labeling of each layer in Figure 6. To verify the superiority of the proposed model, we use both the single CNN model in conventional input mode and the CNN_LSTM [28] model in conventional input mode as the baseline models with excellent performance in the field of PV power forecasting. The parameters of the CNN part of CNN_LSTM are the same as those of the single CNN, the number of filters in the convolutional layer is 64, the kernel size is 2, and the maximum pooling layer kernel size is 2. The number of neurons in the LSTM in CNN_LSTM is 128, and the number of neurons in the FC layer is 256. The number of FC neurons in the single CNN is 64. All the other parameters are the same as those of the proposed model. All the above parameters are finalized using the trial-and-error method.

Table 1. The parameters of the proposed model.

| The Layers for DI_ACNNs | Parameters of Each Layer |
|----------------------------|--|
| Conv1D 11 | Filters: 64, kernel size: 2, stride = 1, activation: 'relu', padding: 'same' |
| Global_MaxPooling1D 11 | - |
| Dropout 11 | Rate = 0.1 |
| Conv1D 12 | Filters: 64, kernel size: 16, stride: 1, activation: 'relu', padding: 'same' |
| GlobalMaxPooling1D 12 | - |
| Dropout 12 | Rate = 0.1 |
| Attention module 1 (Head1) | Neurons: 64, activation: 'relu', kernel_regularizer: l2 (0.0001) |
| Attention module 1 (Head2) | Neurons: 64, activation: 'relu', kernel_regularizer: l2 (0.0001) |
| Multiply 1 | - |
| RepeatVector 1 | 8 |
| Conv1D_21 | Filters: 64, kernel size: 2, stride: 1, activation: 'relu', padding: 'same' |
| GlobalMaxPooling1D 22 | - |
| Conv1D 22 | Filters: 64, kernel size: 16, stride: 1, activation: 'relu', padding: 'same' |
| GlobalMaxPooling1D 22 | - |
| Attention module 2 (Head1) | Neurons: 64, activation: 'relu', kernel_regularizer: l2 (0.0001) |
| Attention module 2 (Head2) | Neurons: 64, activation: 'relu', kernel_regularizer: l2 (0.0001) |
| Multiply 2 | - |
| RepeatVector 2 | 8 |
| Concatenate 3 | - |
| Conv1D 3 | Filters: 64, kernel size: 4, stride: 1, activation: 'relu', padding: 'same' |
| MaxPooling1D 3 | Pool kernel size: 2, stride: 2 |
| flatten | - |
| Fully connected | Neurons: 64, activation: 'relu', kernel_regularizer: l2 (0.0001) |
| Output | Power (t + 1) |
| Others | Epochs: 200; EarlyStopping:monitor: 'mse'; Batch-size: 32 (plant1), 32 (plant2), 48 (plant3); patience: 5; min_delta: 0.0001 |

The first column of the table represents the name of each layer in the proposed model, and the serial number labeled after the name represents the specific location of the layer in the model, which corresponds to the location of the layer that has been labeled with the same serial number in Figure 6.

3.2. Results and Discussion

In the branch ablation experiments, the MAE and RMSE of the univariate input mode, multivariate input mode, and the proposed model are compared, and the experimental results are shown in Table 2. *Input 1_CNN* represents a univariate input mode, which has only historical power sequences as input variables, and *Input 2_CNN* represents a multivariate input mode, which has a mixture of historical power series and weather variable series as the input variables. It can be seen that the forecasting accuracy of the proposed model is higher than that of each input branch alone for the three different PV plants. Figure 7a,b show the percentage boosts in the MAE and RMSE of the proposed models relative to the two single-input models, respectively. Relative to the univariate input mode and the multivariate input mode, the percentage boosts in the MAE for the proposed model are 19.9% and 23.91%, and the percentage boosts in the RMSE are 10.4% and 9.54%, respectively, for plant 1; for plant 2, the percentage boosts in the MAE for the proposed model are 21.63% and 31.5%, and the percentage boosts in the RMSE are 14.13% and 12.3%; for plant 3, the percentage boosts in the MAE for the proposed model are 19.34% and 35.74%, and the percentage boosts in the RMSE are 1.5% and 13.58%, respectively. From the above data, it can be seen that the MAE of the proposed model on the data from the three PV plants has improved to a great extent in general compared with each input branch alone. The RMSEs of the proposed models are also both improved. The above results show that the proposed model deeply integrates the PV data with the network architecture through the dual-input mode, which provides richer feature information to the model than the original data input mode and obtains a higher forecasting accuracy.

Table 2. The forecasting results of the proposed model and the two single-input model.

| Model | Plant 1 (50 MW) | | Plant 2 (20 MW) | | Plant 3 (35 MW) | |
|-------------|-----------------|--------------|-----------------|--------------|-----------------|--------------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| DI_ACNNs | 0.668 | 1.782 | 0.250 | 0.699 | 0.392 | 1.063 |
| Input 1_CNN | 0.834 | 1.989 | 0.319 | 0.814 | 0.486 | 1.079 |
| Input 2_CNN | 0.878 | 1.970 | 0.365 | 0.797 | 0.610 | 1.230 |

The bolded values in the table represent the optimal accuracy values.

