

ISA Transactions

A Self-Attention Integrated Spatiotemporal LSTM Approach to Edge-Radar Echo Extrapolation in the Internet of Radars

--Manuscript Draft--

Manuscript Number:	
Article Type:	VSI:A COMINT P DECMAX AUTOSY
Keywords:	Radar Echo Extrapolation; Self-Attention; Long Short-Term Memory; Spatiotemporal Prediction
Corresponding Author:	Qi Liu CHINA
First Author:	Zhiyun Yang
Order of Authors:	Zhiyun Yang Hao Wu Qi Liu Xiaodong Liu Yonghong Zhang Xuefei Cao
Abstract:	<p>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 networks focus at local features by capturing the spatial features of radar echo variations only, but suffer from rational recognition of holistic radar echo's trajectory, intensity and corresponding dynamics. In this paper, a self-attention integrated LSTM approach is presented by introducing a SAST-LSTM cell into the original model. A self-attentive mechanism is also designed for accurate analysis of the radar echo from both spatial and temporal aspects. Comparative experiments show that the proposed model has better performance on different real radar echo datasets over other recent methods, involving ConvLSTM, PredRNN, MIM, PhydNet and SA-ConvLSTM.</p>

A Self-Attention Integrated Spatiotemporal LSTM Approach to Edge-Radar Echo Extrapolation in the Internet of Radars

Author Details

Name: Zhiyun Yang (#)

Mail: zhiyunyang@nuist.edu.cn

Affiliation: School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology

Address: Nanjing, 210044, China

Name: Hao Wu (#)

Mail: wuhao@nuist.edu.cn

Affiliation: School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology

Address: Nanjing, 210044, China

Name: Qi Liu (*Corresponding Author*)

Mail: qi.liu@nuist.edu.cn

Affiliation: School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology

Address: Nanjing, 210044, China

Name: Xiaodong Liu

Mail: x.liu@napier.ac.uk

Affiliation: School of Computing, Edinburgh Napier University

Address: Edinburgh, EH10 5DT, UK

Name: Yonghong Zhang

Mail: zyh@nuist.edu.cn

Affiliation: School of Automation, Nanjing University of Information Science Technology

Address: Nanjing, 210044, China

Name: Xuefei Cao

Mail: xfcao@xidian.edu.cn

Affiliation: School of Cyber and Information Security, Xidian University

Address: Xi'an, 710071, China

#: Zhiyun Yang and Hao Wu are both first author due to their equal contributions to this paper.

Acknowledgement

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).

Highlights

A Self-Attention Integrated Spatiotemporal LSTM Approach to Edge-Radar Echo Extrapolation in the Internet of Radars

- The self-attention memory mechanism is integrated into to original ST-LSTM.
- The SAM compensates for the limitation of convolution in the receptive field.
- Extra memory G stores global features to prevent the accumulation of errors.

A Self-Attention Integrated Spatiotemporal LSTM Approach to Edge-Radar Echo Extrapolation in the Internet of Radars

ARTICLE INFO

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 networks focus on local features by capturing the spatial features of radar echo variations only, but suffer from rational recognition of holistic radar echo's trajectory, intensity and corresponding dynamics. In this paper, a self-attention integrated LSTM approach is presented by introducing a SAST-LSTM cell into the original model. A self-attentive mechanism is also designed for accurate analysis of the radar echo from both spatial and temporal aspects. Comparative experiments show that the proposed model has better performance on different real radar echo datasets over other recent methods, involving ConvLSTM, PredRNN, MIM, PhyDNet and SA-ConvLSTM.

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 kilometre-level precipitation intensity forecasts in a local area within a relatively short period of time such as 0-2 hours [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 include 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, 10, 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 processes of optical flow estimation and echo extrapolation are separated, so it is difficult to determine the best parameters of the motion

field to obtain the best extrapolation effect, resulting in the limitations of existing optical flow based methods.

