e-Prime - Advances in Electrical Engineering, Electronics and Energy 6 (2023) 100315

Contents lists available at ScienceDirect

e-Prime - Advances in Electrical Engineering, Electronics and Energy

journal homepage: www.elsevier.com/locate/prime

Novel solar forecasting scheme modelled by mixer dual path network and based on sky images

Tongsen Zhu [a], Xuan Jiao [1,b], Xingshuo Li [2,*,a], Xuening Yin [a], Yang Du [2,c], Shuye Ding [2,a], Weidong Xiao [3,b]

[a] Nanjing Normal University, Nanjing, China
[b] The University of Sydney, Sydney, Australia
[c] James Cook University, Cairns, Australia

ARTICLE INFO

Keywords:
Deep learning (DL)
Solar forecasting
Dual path network (DPN)
Convmixer architecture

ABSTRACT

The prediction of global horizontal irradiance has become an effective technique to address the intermittence issue of photovoltaic (PV) power generation. This article proposes a novel deep neural network(DNN), named Mixer Dual Path Network (Mixer-DPN), for promising solar forecasting. It shares common features of cloud images and maintains the flexibility to explore new features through dual-path architecture by combining the Mixer layer and Dual Path Network. Therefore, the proposed model can provide more accurate prediction results compared to the classical DNN-based predictors. Moreover, the proposed model shows a faster convergence speed and smaller model size, which makes it suitable for a practical global horizontal irradiance. The merits of the proposed model are verified by testing it with the data from National Renewable Energy Laboratory comparing it with other DNN-based prediction models. Studies have shown that the new model has achieved excellent results in MSE, MAE and other indicators, and the R2 prediction accuracy rate has increased by 14% compared with the baseline model.

1. Introduction

The increasing penetration of solar power generation is challenging grid stability due to the intermittence, volatility, and uncertainty [1]. One solution is the bulk energy storage to mitigate the intermittent power, which is a costly solution due to the limit of rechargeable battery technologies. Thus, the latest studies focus on solar forecasting to enhance stable operation of power systems without significant expense on energy storage systems, which can bring long-term economic benefits to the development of photovoltaic power generation.

Photovoltaic prediction techniques can be classified according to the training data or the DNN's architecture. Most of the current research forecasts are based on the environmental data [2,3].The article [4] used 5 min power data gathered from 11 utility-scale PV plants in Texas, with ratings from 5 to 100 MW. The dataset include two-plus years from November 2016 to December 2018. The article [5] uses three hourly irradiance datasets measured using meteorological ground stations over a 6-year period from the U.S. NERL web site and a 6 year minute-by-minute irradiance dataset from the Baseline Surface Radiation Network (BSRN). PV power data on the Alice Springs photovoltaic power system of DKASC and averaged hourly values for the original 5 min resolution data were used by Qu et al. [6].

The training can be sourced from the spatio-temporal numerical data and image data. The spatio-temporal numerical data is usually sampled by a dense net composed of many solar irradiation sensors, so these forecasting schemes can infer the situation of the whole PV plant and support more flexible control methods [7,8]. The paper [9] proposes a novel deep learning framework, called PV-Net, for short-term forecasting of photovoltaic (PV) energy production1. The results show that PV-Net achieves superior prediction accuracy and consistency in terms of four performance measure. The paper [10] proposes a deep learning based hybrid method consists of a deep time-series clustering (DTC) and a feature attention based deep forecasting (FADF)2. The DTC groups the solar irradiance data into multiple clusters based on the latent features.

* Corresponding author.
E-mail address: xingshuo.li@njnu.edu.cn (X. Li).
[1] Student Member, IEEE
[2] Member, IEEE
[3] Senior Member, IEEE

https://doi.org/10.1016/j.prime.2023.100315
Received 14 May 2023; Received in revised form 25 September 2023; Accepted 9 October 2023
Available online 15 October 2023

The FADF assigns different importance to different features using a feature attention sub-network and GRU network to forecast global horizontal irradiance (GHI) The method reduces the root mean square error by 11.88% and 12.65% for its datasets, respectively. Due to the expensive sensor network, It is only suitable for large-scale photovoltaic power stations and distributed grid system with high-proportion solar power. The scattered detection devices can also bring additional issues in specific engineering applications, which makes it even more challenging to design solar forecasting models.

Recently, the fast development of computer vision technology inspires researchers to put more effort into predictors trained by image data. Various forecasting schemes have been proposed based on the cloud images taken by either satellite or ground-based sky cameras [11, 12]. A hybrid solar forecasting method using satellite visible images and modified convolutional neural networks (CNNs) proposed to predict the global horizontal irradiance has been demonstrated that the proposed method provides more precise forecasting results [13]. Unsupervised clustering (UC)-based short-term solar forecasting methods, whose predictions are done by a two-layer machine learning framework based on multiple models (M3), have been shown to outperform non-UC models by about 20% for UC-based models, and also outperform single-algorithm machine learning models by about 20% for M3-based models [14]. The satellite images covering a larger area is usually more suitable for regional PV plant forecasting. The ground-based sky camera can provide high-resolution local pictures, which can be applied to central PV plants.

The ability to handle big data and complex non-linear correlations makes deep learning neural networks the most promising approach to PV forecasting. Though various DNNs have been applied to solar prediction, [12] concludes that the CNN (Convolutional Neural Network) model has higher accuracy and smaller mapping errors range than LSTM (Long Short-Term Memory) mapping in general. The paper also proposes a novel method to eliminate the negative influences brought by the solar zenith angle while retaining useful information, and the modified CNN was used to extract the cloud cover factors in visible satellite images.In [14], a multi-model framework based on unsupervised clustering is developed to perform short-term global horizontal irradiance forecasting and also achieves preliminary results. Additionally, [15] trains deep CNNs to automatically learn the relationship between the historical irradiation and different cloud patterns in the image instead of using physical models or explicitly extracting features, which makes great progress in applying deep learning.

Nevertheless, most of the current research applies the existing DNN models and few of them improve the model architecture based on the solar data's unique features. Since the previous photovoltaic prediction model did not combine with the new feature extraction technology, previous research usually requires a significant amount of training data and training episodes for acceptable prediction results. This problem caused them to need more data collection and computer power.

Therefore, this paper proposes a novel solar prediction model that can accurately predict the global horizontal irradiation(GHI) based on the input cloud image data. The proposed predictor organically combine novel dual-path networks technology and photovoltaic prediction. The dual-path networks can share common features while maintaining the flexibility to explore new features through the dual-path architectures and model the cloud-irradiation relationships based on the hybrid feature extraction. Moreover, the proposed model adopts the convMixer layers to separate the spatial and channel dimensions, which significantly improves the feature's diversity and thus makes the prediction results outperform the traditional CNN network. The convMixer has also been proved to benefit the residual calculations. The proposed model is tested by the image data provided by National Renewable Energy Laboratory (NREL) and compared with other popular DNN-based prediction models.