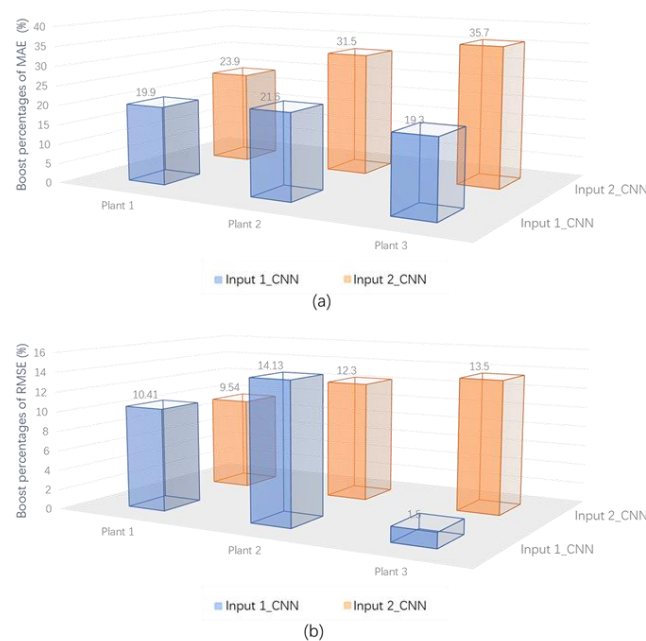


Figure 7. Percentage boosts in MAE (a) and RMSE (b) of the proposed models relative to the two single-input modes in the branch ablation experiments.

In the dual-scale convolution ablation experiment, we separately convolve the input data using the convolutional structures of two scales and then compare their forecasting results with those of the proposed model (dual-scale convolution), and the experimental results are shown in Table 3. *DI_ACNNs (kernel 1)* represents the model with a small-scale (kernel size = 2) convolution structure, and the size of convolution kernel 1 is 2, which is mainly used to extract the short-time dependencies of the adjacent elements of the input sequence; *DI_ACNNs (kernel 2)* represents the model with a large-scale (kernel size = 16) convolution structure, and the size of convolution kernel 2 is 16, which is mainly used to extract the long-time dependencies between the input sequence elements of the input sequence. It can be seen that the forecasting accuracy of the proposed model is higher than that of the model with a single-scale convolutional structure for three different.

Table 3. The forecasting results comparison of the proposed model and the models with single-scale convolutional structure.

| Model | Plant 1 (50 MW) | | Plant 2 (20 MW) | | Plant 3 (35 MW) | |
|--------------------|-----------------|--------------|-----------------|--------------|-----------------|--------------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| DI_ACNNs | 0.668 | 1.782 | 0.250 | 0.699 | 0.392 | 1.063 |
| DI_ACNN (kernel 1) | 0.758 | 1.868 | 0.366 | 0.806 | 0.454 | 1.099 |
| DI_ACNN (kernel 2) | 0.717 | 1.823 | 0.300 | 0.740 | 0.398 | 1.080 |

The bolded values in the table represent the optimal accuracy values.

PV plants. Figure 8a,b show the percentage boosts in the MAE and RMSE of the proposed models relative to the models with a single-scale convolutional structure, respectively. Relative to the model with a small-scale convolution structure and the model with a large-scale convolution structure, the percentage boosts in the MAE for the proposed model are 11.87% and 6.83%, and the percentage boosts in the RMSE are 4.6% and 2.25%, respectively, for plant 1; for plant 2, the percentage boosts in the MAE for the proposed model are 31.69% and 16.67%, and the percentage boosts in the RMSE are 13.28% and 5.5%; for plant 3, the percentage boosts in the MAE for the proposed model are 13.66% and 1.5%, and the percentage boosts in the RMSE are 3.3% and 1.6%, respectively. The above results show that the dual-scale convolutional structure deeply fuses the PV data with the CNN feature extraction mechanism, which extracts both the short-term dependencies and long-term dependencies among the sequence elements, improves the feature extraction capability of the CNN, and enables the whole model to obtain more accurate forecasting results.

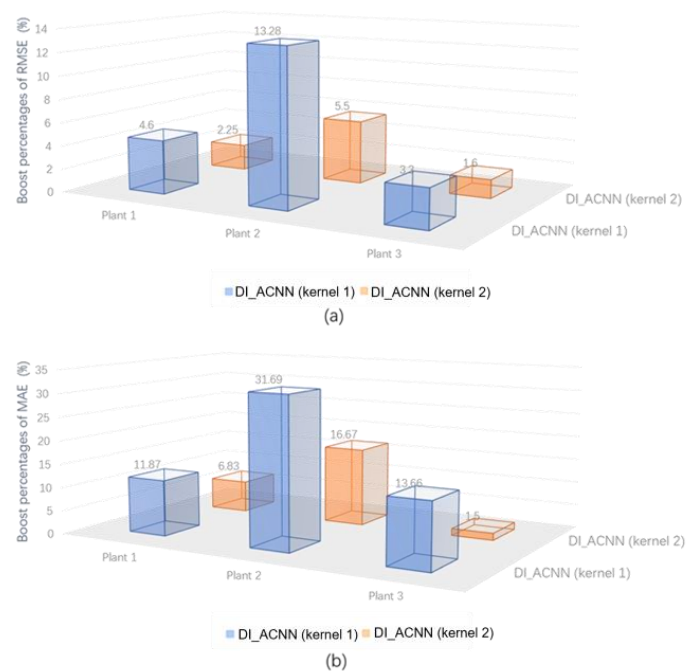


Figure 8. Percentage boost in MAE (b) and RMSE (a) of the proposed models relative to the models with single-scale convolutional structure in the dual-scale convolution ablation experiment.

