ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

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

Research article

A self-attention integrated spatiotemporal LSTM approach to edge-radar echo extrapolation in the Internet of Radars

Zhiyun Yang [a,1], Hao Wu [a,1], Qi Liu [a,*], Xiaodong Liu [b], Yonghong Zhang [c], Xuefei Cao [d]

[a] School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[b] School of Computing, Edinburgh Napier University, Edinburgh, EH10 5DT, UK
[c] School of Automation, Nanjing University of Information Science Technology, Nanjing, 210044, China
[d] School of Cyber and Information Security, Xidian University, Xi'an, 710071, China

ARTICLE INFO

Article history:
Received 30 November 2021
Received in revised form 29 June 2022
Accepted 29 June 2022
Available online xxxx

Keywords:
Radar echo extrapolation
Self-attention
Long short-term memory
Spatiotemporal prediction

ABSTRACT

In recent years, the number of weather-related disasters significantly increases across the world. As a typical example, short-range extreme precipitation can cause severe flooding and other secondary disasters, which therefore requires accurate prediction of extent and intensity of precipitation in a relatively short period of time. Based on the echo extrapolation of networked weather radars (i.e., the Internet of Radars), different solutions have been presented ranging from traditional optical-flow methods to recent deep neural networks. However, these existing networks focus on local features of echo variations to model the dynamics of holistic radar echo motion, so it often suffers from inaccurate extrapolation of the radar echo motion trend, trajectory, and intensity. To address the problem, this paper introduces the self-attention mechanism and an extra memory that saves global spatiotemporal feature into the original Spatiotemporal LSTM (ST-LSTM) to form a self-attention Integrated ST-LSTM recurrent unit (SAST-LSTM), capturing both spatial and temporal global features of radar echo motion. And several these units are stacked to build the radar echo extrapolation network SAST-Net. Comparative experiments show that the proposed model has better performance on different real world radar echo datasets over other recent methods.

1. Introduction

Extreme precipitation can trigger flooding and waterlogging in cities or mudslides and landslides in mountainous areas, which cause significant human and economic loss. The ability to forecast high-intensity precipitation in a short period of time can greatly facilitate damage reduction and even prevention. Therefore, precipitation nowcasting has always been a critical and difficult challenge, especially concerning kilometer-level precipitation intensity forecasts in a local area within a relatively short period of time such as 0-2 h [1]. Traditional Numerical Weather Forecast (NWP) methods rely on mathematical and physical models and require a large amount of computing resources, which have been commonly used for medium and long-term precipitation forecasting, but not suitable for precipitation nowcasting due to low accuracy and "spin-up" problems [2,3]. Modern precipitation nowcasting mainly manipulates echo extrapolation of networked radars, which is also known as Radar Network Composite, or Internet of Radars (IoR) when echo images and extrapolation results are remotely stored and exchanged [4].

Traditional radar echo extrapolation methods can be divided into cross-correlation methods, centroid tracking methods and optical flow-based methods. The cross-correlation methods [5,6] can only capture the direction of motion of individual rainfall clouds, but hardly capture the large-scale motion of the whole weather system. The centroid tracking methods [7,8] are suitable for echoes of high intensity. When the echoes are split, the accuracy of tracking and prediction will be reduced. The optical flow-based methods [9–11] use the variation of the image sequence in the temporal domain to calculate the optical flow correlation between adjacent input frames, as well as the motion field for extrapolation. However, it is arbitrary for these methods to assume that the brightness of radar echo is constant. In addition, the optical flow estimation step separates from the extrapolation step. Therefore, determining the best parameters of the motion field to obtain the best extrapolation results becomes difficult, resulting in the limitations these methods.

Recently, artificial neural networks have been widely used and achieved outstanding performances in fields such as computer

* Corresponding author.
E-mail addresses: zhiyunyang@nuist.edu.cn (Z. Yang), wuhao@nuist.edu.cn (H. Wu), qi.liu@nuist.edu.cn (Q. Liu), x.liu@napier.ac.uk (X. Liu), zyh@nuist.edu.cn (Y. Zhang), xfcao@xidian.edu.cn (X. Cao).

[1] Zhiyun Yang and Hao Wu are both first author due to their equal contributions to this paper.

https://doi.org/10.1016/j.isatra.2022.06.046
0019-0578/

vision [12], edge computing [13–18], medical diagnosis [19,20], anomaly detection [21–23], data mining [24–27], algorithm optimization [28–31], and time series prediction [32–34], etc. This has attracted widespread attention from researchers in the field of weather forecasting, and they began to apply these networks to radar echo extrapolation. Radar raw data retrieved from a weather radar can further generate radar combined reflectivity maps, which are arranged in the chronological order. Due to its similarity to video frames, radar echo extrapolation can be regarded as a spatiotemporal sequence prediction problem, where n future radar echo maps are predicted from the input m maps. Shi et al. [35] replaced the full connection in the original LSTM unit with convolution and proposed a ConvLSTM model. It was applied to predict radar echo with the observed echo maps in Hong Kong. Base on ConvGRU model, a TrajGRU unit [36] implemented by variable convolution was proposed to effectively learn the spatial changes of recurrent connections. PredRNN [37] and its variant PredRNN++ [38] were constructed by the ST-LSTM and Causal LSTM models, respectively, where the features in the top layer of network at the previous time step are conveyed to the first layer of the current time step along a zigzag route, enabling the network to capture short- and long-term features at the same time. In addition, the classic U-Net network was also employed for precipitation nowcasting [39]. Existing deep neural networks being applied to radar echo extrapolation are mainly based on Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), or the combination of multiple deep models. These methods are affected by the receptive field of the convolution layer [40], so they have limitations in capturing long-range spatial features, making it difficult to model the dynamics of complex objects like radar echoes that are constantly moving, merging and dissipating, as therefore causing inaccurate or wrong prediction results [41].

In recent years, the self-attention mechanism has been a hot topic in deep learning. There have been many studies in natural language processing [42,43] and computer vision [44,45]. Compared to the CNN structure, the mechanism can efficiently capture global dependencies and derive information from past aggregated features, enhancing the ability to recognize complex motion objects.

In order to alleviate the impact of problems with the existing networks so as to improve the accuracy of radar echo extrapolation, this paper proposes a Self-Attention Integrated Spatiotemporal LSTM (SAST-LSTM) recurrent unit, in which the global spatiotemporal features are captured based on the self-attention mechanism and are saved in a dedicated global memory to avoid the error accumulation. To better perform radar echo extrapolation tasks, the recurrent units are stacked to build the four-layer extrapolation network SAST-Net. According to the results, the proposed work shows improved performance over the other models for radar echo extrapolation tasks.

The rest of this paper is organized as follows: Section 2 provides a brief overview of related research on radar echo extrapolation using deep neural networks. Section 3 introduces the self-attention mechanism and self-attention memory module. Section 4 describes in detail the working process of the proposed SAST-LSTM and the extrapolation network SAST-Net. Section 5 shows the conducted experiments and obtained results. Finally, the conclusion of the experiment and future work plan are given in Section 6.

2. Related work

Radar echo extrapolation tasks can be regarded as a kind of spatiotemporal sequence prediction problem. At present, a large number of researchers have used deep neural networks to conduct researches. RNN is used to capture temporal features, while CNN is generally used to capture spatial features.

Most CNN-based extrapolation models were inspired by methods in the fields of image classification and semantic segmentation, the models in these fields are further optimized and introduced into the radar echo extrapolation task. Klein et al. [46] proposed a dynamic convolution layer. The kernel size changes dynamically according to the input data, and used this structure for short-term weather forecasting. Zhuang et al. [47] designed a ST-CNN network to mine precipitation precursor information from data for extreme precipitation prediction. Agrawal et al. [39] used U-Net to predict the rainfall status in the next hour. Inspired by U-Net and SegNet, RainNet [48] is proposed, which predicts the precipitation intensity in Germany 5 min in advance. Fernandez et al. [49] proposed a Broad-UNet, which has asymmetric parallel convolution components, so that this model is able to combine multi-scale features for nowcasting. Han et al. [50] used a newly designed loss function to train U-Net to predict the radar echo images in northern China 30 min in advance.

RNN-based models are mostly used in combination with convolution layers. Shi et al. [35] proposed ConvLSTM, a landmark deep neural network in the field of radar echo extrapolation. It combines convolution operation with traditional LSTM to enhance model's ability to capture spatial features. The predicted radar echo intensity is converted into the precipitation intensity through the Z-R relationship, and the precipitation nowcasting result is obtained. Shi et al. expanded the structure of ConvLSTM by using a location-varying connection structure and proposed a TrajGRU unit [36]. This model obtained better extrapolation results than previous methods. Wang et al. [37] proposed a PredRNN, which is built by the newly designed Spatiotemporal LSTM (ST-LSTM), which introduced an additional memory unit into the original LSTM, enabling the model to capture temporal and spatial features at the same time. The network adds zigzag connections to convey captured features that improves the ability to model complex objects. To address the gradient disappearance problem in previous methods, Wang et al. [38] proposed a PredRNN++ model. It is built by several Causal LSTM units and a gradient highway (GHU). The Causal LSTM unit with the double cascade mechanism increases the model's ability to capture short-term features. The GHU enables the model to effectively convey gradient information. Wang et al. [51] proposed a Memory In Memory (MIM) model. It innovatively uses the differential signal between adjacent recurrent states to model the stationary and non-stationary characteristics of spatiotemporal dynamics. This model has powerful generalization capabilities for different tasks in multiple domains. Guen et al. [52] proposed a dual-branch network PhyDNet, which uses PhyCell to capture physical features such as the position objects, and ConvLSTM unit to capture appearance, texture and other residual features. Finally, the two types of information are merged to model dynamics.

