

RESEARCH ARTICLE



Study on Nowcasting Method of Severe Convective Weather Based on SA-PredRNN++

Qiongying Xue¹ and Fei Fang^{2,*}

¹*School of Software, Nanjing University of Information Science and Technology, China*

²*NUIST Reading Academy, Nanjing University of Information Science and Technology, China*

Abstract: Severe convective weather, characterized by short-term intense precipitation, thunderstorms, and strong winds, poses significant threats to human life and property. Therefore, accurate and efficient prediction of severe convective weather is crucial for disaster prevention. Currently, utilizing deep learning for radar echo extrapolation stands as the primary method for forecasting severe convective weather. We propose a predictive recurrent neural network model that integrates a self-attention mechanism, specifically designed for radar echo extrapolation in severe convective weather forecasting. The self-attention mechanism offers the advantage of being lightweight, as it does not substantially increase the model parameters. Additionally, it facilitates global attention extraction, thereby enhancing the model's accuracy to some extent. By utilizing radar echo images from the previous hour as input, the model undergoes self-learning to achieve the best forecast for radar echo extrapolation in the subsequent two hours. Research findings demonstrate that our model outperforms other models in accurately predicting severe convective weather within this two-hour timeframe.

Keywords: severe convective weather, deep learning, radar echo extrapolation, predictive recurrent neural network

1. Introduction

Intense convective weather events can have catastrophic impacts. For example, during the period from July 18th to August 2021, Zhengzhou, China, experienced heavy rainfall, leading to flash floods and significant disruptions to transportation systems. As of 6:00 PM on August 1st, 2021, this extreme weather event had affected a staggering 1.8849 million people and resulted in 292 fatalities [1]. In China, existing weather forecasting methods lack adequate prediction techniques for medium- and small-scale intense convective weather events [2]. Additionally, there is an urgent demand within the meteorological industry for accurate and timely forecasting of intense convective weather events. Therefore, finding effective and accurate methods to forecast intense convective weather has become a challenging task and a key focus area within meteorological research. Nowcasting is a technique used in weather forecasting aimed at providing real-time updates on weather conditions.

Zhang and Yang [3], Yu and Zheng [4] highlighted the advancements made in both research and operations related to severe convective weather. Specifically, there are two primary approaches for severe convective nowcasting: one being the extrapolation technology based on radar echoes and the other being the utilization of numerical weather forecast models [5]. In recent years, deep learning has emerged as a significant approach for the prediction of severe convective weather [6], and significant progress has been made in the study of deep learning-based severe convective weather forecasting, both domestically and internationally [7]. Hence, we aim

to employ deep learning techniques for extrapolating radar echoes to enhance the accuracy of severe convection nowcasting. Lu et al. [8] investigated a model for recognizing heavy precipitation weather in severe convective weather conditions, utilizing physical parameters and the deep learning model DBNs. Zhou et al. [9] conducted short-term lightning forecasting by employing a deep learning fusion of a semantic segmentation model and multisource observation data. Lee et al. [10] devised a predictive model based on deep learning methods to anticipate the potential impact of a rainstorm on an area before it transpires.

Currently, widely used deep learning models in severe convective nowcasting include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN), and their variants. Sherstinsky [11] provided an overview of RNNs and Long Short-Term Memory Networks (LSTM). Han et al. [12] effectively employed RNNs for nowcasting, while Shi et al. [13] introduced Convolutional Long Short-Term Memory Networks (ConvLSTM) for precipitation nowcasting. Additionally, Shi et al. [13] enhanced the model by incorporating a Gated Recurrent Unit to create the Convolutional Gated Recurrent Unit (ConvGRU), which has fewer parameters. Guo et al. [5] applied the ConvGRU model to extrapolation experiments and achieved significant progress in forecast accuracy. Wang et al. [14] proposed a Spatiotemporal Long Short-Term Memory (ST-LSTM) and later constructed a Memory in Memory (MIM) network for precipitation nowcasting, both of which outperform ConvLSTM. Tian et al. [15] introduced the Generative Adversarial Convolutional Gate Recurrent Unit (GA-ConvGRU) model, effectively addressing the limitations associated with fuzzy extrapolation images. Lin et al. [16] enhanced the ConvLSTM model by combining it with a self-attention mechanism to capture

*Corresponding author: Fei Fang, NUIST Reading Academy, Nanjing University of Information Science and Technology, China. Email: 202183710004@nuist.edu.cn

extended spatiotemporal relationships. Adewoyin et al. [17] developed the Temporal Recurrent U-Net (TRU-NET), which has more parameters, longer training times, and lower nowcasting capability for severe convection.

Based on the previous research, CNN models focus more on spatial information in severe convective forecasting, while traditional RNN or LSTM models face challenges in capturing long-term data and are prone to the issue of vanishing gradients. To improve the accuracy and efficiency of model training, we chose the PredRNN++ model as our network architecture. This model accurately predicts regions with detectable echo signatures. To further enhance performance, we introduced the ProbSparse self-attention mechanism and utilized a weighted loss function during training.