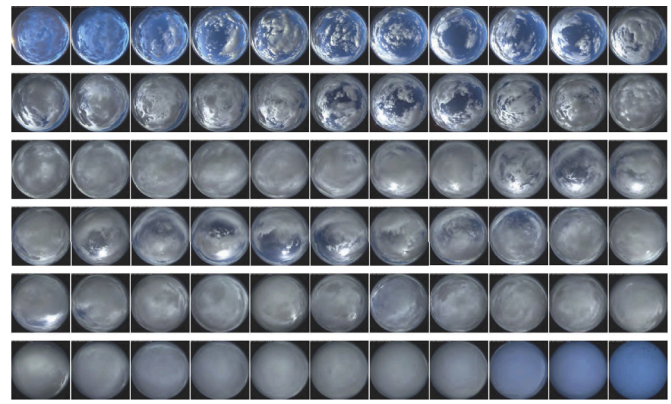

Fig. 1. Cloud map data on March 1st, 2020 [16].

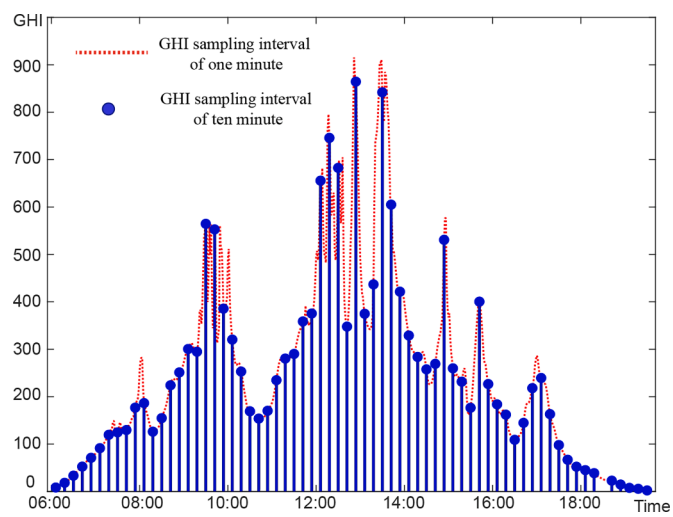

Fig. 2. GHI data on March 1st, 2020.

2. Data description

The data comes from a renewable resource climatology of NREL (National Renewable Energy Laboratory), at Golden, CO, USA. It can provide diversely meteorological data [16], such as wind speed, temperature, etc. Specifically, a Solar Light's Model 501 Series Radiometer (501A UVA) is used to capture images with the cloud cover values recorded every 10 min. The 501A UVA wavelength ranges from 315 nm in W/m^2 and the temperature is controlled at 25 °C.

In this paper, The proposed DNN-based preditor can automatically process the information of cloud and sun, and directly map the irradiation at the corresponding time while bypassing the intermediate link. The images available from January 2020 to December 2020 are used for training and validation. The images available in July 2021 are used for testing. As shown in Fig. 1, the sky imaging instrument takes pictures every 10 min, which mainly shows the distribution of clouds and the sun's position. The forecasting scale of the GHI is 0 to 1200. It is worth noting that the sun, which in old weather analysis and processing systems was often identified as a large cloud, can now be plotted by maximum processing. In this paper, DNN is used to automatically process the information of cloud and sun, and directly map the irradiation at the corresponding time, bypassing the intermediate link,such as identifying cloud thickness and cloud moving speed. Because computers do not need to think according to human rationality in deep learning, which will greatly reduce the unnecessary computational power and design costs.

Fig. 2 illustrates the data of GHI on the test day. The red dash line

Table 1
Data related information.

Parameter	Value
Total sky imager	$TSI-16$
Image format	RGB
Total solar radiation sensor	$CMP22$
The range of GHI	$0-1000W/m^2$
Number of training set pictures	20044
Number of testing set pictures	5012

indicates the solar irradiance variation recorded every minute. The blue spot shows the corresponding GHI for the cloud image taken per 10 min.

Therefore, the dataset $D = \{(x_{t0}, y_{t0})\ldots,(x_{ti}, y_{ti})\ldots,(x_{tn}, y_{tn})\}$ is constructed by mapping the images with the GHI values based on their time stamps, where the subscript ti is the i'th time stamp when its maximum is tn, x_{ti} and y_{ti} refer to the input cloud image and output GHI, respectively. The image data are on a shape of W × H × C representing the width, height, and the number of color channels. Before loading the images into the model, the original 1500 × 1500 × 3 images are converted to 224 × 224 × 3 to reduce computational burden and speed up convergence. Moreover, all the image data will be normalized before fed to the model to avoid gradient explosion and vanishment issues. More details of the training data are summarised in Table 1.

3. Methodology for solar forecasting model based proposed deep learning

The quality of the prediction model largely determines the quality of the prediction effect. Relatively, the more complex the structure, the better the prediction effect. However, too complex models will slow down the calculation speed. Therefore, the PV prediction model generally uses a relatively simple AI model to avoid the model being too large. In order to balance the prediction speed and the prediction effect, we use a dual-path structure to improve the complexity of the model while minimizing the calculation time. The wall clock time from data acquisition to output prediction is 11 s on the computer platform equipped with the GPU of RTX 2080ti and the CPU of Inter 10700KF. In addition, for further reducing the size of the model, we make improvements on the original dual-channel structure, transforming the traditional convolution layer into a mixed convolution layer.

3.1. Deep convolutional neural netorks

Deep CNNs outweigh other shallow models in feature extraction, image analysis and classification. Therefore, the proposed DNN forecasting model consists of many CNN layers to mine the information from the input cloud images.

Moreover, the CNN layers comprise convolutional layers, activation layers, and pooling layers. The convolution layer has a receptive field, which can capture the local and detailed information of the image. Thus, each pixel of the output image only gets the result of a small range of values from the input image. The receptive field is increased with the layers to capture more complex and abstract information from the image. After the operation of multiple convolution layers, the abstract representations of the image at different scales are finally extracted. Besides, the activation layer can introduce nonlinear factors into neurons so that the DNN can approach any nonlinear function arbitrarily. Thus, DNNs can be applied to fitting many nonlinear models.

The pooling layer can replace a certain image area with a value, such as the maximum for maximum pooling or the average value for mean pooling. In addition to reducing the image size, other advantages of down sampling include translation and rotation in variance because the output value calculated from a region of the image is not sensitive to translation and rotation.

Finally, a fully connected layer, known as a dense layer in CNN, has to be cascaded to transfer the learned distributed feature representation into one specific value. Thus, regression calculation can be done for solar irradiance forecasting. Though the classical deep CNNs can make predictions, the accuracy of the results is relatively lower without additional optimization.