In the AM ablation experiments, the MAE and RMSE of the model with the two-headed attention module removed (DI_CNNs) and the proposed model are compared, and the experimental results are shown in Table 4, from which it can be seen that the forecasting accuracy of the proposed model is higher than that of the model without the attention structure for three different PV plants. For plant 1, the proposed model has an MAE percentage boost of 8.49%, while the RMSE has a negative percentage boost of 0.1%. For plant 2, the MAE percentage boost of the proposed model is 4.2%, and the RMSE percentage boost is 3.8%; for plant 3, the MAE percentage boost of the proposed model is 3.7%, and the RMSE percentage boost is 1.1%. From the above data, it can be seen that the proposed model has an improved forecasting accuracy for all three PV plant datasets compared to the model with the two-headed attention module removed. Plant 1 has the most significant MAE boost, while the RMSE is basically not improved, but the performance is still improved according to a comprehensive view. In conclusion, the introduction of the two-headed attention module improves the learning capability of the network and enhances the performance of the overall model.

Table 4. Forecasting result comparison of the proposed model and the model with the two-headed attention module removed.

| Model | Plant 1 (50 MW) | | Plant 2 (20 MW) | | Plant 3 (35 MW) | |
|----------|-----------------|--------------|-----------------|--------------|-----------------|--------------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| DI_ACNNs | 0.668 | 1.782 | 0.250 | 0.699 | 0.392 | 1.063 |
| DI_CNNs | 0.730 | 1.780 | 0.261 | 0.706 | 0.407 | 1.075 |

The bolded values in the table represent the optimal accuracy values.

Table 5 shows the MAE, RMSE, and R^2 of the forecasting results for the proposed model and the baseline models with conventional input patterns (CNN and CNN_LSTM) on the data from the three plants, where the prefix “Multivar” represents the multivariate input mode, and the prefix “Univar” represents the univariate input mode, from which it can be seen that the proposed model achieves a higher forecasting accuracy than all the baseline models for all three plants. The R^2 of both the proposed model and the baseline model is above 0.98 on the data of the three plants, which indicates that all the models have a better data fitting ability. The maximum R^2 value of 0.99 is obtained for all the models on plant 3, indicating that all the models fit the data best on these data.

Table 5. The forecasting results comparison of the proposed model and the baseline models.

| Model | Plant 1 (50 MW) | | | Plant 2 (20 MW) | | | Plant 3 (35 MW) | | |
|---------------------|-----------------|--------------|-------|-----------------|--------------|-------|-----------------|--------------|-------------|
| | MAE | RMSE | R^2 | MAE | RMSE | R^2 | MAE | RMSE | R^2 |
| DI_ACNNs | 0.668 | 1.782 | 0.98 | 0.250 | 0.699 | 0.98 | 0.392 | 1.063 | 0.99 |
| Multivar_CLSTM [30] | 0.853 | 1.844 | 0.98 | 0.274 | 0.738 | 0.98 | 0.542 | 1.160 | 0.99 |
| Univar_CLSTM | 0.734 | 1.854 | 0.98 | 0.419 | 0.908 | 0.98 | 0.487 | 1.121 | 0.99 |
| Multivar_CNN [30] | 0.858 | 1.818 | 0.98 | 0.284 | 0.742 | 0.98 | 0.481 | 1.122 | 0.99 |
| Univar_CNN | 0.776 | 1.881 | 0.98 | 0.397 | 0.895 | 0.98 | 0.500 | 1.124 | 0.99 |

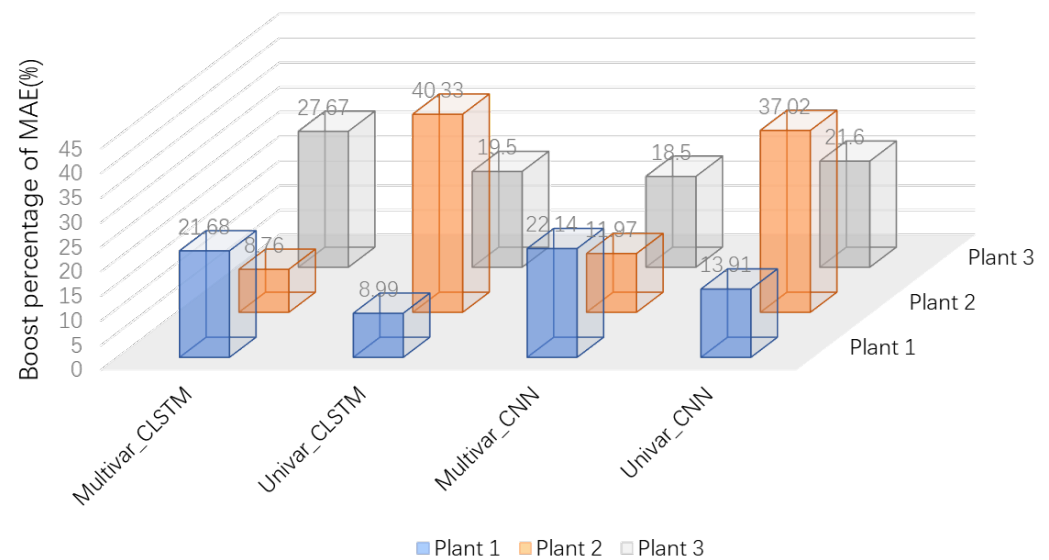
The bolded values in the table represent the optimal accuracy values.