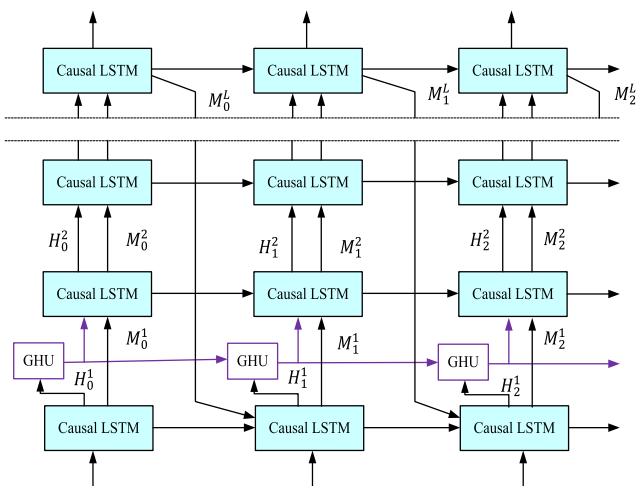
Our contributions can be summarized as follows:

- (1) Addressing the issue of inaccurate forecasts caused by the aging of high-intensity echo wavelengths, we developed a predictive recurrent neural network model with a self-attention mechanism.
- (2) Applying the model to timely forecast severe convective weather using weather radar data.
- (3) Validating our approach on existing datasets, demonstrating superior performance compared to existing deep learning methods.

2. Related Work

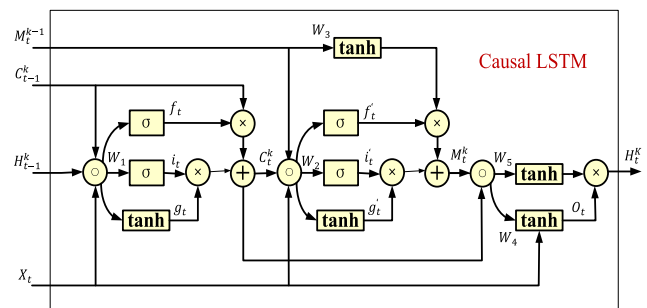
PredRNN++ [18] is a recurrent network designed for spatiotemporal predictive learning. It differs from its predecessor, PredRNN, in several aspects. Firstly, it replaces the ST-LSTM unit with the Causal Long Short-Term Memory (Causal LSTM) unit. Additionally, it introduces the Gradient Highway Unit (GHU) structure in the first and second layer stack. The Causal LSTM unit is a revolutionary recurrent unit that enhances the transition depth between adjacent states. On the other hand, the GHU facilitates the connection between input information, allowing for improved information flow within the model. This architecture allows for high-speed propagation of gradients between the first and second layers, effectively solving the issue of gradient vanishing. It also enables better learning of feature values from previous frames and more efficient utilization of long-term information. Figure 1 explains the structure of PredRNN++.

Figure 1
PredRNN++ model structure diagram



The Causal LSTM introduces additional nonlinear layers in periodic transitions, enhancing feature amplification to better capture abrupt situations caused by short-term changes. It comprises two memory components: temporal memory and spatial memory, as illustrated in Figure 2. The first layer resembles the generic LSTM structure for updating the temporal state C . The second layer mirrors the structure of the first layer, utilizing the updated temporal state C , and the updated spatial state M , from the previous layer. Lastly, the third layer updates the output gate structure, H , using inputs X , C , and M from the previous layer. A Causal LSTM unit takes inputs X , H , C , and M and produces updated H , C , and M outputs. The H value from the last time step is used to predict the generated target sequence.

Figure 2
Causal LSTM structure diagram



where C_t^k represents temporal memory and M_t^k represents spatial memory, where the subscript t represents the step size, and the superscript k denotes the k th hidden layer in the stacked Causal LSTM network. The current network memory is dependent on the last state C_{t-1}^k and is controlled by forget gate f_t , input gate i_t , and input modulation gate g_t . The current spatial memory M_t^k is dependent on M_{t-1}^k at the input Causal LSTM uses a cascade mechanism, and spatial memory is a temporal memory function that passes through another set of gated structures. The updated equation of the k th layer Causal LSTM is:

$$\begin{pmatrix} g_t \\ i_t \\ f_t \end{pmatrix} = \begin{pmatrix} \tanh \\ \sigma \\ \sigma \end{pmatrix} W_1 * [X_t, H_{t-1}^k, C_{t-1}^k] \quad (1)$$

$$C_t^k = f_t \odot C_{t-1}^k + i_t \odot g_t \quad (2)$$

$$\begin{pmatrix} g'_t \\ i'_t \\ f'_t \end{pmatrix} = \begin{pmatrix} \tanh \\ \sigma \\ \sigma \end{pmatrix} W_2 * [X_t, C_t^k, M_{t-1}^k] \quad (3)$$