3.2. Optimized prediction model with MixerDPN architecture

The deep residual network (ResNet) refers that residual unit is added through the short-circuit mechanism. ResNet can directly uses the convolution of stripe= 2 for down sampling and replace the full connection layer with the global average pool layer. Moreover, the number of feature maps can be doubled in ResNet to maintain the complexity of the network layer when the feature map size is reduced by half. Another main advantage of ResNet is that the gradient can flow through the identity function to reach the front layer. It is also the first deep network to apply a hop connection successfully, and each residual

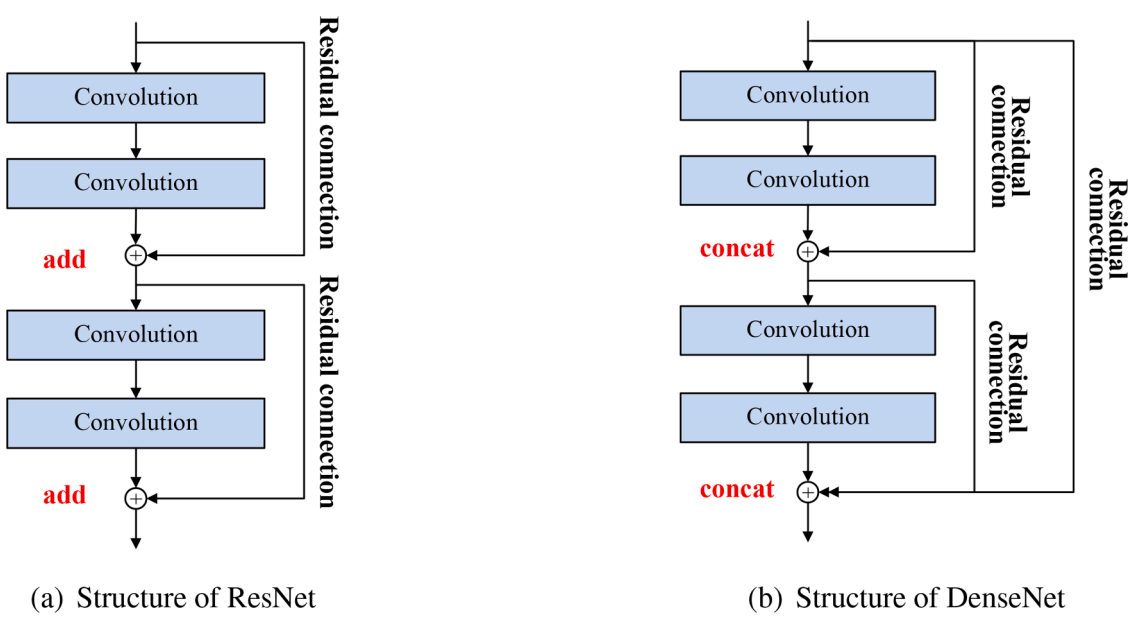

Fig. 3. Differences between ResNet and DenseNet.

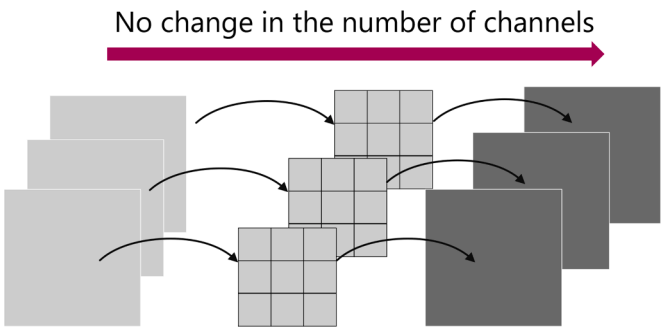

Fig. 4. Depthwise Convolution.

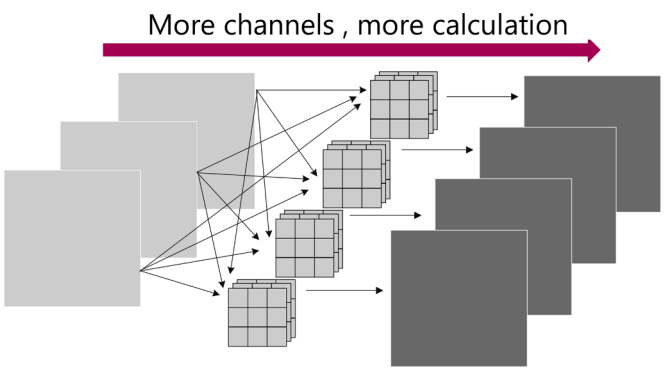

Fig. 5. Pointwise Convolution.

unit is linked with a hop connection. The residual path adds the input features and the output of the block element by element to obtain the final output. However, the superposition of identity mapping and nonlinear transformation output is additive, which destroys the information flow in the network to a certain extent.

The densely connected network (DenseNet) emerges as an improved design based on ResNet. Compared to ResNet, DenseNet proposed a more radical dense join mechanism. In a DenseNet, all the layers are connected to each other. Specifically, each layer will accept all the previous layers as its additional input. Fig. 3(a) shows the connection mechanism of the ResNet network. As a comparison, Fig. 3(b) shows the dense connection mechanism of DenseNet. It can be seen that ResNet is a short circuit connection between the current layer and some previous layers (generally 2-3 layers), and the connection is through element level addition. In comparison, DenseNet will accept all the previously extracted representations.

DenseNet and ResNet are proved to be closely related when DenseNet can be considered as a special ResNet if the connections are shared across layers [11,17–20]. Therefore, ResNet implicitly reuses the features through the residual path while Densenet can better explore new features through this dense connection. As a result, ResNet implicitly reuses features in the images so it is less powerful to explore new features. In contrast, DenseNet can continuously obtain new features but suffers from higher redundancy in the mead time.

To eliminate the limitations of both ResNet and DenseNet, the proposed prediction scheme introduces a simple dual-path architecture to the DNN model. The novel architecture can reuse common features with low redundancy while remaining a densely connected path so that the network can mine new features with more flexibility. To our best knowledge, this architecture has not been introduced to solar forecasting before.

Based on the dual-path network(DPN) structure, the feature extraction method can be extended by increasing convMixer layers [21–23]. This can separate the mixing of spatial and channel dimensions and maintain equal size and resolution with only standard convolutions. A