Figure 9a,b show the percentage boosts in the MAE and RMSE of the proposed models relative to each baseline model, respectively. From the figures, it can be seen that the accuracy of the proposed model is improved over each baseline model, with the MAE boost being larger in aggregate. Compared with the baseline model with the multivariate input mode, the proposed model shows the best forecasting performance on plant 3 with a maximum percentage boost for the MAE of 27.67% and a maximum percentage boost for the RMSE of 8.63%. Compared to the baseline model with a univariate input pattern, the proposed model shows the best forecasting performance on plant 2 with a maximum boost in the MAE of 40.33% and a maximum boost in the RMSE of 23.02%. This also shows that the Univar_CLSTM model exhibits different forecasting performances on different data and exhibits the worst forecasting performance on the data from plant 2, indicating the relatively poor applicability of the Univar_CLSTM model.

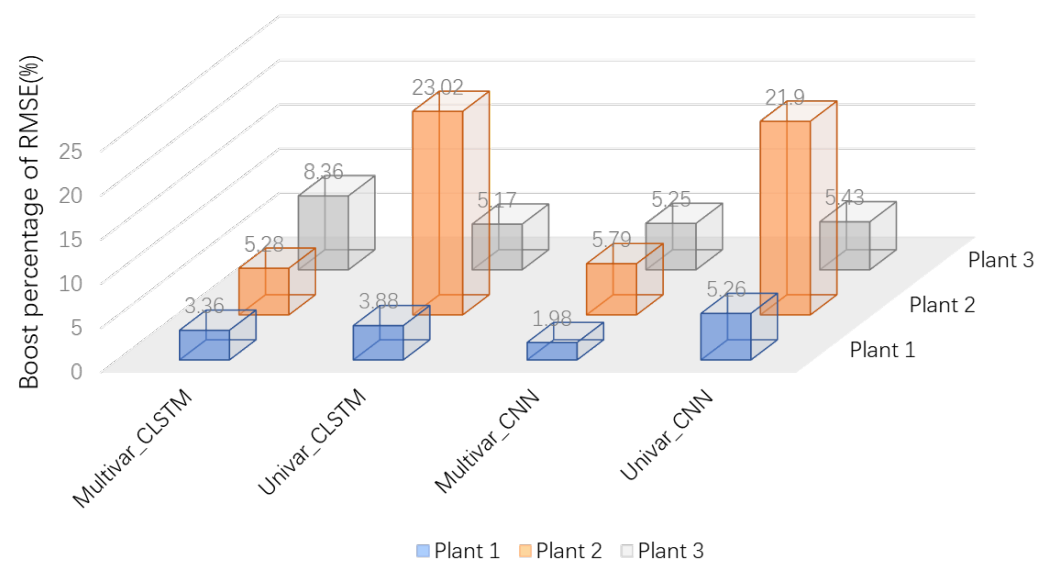
The above RMSE and MAE values are more focused on quantitatively describing the accuracy of the model forecasting. In order to provide a more intuitive description of the ability of the proposed model to fit the actual curve, Figures 10–12 show the forecasting results of the proposed model and the baseline model on the three plants’ data for five consecutive days compared to the actual values. It can be seen that the forecasting results of the proposed model are closer to the actual values for all three plants’ data, indicating that the proposed model has the best ability to fit the data.

Finally, this paper investigates the effect of input data with different meteorological conditions on the proposed model’s forecasting performance. To describe the model’s performance, NMAE and NRMSE are introduced to compare the differences in the model’s performance on different data. Figure 13 shows the NMAE and NRMSE of the proposed model for the three plants. It can be seen that the forecasting performance of the proposed model is the best for plant 3 and the worst for plant 2, and the performance on plant 1 is

very close to that on plant 2 but slightly better than that on plant 2. From the analysis of the meteorological conditions at the three plants in Section 2.1, plants 1 and 3 have a very similar annual mean total irradiance, both higher than plant 2. Plant 3 has a significantly higher annual mean normal direct irradiance than the other two plants, while the annual mean pressure is the lowest of the three plants and the annual mean temperature is the highest of the three plants, with a small fluctuation range. The annual mean total irradiance, horizontal scattered irradiance, and temperature of plant 2 are the lowest among the three plants, and the annual mean air pressure and relative humidity are the highest among the three plants. The differences between the meteorological conditions at plant 2 and the other two plants are still relatively large. In conclusion, the proposed model shows small performance differences in the three plants with different meteorological conditions, demonstrating good applicability.



(a)



(b)

Figure 9. Percentage boost in MAE (a) and RMSE (b) of the proposed models relative to each baseline model in the comparison experiments with the baseline models.