Recently, artificial neural networks have been widely used and achieved outstanding performances in fields such as computer vision [12], edge computing [13, 14, 15, 16, 17, 18], anomaly detection [19, 20, 21], data mining [22, 23, 24, 25], algorithm optimization [26, 27, 28, 29], medical diagnosis [30, 31], and climate prediction [32], 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. A ConvLSTM model was proposed in [33], for example, replacing the full connection in LSTM with convolution to predict radar echo with the observed echo maps in Hong Kong. In [34], a TrajGRU unit implemented by variable convolution was proposed to effectively learn the spatial changes of cyclic connections. PredRNN [35] and its variant PredRNN++ [36] 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 through a zigzag structure, 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 [37]. Existing deep neural networks being applied to radar echo extrapolation are mainly based on Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), etc., or the combination of multiple deep models. These methods are affected by the receptive field and stacked structure of the convolutional layer [38], 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 con-

stantly moving, merging and dissipating, as therefore causing inaccurate or wrong prediction results [39].

In recent years, the Self-Attention mechanism (SA) has been a hot topic in deep learning. There have been many studies in natural language processing [40, 41], and computer vision [42, 43]. Compared to the CNN structure, the SA mechanism can efficiently capture global dependencies and derive information from past aggregated features, enhancing the ability to recognise complex motion objects.

In order to alleviate the impact of problems with the existing networks so as to improve the prediction accuracy of radar echoes, this paper proposes a SAST-LSTM (Self-Attention Integrated Spatiotemporal LSTM) unit based on the SA mechanism, which enables the fusion of current features with global features through a dedicated memory unit. Two practically retrieved radar echo datasets have been prepared for comparative experiments between the SAST-LSTM and other recent published models, including ConvLSTM, PredRNN, MIM, PhyDNet, etc. 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 organised as follows: Section 2 provides a brief overview of related research on radar echo extrapolation using deep neural networks. Section 3 explains the self-attention mechanism, and describes in detail the working process of the proposed SAST-LSTM. Section 4 shows the conducted experiments and obtained results. Finally, the conclusion of the experiment and future work plan are given in Section 5.

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 research. RNN is used to extract temporal features, while CNN is generally used to extract spatial features.

Most CNN-based extrapolation models have been inspired by work 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. [44] proposed a dynamic convolutional layer in which the size of the convolutional kernel changes dynamically according to the input data, and used this structure for short-term weather forecasting. [45] designed a spatiotemporal convolutional neural network (ST-CNN) to mine precipitation precursor information from data for extreme precipitation prediction. [37] used U-Net to predict the rainfall status in one hour. Inspired by U-Net and SegNet, RainNet [46] is proposed, which predicts the precipitation intensity in Germany 5 minutes in advance. [47] proposed Broad-UNet equipped with asymmetric parallel convolution and Atrous Spatial Pyramid Pooling (ASPP) module, which combines multi-scale features for nowcasting. [48] used a newly designed loss function to train U-Net to predict the radar echo images in northern China 30 minutes in advance.

RNN-based models are mostly used in combination with

convolution layers. ConvLSTM [33] is 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 is realized. TrajGRU [34] expanded the structure of ConvLSTM by using a location-varying connection structure and obtained better extrapolation results. PredRNN [35] 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 extracted features that improves the ability to model complex objects. To address the gradient disappearance problem caused by stacking multiple recurrent units in previous networks, PredRNN++ [36] uses the Causal LSTM unit based on the double cascade mechanism to increase the non-linearity of the recurring unit, and a gradient highway (GHU) to effectively convey gradient information. Memory In Memory (MIM) [49] innovatively uses the differential signal between adjacent loop states to model the stationary and non-stationary characteristics of spatiotemporal dynamics, with powerful generalization capabilities for different tasks in multiple domains. [50] proposed a dual-branch network PhyDNet, which uses PhyCell to capture physical features such as the position objects, and ConvLSTM cell 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. [39] proposed the self-attention mechanism based SA-ConvLSTM network, which achieves the best results on several video datasets. [51] proposed CMS-LSTM networks based on multi-scale attention modules for video prediction tasks. [52] improved the classic ST-LSTM cell, and the attention mechanism is used to model the long-range spatiotemporal dependence. [53] added convolutional block attention modules (CBAM) to U-Net, which uses the attention mechanism for both channel and spatial dimensions. [54] proposed a dual attention long short-term memory (IDA-LSTM) for radar echo extrapolation tasks in response to the underestimation of high-intensity radar echo regions by previous models.

