Hindawi
Advances in Meteorology
Volume 2020, Article ID 6484812, 12 pages
https://doi.org/10.1155/2020/6484812

Research Article

Convolutional Residual-Attention: A Deep Learning Approach for Precipitation Nowcasting

Qing Yan,[1] Fuxin Ji,[1] Kaichao Miao,[2] Qi Wu,[1] Yi Xia,[1] and Teng Li[1]

[1]*College of Electrical Engineering and Automation, Anhui University, Hefei 230601, China*
[2]*Anhui Public Meteorological Service Center, Hefei 230031, China*

Correspondence should be addressed to Kaichao Miao; mkc2005@126.com

Received 17 September 2019; Revised 16 December 2019; Accepted 6 January 2020; Published 29 February 2020

Academic Editor: Hiroyuki Hashiguchi

Short-term precipitation forecast in local areas based on radar reflectance images has become a hot spot issue in the meteorological field, which has an important impact on daily life. Recently, deep learning techniques have been applied to this field, and the effect is promoted remarkably compared with traditional methods. However, existing deep learning-based methods have not considered the problem that different areas and channels exert different influence on precipitation. In this paper, we propose to incorporate the multihead attention into a dual-channel neural network to highlight the key areas for precipitation forecast. Furthermore, to solve the problem of excessive loss of global information caused by the attention mechanism, the residual connection is introduced into the proposed model. Quantitative and qualitative results demonstrate that the proposed method achieves the state-of-the-art precipitation forecast accuracy on the radar echo dataset.

1. Introduction

Generally speaking, precipitation forecast refers to providing a very short range (e.g., 0–2 hours) forecast of the rainfall intensity in the local region as accurate as possible based on the radar echo map, rain gauge, or other observation data [1]. A precise weather prediction can be very useful in human life for outdoor activity, traffic condition, early warnings of extreme weather, etc. Due to the inherent complexities of the atmosphere and relevant dynamical processes, the precipitation forecast problem is quite challenging and becomes a hot research topic in meteorology and machine learning community [2].

There are mainly two categories of traditional methods for precipitation prediction. One is the echo extrapolation technique represented by the optical flow method [3–5], as shown in Figure 1. This kind of method estimates convective cloud movements by radar echo maps and predicts the future radar echo maps by Semi-Lagrangian Advection Scheme. However, this method is more suitable for tracking and predicting the echo targets with larger scale and a long life cycle. When the echo happens to split or merge, the accuracy of the prediction will quickly decrease. The other kind of methods are based on the numerical weather prediction [6]. According to the circumstance of the atmosphere and some initial and boundary conditions, the method solves the equations of fluid mechanics and thermodynamics which describe the weather evolution process based on numerical calculation. Then, the future atmospheric motion and weather phenomenon are predicted according to the numerical results. However, this method is limited by the spin-up time; the first two hours of precipitation prediction by the mesoscale numerical model are invalid, especially in the application of nowcasting, which has low accuracy and requires complex physical equation calculation. As a result, it can hardly meet the needs of accurate and real-time in refined prediction [7, 8].

With the development of deep learning methods, some progress has been achieved in precipitation forecast field. Shi et al. [9] proposed the convolutional LSTM (ConvLSTM) to build an end-to-end trainable model for the precipitation forecast problem, which effectively captures spatiotemporal correlations and consistently outperforms the fully connected-Long Short-Term Memory (FC-LSTM) [10].

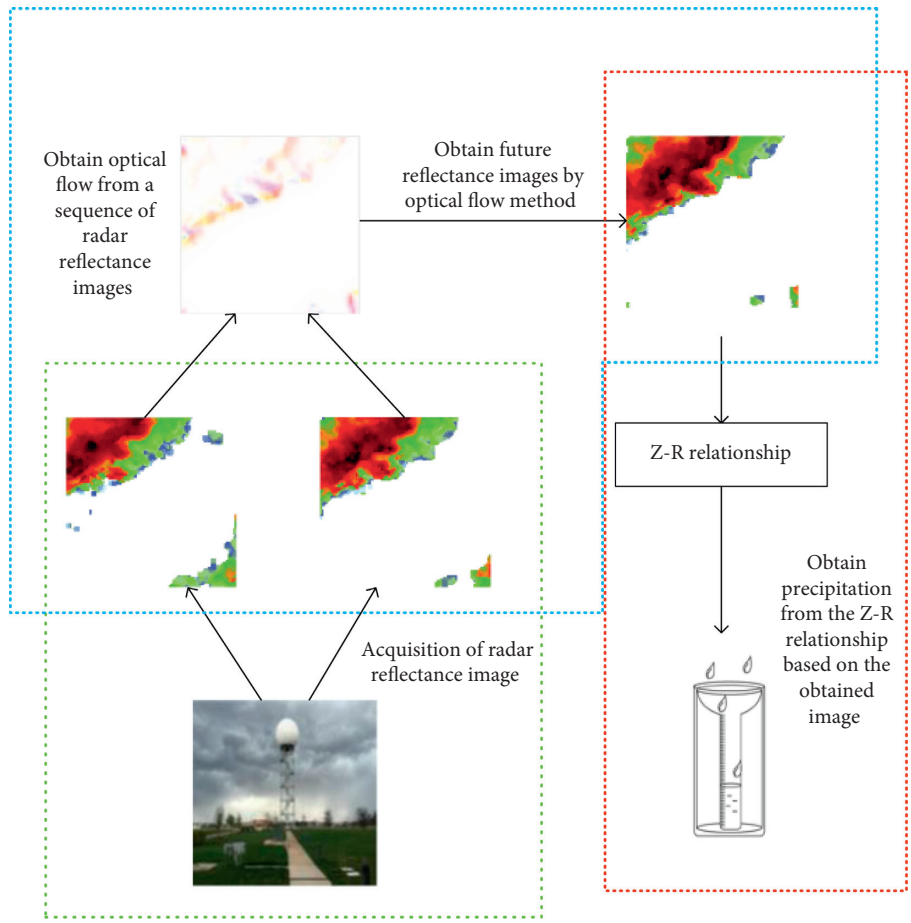

Figure 1: Precipitation prediction process based on the optical flow method.

However, the convolutional recurrence structure in ConvLSTM-based models is location-invariant, while cloud natural motion and transformation (e.g., rotation) are location-variant in general. Shi et al. [1] further improved the method to construct the TrajectoryGRU (TrajGRU) model that can actively learn the location-variant structure for recurrent connections. Both the abovementioned methods have obtained better performance than the traditional optical flow method, but their proposed models are complicated requiring a large number of data for training. Yao et al. [11] proposed a novel method to solve this problem. The precipitation forecast was regarded as a spatial sequence prediction problem according to Taylor Frozen Hypothesis (If the signal pulsation caused by turbulence in the atmosphere is far less than the space variation caused by convection, the cloud cluster tends to shift in space at the local average convection speed. In a short time, there is no sharp change in the shape or reflection intensity and there is a significant spatiotemporal correlation in the flow field.) [11, 12] which is widely applied in meteorology and fluid dynamics. The future radar echo map of the target site was obtained by stitching the radar echo maps based on scale-invariant feature transform (SIFT) key point detection. Then, the radar echo map was fed into the convolutional neural network to get the forecast result. Compared with the previous researches, this method greatly simplified the complex space-time prediction problem by combining machine learning with deep learning. However, there are still many deficiencies to be solved. As well known, precipitation particles in clouds with different heights have different density distribution and scales, so clouds with different heights (1.5 km, 2.5 km, 3.5 km) will have different effects on precipitation. This is an important factor that must be considered in precipitation forecast, which has not been paid enough attention in the previous study.