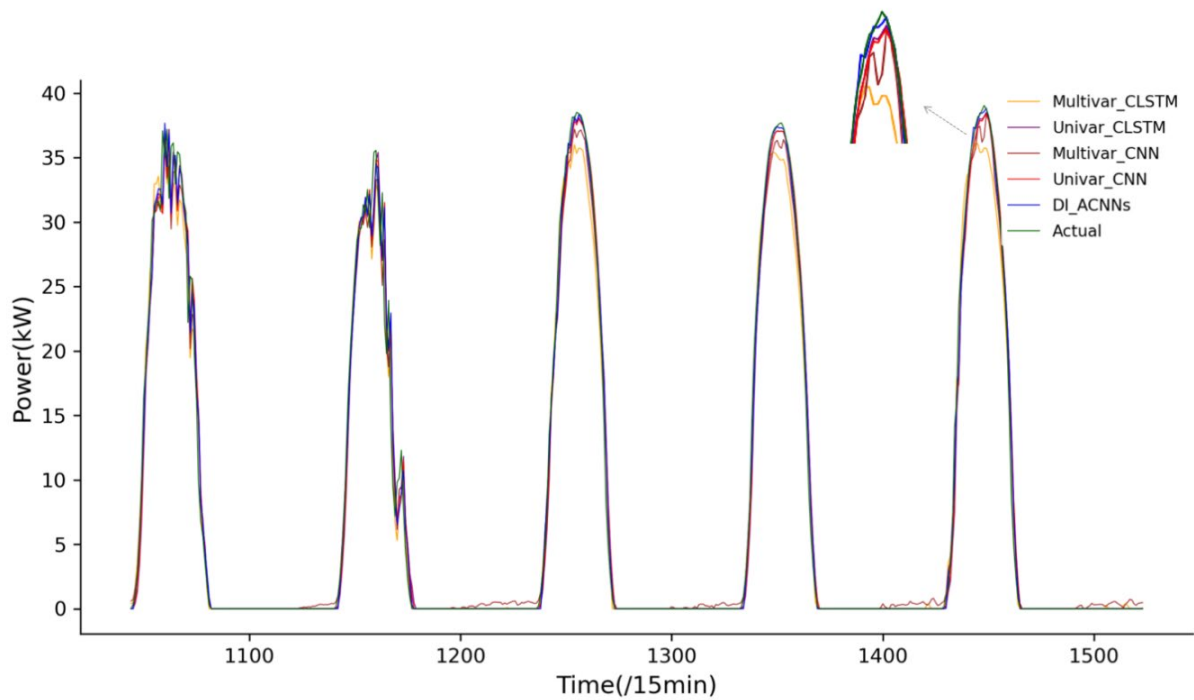


Figure 10. Forecasting results of the proposed model and the baseline models on the plant 1 data for five consecutive days compared to the actual values.

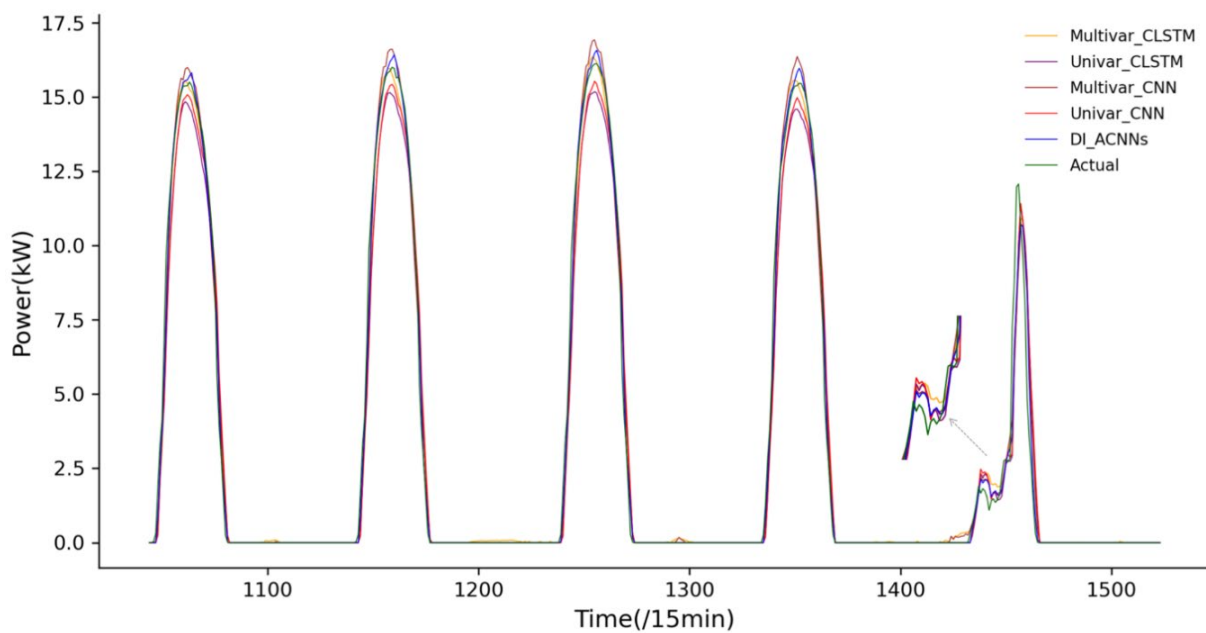


Figure 11. Forecasting results of the proposed model and the baseline models on the plant 2 data for five consecutive days compared to the actual values.