Several research integrated the attention mechanism into the previous network to enhance the prediction performance, and applied improved networks to spatiotemporal sequence prediction tasks. Lin et al. [41] proposed the self-attention mechanism based SA-ConvLSTM network, which achieves the best results on several video datasets. Chai et al. [53] proposed CMS-LSTM networks based on multi-scale attention modules for video prediction tasks. Zhong et al. [54] improved the classic ST-LSTM unit and added the attention mechanism into it, which is used to model the long-range spatiotemporal dependence. Trebing et al. [55] added convolutional block attention modules (CBAM) to U-Net, which uses the attention mechanism for both channel and spatial dimensions. Luo et al. [56] proposed a IDA-LSTM model for radar echo extrapolation tasks in response to the underestimation of high-intensity radar echo regions by previous models.

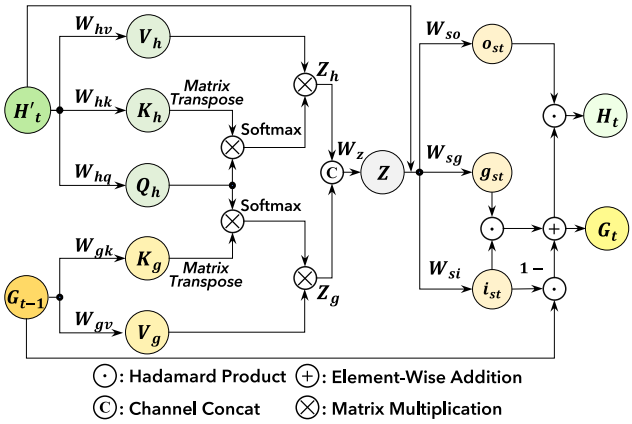

Fig. 1. The structure of Self-Attention Memory (SAM). Left side is the double self-attention mechanism for capturing global features. The right is the gated mechanism similar to GRU for updating global feature memory and hidden states.

3. Preliminary

Self-attention mechanism is a kind of attention mechanism and has been introduced into many deep networks in other fields. This mechanism can be formulated as Eq. (1), where Q is Query, K is Key and V is Value. These three terms are matrices, whose size are all $B \times C \times H \times W$, d_k is the dimension of Key.

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

This mechanism calculates the similarity score of each pair of points in the feature map by performing matrix multiplication on Q with transposed K, and dividing each element of the similarity matrix by $\sqrt{d_k}$ to reduce the influence of variance on the network gradient update. Then it normalizes the result by applying softmax function to get the corresponding weight coefficient matrix. Finally, the obtained matrix is multiplied with V. In this way, the model can capture global features that are important for future prediction, without being limited by the receptive field due to the size of the convolution filter.

To better adapt to prediction tasks, a Self-Attention Memory (SAM) module [41] based on further optimized self-attention mechanism was proposed. Its structure is shown in Fig. 1. The SAM has two self-attention mechanism structure that share the same Query. The global feature memory G and the temporary hidden states H' are the input of this module. Firstly, H' is operated by 1×1 convolution filters W_{hq}, W_{hk}, W_{hv} to obtain Query Q_h, Key K_h and Value V_h. Then the module performs self-attention mechanism operation on H'. The output of this operation is Z_h, which contains global feature newly captured from H'_t. The above process can be formulated as:

$$Z_h = \text{softmax}\left(\frac{Q_h K_h^T}{\sqrt{d_{kh}}}\right)V_h \tag{2}$$

At the same time, G is operated by two 1×1 filters W_{gk}, W_{gv} to get K_g and V_g, respectively. The two feature maps are then performed self-attention mechanism operation together with Q_h. The result is Z_g, which contains features are closely related to future prediction. This can be formulated as:

$$Z_g = \text{softmax}\left(\frac{Q_h K_g^T}{\sqrt{d_{kg}}}\right)V_g \tag{3}$$

Then Z_t, Z_g and original H' are concatenated along the channel for feature fusion as shown in Eq. (4), where $*$ means convolution, $[\cdot]$ means channel concat.

$$Z = W_z * [Z_h, Z_g] \tag{4}$$

The obtained Z and H'_t are transmitted into a gated mechanism similar to that in GRU. The input gate i_{st} and input modulation gate g_{st} are calculated as Eqs. (5) and (6), respectively, where σ is Sigmoid function, W represents convolution filters, b is bias.

$$i_{st} = \sigma\left(W_{s;zi} * Z + W_{s;hi} * H'_t + b_{s;i}\right) \tag{5}$$

$$g_{st} = \tanh\left(W_{s;zg} * Z + W_{s;hg} * H'_t + b_{s;g}\right) \tag{6}$$

The above two gate are then used to update the global feature memory G_t as shown is Eq. (7), where $\odot$ means Hadamard product:

$$G_t = (1 - i_{st}) \odot G_{t-1} + i_{st} \odot g_{st} \tag{7}$$

Similarly, Z and H'_t are used to obtain the output gate:

$$o_{st} = \sigma\left(W_{s;zo} * Z + W_{s;ho} * H_t + b_{s;o}\right) \tag{8}$$

Finally, the module performs Hadamard product operation on output gate o_{st} and global feature memory G to get the updated hidden state H_t:

$$H_t = o_{st} \odot G_t \tag{9}$$

4. Methodology

4.1. Self-attention integrated spatiotemporal LSTM

In order to enhance models' ability to capture the global features of radar echo variations and accurately model their dynamics, so as to obtain accurate extrapolated radar echo maps, this paper introduces SAM into the original Spatiotemporal LSTM (ST-LSTM) to form the Self-Attention Integrated Spatiotemporal LSTM recurrent unit (SAST-LSTM). Its structure is shown in Fig. 2, where t represents the tth time step, l represents the lth layer.

SAST-LSTM is the basic unit for building the stacked radar echo extrapolation network SAST-Net, whose input includes the hidden state H, the temporal memory C, the input state X, the spatiotemporal memory M and the global feature memory G. It should be noted that C is only updated along the temporal dimension, while M and G are updated along the spatial dimension first, and then along the temporal dimension, so they contain the spatiotemporal features of radar echo variations. The specific update routes of the above states and memory cells will be introduced in detail in Section 4.2.

The SAST-LSTM unit performs standard gated mechanism operations like that in ConvLSTM on C^l_{t-1}, H^l_{t-1} and X^l_t. The convolution filters W_f is operated on H^l_{t-1} and X^l_t. The obtained results are performed on element-wise addition operation and Sigmoid function to get the forget gate f_t. This can be formulated as:

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H^l_{t-1} + b_f\right) \tag{10}$$

The convolution filters W_i and W_g are operated on H^l_{t-1} and X^l_t, respectively. Then the unit performs similar operations like Eq. (10) to get the input gate i_t:

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H^l_{t-1} + b_i\right) \tag{11}$$

The input modulation gate g_t can be obtained by the similar approach as well:

$$g_t = \tanh\left(W_{xg} * X_t + W_{hg} * H^l_{t-1} + b_g\right) \tag{12}$$

Then the Hadamard product operation between f_t and C^l_{t-1} is performed to forget the unimportant temporal features transmitted from previous time step. Also, the unit performs Hadamard

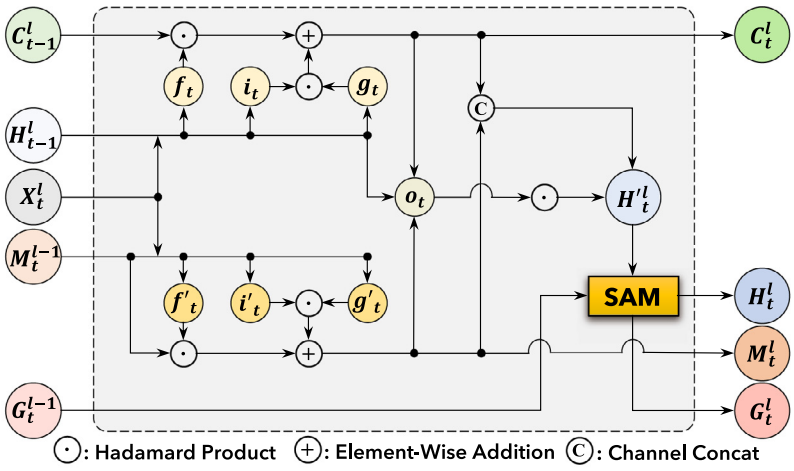

Fig. 2. The structure of a recurrent unit in the proposed SAST-Net. An additional gating mechanism is dedicated to capturing spatiotemporal features, and the SAM is introduced for global feature extraction. In parallel, an extra memory is added to store global features so as to alleviate error accumulation.