Recently, the attention mechanism in deep learning has shown promising results and is widely used in computer vision. It learns a weight matrix to emphasize major features and suppresses inessential features [13]. Vaswani et al [14] proposed a self-multi-head attention which can capture the connections between sequences and resolve long-distance dependencies. Stollenga et al. [15] proposed a deep attention selective network which uses attention to adjust the weight of each convolution filter to achieve image classification. Although attention mechanism can focus on the key areas, some global information may be lost. Addressing this issue, Chu et al. [16] created a multicontext model based on a stacked hourglass network by implementing a global representation of the feature; Wang et al. [17] proposed a nonlocal block for video classification, which considers the contribution of other regions in the

image to the target by introducing a residual link. However, attention mechanism has rarely been used in the field of precipitation prediction.

In this paper, aiming to precipitation forecast, we propose a dual-channel deep learning model, called multihead attention residual convolutional neural network (MAR-CNN). MAR-CNN can distinguish the important height ranges of clouds that exert more impact on precipitation by multihead attention. Meanwhile, it integrates the idea of residual network with multhead attention to reduce the loss of global features. We conducted experiments on the meteorological dataset distributed by the Shenzhen Meteorological Administration in China. The results verified that the proposed MAR-CNN outperforms conventional deep learning methods, such as convolutional attention as well as convolutional multihead attention.

The main contributions of this paper are summarized as follows:

(1) We address the challenges of discovering the key features for precipitation, such as the important areas with great precipitation intensity and the important channels with different heights, by introducing the multiattention convolutional neural network.

(2) We propose to combine multihead attention with a residual connection, which can utilize global and local information synthetically to mitigate the information loss. To the best of our knowledge, this work is the first attempt for precipitation forecast by jointly using residual structure and multiattention mechanism.

2. Materials and Methods

2.1. Research Area. Shenzhen is located in the southern part of the China between 22°27′~22°52′N and 113°46′~114°37′E, which has an area of 2020 km^2 (Figure 2) and belongs to subtropical maritime climate. The annual average temperature is 22.3°C, the highest temperature is 38.7°C, and the lowest temperature is 0.2°C. The rainy season is from April to September every year, with an annual rainfall of 1924.7 mm.

2.2. Data. The radar echo dataset used in this paper is a part of the two-year meteorological radar intensity dataset collected by the Shenzhen Meteorological Bureau. The dataset has 10,000 sets of samples, each containing 60 radar reflection images and the corresponding precipitation in one hour is collected by the ground station shown in Figure 2. Parts of radar reflection images are reported in Figure 3, which are distributed over 15 consecutive time spans, 6 minutes apart, and four different heights, 1 km apart, from 0.5 km to 3.5 km. Each radar reflectivity image is 101×101 pixels corresponding to 101×101 km land surface area. Each pixel records the radar reflectivity factor. The radar reflectivity echo intensity reflects the scale and density of the precipitation particles inside the meteorological target to a certain extent, and thus the relationship between the reflectivity and the precipitation can be established.

Firstly, it demands to predict the radar image above the target site in the future for precise precipitation prediction. Referring to the data processing method of Yao and Li [11], we stitched the original radar images to get the global prospect of the cloud by template matching.

Secondly, with target sites as center, subimages whose size is 41×41 pixels are intercepted from the stitched image. Each subimage corresponds to three height channels from 1.5 km to 3.5 km. Here, the channel of 0.5 km is abandoned because it is too low and contains a lot of noise. So, the dimension of refection images used in subsequent experiment is $41 \times 41 \times 3$. These subimages will be fed into network to extract the image features. Finally, we obtained 8721 subimages. We randomly selected 1000 groups as test sets, and the remaining 7721 groups are used as training sets.

Last, the nonimage features of cloud, such as cloud movement speed information etc., were obtained by the traditional method like scale-invariant feature transform (SIFT) descriptor [18] which was used to find key points in an image. The size of nonimage features extracted from each subimage is 49×1. [11]. These subimages and the corresponding nonimage features are the input of network.

The classification criteria of precipitation grades and the data distribution of datasets are shown in Table 1. These criteria are widely used as a national standard in China.

As seen from Table 1, the data distribution of original samples is unbalanced. In the actual forecast, heavy rainfall events, such as rainstorms, big heavy rain, and extraordinary heavy rain, should be predicted as accurately as we can, because they will cause more threat to society. However, compared to other weather conditions, the proportion of heavy rainfall is very low. Considering this situation and in order to reduce the impact of data imbalance on network training, we performed data enhancement on the heavy rain, big heavy rain, and extraordinary heavy rain of the training sets through the SMOTE algorithm [19]. Simultaneously, we also expanded the light rain and no rain data. The size of enhanced training set is listed in Table 1.

2.3. Methodology

2.3.1. Convolutional Neural Network. CNN was first proposed by LeCun et al. [20] as a feed-forward neural network. The LeNet-5 model shown in Figure 4 is the most typical type of convolutional neural network. It includes an input layer (input), a convolution layer, a pooling layer, a fully connected layer, and an output layer (output).

The essence of convolution and pooling in CNN is similar to the filter to extract data features. By convolution and pooling, the input data are transformed into hidden topological structure features between data. Then, these features are merged in the full connection layer, and the classification or regression result can be completed in output layer.

2.3.2. Multihead Attention. An attention function can be described as mapping a query and a set of key-value pairs to

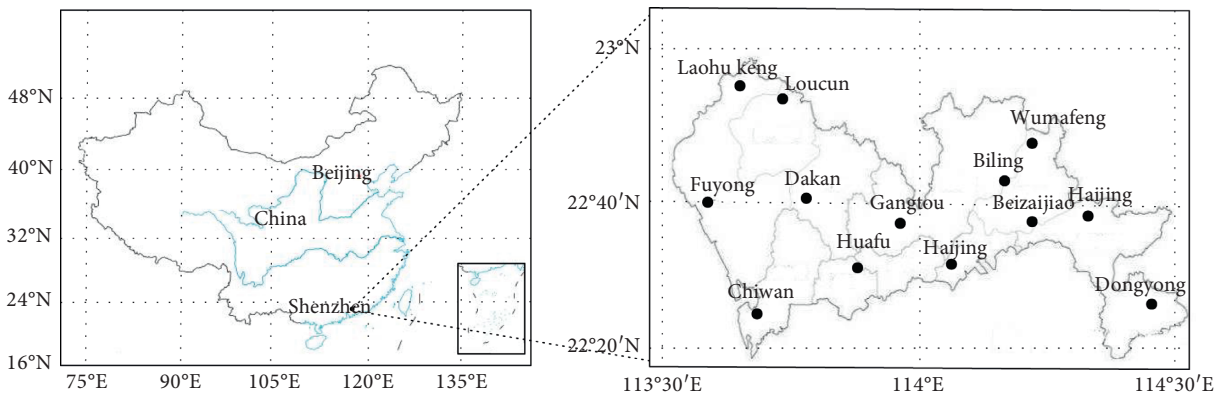

Figure 2: The geographical location of Shenzhen and the spatial distribution of some meteorological stations.