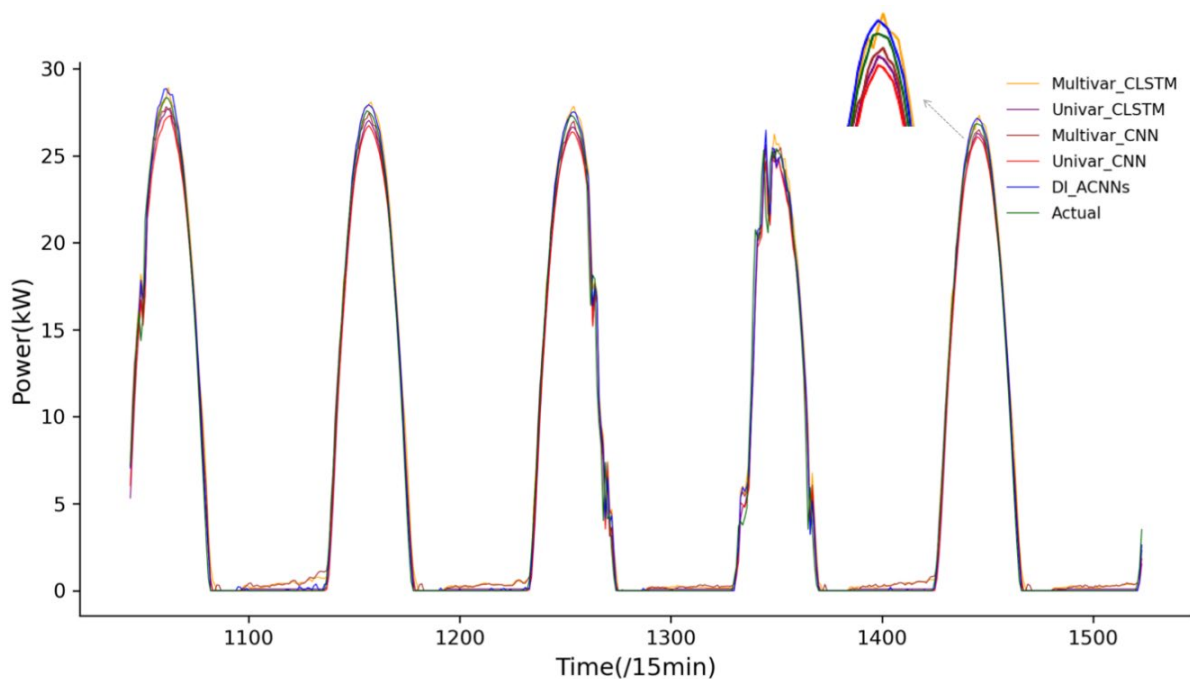


Figure 12. Forecasting results of the proposed model and the baseline models on the plant 3 data for five consecutive days compared to the actual values.

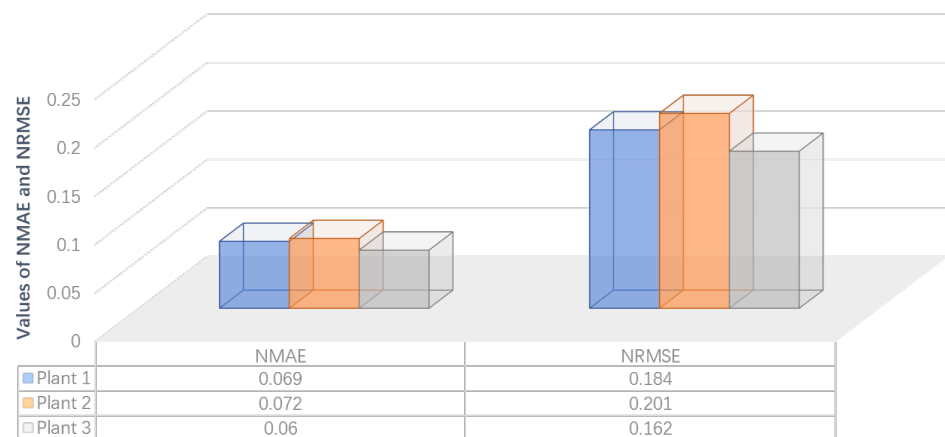


Figure 13. Comparison of NMAE and NRMSE of the proposed model on data from three plants.

4. Conclusions

In this paper, a dual-input CNN model based on a two-head AM incorporating multiple meteorological variables, DI_ACNNs, is proposed for an ultra-short-term PV power forecasting task. Three possible measures to improve the forecasting accuracy are considered innovatively: first, the dual-input pattern of the model is constructed by incorporating meteorological variables and historical electricity, so that the interaction information between the meteorological variables and PV power can be transmitted to the model in a more fine-grained way; then, a two-branch dual-scale convolution structure is designed to deeply integrate the PV data with the feature extraction mechanism of the CNN and to extract the long-term dependency and short-term dependency relationships in the feature sequences, respectively, so as to improve the feature extraction capability of the CNN; finally, a two-head AM is introduced based on the dual-scale convolution structure to focus on the important features in the dual-scale convolution feature map, so as to improve the learning capability of the whole network. These three measures, combined with the design of a multi-CNN parallel cascade hybrid structure, give full play to the

advantages of the CNN algorithm and the layer-by-layer feature extraction capability, so that the model is able to extract finer, more diversified, and deeper nonlinear features and exhibits excellent prediction performance. The specific findings are as follows:

- To demonstrate the effectiveness of the dual-input pattern proposed in this paper, input branch ablation experiments were conducted on data from three plants. The experimental results show that the proposed model in this paper obtains a higher forecasting accuracy compared with the traditional single-input mode, obtaining a maximum MAE percentage boost of 35.74% for plant 3 and a maximum RMSE percentage boost of 14.3% for plant 2.
- To verify the feature extraction capability of the proposed dual-scale convolutional structure, a dual-scale convolutional structure ablation experiment was conducted. The experimental results show that the forecasting accuracy of the proposed model is higher than that of the model with a single-scale convolutional structure on the data from all three plants. The maximum percentage boost in the MAE and the maximum percentage boost in the RMSE are obtained for plant 2, which are 31.69% and 16.67%, respectively.
- To verify the enhancement of the network's learning ability by the two-headed AM proposed in this paper, we conducted an AM ablation experiment. The results show that the forecasting accuracy of the proposed model is higher than that of the model without the attention structure on the data of the three plants. A maximum MAE percentage boost of 8.49% is obtained on plant 1, and a maximum RMSE percentage boost of 3.8% is obtained on plant 2.
- To demonstrate the superiority of the performance of the proposed model, the forecasting results were compared with the baseline models (CNN and CNN_LSTM) with traditional input patterns. The experimental results show that the forecasting accuracy of the proposed model is higher than that of the respective baseline models. The maximum percentage boost in the MAE and the maximum percentage boost in the RMSE are both obtained on plant 2 with 40.33% and 23.02%, respectively.
- The effect of PV plant data with different meteorological conditions on the performance of the proposed model was investigated. The experimental results show that the proposed model has a slight difference in performance on the data of the three plants, showing good applicability.

In this study, in order to give full play to the layer-by-layer feature extraction capability of the CNN while considering deep fusion of the PV data and model architecture, as well as the fusion of the data and CNN feature extraction mechanism, a dual-head AM is introduced to enhance the network learning capability, and a dual-scale parallel cascaded CNN model architecture is integrated and designed to obtain excellent forecasting performance. Practice shows that the proposed method performs well on the ultra-short-term PV power forecasting task. Considering the limitation of CNNs in long-term dependent feature extraction, in the subsequent study, we will consider the impact of different time-scale data resolutions on the model performance and try to group the input data according to different time scale resolutions to construct a model containing more time-scale feature information, with a view to improving the long-term feature extraction capability of the model and applying it to PV power forecasting tasks for longer time scales (short-term and medium- to long-term) so that it can provide reliable technical support for the sustainable development of PV energy.