$$M_t^k = f'_t \tanh \odot (W_3 * M_{t-1}^k) + i'_t \odot g'_t \quad (4)$$

$$o_t = \tanh(W_4 * [X_t, C_t^k, M_t^k]) \quad (5)$$

$$H_t^k = o_t \odot \tanh(W_5 * [C_t^k, M_t^k]) \quad (6)$$

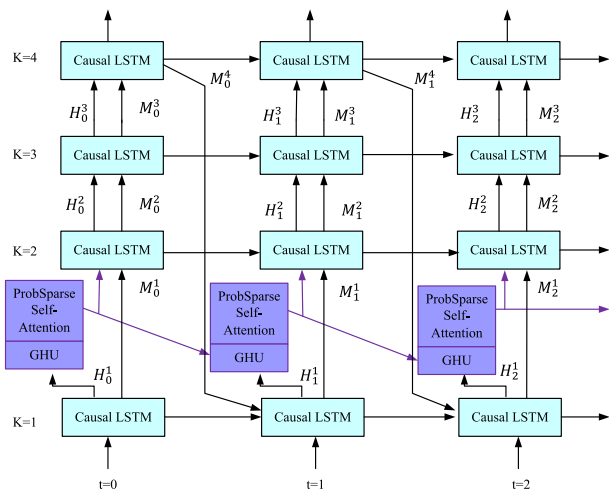
“*” in the above formula represents the convolution operation, “ \odot ” denotes the Hadamard product, “ σ ” refers to the sigmoid function in the activation function, and $W_{1\sim 5}$ represents the convolution filter.

3. Research Methodology

3.1. SA-PredRNN++

Due to the abundance of data and limited availability of target data, predicting severe convective weather has always been a challenge. Additionally, PredRNN++ faces difficulties in extracting deep network features while identifying local dependencies and global features. To address this issue, we introduce the ProbSparse self-attention mechanism [19] into the second layer's GHU and causal LSTM layer. This measure not only enhances the model's capability to extract global features and local dependencies but also maintains the lightweight nature of the model, without increasing its complexity. This improvement aids in model training and enhances prediction accuracy. The new model is named Self-Attention PredRNN++ (SA-PredRNN++), which effectively addresses the issue of gradient disappearance while increasing memory utilization and strengthening the model's long-term modeling capacity. The SA-PredRNN++ structure excels in capturing both short-term and long-term features. Figure 3 illustrates the structure of SA-PredRNN++.

Figure 3
SA-PredRNN++ model structure diagram



The cascaded causal LSTM structure is highly effective; however, it encounters issues such as gradient vanishing during the backpropagation process, especially in scenarios involving periodic motion or frequent occlusion. Additionally, prolonged transmission times may result in unclear features. To alleviate these challenges and ensure the model can continuously and accurately learn frame features, we introduce a GHU, as depicted in Figure 4. The GHU unit is specifically designed to learn the relationship between frame skipping, thereby enabling better feature extraction.

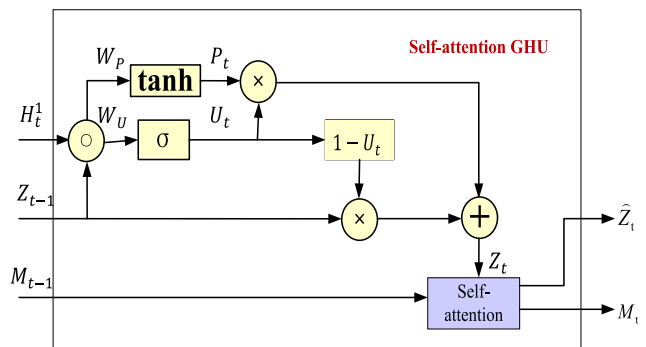
The equation of GHU can be expressed as:

$$P_t = \tanh(W_{px} * X_t + W_{pz} * Z_{t-1}) \quad (7)$$

$$U_t = \sigma(W_{sx} * X_t + W_{sz} * Z_{t-1}) \quad (8)$$

$$Z_t = U_t \odot P_t + (1 - U_t) \odot Z_{t-1} \quad (9)$$

Figure 4
Self-attention GHU structure diagram



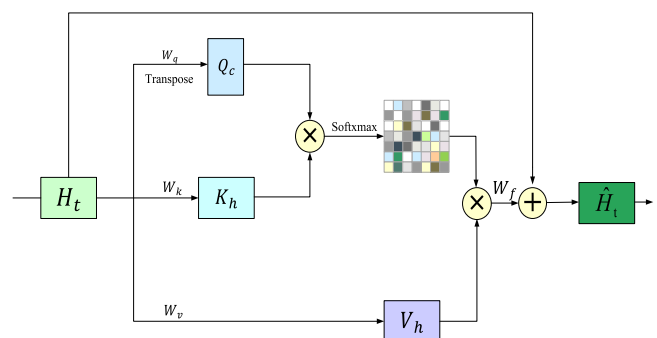
where P_t signifies the input of the transition, U_t signifies the switch gate, and Z_t signifies the hidden state. We can express the above as:

$$Z_t = GHU(X_t, Z_{t-1}) \quad (10)$$

The attention mechanism plays a crucial role in assigning varying weights to different samples based on their unique features. This enables the extraction of relevant information for data analysis and prediction, thereby improving evaluation results and accelerating model training speed. For instance, the ConvSeq2Seq model proposed by Gehring et al. [20] integrates an independent attention module in each decoder layer, combined with CNNs.