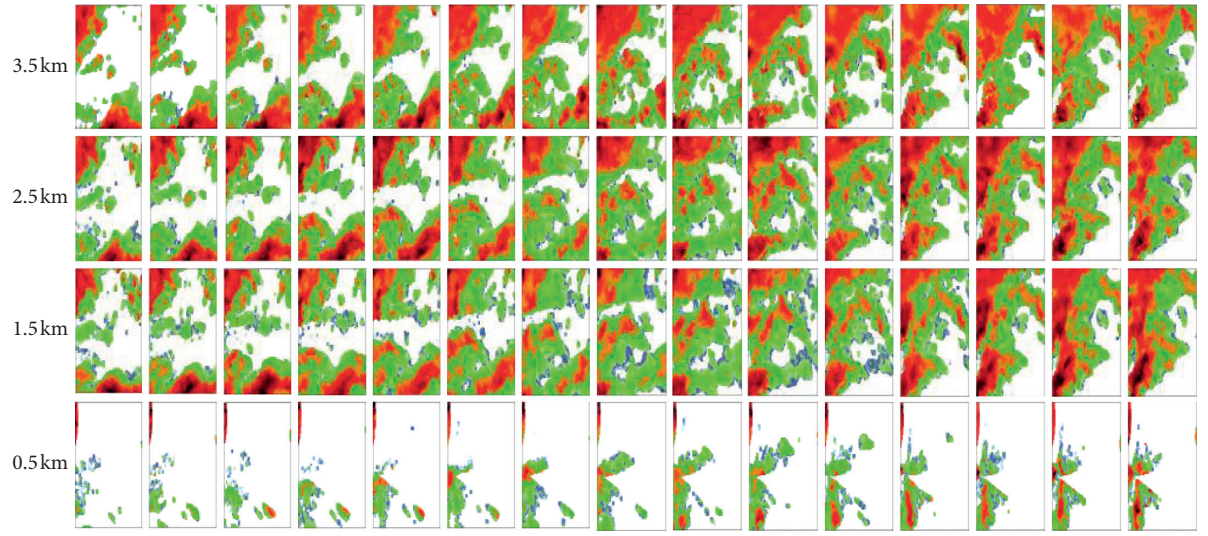

Figure 3: Historical radar images of 15 time spans with 4 different heights.

Table 1: The classification criteria of precipitation grades and the data distribution.

Precipitation level	Precipitation range (mm/h)	Overall sample size	Enhanced training set size	Test set size
No rain	<0.1	257	2097	24
Light rain	0.1–1.5	844	2352	60
Moderate rain	1.6–6.9	2794	2503	291
Heavy rain	7.0–14.9	781	1696	85
Rain storm	15.0–39.9	3275	2846	429
Big heavy rain	40.0–49.9	450	2246	69
Extraordinary heavy rain	≥50.0	320	2224	42

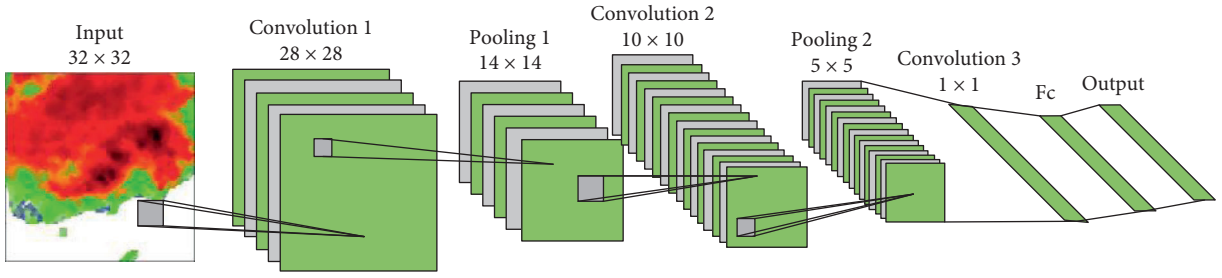

Figure 4: LeNet-5 model.

the output, where the queries, keys, values, and output are all vectors. In general, keys and values are equal. Specifically to our mission, key and value are all the characteristics of the radar reflectivity image extracted by the CNN network and query is the weight matrix that the network needs to learn. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key [14].

Through the attention mechanism, the model can focus on important information for the task [14].

Recently the multihead attention method [14, 21, 22] has been demonstrated to be successful in machine translation and image recognition [23–25]. Compared with the single attention, multihead attention executes the attention mechanism several times in a parallel way. The queries, keys, and values are denoted by Q, K, and V. Then the weights on the values are obtained by

$$\text{Attention}(Q, K, V) = \text{soft}\max\left(\frac{QK}{\sqrt{d_k}}\right)V, \tag{1}$$

$$\underset{j=1,\ldots,}{\text{soft}\max}\left(Z_j\right) = \frac{e^{Z_j}}{\sum_{k=1}^{K} e^{Z_k}}, \tag{2}$$

where d_k is the dimension of K, which will affect the size of the dot product. Equation (2) is a soft max activation function that can be used to normalize weight, where Z is a K-dimensional vector and j represents one of the elements. Through softmax, we can normalize the elements in the vector to 0 to 1, and the sum of the elements is 1.

A multihead attention consists of several parallel heads (layers) of attention which have different sets of trainable parameters; each head performs linear transformation before attention operation to project the three inputs to a lower dimension [25]. Each attention operation is implemented independently, and then the results obtained by concatenating the output of each head. Specifically, the inputs of the multihead attention layer are three sequences of vectors: query $Q \in R^{l_1 \times d_f}$, key $K \in R^{l_2 \times d_f}$, and value $V \in R^{l_2 \times d_f}$. As for i-th head, an attention function is performed asfollows:

$$\text{head}_i = \text{Attention}\left(QW^{Q_i}, KW^{K_i}, VW^{V_i}\right), \tag{3}$$

where $W^{Q_i}, W^{K_i}, W^{V_i} \in R^{d_f \times d_p}$ are used to project the three inputs to a subspace with lower dimension d_p, which is a parameter learned in the model. Then, the output of the multihead attention is produced by

$$\text{Multihead}(Q, K, V) = \text{Concat}\left(\text{head}_1, \text{head}_2, \ldots, \text{head}_h\right)W^0, \tag{4}$$

where h is the number of head and $W^0 \in R^{hd_p \times d_f}$ is a weight matrix [14, 23]. The structure of the multihead attention is depicted in Figure 5.