Author Contributions: Conceptualization, X.R. and Y.L.; data curation, X.R.; formal analysis, X.R., F.Z. and Y.L.; funding acquisition, X.R. and Y.L.; investigation, X.R., F.Z. and Y.L.; methodology, X.R. and F.Z.; project administration, Y.L.; software, X.R. and J.Y.; supervision, X.R., F.Z. and J.Y.; validation, X.R., F.Z., J.Y. and Y.L.; visualization, X.R.; writing—original draft, X.R.; writing—review and editing, X.R., F.Z., J.Y. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, No. 2019YFE0104800 and Inner Mongolia Autonomous Region Key R&D and Achievement Transformation Program Project, No. 2022YFSJ0033. The APC was funded by the Inner Mongolia Autonomous Region Key R&D and Achievement Transformation Program Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The PV site operators require data confidentiality.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Houran, M.A.; Bukhari, S.M.S.; Zafar, M.H.; Mansoor, M.; Chen, W. COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. *Appl. Energy* **2023**, *349*, 121638. [CrossRef]
- International Renewable Energy Agency. Renewable Energy Capacity Statistics. 2022. Available online: <https://www.irena.org/publications/2022/Apr/Renewable-Capacity-Statistics-2022> (accessed on 5 February 2024).
- Li, G.; Ding, C.; Zhao, N.; Wei, J.; Guo, Y.; Meng, C.; Huang, K.; Zhu, R. Research on a novel photovoltaic power forecasting model based on parallel long and short-term time series network. *Energy* **2024**, *293*, 130621. [CrossRef]
- Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144. [CrossRef]
- Jiang, J.; Lv, Q.; Gao, X. The Ultra-Short-Term Forecasting of Global Horizontal Irradiance Based on Total Sky Images. *Remote Sens.* **2020**, *12*, 3671. [CrossRef]
- Alonso-Montesinos, J.; Batlles, F. Solar radiation forecasting in the short- and medium-term under all sky conditions. *Energy* **2015**, *83*, 387–393. [CrossRef]
- David, M.; Ramahatana, F.; Trombe, P.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Sol. Energy* **2016**, *133*, 55–72. [CrossRef]
- Dolara, A.; Leva, S.; Manzoloni, G. Comparison of different physical models for PV power output prediction. *Sol. Energy* **2015**, *119*, 83–99. [CrossRef]
- Wang, H.; Liu, Y.; Zhou, B.; Li, C.; Cao, G.; Voropai, N.; Barakhtenko, E. Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manag.* **2020**, *214*, 112909. [CrossRef]
- Jebli, I.; Belouadha, F.-Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by pearson correlation using machine learning. *Energy* **2021**, *224*, 120109. [CrossRef]
- Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* **2018**, *8*, 689. [CrossRef]
- Zhou, Y.; Zhou, N.; Gong, L.; Jiang, M. Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine. *Energy* **2020**, *204*, 117894. [CrossRef]
- Ma, H.; Zhang, C.; Peng, T.; Nazir, M.S.; Li, Y. An integrated framework of gated recurrent unit based on improved sine cosine algorithm for photovoltaic power forecasting. *Energy* **2022**, *256*, 124650. [CrossRef]
- Peng, X.; Deng, D.; Cheng, S.; Zhan, J.; Huang, J.; Niu, L. Study of the key technologies of electric power big data and its application prospects in smart grid. *Chin. Soc. Electr. Eng.* **2015**, *35*, 503–511. [CrossRef]
- Huang, X.; Li, Q.; Tai, Y.; Chen, Z.; Liu, J.; Shi, J.; Liu, W. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. *Energy* **2022**, *246*, 123403. [CrossRef]
- Vu, B.H.; Chung, I.-Y. Optimal generation scheduling and operating reserve management for PV generation using RNN-based forecasting models for stand-alone microgrids. *Renew. Energy* **2022**, *195*, 1137–1154. [CrossRef]
- Ren, X.; Zhang, F.; Zhu, H.; Liu, Y. Quad-kernel deep convolutional neural network for intra-hour photo-voltaic power forecasting. *Appl. Energy* **2022**, *323*, 119682. [CrossRef]
- Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2014**, arXiv:1312.6114. [CrossRef]
- Zhang, J.; Ling, C.; Li, S. EMG Signals based Human Action Recognition via Deep Belief Networks. *IFAC Pap. Online* **2019**, *52*, 271–276. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the International Conference on Computer Vision, Las Condes, Chile, 11–18 December 2015; pp. 1026–1034.
- Obiora, C.N.; Hasan, A.N.; Ali, A. Predicting Solar Irradiance at Several Time Horizons Using Machine Learning Algorithms. *Sustainability* **2023**, *15*, 8927. [CrossRef]
- Li, P.; Zhou, K.; Lu, X.; Yang, S. A hybrid deep learning model for short-term PV power forecasting. *Appl. Energy* **2020**, *259*, 114216. [CrossRef]
- Gao, M.; Lia, J.; Honga, F.; Longb, D. Day-ahead power forecasting in a largescale photovoltaic plant based on weather classification using LSTM. *Energy* **2019**, *187*, 115838. [CrossRef]
- Zhang, C.; Peng, T.; Nazir, M.S. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. *Electr. Power Syst. Res.* **2022**, *213*, 108796. [CrossRef]

25. Wang, R.; Li, C.; Fu, W.; Tang, G. Deep Learning Method Based on Gated Recurrent Unit and Variational Mode Decomposition for Short-Term Wind Power Interval Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 3814–3827. [[CrossRef](#)] [[PubMed](#)]
26. Díaz-Vico, D.; Torres-Barrán, A.; Omari, A.; Dorransoro, J.R. Deep Neural Networks for Wind and Solar Energy Prediction. *Neural Process. Lett.* **2017**, *46*, 829–844. [[CrossRef](#)]
27. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.W.; Wei, Z.; Sun, G. Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105790. [[CrossRef](#)]
28. Wang, K.; Qi, X.; Liu, H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* **2019**, *251*, 113315. [[CrossRef](#)]
29. Gu, B.; Li, X.; Xu, F.; Yang, X.; Wang, F.; Wang, P. Forecasting and Uncertainty Analysis of Day-Ahead Photovoltaic Power Based on WT-CNN-BiLSTM-AM-GMM. *Sustainability* **2023**, *15*, 6538. [[CrossRef](#)]
30. Agga, A.; Abbou, A.; Labbadi, M.; El Houm, Y. Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. *Renew. Energy* **2021**, *177*, 101–112. [[CrossRef](#)]
31. Kim, T.-Y.; Cho, S.-B. Web traffic anomaly detection using C-LSTM neural networks. *Expert Syst. Appl.* **2018**, *106*, 66–76. [[CrossRef](#)]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, Ł.; Polosukhin, L. Attention is all you need. *Neural Inf. Syst.* **2017**, *30*, 5998–6008. [[CrossRef](#)]
33. Bai, M.; Chen, Y.; Zhao, X.; Liu, J.; Yu, D. Deep attention ConvLSTM-based adaptive fusion of clear-sky physical prior knowledge and multivariable historical information for probabilistic prediction of photovoltaic power. *Expert Syst. Appl.* **2022**, *202*, 117335. [[CrossRef](#)]
34. Qu, J.; Qian, Z.; Pei, Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* **2021**, *232*, 120996. [[CrossRef](#)]
35. Oh, S.L.; Ng, E.Y.; San Tan, R.; Acharya, U.R. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Biol. Med.* **2018**, *102*, 278–287. [[CrossRef](#)] [[PubMed](#)]
36. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.