In recent years, self-attention modules have garnered attention in sequence prediction tasks. The Transformer model, introduced by Vaswani et al. in 2017 [21], utilizes self-attention mechanisms to capture long-range dependencies within sequences. This mechanism facilitates the exchange of information and features between network layers, progressively strengthening spatial dependencies from local to global regions and temporal dependencies from internal to external segments. A schematic representation of the self-attention mechanism module is depicted in Figure 5.

Figure 5
Self-attention structure diagram



To improve the accurate determination of feature positions during the model training process and speed up the model training, we added a ProbSparse Self-Attention mechanism between the GHU and Causal LSTM layers. Among them, Q, K, and V are the input embeddings and multiply a matrix with weights

to obtain query, key, and value. ProbSparse Self-Attention first samples K to obtain K samples and then obtains the value of M for the relationship between each $q_i \in Q$ and K :

$$M(q_i, K) = \max_j \left\{ \frac{q_i k_k^T}{\sqrt{d}} \right\} - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_k^T}{\sqrt{d}} \quad (11)$$

The self-attention mechanism differs from the attention mechanism by incorporating a scaling factor during the attention computation, thereby alleviating overflow concerns arising from large inner products. Within the self-attention mechanism, the correlation between each input sequence is computed using formula (12), where the similarity between a Query and a Key is determined through their dot product. Subsequently, the Softmax function is applied to yield the weights, which are then used to measure the values and generate the output vector of self-attention.

$$A(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (12)$$

3.2. Loss function

In severe convective weather conditions, the radar echo reflectivity values are typically expected to exceed 30 dBZ (basic reflectivity). Shi et al. [22] conducted a study where they graded the rainfall intensity and assigned varying weights to different precipitation zones based on radar reflectivity values. The reflectivity values of each pixel point in the radar echogram generally range from 0 dBZ to 75 dBZ within the effective prediction range and are classified into distinct categories [23]. As shown in formula (13), $w(x)$ represents the weight, and x is the value of each segment dBZ.

$$w(x) = \begin{cases} 1 & x \in [0, 20) \\ 10 & x \in [20, 30) \\ 15 & x \in [30, 40) \\ 30 & x \in [40, 50) \\ 50 & x \in [50, 75) \end{cases} \quad (13)$$

The Mean Squared Error (MSE) is frequently used to represent the overall error in recognition outcomes. The formula for calculation is:

$$MSE = \frac{1}{n} \sum_i^n [X(i) - Y(i)]^2 \quad (14)$$

where X is the standard value, Y is the model identification value, and n represents the total sample count. MSE quantifies the accuracy of severe convection forecasts, with smaller MSE values indicating smaller errors.

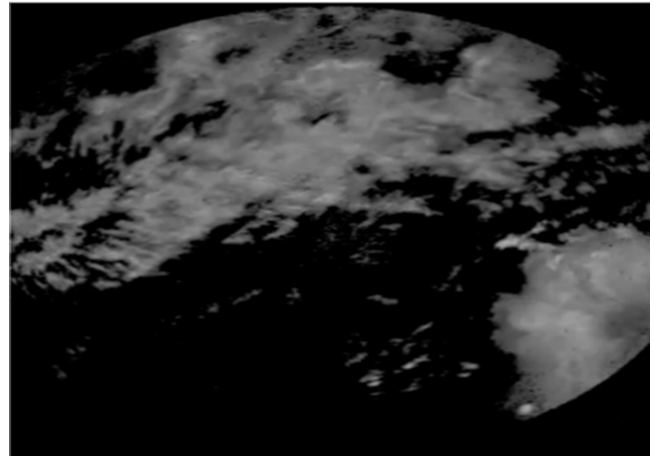
3.3. Dataset

The weather radar dataset we utilized is obtained from the Artificial Intelligence Weather Forecast Innovation Competition organized by the Shanghai Meteorological Bureau in China. Each sample in the radar dataset comprises pixel values organized in a 500×500 format. The dataset spans 3 h, with input data covering 1 h at 6-minute intervals and target data covering 2 h at 12-minute intervals, resulting in a total of 20 instances.

Regarding the horizontal grid point range, the radar data samples have a resolution of 0.01° . Radar echoes undergo quality control via correlation coefficient analysis. The accepted data range is from 15 to 70 dBZ. Any values exceeding 70 dBZ are capped at 70, while values below 15 dBZ or missing values are set to 0.

For training purposes, we utilized a dataset consisting of 40,000 radar map samples, while the testing dataset contained 3,000 radar map samples. An example of a grayscale radar data image can be seen in Figure 6.