The advantage of multihead attention is that it can learn relevant information in different representation subspaces. However, this structure may lose some global information.

2.3.3. Combining Self-Multi-Head Attention with Residual Thought. As well known, it cannot enhance the effect of the network by simply increasing the depth of the network due to the gradient divergence. He et al. [26] proposed a residual network which introduced a shortcut to solve this problem. Moreover, the residual connection avoids the loss of global features to ensure the integrity of the original information [15, 27]. The proposed model, combining multi-head attention and residual connection, is given as follows:

$$\text{REAT}(f, x) = X + f(X), \tag{5}$$

where $f(X) = \text{Multihead}(X, X, X)$ and X is the characteristic of the radar reflectivity image, which will be extracted by the convolution operation in our method. Note that here we adopt the structure of the multihead attention with the same sequence for the query, key, and value, which is named as self-multi-head attention [14]. By this way, each row vector in the feature matrix must be a dot product with all column vectors, which allows the network to capture the spatial structure of the radar reflectance image, so that the correlation of radar reflectivity at different locations is learnt. With the help of the principle of attention mechanism, we hope to highlight the features that contribute more to precipitation by learning the weight matrix query in network, so as to achieve a better mapping relationship between radar reflectivity and precipitation.

So, based on the model incorporating the multihead attention with residual thought, the more comprehensive features fusing global and local information can be studied, which are vital in precipitation forecasting.

2.3.4. Proposed Model Architecture. In this section, we will introduce our model in detail. The goal of our model is to extract the characteristics of radar images by deep network to achieve regression prediction of precipitation. In order to capture the important features of the radar images and grasp the spatiotemporal characteristics of the cloud layers, we designed the following framework drawn in Figure 6 inspired by Yao and Li [11].

As seen from Figure 6, CNN1 is to extract the deep characteristics of radar images and CNN2 is responsible for acquiring the deep features from the nonimage characteristic extracted by the original feature extracting method mentioned above. Finally, the concatenated features, extracted by the two channels, are sent into the fully connected and output layer to obtain the predicted output of the precipitation.

The input images in CNN1 are radar images of the future cloud moment above the target site, distributing over three heights, whose sizes are $41 * 41\ \text{km}^2$. In CNN2, the input is a nonimage feature whose dimension is 49×1. Different from the work of Yao and Li [11], we introduce multihead attention to emphasize the key areas and channels corresponding to precipitation. Furthermore, in order to avoid unnecessary global information loss caused by attention layer, we put to use the residual connection in our multihead attention framework.

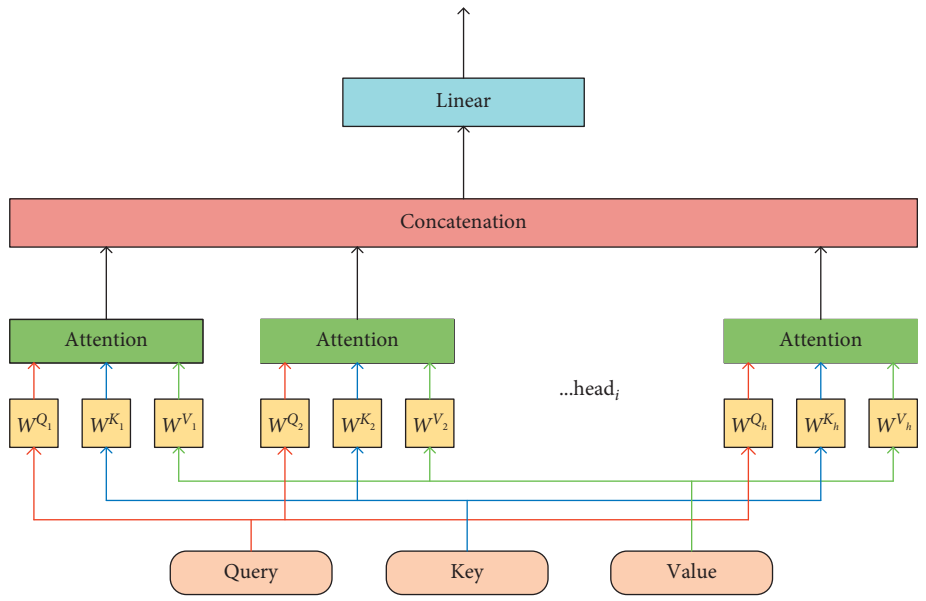

Figure 5: The illustration of multihead attention.

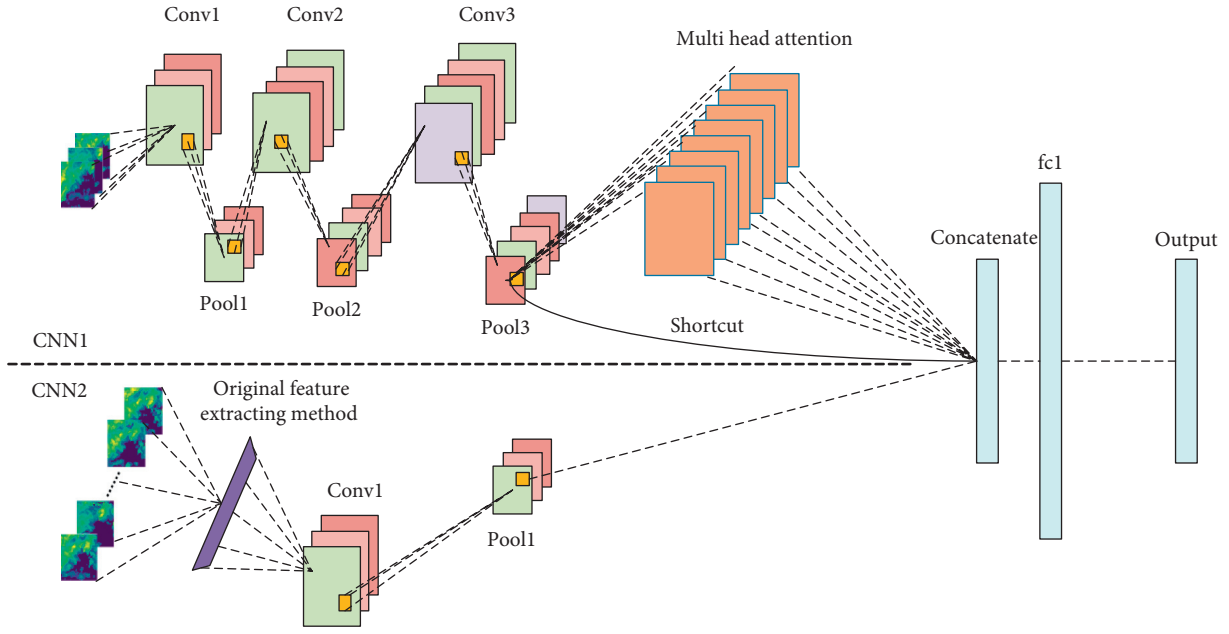

Figure 6: The structure of MAR-CNN.