3. Methodology

3.1. Self-Attention Mechanism

Self-attention mechanism is one kind of attention mechanisms. Compared to the convolutional operation, the self-attention mechanism can focus on global important features by calculating similarity scores, without limiting the receptive field of the network due to kernel size. The structure of self-attention mechanism is shown in Fig. 1, this mechanism

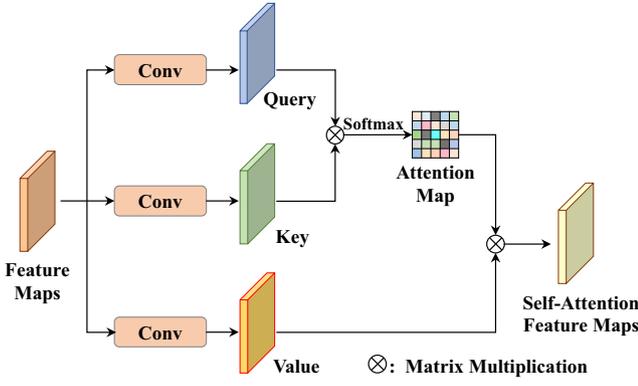


Figure 1: The structure of standard self-attention mechanism.

can be formulated as:

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

The input original feature maps are first operated by 1×1 convolution filters W_q , W_k , W_v to obtain three matrixes: Query, Key, Value, their size is $B \times C \times H \times W$. Then, the matrix transpose operation is performed on Query, and the matrix multiplication with the Key is used to calculate the similarity score, the score for each group of points forms the similarity matrix, and dividing each element of the similarity matrix by $\sqrt{d_k}$ to reduce the influence of variance on the network gradient update, where d_k is the dimension of Key. Then the results are normalized by the Softmax function to obtain the Weight matrix, i.e. the attention map. Finally, the attention map is weighted and summed with Value to obtain the final self-attention maps, which contain the extracted global features. [39] proposes the self-attention memory module based on the standard self-attention mechanism, as shown in Fig. 2. The upper part on the left side of the structure is a standard self-attention module, the output of this part is Z_h . The lower part is also essentially a Self-Attention module, sharing the same Query as the upper half, where the global features extracted by the self-attention mechanism are stored in M . Through the filter W_{mk} and W_{mv} , the Key and Value of the lower half are obtained, and the output Z_m is calculated together with Query. The Z_h and Z_m are concatenated along the channel dimension, and then concatenated with the input feature H_t after the convolution operation. The lower right part updates M with a gated structure similar to GRU. The upper right part updates the feature H_t through the output gate. The entire calcula-

tion process can be formulated as follows:

$$\begin{aligned} Z_h &= \text{softmax} \left(\frac{Q_h K_h^T}{\sqrt{d_{kh}}} \right) V \\ Z_m &= \text{softmax} \left(\frac{Q_h K_m^T}{\sqrt{d_{km}}} \right) V \\ Z &= W_z * [Z_h, Z_m] \\ i'_t &= \sigma (W_{m:zi} * Z + W_{m:hi} * H_t + b_{m:i}) \\ g'_t &= \tanh (W_{m:zg} * Z + W_{m:hg} * H_t + b_{m:g}) \\ M_t &= (1 - i'_t) \odot M_t - 1 + i'_t \odot g'_t \\ o'_t &= \sigma (W_{m:zo} * Z + W_{m:ho} * H_t + b_{m:o}) \\ \hat{H}_t &= o'_t \odot M_t \end{aligned} \quad (2)$$

3.2. SAST-LSTM

Inspired by previous work, this paper proposes a novel recurrent cell called SAST-LSTM, which is a fusion of SAM module and the original ST-LSTM. With the SAM module, the recurrent cell can effectively capture global spatiotemporal variations, and the extracted features are stored in a dedicated attentional memory, which works in concert with the standard temporal memory and spatiotemporal memory to compensate for the previous networks' limitation on the receptive field due to the size of the convolutional kernel, the structure of a single SAST-LSTM cell is shown in Fig. 3, where t represents the t -th time step, and l is the number of layers. X is the input data of the current time step, H is the hidden state, C is the standard temporal memory, M is the spatiotemporal memory, and G is the attention memory.