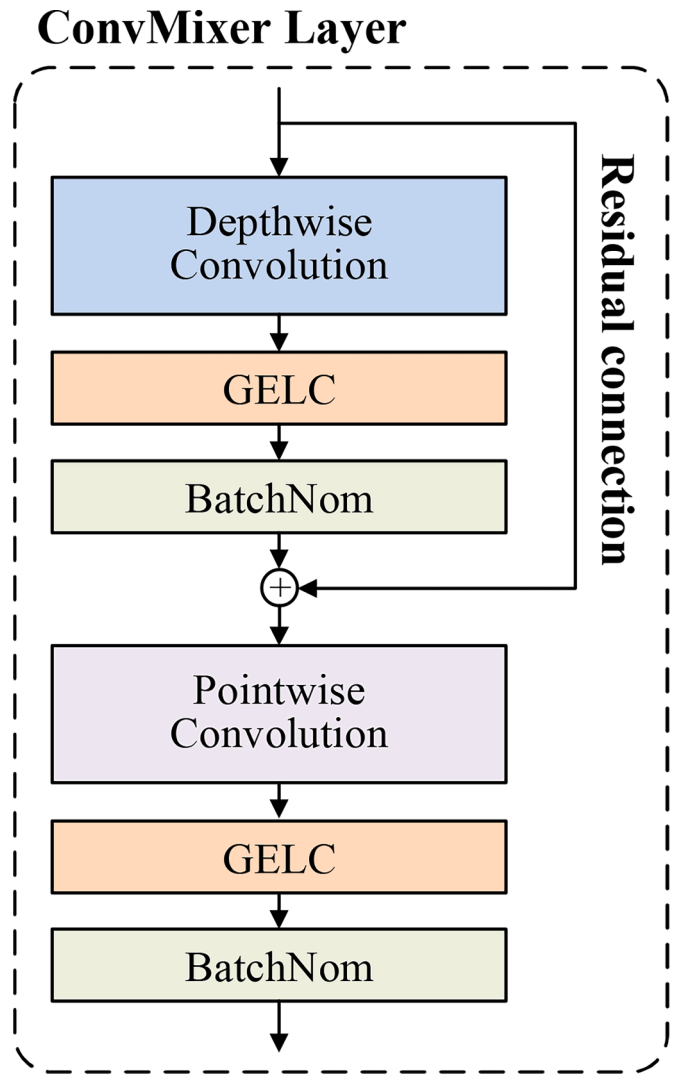

Fig. 6. Structure of convMixer.

large convolution kernel is selected to extract the information from large clouds. It should be noted that each convolution kernel of depthwise revolution is usually responsible for a channel, which means that every channel is convoluted by only one convolution kernel. As shown in Fig. 4, the convolution kernel mentioned above operates each channel of the input picture simultaneously. The number of feature maps after a depthwise revolution is the same as the number of channels in the input layer, so the feature map cannot be extended. Moreover, this operation convolutes each channel of the input layer independently and does not effectively use the feature information of different channels in the same spatial position. To address these issues, pointwise revolutions are needed to combine these feature maps and generate a new one. As shown in Fig. 5, the operation of pointwise convolution is very similar to the classical convolution operation. Its convolution kernel size is $1 \times one \times M$, where M is the number of channels on the upper layer. Therefore, the convolution operation with multiple kernels will weigh and combine the previous map in the depth direction to generate new feature maps.

As shown in the Fig. 6, this paper designs a lightweight mixed residual convolution layer. Firstly, deep separable convolution is used to replace the ordinary convolution in the network. The ordinary convolution is separated on the channel dimension through channel-by-channel convolution, increasing the network width, expanding the range of feature extraction, using point by point convolution to reduce

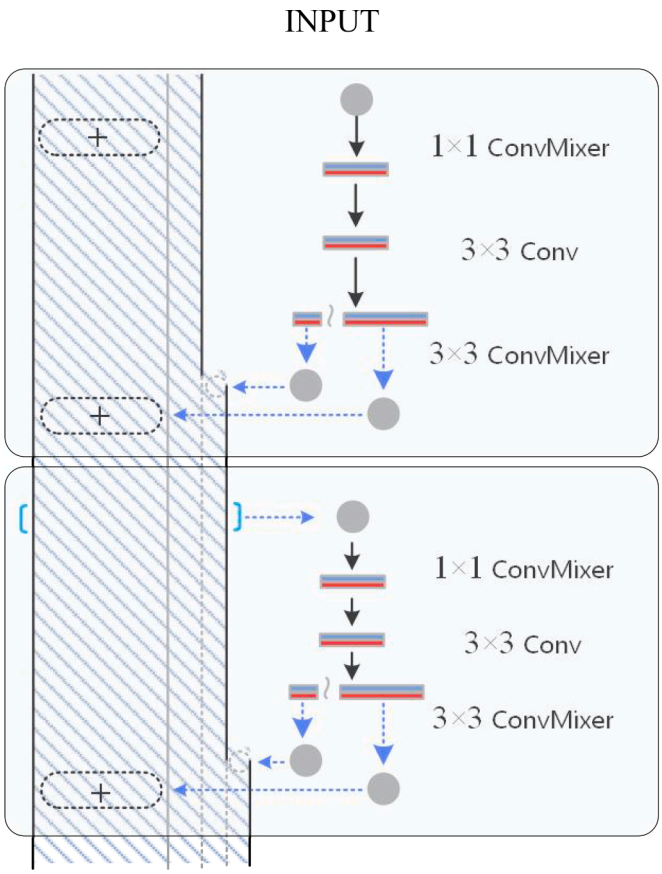

Fig. 7. Structure of Mixer-DPN.

the computational complexity of ordinary convolution operations, and finally adding residual paths to reduce feature information loss. This structure avoids the problem of single feature extraction makes the network more powerful for numerical prediction tasks that require high resolution. And instead of considering both channels and regions in previous ordinary convolution operations, convolution only considers regions first, and then channels, achieving the separation of channels and regions.

The paper [13] compared the performance of three kinds of pooling methods. Different pooling methods are applied to sunny, partly cloudy, and cloudy conditions. The error of max pooling and average pooling under partly cloudy and cloudy days is more significant than that of rank-based average pooling methods. The purpose of a pooling layer is to keep the scale of image features unchanged and remove unimportant redundant features. However, too many pooling layers will result in the loss of characteristic quantity. In DPN model, the size of stride was adjusted to achieve feature dimension reduction while retaining more features. Therefore, we only used a pooling layer for resizing images in the data import module. This paper utilizes ConvMixer to improve the DPN module. As shown in Fig. 7, there are two modules connected from top to bottom. one channel of the model implements feature extraction for residual networks, while another channel implements feature extraction for dense networks. As shown in the Fig. 8, the Sky image data is imported to the model. The model extracts image features by the structure of Mixer-DPN to improve prediction accuracy, and finally obtain the output results through a fully connected layer. As shown in Fig. 9, a portion of the data is returned to the input data in a numerical superposition manner according to the logic of residual operation, while the other portion is returned to the input data in a channel superposition manner according to the logic of a dense network. Finally, all data is packaged and transmitted to the next module.