Figure 6
Grayscale image of radar data



The connection between image gray value and integrated reflectivity:

$$G = \text{round} \left(255 * \frac{\text{dbz}}{70} \right) \quad (15)$$

4. Experiments

4.1. Model monitoring metrics

In meteorology, Probability of Detection (POD), Critical Success Index (CSI), and False Alarm Ratio (FAR) are widely used to assess severe convective weather forecast outcomes. POD and CSI are utilized to measure the likelihood and accuracy of severe convective weather events. Higher values of POD and CSI indicate more accurate predictions of severe convective weather events. Conversely, lower FAR values indicate a lower false alarm rate in model predictions. Here N_1 , N_2 , and N_3 represent the number of hits, false alarms, and misses, respectively.

$$POD = \frac{N_1}{N_1 + N_3} \quad (16)$$

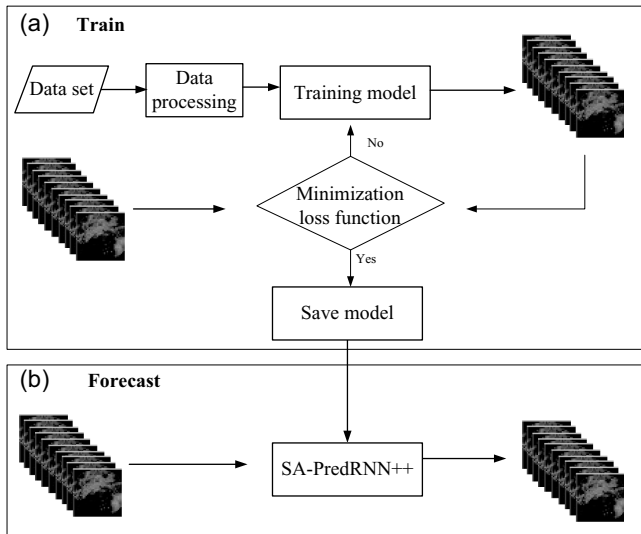
$$CSI = \frac{N_1}{N_1 + N_2 + N_3} \quad (17)$$

$$FAR = \frac{N_2}{N_1 + N_2} \quad (18)$$

4.2. Experimental scheme

Figure 7 depicts the model's initial phase, which includes the training process. The model receives input at time t , which consists of 10 radar echo data. Subsequently, the output of the model is a future map of 10 radar echoes, which is generated by analyzing all the elements at each moment. A loss function is then computed by summing the average loss function value of each frame of the radar echo forecast map. In the training process, the objective is to