product operation between i_t and g_t to obtain the captured features. The result of these two steps is combined by performing element-wise addition operation. The result is used to update temporal memory C_t^l. The above process can be formulated as:

$$C_t^l = f_t \odot C_{t-1}^l + i_t \odot g_t \tag{13}$$

The recurrent unit also performs gated mechanism operations on spatiotemporal memory M_t^{l-1} and input state X_t^l to capture spatiotemporal features from the input state at current time step. Another group of convolution filters W'_f, W'_i, and W'_g act on M_t^{l-1} and X_t^l to obtain f'_t, i'_t, and g'_t. These three gates are obtained in a similar way to Eqs. (10)–(12):

$$f'_t = \sigma\left(W'_{xf} * X_t + W_{mf} * M_t^{l-1} + b'_f\right) \tag{14}$$

$$i'_t = \sigma\left(W'_{xi} * X_t + W_{mi} * M_t^{l-1} + b'_i\right) \tag{15}$$

$$g'_t = \tanh\left(W'_{xg} * X_t + W_{mg} * M_t^{l-1} + b'_g\right) \tag{16}$$

By performing operations similar to the upper gated mechanism, the newly captured spatiotemporal features are saved in M_t^l. This above process can be formulated as:

$$M_t^l = f'_t \odot M_t^{l-1} + i'_t \odot g'_t \tag{17}$$

Next, the input state X_t^l, the hidden state H_{t-1}^l, the updated time memory C_t^l and the updated spatiotemporal memory M_t^l are operated by convolution filter W_o to obtain the output gate o_t:

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1}^l + W_{mo} * M_t^l + b_o\right) \tag{18}$$

Memory C_t^l and memory M_t^l are concatenated along the channel dimension, the concatenated result is then operated by o_t to obtain the temporary hidden state $H_t'^l$:

$$H_t'^l = o_t \odot \tanh\left(\left[M_t^l, C_t^l\right]\right) \tag{19}$$

Finally, the temporary hidden state $H_t'^l$ and the global feature memory G_t^{l-1} are used as the input of SAM to capture global features. The obtained global feature map is stored in the memory G. The final hidden state H_t^l is obtained at the same time. The process can be formulated as:

$$H_t^l, G_t^l = \mathrm{SAM}\left(H_t'^l, G_{t-1}^l\right) \tag{20}$$

4.2. Radar echo extrapolation network SAST-Net

Following the settings in previous works [37,38], this paper stacks the SAST-LSTM recurrent units to build a four-layer radar echo extrapolation network SAST-Net. It has been demonstrated in previous work that stacking four layers of recurrent units can increase the depth of the network to capture refined features. In this way, models are better to model abrupt variation of radar echoes, while avoiding the gradient vanishing problem. The overall architecture of SAST-Net is shown in Fig. 3. Each column in the figure shows the network at a single time step, while each row represents a single layer of the network. The horizontal direction represents the temporal domain, and the vertical direction represents the spatial domain. All hidden states and memory cells are initialized with zero matrices at the first time step.

Taking the network at the tth time step as an example, it can be clearly seen from the figure the SAST-LSTM unit in the first layer takes the input radar echo map I_t at the tth time step as the input state X_t^1. For units in other layers of the network, the hidden state H_t^{l-1} output by the previous layer is used as the input state. In particular, the hidden state H_t^4, which is the output of the fourth-layer unit, is operated by a 1×1 convolution filter so as to obtain the extrapolated radar echo image $\widehat{I}_{t+1}$ at current time step. In horizontal direction, the updated temporal memory C_t^1 and the hidden state H_t^1, which are the output of current unit, are transmitted to the unit in the same layer at the next time step. The blue arrows in the figure indicate the update route of temporal memory M and global feature memory G. Different from temporal memory C, they are first updated in the vertical direction, and memory the local and global features in spatial domain captured by units layer by layer. Then M_t^4 and G_t^4 are sent to the first layer at next time step. These two memories are updated with the new features captured from the new input state X_{t+1}^1. Therefore, M and G also contain temporal features. The stacked structure follows the above-mentioned update route to complete the modeling of radar echo dynamics and the extrapolation of future echo maps.

5. Experiments

5.1. Experimental setup

To show the performance improvement of the proposed SAST-Net, this model and some other representative methods are used

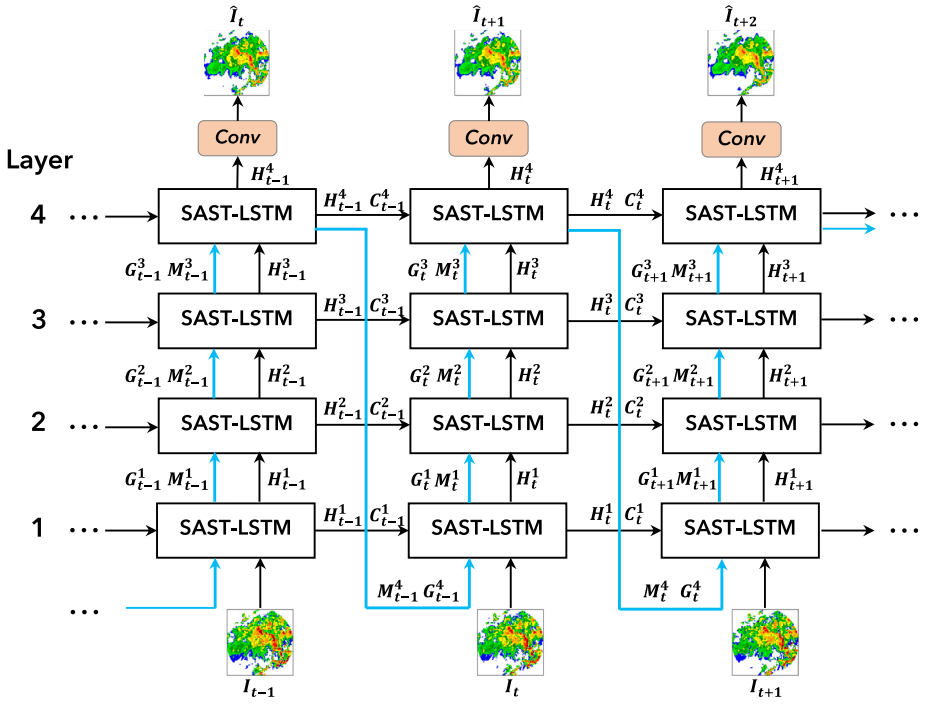

Fig. 3. The overall architecture of SAST-Net built by stacking four-layer SAST-LSTM recurrent units. The blue arrows are depicting the zigzag route, where the global feature memory G and the spatiotemporal memory M transmit and update. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to perform extrapolation tasks on two different real world radar echoes datasets.

The first dataset is the Guangzhou Station dataset, in which the data was observed by a CINRAD-SA Doppler weather radar in Guangzhou during rainy season (May to August) from 2012 to 2014. The training set, validation set, and test set has 4570, 600 and 605 sequences, respectively. Each sequence has 20 radar echo maps arranged in chronological order. The size of each map is 500 $\times$ 500, and the spatial resolution is 1 km. To reduce the computational complexity, the original maps are resized to 100 $\times$ 100. The observation interval of every two adjacent radar echo maps is 6 min. In this experiment, all models predict the last 10 echo maps based on the first 10 maps.

The second dataset is the CIKM 2017 dataset, a publicly available dataset contains Doppler weather radar echo maps, which were observed by the Meteorological Bureau of Shenzhen during two consecutive years. The training set, validation set, and test set has 8000, 4000 and 2000 sequences, respectively. Each sequence in the dataset contains 15 maps. The original size of each echo map is 101 $\times$ 101, and each point in the grid represents an 1 km $\times$ 1 km area. To facilitate the model to divide the map into patches, the right and lower sides of the map are filled with zeros, so that the size of new obtained image is 104 $\times$ 104. All models receive the first 5 maps to extrapolate the next 10 maps in this experiment.

The points with weak relationship to precipitation events (dBZ $<$ 10) in the above two datasets were replaced by 0, so that the value of points represent radar echo in the map ranges from 10 to 70. The value of points represent background is 0. All radar echo maps are converted into images by Eq. (21):

$$\text{pixel\_value} = \left\lfloor 255 \times \left(\frac{\text{dBZ\_value}}{70} \right) + 0.5 \right\rfloor \tag{21}$$

Representative models used for comparative experiments include ConvLSTM, PredRNN, PredRNN++, Memory in Memory (MIM), PhyDNet, and SA-ConvLSTM. The residual branch of PhyDNet and other models are all four-layers stacked structure like the proposed SAST-Net. The recurrent units in these models have 64 feature maps. All models are trained with the Adam optimizer [57]. Scheduled sampling [58], layer normalization [59] and early stopping training strategies are also used. The initial learning rate and the batch size is set to 0.001 and 4, respectively. The mean squared error (MSE) is used as the loss function. All experiments are conducted on NVIDIA RTX 3090.