3.3. Evaluation metrics

To evaluate the forecasting model performance and improve the training process, this paper cites several common loss functions and adjusts the original loss function based on the photovoltaic characteristics. Mean-square error (MSE) is a commonly used loss function, which is a smooth, continuous, differentiable, and easy-to-use gradient descent algorithm [24]. In addition, the gradient of MSE decreases with the decrease of the error, which is conducive to the convergence of the function. The function can obtain the minimum value quickly if the learning factor is fixed. If there are outliers in the sample, MSE will assign a higher weight to the outliers, which is sensitive to outliers and greatly affected by them. The MSE formula is as follows:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{1}$$

Compared to MSE, MAE takes the advantage of being less sensitive to the outliers [25]. Because MAE calculates the absolute value of the error, the penalty is fixed for any size difference. No matter the input value, it has a stable gradient and will not lead to the gradient explosion problem, so it has a relatively robust solution. Although the MAE curve is continuous, it is not differentiable at $x = 0$. And the MAE gradient is equal in most cases, which means that even for a small loss value, the gradient is large. This is not conducive to function convergence and model learning. The MAE formula is as follows:

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{2}$$

R-square (R^2) represents the change degree the model can fit in proportion to the change degree of real data [26]. R^2 ranges from minus

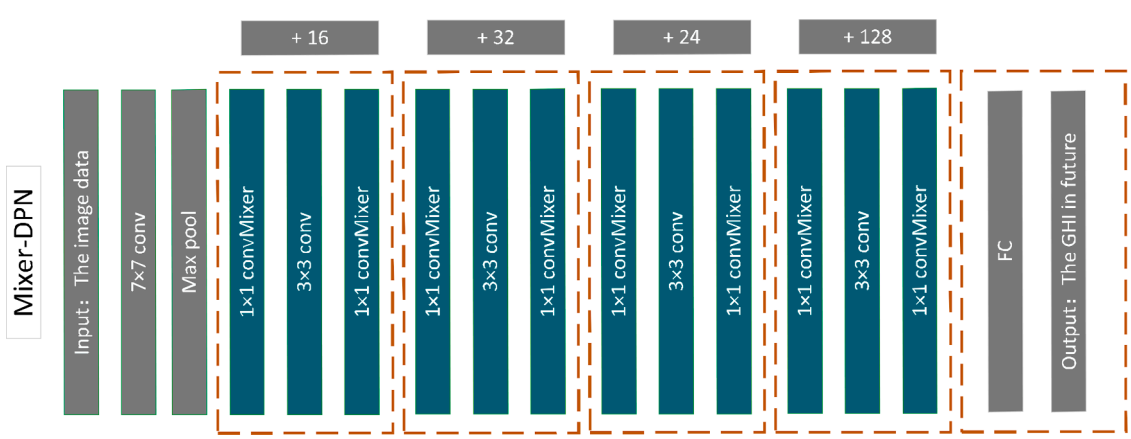

Fig. 8. The progress of the model.

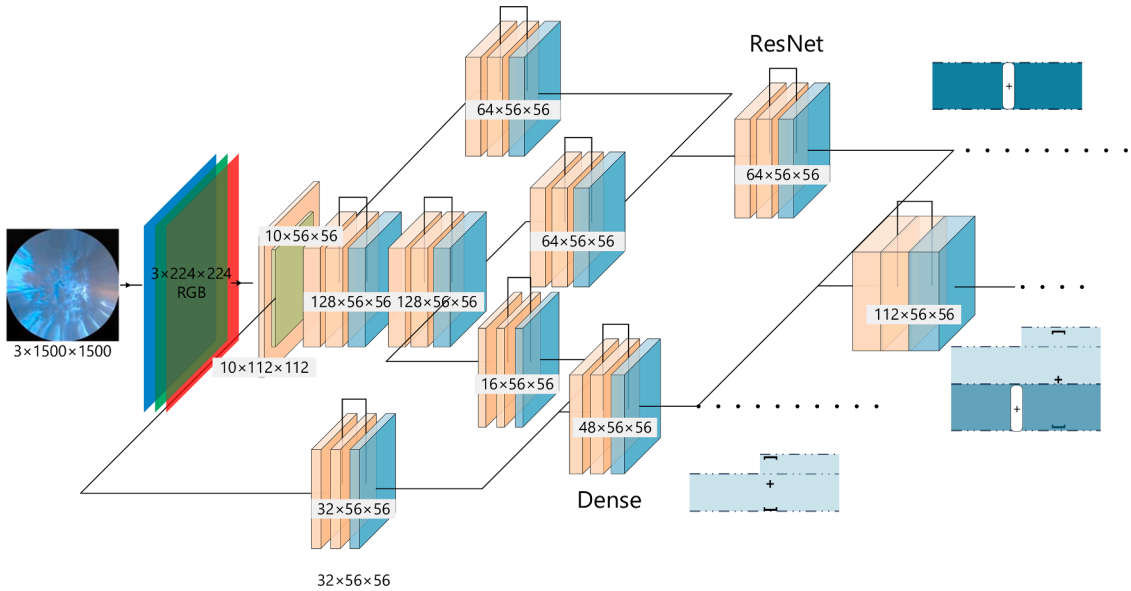

Fig. 9. Data in Mixer-DPN.

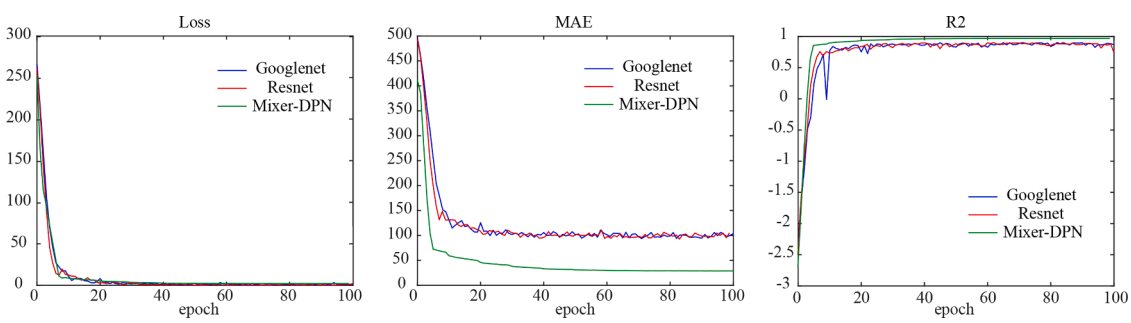

Fig. 10. Loss functions of train.

infinity to 1, usually 0 to 1. When the prediction error is significant and the predicted value is large, R^2 is far less than 0. The general idea is that a bigger R^2 is better. Because in the best case, the Square of the residual is 0, and R minus Square is 1. The R^2 formula is as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \tag{3}$$

Due to the special properties of photovoltaic power generation, the change of irradiance under high illumination is more likely to cause the unbalance of active power. Therefore, we modified the function formula of MSE, and the new loss function improved the sensitivity of the training process to the frequent changes of photovoltaic to a certain extent. The Loss formula is as follows:

$$\text{Loss} = \frac{\sum_{i=1}^{n} y_i (y_i - \hat{y}_i)^2}{n(y_{\max} - y_{\min})} \tag{4}$$

4. Experiment results and analysis

The experiments are implemented under the PyTorch DL framework with NVIDIA 2080Ti GPUs on a server. The models are trained and tested on the Golden dataset for 50 epochs, during which the test set is isolated from the training process. The learning rate is initially set to 0.003 and automatically tuned by a Plateau learning rate scheduler that reduces the learning rate by a factor of 0.5 when the validation loss has stopped decreasing for over 50 epochs. Adam optimizer is used to find