2.4. Performance Assessment. To assess the performance of the forecasting approaches, three statistics criteria are used in this paper. The definition of these criteria is summarized in Table 2, where $\widehat{X}_i$ ($\widehat{X}$) is the predictive precipitation value, X_i (X) is the actual precipitation value, and Var represents the variance.

3. Results and Discussion

In this section, the proposed dual channel MAR-CNN model is used to predict precipitation in the next hour. To evaluate the performance of the proposed algorithm, we used the enhanced training set and the test set in Table 1 for model training and test. We implemented the proposed model based on TensorFlow. We compared the proposed dual channel MAR-CNN with existing algorithms, including dual-channel convolutional attention model, dual-channel convolutional model, single-channel CNN model (baseline model), and traditional machine learning algorithms including GBDT [28] and SVM [12]. We give details of these models in Figure 7, and the parameter settings are shown in Table 3. In addition, all the algorithms are implemented in Anaconda3 software on a computing server with one NVIDIA TITAN 1080ti GPU.

3.1. Result of MAR-CNN. It is noted that the main parameter in our proposed model is the number of heads. In

Table 2: Definition of evaluation criteria.

Evaluation criteria	Formula
Root-mean-squared error	$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} (\widehat{X}_i - X_i)^2 / n}$
Explained variance score	$\mathrm{Explained\_variance}(X, \widehat{X}) = 1 - (\mathrm{Var}\{X - \widehat{X}\}/\mathrm{Var}\{X\})$

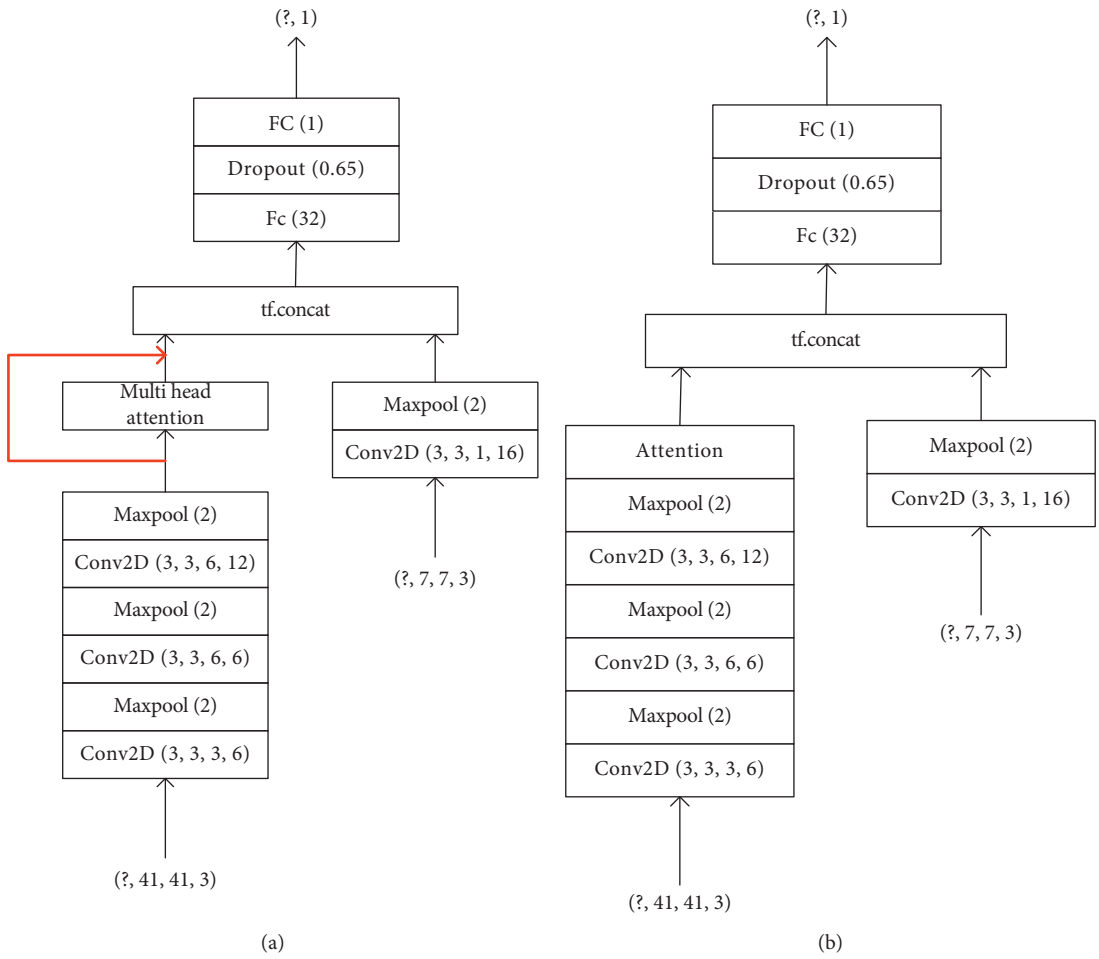

Figure 7: Continued.