Due to the privacy concerns of weather radar data, this paper uses a distributed cloud platform based on Java EE and ICE (Internet Communications Engine) technology. Researchers use the SSL-encrypted Internet of Weather Radars transmission network to read and manage radar data. The platform accesses the Java Web API through the ICE framework, realizing model training, testing, log viewing and other functions. ICE is similar to socket technology and is responsible for handling the underlying communication programming. The service interfaces written in the SLICE (Specification Language for ICE) language, which decouples the client end from the server end, and the two ends can use different programming languages to keep the programming style consistent. The communication process of the ICE framework is shown in Fig. 4. In addition, the communication channel is configured with security protocols such as RSA, which can guarantee the security of communication during the training process.

5.2. Evaluation metrics

To evaluate the quality of extrapolated echo maps by each model, this paper selects Critical Success Index (CSI) and Heidke Skill Score (HSS), which are often used in meteorological area as the evaluation metrics for forecast results. The CSI focuses on the correspondence between precipitation events predicted by the model and the truly observed precipitation events, that is, the probability that the model successfully predicts the precipitation events. The higher the CSI score, the better the performance of the model. The HSS focuses on the fraction of correct predictions after eliminating those predictions, which would be correct due

Table 1
The quantitative comparison of all models in terms of CSI and HSS on **Guangzhou Station** dataset in the comparative experiment.

Model	CSI ↑					HSS ↑				
	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average
ConvLSTM	0.6738	0.6210	0.4807	0.1897	0.4913	0.7343	0.7212	0.6196	0.3081	0.5958
PredRNN	0.6956	0.6391	0.5086	0.2268	0.5175	0.7505	0.7360	0.6456	0.3584	0.6226
PredRNN++	0.6965	0.6328	0.4871	0.2135	0.5075	0.7522	0.7314	0.6263	0.3411	0.6128
MIM	0.6937	0.6373	0.5193	0.2524	0.5257	0.7487	0.7343	0.6547	0.3905	0.6320
PhyDNet	0.6683	0.5673	0.4154	0.1678	0.4547	0.7223	0.6723	0.5552	0.2774	0.5568
SA-ConvLSTM	0.6889	0.6222	0.5027	0.2227	0.5091	0.7423	0.7197	0.6378	0.3509	0.6127
SAST-Net	**0.6982**	**0.6499**	**0.5294**	**0.2680**	**0.5364**	**0.7544**	**0.7451**	**0.6637**	**0.4101**	**0.6433**

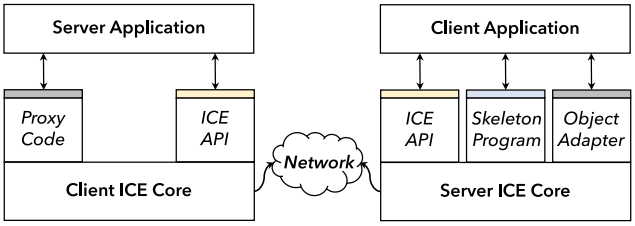

Fig. 4. The communication process between server and client based on an ICE framework. The ICE can separate both sides (Client and Server) into different runtime/development environments. It can also establish a secure communication tunnel using RSA or other cryptography algorithms.

purely to random chance. The higher HSS score indicates better extrapolation performance of the model.

The CSI and HSS scores of each model are calculated as follows. Given a dBZ threshold τ, for every pair of points with the same coordinates in extrapolated map and the corresponding observed echo map, if the point's value in extrapolated map and the corresponding point's value in observed map are both greater than τ, this point pair belongs to the True Positive (TP) class. If the point's value in extrapolated map is greater than τ while the corresponding point's value in observed map is smaller than τ, the pair belongs to the False Positive (FP) class. If the point's value in extrapolated map is smaller than τ while the corresponding point's value in observed map is greater than τ, the pair belongs to the False Negative (FN) class. If the point's value in extrapolated map and the corresponding point's value in observed map are both smaller than τ, this point pair belongs to the True Negative (TN) class. Finally, counting the number of point pair in each class. The CSI score and HSS score of each model are calculated following Eqs. (22) and (23). In the experiments, threshold τ is set to 10, 20, 30 and 40 dBZ.

$$\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{22}$$

$$\mathrm{HSS} = \frac{2 \times (\mathrm{TP} \times \mathrm{TN} - \mathrm{FN} \times \mathrm{FP})}{(\mathrm{TP} + \mathrm{FN}) \times (\mathrm{FN} + \mathrm{TN}) + (\mathrm{TP} + \mathrm{FP}) \times (\mathrm{FP} + \mathrm{TN})} \tag{23}$$

5.3. Comparative experiments

Table 1 shows the quantitative evaluation results of all models on Guangzhou Station dataset. It can be seen that the proposed SAST-Net obtained the best CSI and HSS scores for all thresholds. When the threshold τ is set to relatively low values (10 and 20 dBZ), SAST-Net shows slight improvements over other models. For relatively high thresholds (30 and 40 dBZ), considering the points with relatively high dBZ values are strongly related to heavy precipitation, so it is quite important for the models to make more accurate prediction under these two thresholds. The CSI-30 score, and CSI-40 score of SAST-Net is 1.94% and 6.18% higher than second-ranked model MIM. In terms of HSS score, it improves by 1.37% when τ is set to 30 and improves by 5.02% when τ is set to 40. The above results show that the proposed SAST-Net has better extrapolation performance for radar echo, especially that of higher intensity, compared with the existing methods.

To evaluate the model's robustness for high-intensity echoes extrapolation, the CSI and HSS scores of all models at each lead time stamp on the Guangzhou Station dataset when the threshold is set to 30 and 40 are drawn in Fig. 5. The lead time is the observation interval between the echo map and the last input map. As can be seen from the figure, SAST-Net's curves is always at the top of all curves, indicating that it has achieved the highest score at all time steps. Its slope is also smaller than that of the other models, showing its more stable performance than the existing methods. MIM's curve is very close to that of SAST-Net in the first several time steps, but the distance between them becomes larger and larger as time goes. The reason is that the lack of the memory, which can store global features, making it difficult to model the dynamics of radar echo when performing long-term extrapolation tasks. It is noted that the curve of SA-ConvLSTM is unsmooth, because it lacks memory to store spatiotemporal features, so the curve fluctuates greatly.

To assess the quality of extrapolated radar echo maps, Fig. 6 shows the visualization of all models' extrapolated maps for one test sequence from the Guangzhou Station dataset. The mapping relationship between dBZ value and color of the radar echo is shown in the color bar on the right. The first two rows are the observed maps from the sequence, and all models extrapolate maps in the second row based on the maps in the first row. The following rows are the extrapolated results of each model. The echo maps in one column have the same observation time stamp. The more similar the extrapolated map is to the corresponding observed map in the second row, indicating the better the performance of the model.

The sequence shows that the high-intensity echo moving to the right side of the map. The red echo area highlighted by the red box in the figure is closely related to the potential severe convective weather, so it is important for the models to make accurate extrapolation of this part. ConvLSTM predicts the rough outline of echoes, but the yellow and orange areas in the middle and lower of the map showed a trend of dissipating, indicating that the model fails to correctly predict the echoes with high intensity value. PredRNN predicts the yellow area in the lower of the map better, but it can be seen that the echo intensity of the upper left yellow area is underestimated. The same problem also appears in the extrapolated maps of PredRNN++. The yellow and orange areas on the right of maps are quite different from those in observed maps. PhyDNet predicts that the echoes are in dissipating, so in the last few pictures, the whole area is significantly smaller than that in observed maps, and the error of the yellow and orange areas inside is larger, which results in its lowest CSI and HSS scores. The extrapolated maps of MIM and SA-ConvLSTM are more similar to the observation compared with the above-mentioned methods, but it is worth noting that the red area marked by the red box in these two models' extrapolated