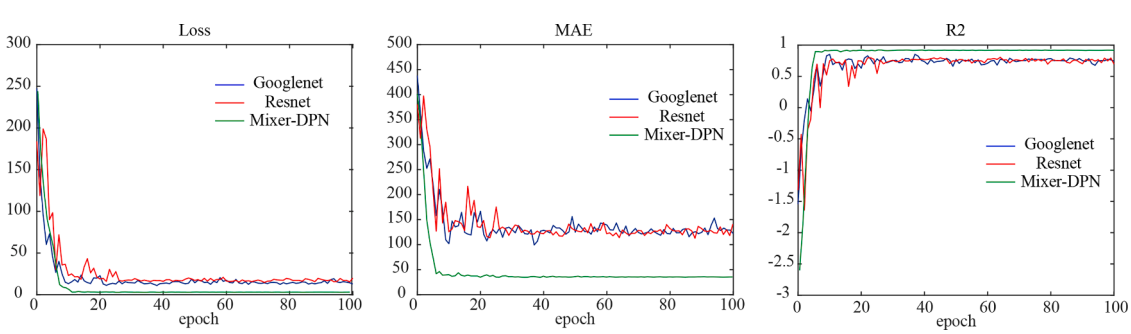

Fig. 11. Loss functions of test.

Table 2
Predicted performance.

Model	Loss	MAE	R2	Score	Size
Googlenet	24.676	124.168	0.784	1.49	43,745KB
Resnet	42.041	157.045	0.633	1.00	22,069KB
MixerDPN	9.135	63.472	0.921	2.32	13,183KB

the global minima. The train progress is shown in the Fig. 14 To study the proposed prediction model's effectiveness, we compared the proposed prediction model's performance with ResNet and GoogleNet, which are well-known and mature DNN structures. The comparison results of the loss function approach using different methods are provided in Figs. 10 and 11.

GoogleNet model is a classical single channel model, increasing network complexity by increasing width [27]. The loss value of the train is stabilized at around 1, while the test's loss is about 15, which means that the model fails to extract effective feature quantities. Meanwhile, the value of MAE and R^2 show the same trend. The over-fitting is caused by the inappropriate architecture of the NN. Although the model can obtain the appropriate fitting effect during training with the help of a large number of parameter adjustments, the feature quantity with good universality has not been extracted. The same question arises in the ResNet model.

Compared with the previous two, the Mixer-DPN model achieves good results with its dual-channel structure. Although the model has been overfitted after 50 epochs, the training loss function curve is close to the test loss function curve. At the same time, it also performs well in the numerical value and stability of the loss function. Although the final training loss is higher than the first two models, the test loss function value is only three, which is much smaller than the value of 15 in the other model.

In summary, Mixer-DPN performs better than other methods in all indicators, and the convergence speed is also higher than other methods.

To intuitively compare the effect of the model, baseline model score provide an acceptable minimum standard for performance. The baseline model results provide a scenario where all the other models trained based on the data can be evaluated [15]. In this paper, ResNet model is selected as the baseline model, and the training results of many other models are provided as a reference. The R^2 and MAE of baseline model are marked as R_p^2 and MAE_p. To compare the results intuitively and reliably, MAE and R^2 were selected to obtain the prediction effect score. The score is defined as

$$\frac{R^2}{R_p^2}\left(2-\frac{MAE}{MAE_p}\right) \tag{5}$$

In order to observe the intuitive prediction performance, we also selected data from five consecutive days, which are not included in the training set and test set, to test the prediction performance.

In Table 2, the value of Loss, MAE, and R^2 shows the relative sizes of the prediction error. The Loss's value is 9.135 far smaller than the 24.676 and 42.041 which proves that the error of MixerDPN is less than other models. Comparing the mse and R^2 values of the three models can also demonstrate MixerDPN's accuracy in prediction. The score directly shows the prediction performance of the model. The higher the score, the better the prediction effect. Because the model of Resnet is chosen as baseline, the score of Resnet is 1. The score of MixerDPN is 2.32 far more than others. Finally, the size of each model is also listed in the table. It can be seen that although mixer DPN is a dual channel model, depending

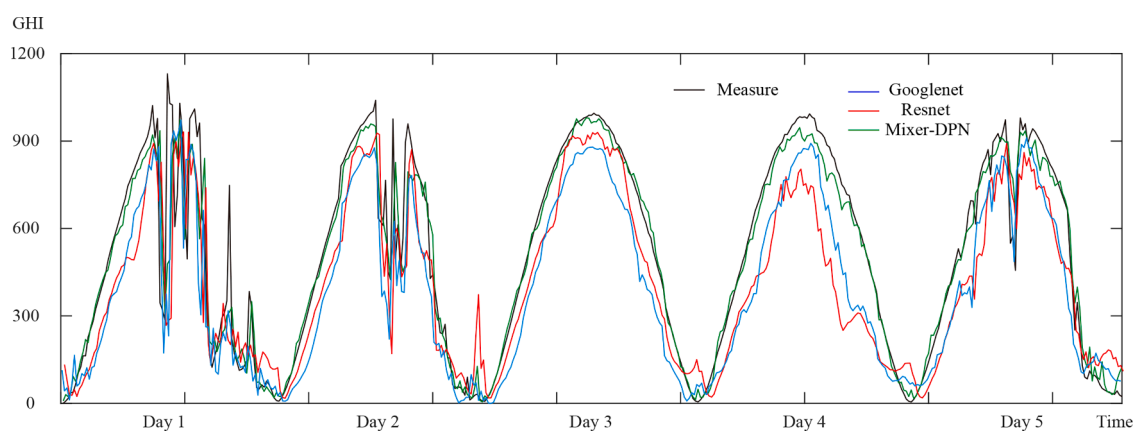

Fig. 12. Compare of Prediction.

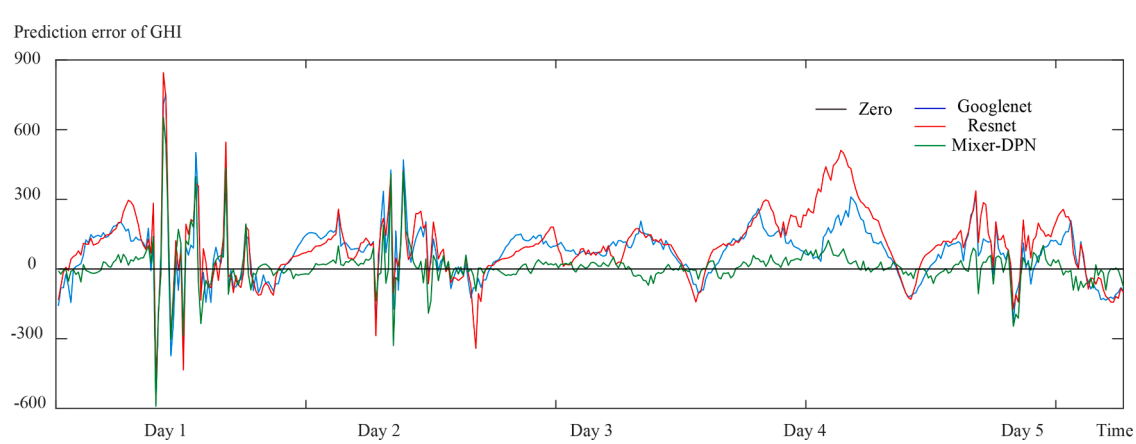

Fig. 13. Error of Prediction.