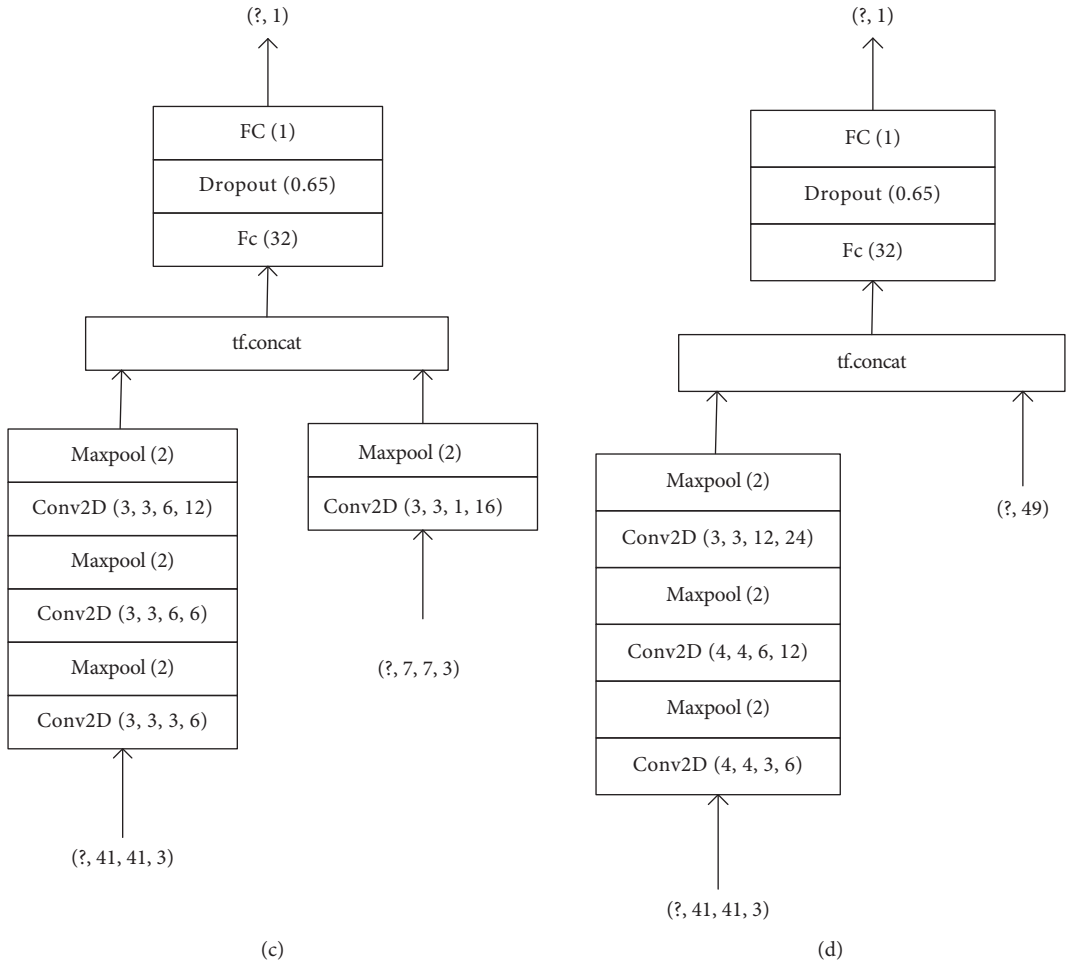

Figure 7: Structure diagram of popular models for precipitation prediction. (a) MAR-CNN. (b) Dual-channel convolutional attention model. (c) Dual-channel convolutional model. (d) Single convolutional model.

Table 3: Short-term precipitation prediction model parameter setting.

Parameter name	Parameter value
Learning rate	0.001
Maximum number of iterations	50000
Number of heads	12
Batch size of training set	128

Table 4: The effect of the number of heads in multihead attention on the experimental results.

Number of heads	6	8	9	10	12	13	14
RMSE (no res) ↓	9.57	9.07	8.57	8.62	8.43	8.46	8.55
RMSE (residual) ↓	8.15	8.09	8.13	8.10	7.90	8.04	8.24
EVS (no res) ↑	0.70	0.69	0.72	0.70	0.70	0.70	0.70
EVS (residual) ↑	0.75	0.764	0.765	0.76	0.77	0.75	0.74

order to evaluate the influence of this parameter on performance, we conducted an experiment with different numbers of heads of attention, and the results are shown in Table 4.

It can be seen from Table 4 that the number of heads in multihead attention has a great influence on the index of RMSE. When the number of heads is less than 12, the RMSE gradually decreases as the number of heads increases, and it reaches the best result when the number of heads is 12. After adding a residual connection to multihead attention, the trend of RMSE changing is the same. As for EVS (higher value means better, and the value of 1 is perfect), in the cases of with residual connections and without residual connections, there is little fluctuation in their performance. Further, it is worth noting that the model with residual connection performs better in general. Because both networks achieved the best performance when the number of heads is 12, we set this parameter to 12 in the following experiments. To observe the effect of the residual connection, we draw the loss curve in the training process in Figure 8.

Just as we can observe from Figure 8, despite of the change of the number of heads, the residual connection consistently leads the result to be better. Futhermore, it makes the model converge faster and more stable than the normal multihead attention.

It is well known that in the radar reflectance image, different colors represent different reflectance values. In general, the bright color represents the large reflectance

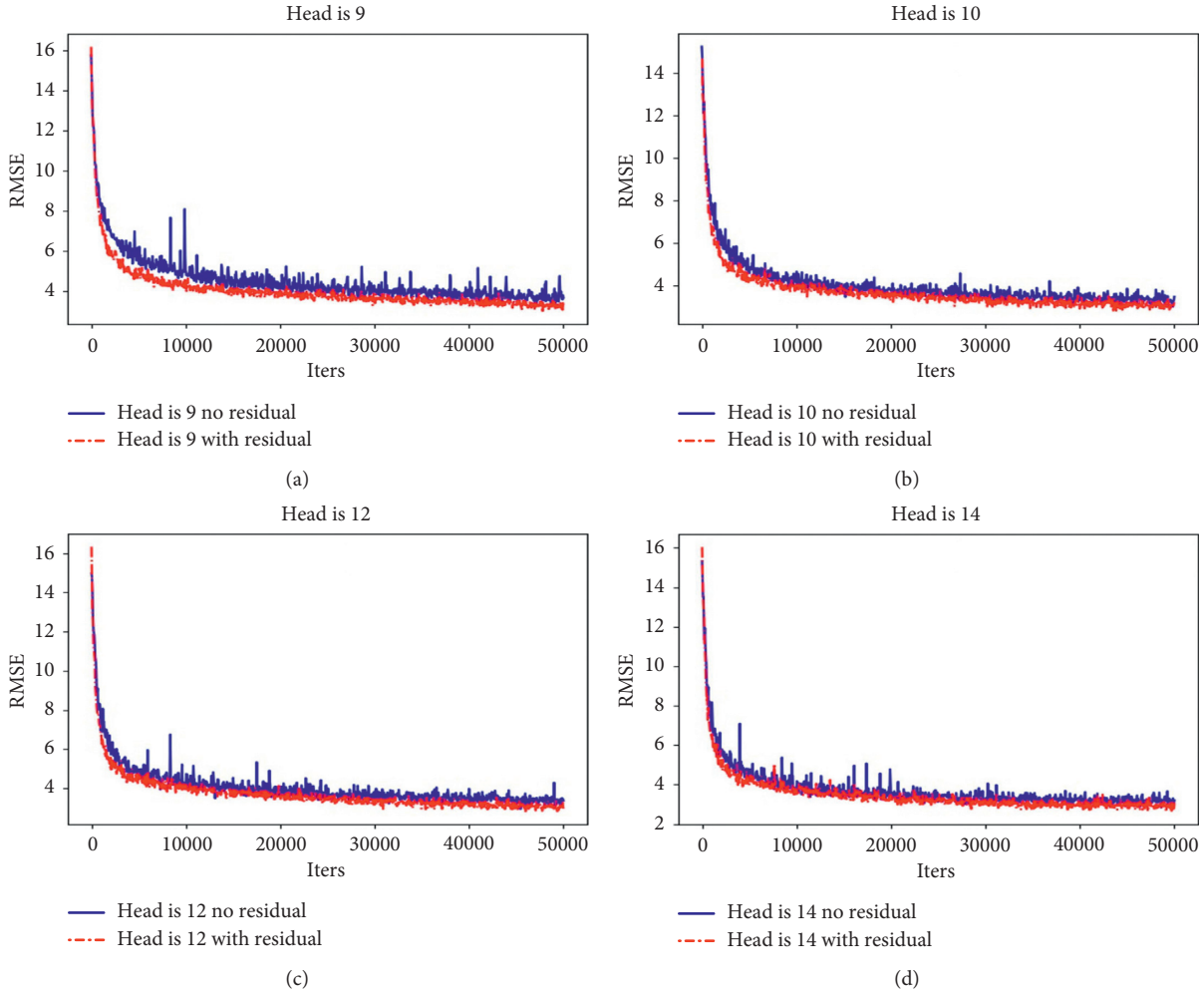

Figure 8: Training loss curve.

value, such as red and scarlet, which indicates that the corresponding precipitation or the probability of precipitation in this area is greater. To verify whether our attention model really captures these more potential precipitation areas in the images, we drew the attention heat maps and compared them with the original images in Figure 9.