Figure 7
Training process and forecasting process

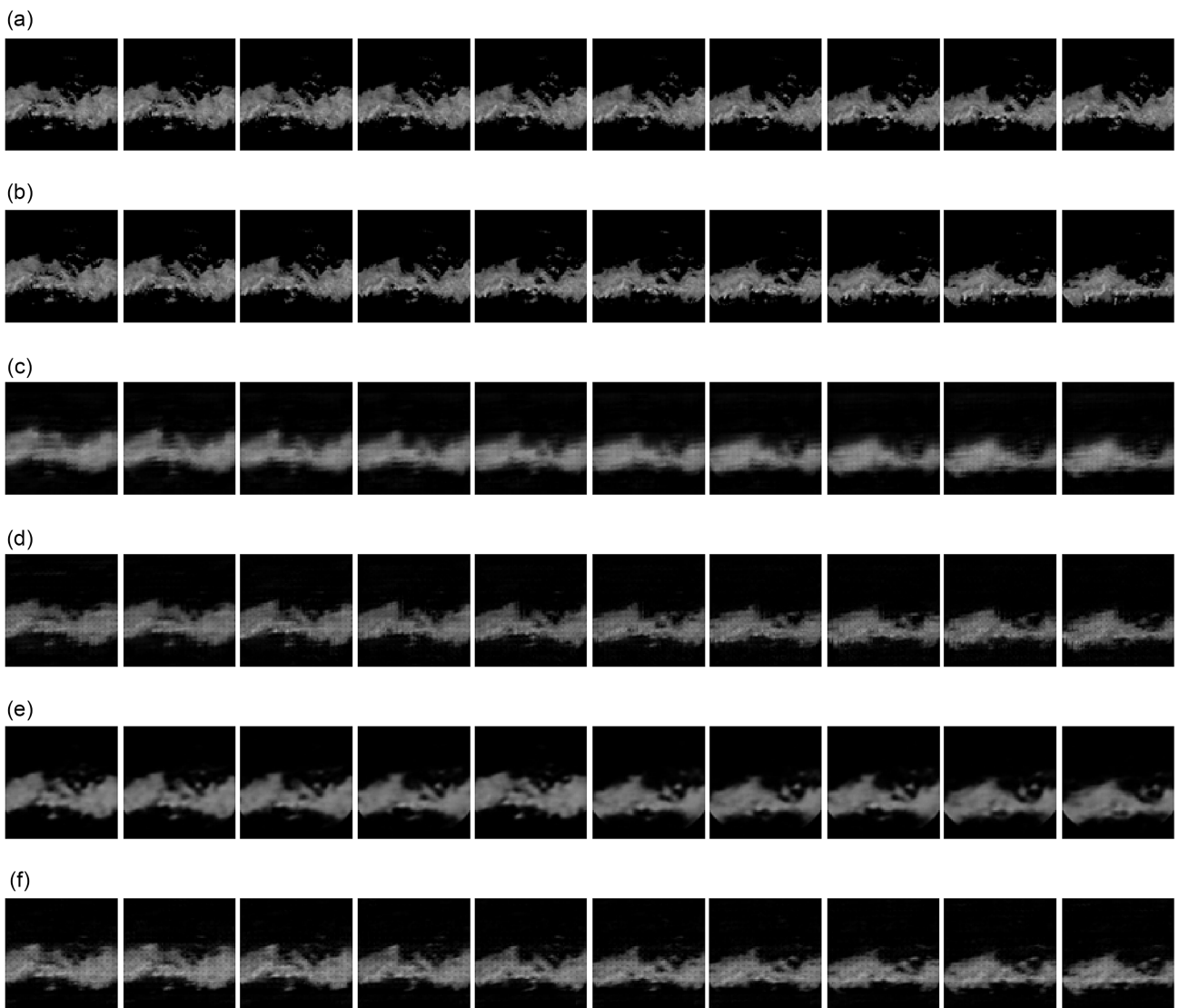


minimize the loss function value. To accomplish this, the backpropagation algorithm, specifically Adam, is employed for training and learning purposes. The training process parameter specifications are as follows: the initial learning rate is set to 10^{-4} , the learning rate penalty factor is 0.5, the batch size is 30, and the required number of iterations is 30.

4.3. Experimental results

Figure 8 shows the forecast visualization results of radar echograms, where group (a) represents the time series of radar echograms obtained over one hour. (b) Panel is the real output image sequence (over the next 2 h). (c), (d), (e), and (f) are four sets of experimental results, respectively, showing the forecast output of the ConvLSTM, PredRNN++, MIM, and SA-PredRNN++ algorithms for the next 2-hour radar echo map sequence. Radar echoes constantly undergo accumulation, deformation, and dissipation. In Figure 8, a visual comparison is presented between the 2-hour radar live data and the model forecast. The experimental results demonstrate that the algorithm produces clearer outcomes, effectively capturing the spatial and temporal features of radar maps at various elevations. However, there is

Figure 8
Comparison of radar echo images of severe convection near forecast results of various models



room for improvement in capturing fine-grained features in later frames and scattered areas. Consequently, the model utilizes radar map data to forecast rainfall amounts for the next 2 h with heightened precision. Upon visual comparison of forecasted images, it's evident that while the ConvLSTM model's predictions appear blurry, the PredRNN++ model yields clearer predictions. However, both models encounter challenges in accurately capturing regions with high radar echo intensity.

While the MIM model excels in capturing areas with high radar echo reflectivity, it may exhibit some blurriness in subsequent frames and struggle to accurately depict features in sparse regions. Overall, experimental results suggest that the MIM model provides a more precise representation of regions with high radar echo reflectivity. However, improvements are still needed to capture finer features in subsequent frames and sparse areas.

The prediction of severe convective weather relies on binarizing grid data, where we select radar reflectivity levels of 40 dBZ and 50 dBZ. Additionally, we evaluate the extrapolation performance of the models using grid data, employing three evaluation metrics (CSI, POD, and FAR), with results presented in Table 1. The ConvLSTM, PredRNN++, MIM, and SA-PredRNN++ models are tested on the evaluation set, predicting the next 10 frames based on the preceding 10 frames. These models are fairly compared in terms of experimental settings and hyperparameters.

Table 1
A comparison for predicting 10 frames on the subsets of the radar data set

Threshold	Model	CSI	POD	FAR
40	ConvLSTM	0.279	0.306	0.593
	PredRNN++	0.278	0.638	0.669
	MIM	0.298	0.516	0.589
	SA-PredRNN++	0.384	0.674	0.527
50	ConvLSTM	0.152	0.269	0.740
	PredRNN++	0.195	0.294	0.632
	MIM	0.171	0.253	0.651
	SA-PredRNN++	0.233	0.337	0.575

As depicted in Table 1, the SA-PredRNN++ model outperforms the other models and significantly reduces the model's parameters, showcasing the efficacy of the self-attention mechanism. However, ConvLSTM does not perform well with either the 40 or 50 threshold. Although ConvLSTM is proficient in handling time series problems and focuses on convolution for spatial feature extraction, it heavily relies on convolutional layers to capture spatial relationships and is unable to extract long-term features.