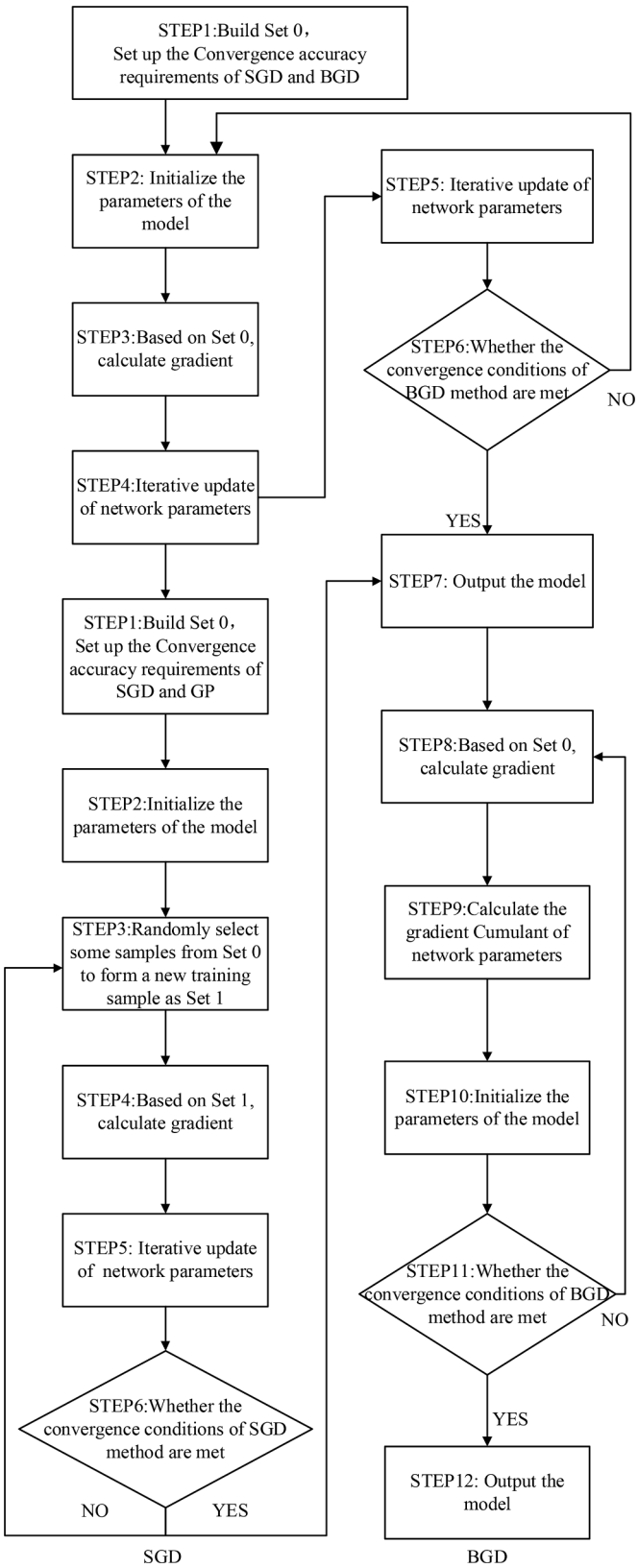

Fig. 14. The progress of the training.

on model fusion, the size of the model is smaller.

The prediction performances are compared in Fig. 12 and the error of prediction are compared in Fig. 13 The selected cloud image data is from July 6th, 2021, to July 10th, 2022, including various meteorological conditions. It is obvious that the proposed mixer-DPN model leads to a higher prediction accuracy. However, all the models make relatively less inaccurate predictions under low irradiance conditions. This is a common phenomenon since the sampling of cloud layer by ground-based cloud image is inaccurate with low solar light. Since the power generation capacity of photovoltaic equipment is limited and the resulting fluctuation is also insignificant, this deficiency can be ignored.

5. Conclusion

This article proposes a novel DNN-based solar forecasting model that can accurately map the real-time ground-based cloud images to the solar irradiance. The model is constructed based on the advanced Mixer-DPN architecture so it can better extract the features from the previous network layers. Moreover, The model is proved to be more suitable for mining information from cloud images since it can share common extracted features while maintaining the flexibility to explore new representations. The proposed method can also separate the spatial and channel dimensions, which significantly improves the feature's diversity. The optimizations make the proposed forecaster outperform other popular DNN models at both prediction accuracy and convergence speed. Studies have shown that the new model has achieved excellent results in MSE, MAE and other indicators, and the R2 prediction accuracy rate has increased by 14% compared with the baseline model. At the same time, the model size has been reduced by 8 MB. The proposed solar forecasting model can be applied to distributed or regional PV power plants, which benefits energy management and economic dispatch in the smart grid.

Ethical Procedure

The research meets all applicable standards with regard to the ethics of experimentation and research integrity, and the following is being certified/declared true.

As an expert scientist and along with co-authors of concerned field, the paper has been submitted with full responsibility, following due ethical procedure, and there is no duplicate publication, fraud, plagiarism.

Declaration of Competing Interest

None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.

It is to specifically state that "No Competing interests are at stake and there is No Conflict of Interest" with other people or organizations that could inappropriately influence or bias the content of the paper.

Data availability

The data that has been used is confidential.

Funding information

Natural Science Research Project of Jiangsu Higher Education Institutions. Funding number: 23KJB470019.

References

[1] D. Yang, J. Kleissl, C.A. Gueymard, H.T.C. Pedro, C.F.M. Coimbra, History and trends in solar irradiance and PV power forecasting: a preliminary assessment and review using text mining, Sol. Energy 168 (2018) 60–101.
[2] K. Doubleday, S. Jascourt, W. Kleiber, B.-M. Hodge, Probabilistic solar power forecasting using Bayesian model averaging, IEEE Trans. Sustain. Energy 12 (1) (2020) 325–337.
[3] F. Rodríguez, I.n. Azcárate, J. Vadillo, A. Galarza, Forecasting intra-hour solar photovoltaic energy by assembling wavelet based time-frequency analysis with