In Figure 9, the six images above are original radar images distributing over three heights and the six images below are the corresponding heat maps. In the heat maps, the green areas have the largest weight, followed by blue. That is, the brighter the color, the greater the weight applied to the attention. We can discover that the green areas with a larger weight in the heat map correspond well to the red areas with a larger reflectance value in the radar image. This phenomenon illustrates that our attention mechanism can highlight the important areas in the radar reflectivity images exactly. Furthermore, based on the observation, we found that the heat map better matches the high reflectance area in the original image. That is to say the self-attention which can capture the inside characteristic of the input sequence thinks that the clouds with height of 2.5 km has a greater impact on precipitation. Obviously, in precipitation forecasting, the ability to detect key areas is very vital. Benefiting from this ability offered by self-attention, our model achieved promising results.

3.2. Comparison with Existing Models. In the comparative experiments, we used the same parameter settings for all models as listed in Table 3. The inputs to each model contain original radar reflectivity image information and nonimage information of cloud. We report the comparative results of these models in Table 5.

From Table 5, it can be seen that all deep learning algorithms perform better than GBDT. However, the SVM is better than single-channel CNN (baseline). The accuracy of the dual-channel network model is higher than that of the single channel (baseline model), and the dual-channel CNN with attention is comparable to the dual-channel CNN. However, the promotion of forecasting is not so obvious.

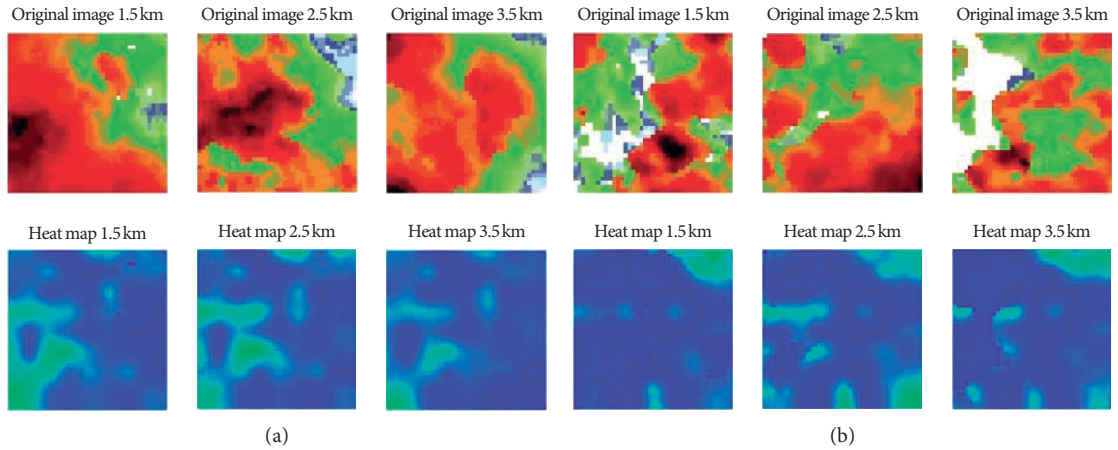

Figure 9: Original images and corresponding attention heat maps.

Table 5: Experimental results of different algorithms.

Model	RMSE	EVS
Dual-channel MAR-CNN	7.90	0.77
Dual-channel CNN with attention	9.46	0.70
Dual-channel CNN	9.48	0.70
Single-channel CNN (baseline)	10.29	0.63
GBDT	11.86	0.47
SVM	9.80	0.70

Apparently, the proposed model has a more exciting result. For advantages in local key feature extraction and global information retention, our dual-channel MAR-CNN model achieved the best precipitation forecast effect.

In addition, in order to analyse the prediction effects of the models under different precipitation levels, we compared the prediction results of the models with the ground-truth observations according to the precipitation level. Furthermore, we calculated their average values, respectively, as shown in Table 6.

Observed from Table 6, for the case of no rain, our MAR-CNN and dual-channel CNN give relatively accurate prediction results. The average prediction of MAR-CNN and dual-channel CNN is less than 0.1 mm/h, which is consistent with the actual level in meteorology (No rain<0.1 mm/h). The result of dual channel CNN with attention is slightly different from the actual precipitation situation. For the case of light rain, the average predicted value of MAR-CNN is close to the average of the observed values, and the predicted precipitation level is in accordance with the actual level (light rain, 0.1–1.5 mm/h). Neglecting the deviation of the predicted value provisionally, we find the precipitation level predicted by these models only deviates from the real situation for one grade, except for GBDT. For the case of moderate rain, MAR-CNN has the smallest predicted error, and the predictive levels of MAR-CNN, dual-channel CNN with attention, and SVM all match with the actual level (moderate rain, 1.6–6.9 mm/hour). For the case of heavy rain and rainstorm, what surprised us is that the error between the predicted mean values of all models including GBDT and the observed value is acceptable, and the predicted precipitation levels are also same as the actual level (heavy rain, 7.0–14.9 mm/h; rainstorm, 15.0–39.9 mm/h). Additionally, for the heavy rain and rainstorm, dual-channel CNN and MAR-CNN achieved the best prediction results, respectively. For the case of big heavy rain and extraordinary heavy rain, although the model we proposed does not have obvious superiority on the predicted average value, the prediction level is very close to the actual level. Especially for the extraordinary heavy rain, predicted precipitation level computed by MAR-CNN's equals to the actual observation (big heavy rain, 40.0–49.9 mm/h; extraordinary heavy rain, ≥50.0 mm/h), which has an important guiding significance for the accurate release of disaster warning.

According to the analysis above, in general, our MAR-CNN achieves accurate rainfall forecast which is consistent with the actual observation when the precipitation level is below the rainstorm level. As the precipitation level continues to increase, the forecast precipitation given by MAR-CNN does not match the actual situation very well, but it is still the best compared with other methods. These analyses illustrate that the residual connection on multihead attention we designed in MAR-CNN can highlight the key areas in the radar reflectivity image while retaining the global information of the image, so that the model achieved the best prediction effect.

Next, we plotted the error curves of models for each precipitation level in Figure 10.

It can be seen from Figure 10, as the precipitation level increases, the error of all models becomes larger. However, all the deep learning-based models perform better than the traditional method GBDT, which illustrates that the deep learning model can better capture important features in radar reflectivity images. For dual-channel CNN with attention, dual-channel CNN, single-channel CNN (Baseline), and SVM, although their effects are significantly improved compared with GBDT, the proposed method has the best performance almost at all levels. Since the residual connection decreases the loss of global information caused by attention, our model MAR-CNN not only highlights the information that affects precipitation in the image but also preserves the global information of radar reflectivity images

Table 6: Predicted average of different models under different precipitation levels.