The cell receives these parameters as input and updates C after two symmetric gating mechanisms. The temporary hidden state T_t^l is then obtained through the output gate, which, together with G_{t-1}^l are used as the input of the SAM module to capture the global dependencies. The extracted global features are stored in G_t^l , and the final hidden state H_t^l is obtained. The memory update process in SAST-LSTM can be formulated as Equation 3.

SAST-LSTM cells are stacked to build a four-layer architecture called SAST-Net, as shown in Fig. 4. The cell at the bottom layer receives the input data X_t . The spatiotemporal memory cell M_t^l , which stores higher-order features, is updated layer by layer in a zigzag route, while the hidden state H_t is conveyed to the next cell in both the horizontal and vertical directions. Memory M_t^l that stores coarse local features and G_t^l that stores global features, are conveyed horizontally to the next cell. Under the action of these three types of memory, the SAST-LSTM can receive more information and dependencies to model complex dynamics such

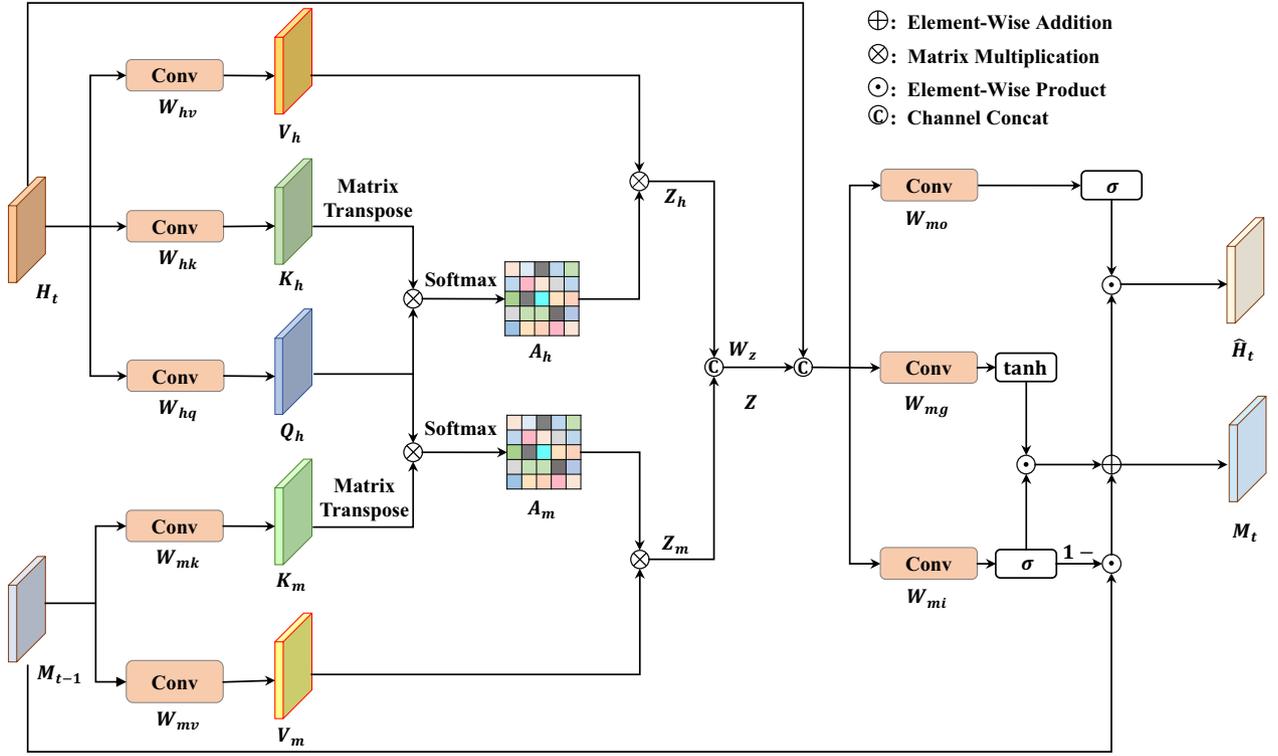


Figure 2: The structure of self-attention memory module (SAM).

as radar echo more accurately.