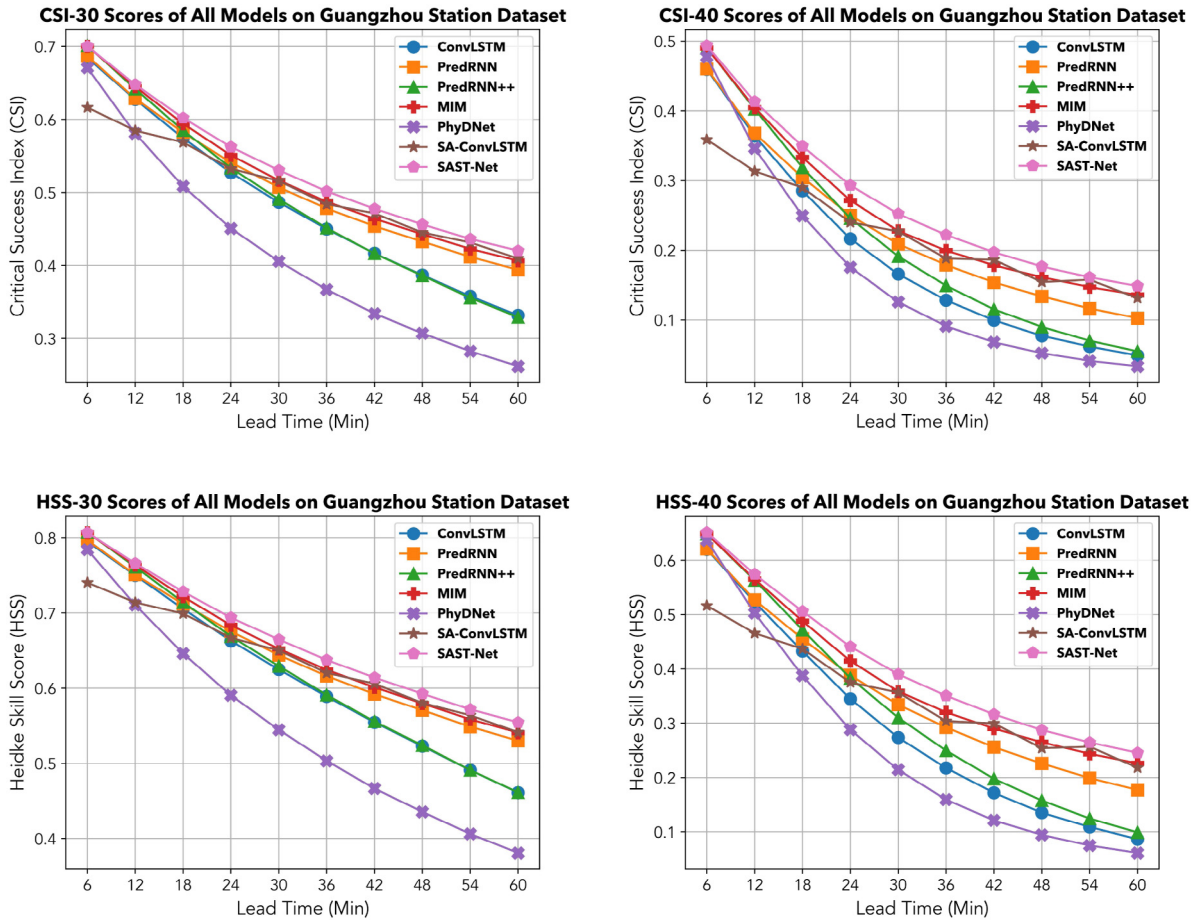

Fig. 5. The CSI and HSS scores of all models at each lead time stamp on the **Guangzhou Station** dataset when the threshold τ is set to 30 and 40 in the comparative experiment. The curves of proposed SAST-Net are smoother and always at the top of other models', indicating its better robustness and extrapolation performance than comparison models.

Table 2
The quantitative comparison of all models in terms of CSI and HSS on **CIKM 2017** dataset in the comparative experiment.

Model	CSI ↑					HSS ↑				
	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average
ConvLSTM	0.6809	0.4070	0.2075	0.0831	0.3446	0.6673	0.4844	0.3306	0.1525	0.4087
PredRNN	0.6699	0.3933	0.2200	0.0969	0.3450	0.6632	0.4653	0.3397	0.1746	0.4107
PredRNN++	0.6676	0.3930	0.2068	0.1026	0.3425	0.6533	0.4779	0.3287	0.1847	0.4112
MIM	0.6832	0.4028	0.2360	0.1097	0.3579	0.6679	0.4779	0.3607	0.1954	0.4255
PhyDNet	0.6321	0.3597	0.1588	0.0885	0.3098	0.5971	0.4106	0.2411	0.1606	0.3524
SA-ConvLSTM	0.6857	0.3925	0.2159	0.1055	0.3499	0.6720	0.4845	0.3412	0.1892	0.4217
SAST-Net	**0.6868**	**0.4100**	**0.2515**	**0.1264**	**0.3687**	**0.6722**	**0.4847**	**0.3838**	**0.2228**	**0.4409**

maps, which indicates possible severe precipitation events, gradually dissipates with time, which may lead to the failure to take timely responses to heavy precipitation, and thus this problem is fatal. For the SAST-Net proposed in this paper, the boundaries of echo in the extrapolated maps are almost consistent with the observed maps. With the help of its spatiotemporal structure and self-attention mechanism, the model is able to capture and learn features, so as to model the dynamics of radar echo more accurately. The disappearance problem of the red area for MIM and SA-ConvLSTM is also alleviated.

To evaluate the general applicability of the proposed SAST-Net, comparative experiments were also conducted on the CIKM 2017 dataset. Table 2 shows the quantitative results of all models, SAST-Net also obtains the best CSI and HSS scores under all four thresholds on this dataset. In particular, compared with the second-rank model MIM, its CSI-30 and CSI-40 score improve by 6.57% and 17.76%, respectively. Its HSS-30 and HSS-40 score improve by 6.40% and 14.02%, respectively.

Similarly, Fig. 7 shows the CSI and HSS scores of all models on the CIKM 2017 dataset at each lead time stamp when the threshold is set to 30 and 40. It can be seen that the curves of SAST-Net are still ahead of other models' at every time step. The distances between its and other models' increase over time.

Fig. 8 is the visualization of all models' extrapolated maps for one test sequence from the CIKM 2017 dataset. In this sequence, the echoes move downward, and the two orange areas in the middle of the map gradually merge. The red box is used to mark the high-intensity echoes that need to be focused on the sequence as well. In the extrapolated maps of ConvLSTM, the yellow area tends to shrink and is scattered into several small areas of echoes, indicating that the model cannot predict the motion trend of radar echo correctly. The result of PredRNN suffers from the problem of underestimating the intensity of echoes, so the yellow

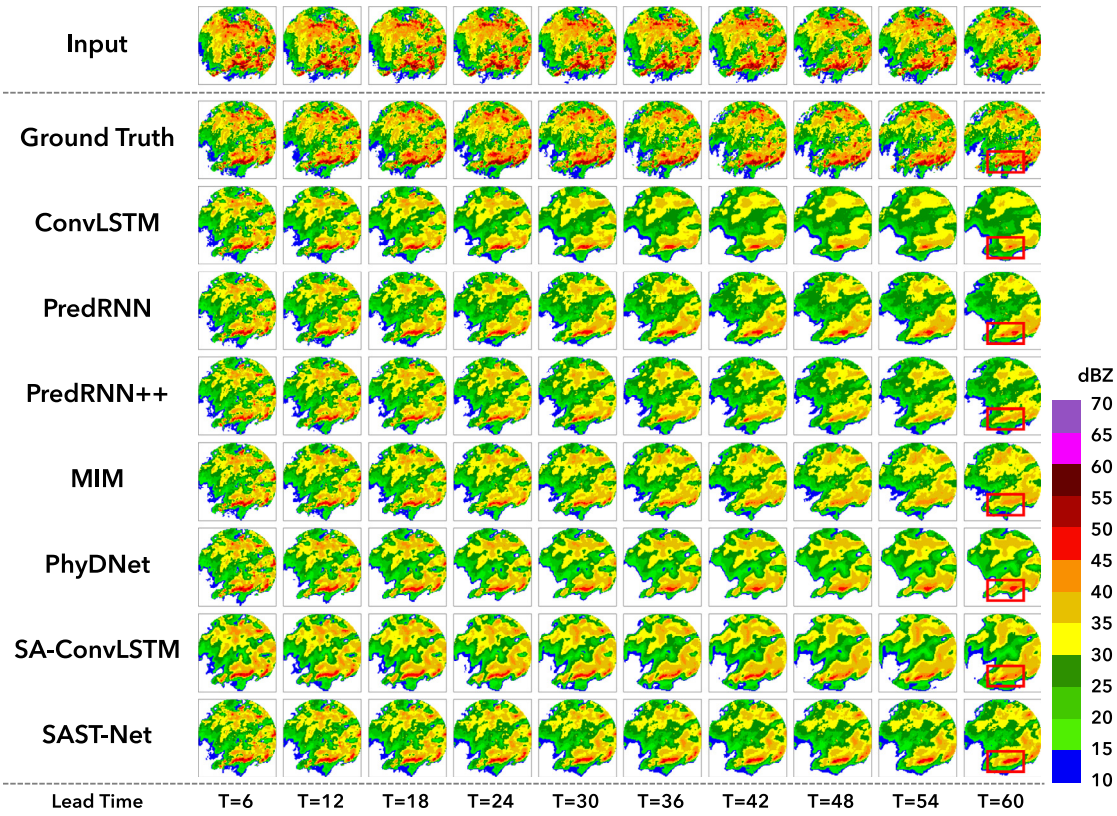

Fig. 6. The visualization of all models' extrapolated radar echo maps for one test sequence from the **Guangzhou Station** dataset in the comparative experiment. Both the boundary and intensity of the red striped echo area highlighted by the red box in SAST-Net's extrapolated map are most similar to the Ground Truth, while this area in other models' maps suffer from the problem of dissipation or inaccurate position. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
The quantitative evaluation results of the four models on **Guangzhou Station** dataset in the ablation study.