deep learning neural networks, Int. J. Electr. Power Energy Syst. 137 (2022) 107777.
[4] S. Sharda, M. Singh, K. Sharma, RSAM: robust self-attention based multi-horizon model for solar irradiance forecasting, IEEE Trans. Sustain. Energy 12 (2) (2020) 1394–1405.
[5] B. Gao, X. Huang, J. Shi, Y. Tai, J. Zhang, Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks, Renew. Energy 162 (2020) 1665–1683.
[6] J. Qu, Z. Qian, 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 232 (2021) 120996.
[7] K. Chen, Z. He, K. Chen, J. Hu, J. He, Solar energy forecasting with numerical weather predictions on a grid and convolutional networks. 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), IEEE, 2017, pp. 1–5.
[8] I.-I. Prado-Rujas, A. García-Dopico, E. Serrano, M.S. Pérez, A flexible and robust deep learning-based system for solar irradiance forecasting, IEEE Access 9 (2021) 12348–12361.
[9] M. Abdel-Basset, H. Hawash, R.K. Chakrabortty, M. Ryan, PV-Net: an innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production, J. Clean. Prod. 303 (2021) 127037.
[10] C.S. Lai, C. Zhong, K. Pan, W.W.Y. Ng, L.L. Lai, A deep learning based hybrid method for hourly solar radiation forecasting, Expert Syst. Appl. 177 (2021) 114941.
[11] P. Khorramshahi, A. Kumar, N. Peri, S.S. Rambhatla, J.-C. Chen, R. Chellappa, A dual-path model with adaptive attention for vehicle re-identification. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6132–6141.
[12] F. Wang, Z. Zhang, H. Chai, Y. Yu, X. Lu, T. Wang, Y. Lin, Deep learning based irradiance mapping model for solar PV power forecasting using sky image. 2019 IEEE Industry Applications Society Annual Meeting, IEEE, 2019, pp. 1–9.
[13] Z. Si, Y. Yu, M. Yang, P. Li, Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks, IEEE Trans. Ind. Appl. 57 (1) (2020) 5–16.
[14] C. Feng, M. Cui, B.-M. Hodge, S. Lu, H.F. Hamann, J. Zhang, Unsupervised clustering-based short-term solar forecasting, IEEE Trans. Sustain. Energy 10 (4) (2018) 2174–2185.
[15] H. Wen, Y. Du, X. Chen, E. Lim, H. Wen, L. Jiang, W. Xiang, Deep learning based multistep solar forecasting for PV ramp-rate control using sky images, IEEE Trans. Ind. Inf. 17 (2) (2020) 1397–1406.
[16] S.T. Andreas A, NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (data); NREl report no. da-5500-56488, 10.5439/1052221.
[17] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, J. Feng, Dual path networks, Adv. Neural Inf. Process. Syst. 30 (2017).
[18] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481.
[19] Z. Allen-Zhu, Y. Li, What can resnet learn efficiently, going beyond kernels? Adv. Neural Inf. Process. Syst. 32 (2019).
[20] R. Ranftl, A. Bochkovskiy, V. Koltun, Vision transformers for dense prediction. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12179–12188.
[21] I.O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, et al., MLP-mixer: an all-MLP architecture for vision, Adv. Neural Inf. Process. Syst. 34 (2021) 24261–24272.
[22] A. Trockman, J.Z. Kolter, Patches are all you need?, arXiv preprint arXiv:2201.09792(2022).
[23] P. Schulam, S. Saria, Can you trust this prediction? Auditing pointwise reliability after learning. The 22nd International Conference on Artificial Intelligence and Statistics, PMLR, 2019, pp. 1022–1031.
[24] V.A. Nguyen, S. Shafieezadeh-Abadeh, D. Kuhn, P. Mohajerin Esfahani, Bridging Bayesian and minimax mean square error estimation via Wasserstein distributionally robust optimization, Math. Oper. Res. 48 (1) (2023) 1–37.
[25] T.O. Hodson, Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, Geosci. Model Dev. 15 (14) (2022) 5481–5487.
[26] C. Ma, L. Wu, L. Ying, The multiscale structure of neural network loss functions: the effect on optimization and origin, arXiv preprint arXiv:2204.11326(2022).
[27] P.S. Satya Sreedhar, N. Nandhagopal, Classification similarity network model for image fusion using ResNet50 and GoogLeNet, Intell. Autom. Soft Comput. 31 (3) (2022).

Tongsen Zhu received the MS degree in electrical engineering from Nanjing Normal University, Nanjing, China, in 2023. His current research interests include solar power forecasting, photovoltaic power systems, machine learning, and image processing.

Xuan Jiao received the PhD degree in electrical engineering from University of Sydney, Camperdown, NSW, Australia in 2023. His current research interests include photovoltaic power systems, deep learning (artificial intelligence), graph theory, maximum power point trackers, power engineering computing, power generation control, power grids, recurrent neural nets, search problems.

Xingshuo Li (Member, IEEE) received the BS degree in computer science from Zhengzhou University, Zhengzhou, China, in 2012, and the MS degree in sustainable energy technology (with distinction) from Xi'an Jiaotong-Liverpool University, Suzhou, China, in 2015, and the PhD degree from the University of Liverpool, Liverpool, U.K., in 2019. He is currently a Lecturer with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China. His research interest includes renewable energy technology and distributed generation, especially ancillary service, power forecasting, and fault diagnosis in Photovoltaic systems.

Xuening Yin is studying in electrical engineering at Nanjing Normal University, Nanjing, China. Her current research interests include solar power forecasting, photovoltaic power systems, machine learning.

Yang Du (Senior Member, IEEE) received the PhD degree in electrical engineering from The University of Sydney, Sydney, NSW, Australia, in 2013. From 2013 to 2014, he was with the Masdar Institute of Science and Technology, Abu Dhabi, UAE, as a Postdoctoral Research Fellow. From 2014 to 2018, he was a Lecturer with Xi'an Jiaotong-Liverpool University, Suzhou, China, where he maintains an honorary position. He joined James Cook University, Cairns, QLD, Australia, in 2019. He was a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2018. He is an Associate Editor for the IET Renewable Power Generation. He was featured in the 2022 World's Top 2% Scientists List published by Stanford University, Stanford, CA, USA.

Shuye Ding was born in China. He received the BS, MS, and PhD degrees in electrical machinery and appliance from the Harbin University of Science and Technology, Harbin, China, in 2001, 2004, and 2008, respectively. He is currently a Professor with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China. He is the author or co-author of more than 80 published peer-reviewed papers and holds more than ten patents. His research Author Bio(s) with Photo(s) interests include synthesis physical fields of large electrical machines and theoretical study of special electrical machines.

Weidong Xiao (Senior Member, IEEE) received the master's and PhD degrees in electrical engineering from The University of British Columbia, Vancouver, BC, Canada, in 2003 and 2007, respectively.,He is currently an Associate Professor with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW, Australia. From 2010 to 2016, he was with the Masdar Institute of Science and Technology, United Arab Emirates. In 2010, he was a Visiting Scholar with the Massachusetts Institute of Technology (MIT), Cambridge, USA, where he worked on the power interfaces for PV power systems. Before his academic career, he was an R&D Engineering Manager with MSR Innovations Inc., Canada, focusing on integration, research, optimization, and design of photovoltaic power systems. His research interest includes photovoltaic power systems, power electronics, dynamic modeling, control engineering, dc systems, and industrial applications.