$$\begin{aligned}
 g_t &= \tanh(W_{xg} * X_t + W_{hg} * H_{t-1}^l + b_g) \\
 i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1}^l + b_i) \\
 f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1}^l + b_f) \\
 C_t^l &= f_t \odot C_{t-1}^l + i_t \odot g_t \\
 g'_t &= \tanh(W'_{xg} * X_t + W'_{mg} * M_{t-1}^{l-1} + b'_g) \\
 i'_t &= \sigma(W'_{xi} * X_t + W'_{mi} * M_{t-1}^{l-1} + b'_i) \\
 f'_t &= \sigma(W'_{xf} * X_t + W'_{mf} * M_{t-1}^{l-1} + b'_f) \\
 M_t^l &= f'_t \odot M_{t-1}^{l-1} + i'_t \odot g'_t \\
 o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1}^l + W_{mo} \odot M_t^l + b_o) \\
 T_t^l &= o_t \odot \tanh(M_t^l) \\
 H_t^l, G_t^l &= \text{SAM}(T_t^l, G_{t-1}^l)
 \end{aligned}$$

(3)

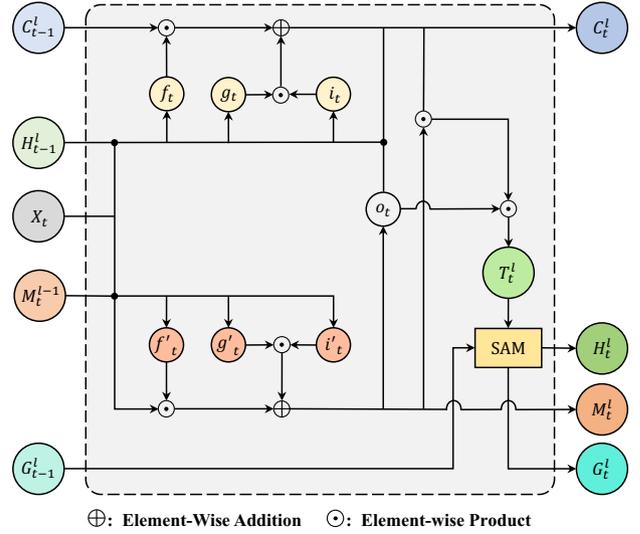


Figure 3: The structure of SAST-LSTM cell.

4. Experiments

4.1. Experimental Setup

In this section, the SAST-LSTM proposed in this paper is evaluated on two different real world radar echo datasets, Guangzhou Station dataset and CIKM 2017 dataset.

The radar echo data in Guangzhou Station dataset is collected from a CINRAD-SA Doppler weather radar in Guangzhou. The dataset contains radar echo data during the rainy season

(May to August) from 2012 to 2014, the interval of every two radar echo maps is 6 minutes, so the radar can generate 10 radar echo maps in one hour. The original size of each radar echo map is 500×500 , and the spatial resolution is 1km. To reduce the computational pressure, the original echo map is resized to 100×100 by the method of max pooling. The model predicts the next 10 frames based on 10 input frames in the experiments.