Model	CSI ↑					HSS ↑				
	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average	$\tau = 10$	$\tau = 20$	$\tau = 30$	$\tau = 40$	Average
Baseline Model	0.6956	0.6391	0.5086	0.2268	0.5175	0.7505	0.7360	0.6456	0.3584	0.6226
SAST-Net wo.SA	0.6919	0.6418	0.5177	0.2303	0.5204	0.7476	0.7375	0.6530	0.3627	0.6252
SAST-Net wo.G	0.6953	0.6419	0.5229	0.2536	0.5284	0.7513	0.7382	0.6577	0.3919	0.6348
SAST-Net	**0.6982**	**0.6499**	**0.5294**	**0.2680**	**0.5364**	**0.7544**	**0.7451**	**0.6637**	**0.4101**	**0.6433**

and orange areas marked by the red box are significantly smaller in area than they actually are. The PredRNN++, MIM and SA-ConvLSTM also fail to correctly predict high-intensity echoes, with the yellow and orange areas in the red box scattered or dissipated. The extrapolated maps of PhyDNet are far from the observed maps and suffer from distortion problem. The SAST-Net proposed in this paper extrapolate the holistic variation of echoes better, and the parts marked in red box, which need to be focused on are closest to the actual observations compared with the other models.

5.4. Ablation study

To demonstrate the effectiveness of the proposed structure, ablation study for the SAST-LSTM unit, which constitutes the SAST-Net, are conducted on the Guangzhou Station dataset. The self-attention mechanism and the global feature memory were added to the SAST-LSTM respectively. Thus, the experimental subjects include the baseline model PredRNN built by ST-LSTM units, SAST-Net without the self-attention mechanism (SAST-Net wo. SA), SAST-Net without global feature memory (SAST-Net wo. *G*), and the standard SAST-Net. The quantitative results of all models are shown in Table 3.

After adding self-attention mechanism and global feature memory to the original ST-LSTM respectively, the scores of the newly obtained SAST-Net without *G* and SAST-Net without SA, although slightly lower than the baseline model when τ = 10, but they show great improvements in their scores at higher thresholds as well as in average scores, which is of significance for destructive heavy precipitation forecasting. The performance improvement of SAST-Net without SA is smaller than that of SAST-Net without *G*, which shows that the self-attention mechanism has a significant contribution to model the holistic radar echo motion. The best results are obtained by the standard SAST-Net proposed in this paper, which is attributed to the combined effect of the two new components mentioned above.

Fig. 9 is the visualization of radar echo maps extrapolated by different models for a sequence from Guangzhou Station dataset in the ablation study. This sequence shows the echo in the middle moves towards the right side of the map, and the generation of high-intensity echo in the lower side of the map, as highlighted by two red boxes in the figure. The baseline model PredRNN fails

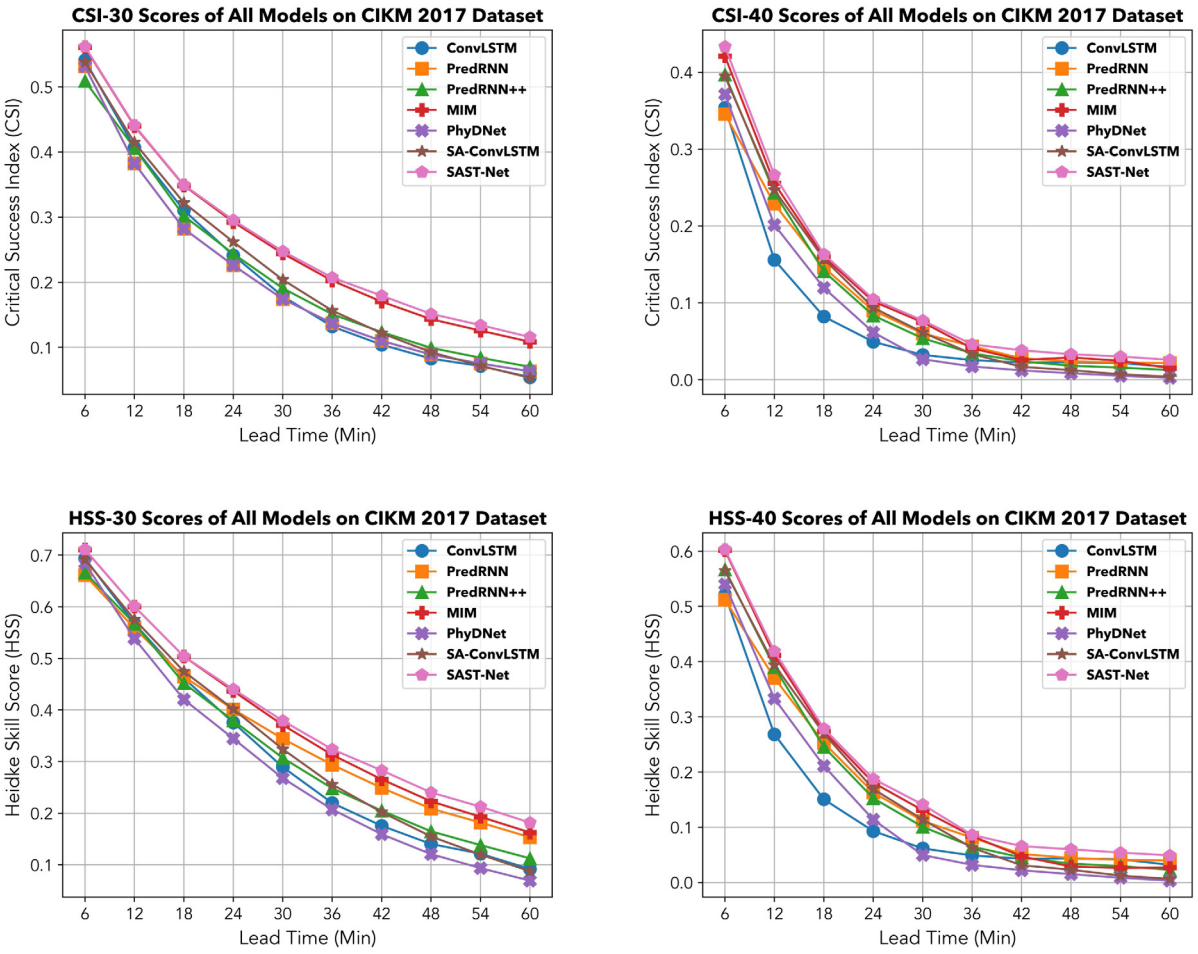

Fig. 7. The CSI and HSS scores of all models at each lead time stamp on the **CIKM 2017** dataset when the threshold τ is set to 30 and 40 in the comparative experiment. The curves of the proposed SAST-Net are above other models and decline slower, showing its more stable performance for long-range extrapolation.

to correctly predict the motion trend of radar echoes. The echo area marked by red boxes showed the occurrence of dissipation and underestimation of intensity. The SAST-Net without SA but with the extra memory can remember more features of echo variations, enabling the model to accurately predict the rough boundary of echo areas. But the intensity of echoes marked by red boxes is underestimated. The SAST-Net without G captures the global features of radar echo motion with the help of the self-attention mechanism, so it successfully predicts the echo in the middle of map. However, due to lack of global feature memory, prediction errors started to occur and accumulated over time from lead time $T = 30$, leading to an inaccurate extrapolation of echo variations in the top left of the map, as well as the underestimation of echo intensity highlighted by the lower red box. The standard SAST-Net not only successfully predicts the position of the echoes in the maps, but also make accurate extrapolations in which both the boundary and intensity of the echo areas marked in red boxes are most similar to those in the corresponding observed maps. This owing to its ability to capture and learn global features of the radar echo motion by using the self-attention mechanism, and to avoid the accumulation of errors with the help of global feature memory.

5.5. Result analysis

From the results of comparative experiments and ablation study conducted on two different real-world radar echo datasets, it can be clearly seen the effectiveness of SAST-LSTM's components and SAST-Net's improved extrapolation performance over other models.

Unlike traditional video prediction tasks, radar echo extrapolation is more challenging as, echoes may dissipate or aggregated together rather than just do their own motion, which requires the model to infer future maps from a global view.

Those recent models are only able to capture local features of the echo variations by performing convolution operations but cannot model the holistic motion of the radar echo, so their extrapolated maps often have echo-intensity underestimation problem. The second-ranked MIM model alleviates this problem to some extent by decomposing the dynamics into non-stationary and stationary states, but it still faces the inaccurate echo-motion modeling problem, which is shown in Fig. 8. The orange area gradually dissipates. In addition, the model may face the error accumulation problem due to its lack of additional memory, leading to a rapid decline in its prediction performance over time.

The SAST-Net, which built by SAST-LSTM units, can better capture global features of echo variations by integrating the self-attention mechanism, improving the accuracy in predicting holistic radar motion. In addition, the additional memory added to the model alleviate error accumulation problem, so the model has more stable performance when perform long-range extrapolation. With the collaborative work of these two components, the dynamic modeling ability of the SAST-Net for radar echoes can be improved, and more accurate extrapolated maps can be obtained.