For the accurate forecast of severe convective weather based on radar echo extrapolation, each image frame contains abundant spatial information, with closely interconnected data points. However, it still cannot capture long-term dependencies effectively. Although PredRNN++ and MIM have made further improvements on ConvLSTM, they still fall short in capturing deep network features due to deficiencies in feature extraction of images and excessive parameters. Despite this, the PredRNN++ model shows better performance according to the POD value, achieving 0.638 and 0.400 at the 10th frame and threshold values of 40 and 50, respectively. However, when compared with other evaluation indicators, the SA-PredRNN++ model outperforms at a threshold of 40 with CSI and POD reaching best results of 0.308 and 0.553, respectively, and FAR also performing better at 0.588 at the 10th frame. Similarly, at

a threshold of 50, the SA-PredRNN++ model achieves the best results for CSI and POD at 0.195 and 0.294 at the 10th frame, respectively, with FAR also showing better performance at 0.632. In conclusion, the SA-PredRNN++ model in this study has demonstrated better performance compared to other models. Figure 9 illustrates the superior CSI value of the SA-PredRNN++ model in predicting the performance of 10 frames. As the forecast time increases, the CSI value decreases for all models, with a slower decrease for our model. Figure 10 shows a comparison between the two-hour SA-PredRNN++ approach forecasts (b) and radar observations (a). It can be seen that SA-PredRNN++ can have a good prediction of the echo drop zone.

Figure 9
Frame-wise analyses of the next 10 generated radar maps

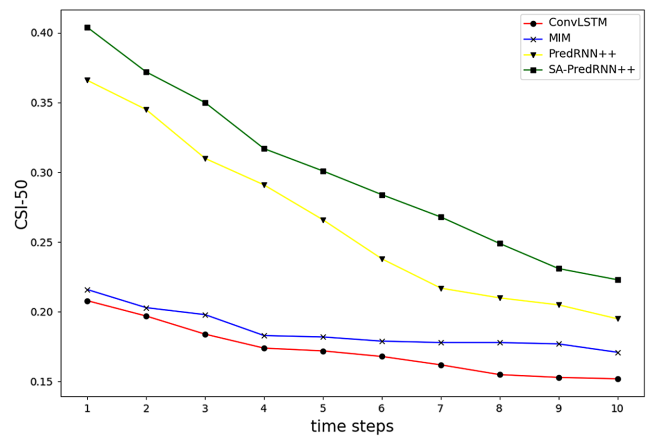
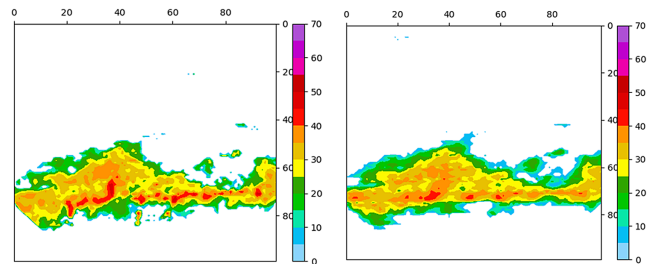


Figure 10
Comparison between radar observation and nowcast by SA-PredRNN++ model



The ablation study in Table 1 demonstrates that integrating the ProbSparse self-attention mechanism into the PredRNN++ model, along with GHU units and a second layer of causal LSTM, positively impacts experimental results. Setting the threshold to 40 increases the PredRNN++ score from 0.278 to 0.384. These findings indicate that the self-attention mechanism within the model enhances connectivity in spatiotemporal memory compared to models lacking this mechanism. Additionally, Table 1 also suggests that pairing GHU units with the self-attention mechanism consistently enhances model performance. In Table 2, various network variants are discussed to illustrate different placements for the self-attention mechanism within the model. As a control experiment, we relocated the self-attention mechanism between the second and third layers of the Causal LSTM, resulting in a CSI

Table 2

The slot of the self-attention mechanism is located by the index (k_1, k_2) of the causal LSTM to which it is attached

LOCATION	k_1, k_2	CSI	POD	FAR
BOTTOM(SA-PredRNN++)	1,2	0.384	0.674	0.527
MIDDLE	2,3	0.317	0.496	0.531
TOP	3,4	0.382	0.544	0.537

score of 0.317. Placing the self-attention mechanism between the third and fourth layers increased the CSI score to 0.382. This positioning indicates that configuring the unit behind the GHU unit was most effective, as it enables the GHU unit to prioritize long-term highway features, short-term deep transition path features, and spatial features extracted from the current input frame. The self-attention mechanism facilitates enhanced feature extraction.

5. Conclusion

We have provided a detailed introduction to our deep learning model SA-PredRNN++. Our model utilizes radar echo data to extrapolate variations of the reflectivity factor in both time and space, enabling the forecasting of severe convective weather. The SA-PredRNN++ model integrates improved GHU units and incorporates a self-attention mechanism into its second layer. This allows us to capture spatial information of meteorological elements such as the reflectivity factor throughout the entire time series, thereby accelerating the convergence of the model. Furthermore, our model employs a unique neural network framework and a weighted loss function based on weight construction, enhancing its predictive capabilities for regions with strong echo and enabling more accurate forecasting of severe convective weather. Due to its exceptional ability to capture both temporal and spatial features, our model is highly suited for a wide range of meteorological forecasting tasks. Experimental results demonstrate that our model is remarkably precise in forecasting the position, contour, distribution, and progression of severe convective weather.