Model	No rain	Light rain	Moderate rain	Heavy rain	Rainstorm	Big heavy rain	Extraordinary heavy rain
Observation	0	1.13	3.47	10.14	26.87	44.03	58.97
Dual-channel MAR-CNN	0.06	0.16	2.30	12.34	24.80	32.99	50.23
Dual-channel CNN with attention	2.50	3.27	6.76	11.24	25.60	32.64	47.38
Dual-channel CNN	0	0.01	0.97	9.23	24.21	28.03	49.21
Single-channel CNN (baseline)	5.09	3.99	7.34	12.52	23.91	32.20	48.33
GBDT	7.55	8.24	11.09	12.63	21.53	25.47	27.31
SVM	4.06	2.95	6.15	11.38	22.48	30.01	46.45

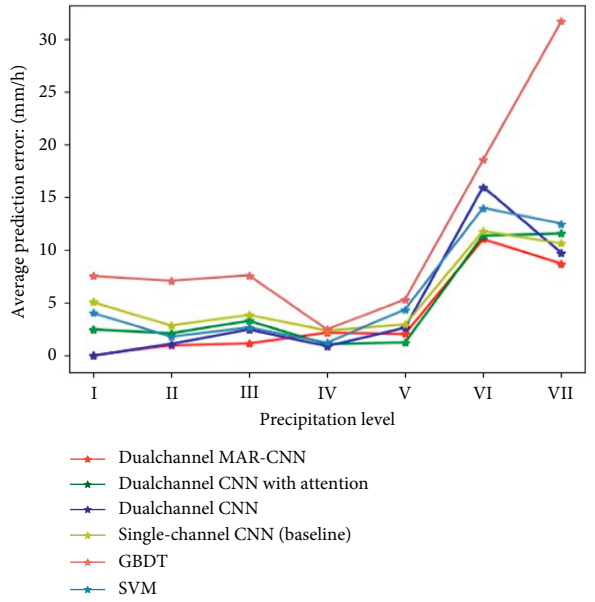

Figure 10: Error curves of each model under different precipitation levels (I~VII matches to the precipitation level mentioned in Table 1, from no rain to extraordinary heavy rain).

better than the original attention. The prediction results achieved by MAR-CNN are more robust in all levels of precipitation.

4. Conclusions

This paper proposed a dual-channel multihead attention model combined with a residual connection based on deep learning. Extensive experiments validated that, by adding multihead attention to CNN, the model can extract the local spatial feature of radar reflectivity images precisely. At the same time, the residual connection introduced can well retain the global information based on attention. The results showed that both global and local features are of great significance for precipitation prediction. Moreover, the second channel in the proposed dual-channel network can effectively extract information of the moving speed, size, etc. of the cloud. Compared with other algorithms, the proposed model has better prediction performance. Moreover, as demonstrated in experiments, the training convergence of the model is fast and stable. As a result, the proposed two-channel MAR-CNN model provides a new effective scheme for the spatiotemporal characteristics extraction in precipitation forecasting.

Data Availability

In our research, the data were published by CIKM AnalytiCup 2017 with the link https://tianchi.aliyun.com/dataset/?spm=5176.12281905.0.0.358b5699fDHRjX.

Disclosure

Qing Yan and Fuxin Ji are the first authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Qing Yan and Fuxin Ji contributed equally to the paper.

Acknowledgments

This work was supported by the National Science Foundation of China (no. 61602002) and the Anhui Provincial Natural Science Foundation (grant no. 1808085MF209).

References

[1] X. Shi, Z. Gao, L. Leonard et al., "Deep learning for precipitation forecast: a benchmark and A new model," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, 2017.

[2] J. Sun, M. Xue, J. W. Wilson et al., "Use of NWP for nowcasting convective precipitation: recent progress and challenges," *Bulletin of the American Meteorological Society*, vol. 95, no. 3, pp. 409–426, 2014.

[3] W. C. Woo and W. K. Wong, "Application of optical flow techniques to rainfall nowcasting," in *Proceedings of the the 27th Conference on Severe Local Storms*, Madison, WI, USA, November 2014.

[4] Z. Liu, Q. He, and J. Luo, "Spatial angular compounding with affine-model-based optical flow for improvement of motion estimation," *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 66, no. 4, pp. 701–716, 2019.

[5] H. Guo and M. Chen, "High convection high resolution proximity forecasting experiments based on deep learning," *Journal of Meteorology*, 2019.

[6] M. Chen and X. Yu, "Development and research progress of convective weather nowcasting technology," *Journal of Applied Meteorology*, vol. 6, pp. 115–127, 2004.

[7] M. L. Weisman, C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, "Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model," *Weather & Forecasting*, vol. 23, no. 3, pp. 407–437, 2010.

[8] D. Chen and J. Xue, "The present situation and Prospect of operational model of numerical weather forecast," *Journal of Meteorology*, no. 5, pp. 112–122, 2004.

[9] X. Shi, Z. Chen, H. Wang et al., "Convolutional LSTM network: a machine learning approach for precipitation forecast," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] Y. Yao and Z. Li, "Short-term precipitation forecasting based on radar reflectivity images," in *Proceedings of the CIKM AnalytiCup*, Singapore, November 2017.

[12] T. H. Shin, B. Manavalan, and G. Lee, "PVP-SVM: sequence-based prediction of phage virion proteins using a support vector machine," *Frontiers in Microbiology*, vol. 9, p. 476, 2018.

[13] C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1179–1188, Salt Lake City, UT, USA, June 2018.

[14] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," 2017, https://arxiv.org/abs/1706.03762.

[15] M. Stollenga, J. Masci, F. Gomez et al., "Deep networks with internal selective attention through feedback connections," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[16] X. Chu, W. Yang, W. Ouyang et al., "Multi-context attention for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[17] X. Wang, R. Girshick, A. Gupta et al., "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.

[18] E. N. Mortensen and H. Deng, "A SIFT descriptor with global context," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, San Diego, CA, USA, June 2005.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2011.

[20] Y. Lecun, Y. L. Bottou, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[21] A. Parikh, O. Täckström, D. Das et al., "A decomposable attention model for natural language inference," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, USA, November 2016.

[22] R. Paulus, C. Xiong, and R. Socher, "A deep reinforced model for abstractive summarization," 2017, https://arxiv.org/abs/1705.04304.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[24] H. Zheng, J. Fu, M. Tao et al., "Learning multi-attention convolutional neural network for fine-grained image recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October 2017.

[25] T. R. Chiang, C. W. Huang, S. Y. Su, and Y.-N. Chen, "Learning multi-level information for dialogue response selection by highway recurrent transformer," Article ID 101073, 2019.

[26] K. He, X. Zhang, S. Ren et al., "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[27] F. Wang, M. Jiang, C. Qian et al., "Residual attention network for image classification," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164, Honolulu, HI, USA, July 2017.

[28] O. Turan, A. Sachdeva, R. J. Poole et al., "Laminar natural convection of power-law fluids in a square enclosure with differentially heated sidewalls subjected to constant wall heat flux," *Journal of Non-newtonian Fluid Mechanics*, vol. 166, no. 17-18, pp. 1049–1063, 2012.