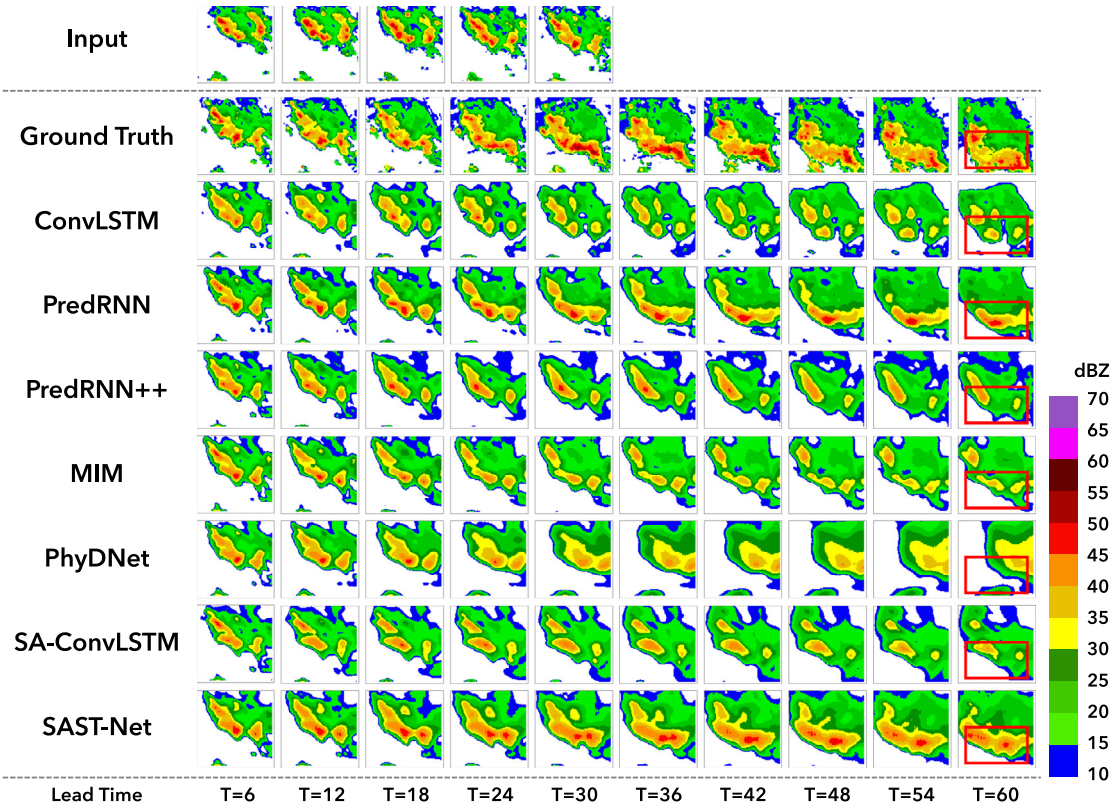

Fig. 8. The visualization of all models' extrapolated radar echo maps for one test sequence from the CIKM 2017 dataset in the comparative experiment. The proposed SAST-Net can predict the yellow and orange areas marked in the red box, as well as sporadic red areas related to possible severe convective weather events. The extrapolated maps of the other models differ significantly from the observed maps and suffer from echo dissipation or even distortion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

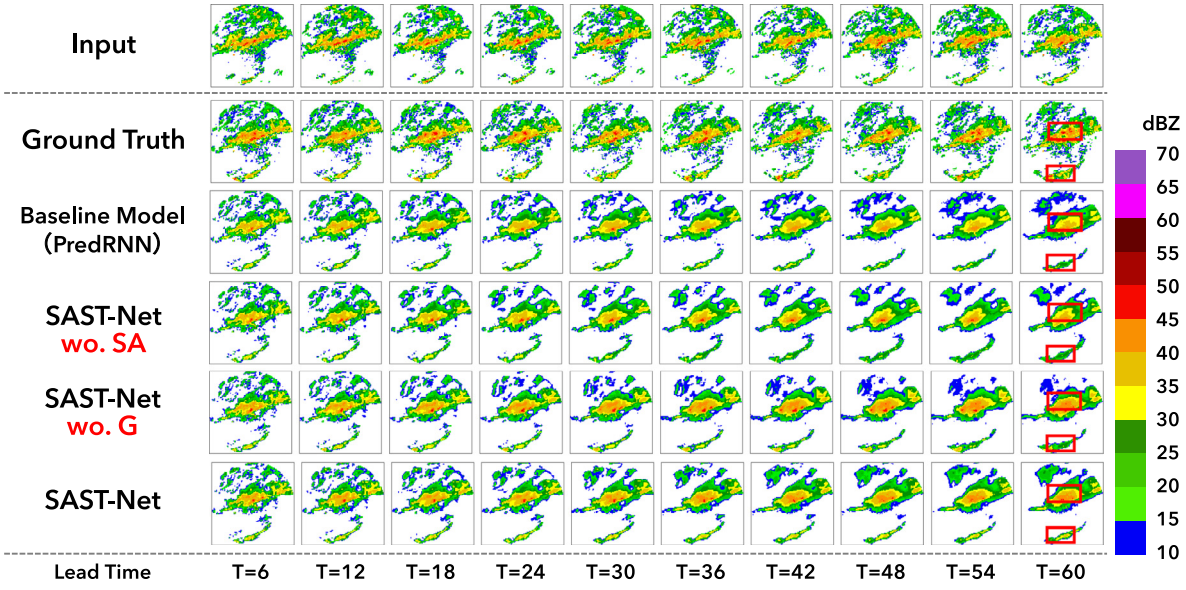

Fig. 9. The visualization of the four models' extrapolated radar echo maps for one test sequence from the **Guangzhou Station** dataset in the ablation study. The standard SAST-Net captures and learns the global features of the radar echoes with the help of SA and global feature memory, so as to alleviate error accumulation and achieve accurate extrapolation of both the motion trends and intensity of two echo areas marked by the red boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusion

This paper proposes the SAST-LSTM, a spatiotemporal recurrent unit integrates the self-attention mechanism, which enable it to capture and learn global spatiotemporal features, and an extra global feature memory G is added to avoid error accumulation with time goes. With the help of the above two components, the model can accurately model the dynamics of radar echo

motion. In order to perform radar echo extrapolation tasks, the SAST-LSTM unit are stacked to build a four-layer radar echo extrapolation network SAST-Net. The experiments conducted on two different real world radar echo datasets demonstrate the improvements of the proposed SAST-Net on performing radar echo extrapolation tasks compared with other recent methods. In the future, the authors will seek solutions to further improve the clarity of extrapolated radar echo maps.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has received funding from the Key Laboratory Foundation of National Defence Technology under Grant 61424010208, National Natural Science Foundation of China (No. 62002276, 41911530242 and 41975142), 5150 Spring Specialists (05492018012 and 05762018039), Major Program of the National Social Science Fund of China (Grant No. 17ZDA092), 333 High-Level Talent Cultivation Project of Jiangsu Province (BRA2018332), Royal Society of Edinburgh, UK and China Natural Science Foundation Council (RSE Reference: 62967_Liu_2018_2) under their Joint International Projects funding scheme and basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20191398 and BK20180794).