However, it is important to note the limitations of the SA-PredRNN++ model. The model heavily relies on historical data for parameter calibration, and in this study, the available dataset lacked a sufficient number of instances of severe convective weather. To address this issue, the dataset was filtered to include only data above 30 dBZ and underwent noise reduction and rotation operations. The experimental results show an improvement in output precision. However, due to the limitations of available data, the accuracy of predicting intense echo areas noticeably decreases as the prediction duration increases. Future improvements to the model should focus on optimizing the construction process, such as integrating numerical simulation data at each time point to correct transmission errors. Additionally, integrating other meteorological features such as temperature and wind field information will contribute to enhancing the accuracy of nowcasting severe convection.

Recommendations

In the future, the prediction of meteorological disasters is a crucial area of research, playing a significant role in safeguarding lives and property, as well as promoting social stability and development. We urge for greater attention to be given to this field by the general public.

Funding Support

This work was funded by the National Natural Science Foundation of China (Grant No.42075007), the Open Grants of the State Key Laboratory of Severe Weather (No.2021LASW-B19), and the Open Project of the Provincial Key Laboratory for Computer Information Processing Technology under Grant KJS1935, Soochow University.

Ethical Statement

No research involving either human or animal subjects was conducted by any of the authors of this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data used in this study is a competition dataset, which is not an open source at this time. It is available from the corresponding author upon reasonable request.

References

- [1] Su, A. F., Lü, X. N., & Cui, L. M. (2021). The basic observational analysis of “7.20” extreme rainstorm in Zhengzhou. *Torrential Rain and Disasters*, 40(05), 445–454.
- [2] Cui, C. G., Du, M. Y., Xiao, Y. J., Lai, A. W., Li, H. L., Wang, Z. B., . . . , & Kang, Z. P. (2021). Study on the technique of data assimilation and nowcasting of severe convective weather. *Meteorological Monthly*, 47(08), 901–918.
- [3] Zhang, X. L., & Yang, B. (2018). Development of operations on forecasting severe convective weather in China. *Advances in Meteorological Science and Technology*, 8(03), 8–18.
- [4] Yu, X. D., & Zheng, Y. G. (2020). Advances in severe convective weather research and operational service in China. *Acta Meteorological Sinica*, 78(03), 391–418.
- [5] Guo, H. Y., Chen, M. X., & Han, L. (2019). High resolution nowcasting experiment of severe convections based on deep learning. *Acta Meteorological Sinica*, 77(04), 715–727.
- [6] Zhou, K. H., Zheng, Y. G., Han, L., & Dong, W. S. (2021). Advances in application of machine learning to severe convective weather monitoring and forecasting. *Meteorological Monthly*, 47(03), 274–289.
- [7] Fang, W., Xue, Q. Y., Shen, L., & Sheng, V. S. (2021). Survey on the application of deep learning in extreme weather prediction. *Atmosphere*, 12(6), 661.
- [8] Lu, Z. Y., Ren, Y. M., & Sun, X. L. (2018). Recognition of short-time heavy rainfall based on deep learning. *Journal of Tianjin University*, 51(02), 111–119.
- [9] Zhou, K. H., Zheng, Y. G., & Wang, T. B. (2021). Very short-range lightning forecasting with NWP and observation data: A deep learning approach. *Acta Meteorological Sinica*, 79(01), 1–14.
- [10] Lee, K., Choi, C., Shin, D. H., & Kim, H. S. (2020). Prediction of heavy rain damage using deep learning. *Water*, 12(7), 1942.
- [11] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306.

- [12] Han, F., Long, M. S., Li, Y. A., Xue, F., & Wang, J. M. (2019). The application of recurrent neural network to nowcasting. *Journal of Applied Meteorological Science*, 30(01), 61–69.
- [13] Shi, X. J., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.
- [14] Wang, Y. B., Long, M., Wang, J., Gao, Z., & Yu, P. S. (2017). PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 879–888.
- [15] Tian, L., Li, X., Ye, Y., Xie, P., & Li, Y. (2019). A generative adversarial gated recurrent unit model for precipitation nowcasting. *IEEE Geoscience and Remote Sensing Letters*, 17(4), 601–605.
- [16] Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C. (2020). Self-attention convlstm for spatiotemporal prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 11531–11538.
- [17] Adewoyin, R. A., Dueben, P., Watson, P., He, Y., & Dutta, R. (2021). TRU-NET: A deep learning approach to high resolution prediction of rainfall. *Machine Learning*, 110(8), 2035–2062.
- [18] Wang, Y., Gao, Z., Long, M., Wang, J., & Yu, P. S. (2018). PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *International Conference on Machine Learning*, 5123–5132.
- [19] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12), 11106–11115.
- [20] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. In *International Conference on Machine Learning*, 1243–1252.
- [21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . , & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
- [22] Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in Neural Information Processing Systems*, 30, 5622–5632.
- [23] Huang, X. Y., Ma, Y. R., & Hu, S. M. (2021). Extrapolation and effect analysis of weather radar echo sequence based on deep learning. *Acta Meteorologica Sinica*, 79(5), 817–827.

How to Cite: Xue, Q., & Fang, F. (2024). Study on Nowcasting Method of Severe Convective Weather Based on SA-PredRNN++. *Journal of Data Science and Intelligent Systems*. <https://doi.org/10.47852/bonviewJDSIS42022197>