References

[1] Li L, Chen S, Mai X-F. Sub-pixel precipitation nowcasting over Guangdong Province using optical flow algorithm. In: 2017 IEEE international geoscience and remote sensing symposium (IGARSS). 2017, p. 4638–41.
[2] Mecklenburg S, Joss J, Schmid W. Improving the nowcasting of precipitation in an alpine region with an enhanced radar echo tracking algorithm. J Hydrol 2000;239(1):46–68.
[3] Chu Q, Xu Z, Chen Y, Han D. Evaluation of the ability of the Weather Research and Forecasting model to reproduce a sub-daily extreme rainfall event in Beijing, China using different domain configurations and spin-up times. Hydrol Earth Syst Sci 2018;22(6):3391–407.
[4] Akan OB, Arik M. Internet of radars: Sensing versus sending with joint radar-communications. IEEE Commun Mag 2020;58(9):13–9. http://dx.doi.org/10.1109/MCOM.001.1900550.
[5] Liang Q, Feng Y, Deng W, Hu S, Huang Y, Zeng Q, Chen Z. A composite approach of radar echo extrapolation based on TREC vectors in combination with model-predicted winds. Adv Atmos Sci 2010;27(5):1119–30.
[6] Rinehart R, Garvey E. Three-dimensional storm motion detection by conventional weather radar. Nature 1978;273(5660):287–9.
[7] Handwerker J. Cell tracking with TRACE3D—A new algorithm. Atmos Res 2002;61(1):15–34.
[8] Johnson J, MacKeen PL, Witt A, Mitchell EDW, Stumpf GJ, Eilts MD, Thomas KW. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. Weather Forecast 1998;13(2):263–76.
[9] Woo W-c, Wong W-k. Operational application of optical flow techniques to radar-based rainfall nowcasting. Atmosphere 2017;8(3):48.
[10] Sakaino H. Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow. IEEE Trans Geosci Remote Sens 2013;51(5):3023–36.
[11] Li P, Lai ES. Applications of radar-based nowcasting techniques for mesoscale weather forecasting in Hong Kong. Meteorol Appl 2004;11(3):253–64.
[12] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Comput Intell Neurosci 2018;2018:7068349.
[13] Xu X, Zhang X, Liu X, Jiang J, Qi L, Bhuiyan MZA. Adaptive computation offloading with edge for 5G-envisioned internet of connected vehicles. IEEE Trans Intell Transp Syst 2021;22(8):5213–22.
[14] Xu X, Fang Z, Zhang J, He Q, Yu D, Qi L, Dou W. Edge content caching with deep spatiotemporal residual network for IoV in smart city. ACM Trans Sensor Netw 2021;17(3):1–33.
[15] Xu X, Fang Z, Qi L, Zhang X, He Q, Zhou X. Tripres: Traffic flow prediction driven resource reservation for multimedia iov with edge computing. ACM Trans Multimed Comput Commun Appl 2021;17(2):1–21.
[16] Xu X, Li H, Xu W, Liu Z, Yao L, Dai F. Artificial intelligence for edge service optimization in internet of vehicles: A survey. Tsinghua Sci Technol 2022;27(2):270–87.
[17] Malek YN, Najib M, Bakhouya M, Essaaidi M. Multivariate deep learning approach for electric vehicle speed forecasting. Big Data Min Anal 2021;4(1):56–64.
[18] Bi R, Liu Q, Ren J, Tan G. Utility aware offloading for mobile-edge computing. Tsinghua Sci Technol 2021;26(2):239–50.
[19] Bhardwaj N, Sharma P. An advanced uncertainty measure using fuzzy soft sets: Application to decision-making problems. Big Data Min Anal 2021;4(2):94–103.
[20] Wei D, Ning H, Shi F, Wan Y, Xu J, Yang S, Zhu L. Dataflow management in the internet of things: Sensing, control, and security. Tsinghua Sci Technol 2021;26(6):918–30.
[21] Wang W, Wang Z, Zhou Z, Deng H, Zhao W, Wang C, Guo Y. Anomaly detection of industrial control systems based on transfer learning. Tsinghua Sci Technol 2021;26(6):821–32.
[22] Alaoui EAA, Tekouabou SCK, Hartini S, Rustam Z, Silkan H, Agoujil S. Improvement in automated diagnosis of soft tissues tumors using machine learning. Big Data Min Anal 2021;4(1):33–46.
[23] Djenouri Y, Belhadi A, Srivastava G, Ghosh U, Chatterjee P, Lin JC-W. Fast and accurate deep learning framework for secure fault diagnosis in the industrial internet of things. IEEE Internet Things J 2021;1.
[24] Liu Y, Song Z, Xu X, Rafique W, Zhang X, Shen J, Khosravi MR, Qi L. Bidirectional GRU networks-based next POI category prediction for healthcare. Int J Intell Syst 2022;37(7):4020–40.
[25] Qi L, Song H, Zhang X, Srivastava G, Xu X, Yu S. Compatibility-aware web API recommendation for mashup creation via textual description mining. ACM Trans Multimedia Comput Commun Appl 2021;17(1s).
[26] Ahmed U, Srivastava G, Lin JC-W. Reliable customer analysis using federated learning and exploring deep-attention edge intelligence. Future Gener Comput Syst 2022;127:70–9.
[27] Li H, Liu J, Liu RW, Xiong N, Wu K, Kim T-h. A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis. Sensors 2017;17(8):1792.
[28] Ren Z, Liu Y, Shi T, Xie L, Zhou Y, Zhai J, Zhang Y, Zhang Y, Chen W. AIPerf: Automated machine learning as an AI-HPC benchmark. Big Data Min Anal 2021;4(3):208–20.
[29] Shao Y, Lin JC-W, Srivastava G, Guo D, Zhang H, Yi H, Jolfaei A. Multi-objective neural evolutionary algorithm for combinatorial optimization problems. IEEE Trans Neural Netw Learn Syst 2021;1–11.
[30] Tong Z, Ye F, Yan M, Liu H, Basodi S. A survey on algorithms for intelligent computing and smart city applications. Big Data Min Anal 2021;4(3):155–72.
[31] Fang W, Yao X, Zhao X, Yin J, Xiong N. A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms. IEEE Trans Syst Man Cybern Syst 2018;48(4):522–34. http://dx.doi.org/10.1109/TSMC.2016.2606400.
[32] Xiao Y, Yin H, Zhang Y, Qi H, Zhang Y, Liu Z. A dual-stage attention-based Conv-LSTM network for spatio-temporal correlation and multivariate time series prediction. Int J Intell Syst 2021;36(5):2036–57.
[33] Liu Y, Li D, Wan S, Wang F, Dou W, Xu X, Li S, Ma R, Qi L. A long short-term memory-based model for greenhouse climate prediction. Int J Intell Syst 2022;37(1):135–51.
[34] Xiao Y, Yin H, Duan T, Qi H, Zhang Y, Jolfaei A, Xia K. An Intelligent prediction model for UCG state based on dual-source LSTM. Int J Mach Learn Cybern 2021;12(11):3169–78.
[35] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-c. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Proceedings of the 28th international conference on neural information processing systems. 2015; p. 802–810.
[36] Shi X, Gao Z, Lausen L, Wang H, Yeung D-Y, Wong W-k, Woo W-c. Deep learning for precipitation nowcasting: A benchmark and a new model. In: Proceedings of the 31st international conference on neural information processing systems. 2017; p. 5622–5632.
[37] Wang Y, Long M, Wang J, Gao Z, Yu PS. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In: Proceedings of the 31st international conference on neural information processing systems. 2017; p. 879–888.
[38] Wang Y, Gao Z, Long M, Wang J, Philip SY. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In: Proceedings of the 35th international conference on machine learning. 80, PMLR; 2018, p. 5123–32.
[39] Agrawal S, Barrington L, Bromberg C, Burge J, Gazen C, Hickey J. Machine learning for precipitation nowcasting from radar images. 2019, arXiv preprint arXiv:1912.12132.

[40] Luo W, Li Y, Urtasun R, Zemel R. Understanding the effective receptive field in deep convolutional neural networks. In: Proceedings of the 30th international conference on neural information processing systems. 2016; p. 4905–4913.
[41] Lin Z, Li M, Zheng Z, Cheng Y, Yuan C. Self-attention convlstm for spatiotemporal prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34. 2020; p. 11531–11538.
[42] Kim J, El-Khamy M, Lee J. T-GSA: Transformer with Gaussian-weighted self-attention for speech enhancement. In: ICASSP 2020 - 2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2020, p. 6649–53.
[43] Han KJ, Prieto R, Ma T. State-of-the-art speech recognition using multi-stream self-attention with dilated 1D convolutions. In: 2019 IEEE Automatic speech recognition and understanding workshop (ASRU). 2019, p. 54–61.
[44] Wu Y, Ma Y, Liu J, Du J, Xing L. Self-attention convolutional neural network for improved MR image reconstruction. Inform Sci 2019;490:317–28.
[45] Zhao H, Jia J, Koltun V. Exploring self-attention for image recognition. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020, p. 10073–82.
[46] Klein B, Wolf L, Afek Y. A dynamic convolutional layer for short rangeweather prediction. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). 2015, p. 4840–8.
[47] Zhuang W, Ding W. Long-lead prediction of extreme precipitation cluster via a spatiotemporal convolutional neural network. In: Proceedings of the 6th international workshop on climate informatics. 2016; p. 128–131.
[48] Ayzel G, Scheffer T, Heistermann M. RainNet v1. 0: a convolutional neural network for radar-based precipitation nowcasting. Geosci Model Dev 2020;13(6):2631–44.
[49] Fernández JG, Mehrkanoon S. Broad-UNet: Multi-scale feature learning for nowcasting tasks. Neural Netw 2021;144:419–27.
[50] Han L, Liang H, Chen H, Zhang W, Ge Y. Convective precipitation nowcasting using U-net model. IEEE Trans Geosci Remote Sens 2022;60:1–8.
[51] Wang Y, Zhang J, Zhu H, Long M, Wang J, Yu PS. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2019, p. 9146–54.
[52] Le Guen V, Thome N. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020, p. 11471–81.
[53] Chai Z, Yuan C, Lin Z, Bai Y. CMS-LSTM: Context-embedding and multi-scale spatiotemporal-expression LSTM for video prediction. 2021, arXiv preprint arXiv:2102.03586.
[54] Zhong S, Zeng X, Ling Q, Wen Q, Meng W, Feng Y. Spatiotemporal convolutional LSTM for radar echo extrapolation. In: 2020 54th asilomar conference on signals, systems, and computers. 2020, p. 58–62.
[55] Trebing K, Stańczyk T, Mehrkanoon S. SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. Pattern Recognit Lett 2021;145:178–86.
[56] Luo C, Li X, Wen Y, Ye Y, Zhang X. A novel LSTM model with interaction dual attention for radar echo extrapolation. Remote Sens 2021;13(2):164.
[57] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.
[58] Bengio S, Vinyals O, Jaitly N, Shazeer N. Scheduled sampling for sequence prediction with recurrent neural networks. 2015, arXiv preprint arXiv:1506.03099.
[59] Ba JL, Kiros JR, Hinton GE. Layer normalization. 2016, arXiv preprint arXiv:1607.06450.