The CIKM 2017 dataset is a two-consecutive-year Doppler

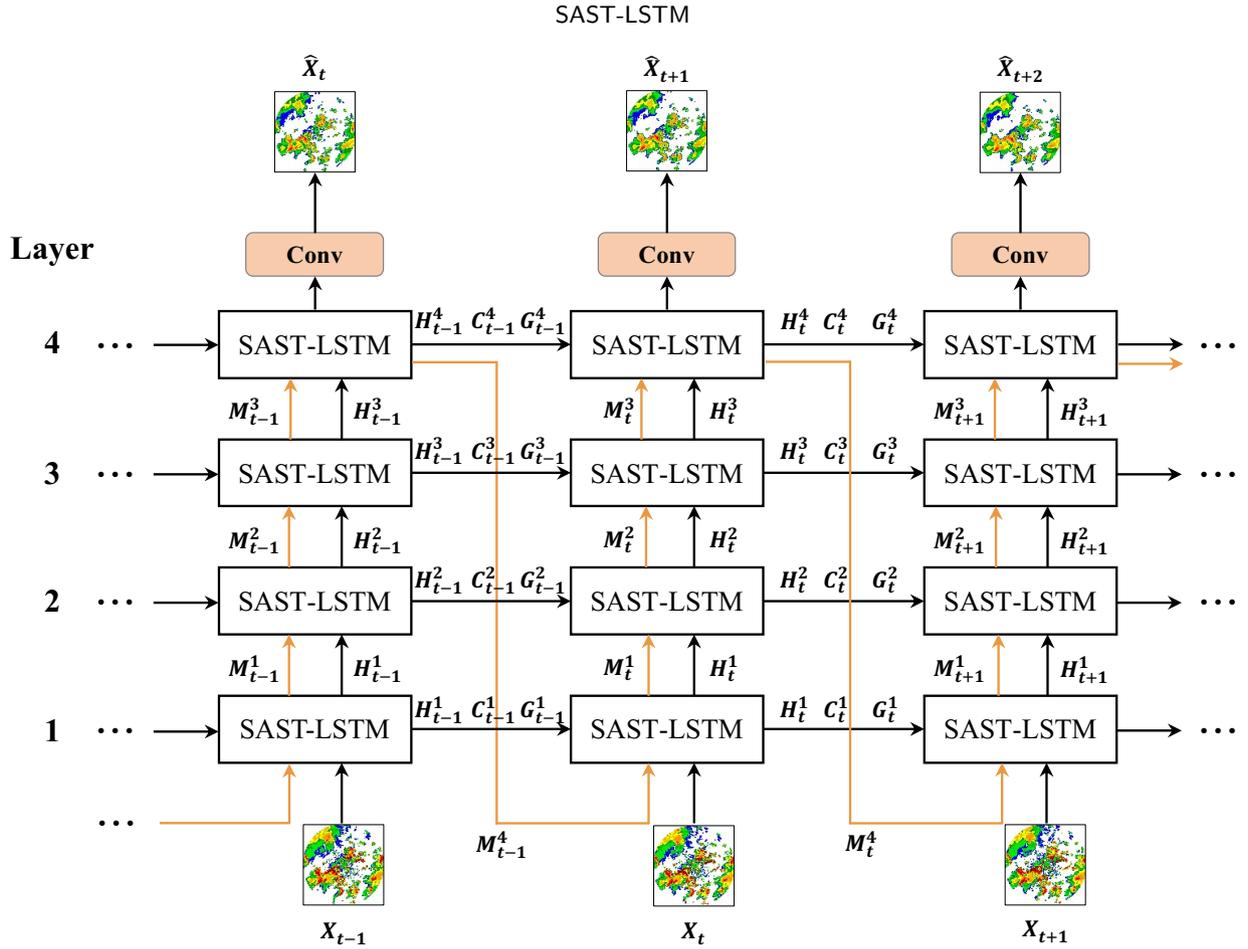


Figure 4: The overall architecture of SAST-Net built by stacking SAST-LSTM cell.

Table 1
Statistics of training, validation and test sets.

Dataset	Training	Validation	Test
Guangzhou Station	4,570	602	603
CIKM 2017	8,000	2,000	4,000

weather radar echo image dataset collected by the Shenzhen Meteorological Bureau. The size of each image in dataset is 101×101 , and each 1×1 pixel represents a 1×1 km square area. To facilitate the division of patches, we fill in zeros at the bottom and right side of the original image, so that the new image size is 104×104 . Each sequence of CIKM 2017 contains 15 frames, with the first 5 frames used for input and the last 10 frames used for prediction. The number of sequences for the training set, validation set and test set of two datasets is shown in Table. 1.

In experiments, we selected ConvLSTM, PredRNN, PredRNN++, Memory in Memory (MIM), PhyDNet and SA-ConvLSTM as comparison models. The residual branch of PhyDNet and the other five models all use a four-layer architecture, and the number of feature maps in each layer is 64. All models are trained using with Adam optimizer [55], using scheduled sampling [56], layer normalization [57] and early stopping training strategies. The initial learning rate

is 0.001, mini batch size is 4, and the mean squared error (MSE) is used as the loss function.

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. 5. In addition, the communication channel is configured with security protocols such as RSA, which can guarantee the security of communication during the training process.

4.2. Evaluation Metrics

In this paper, Critical Success Index (CSI) and Heidke Skill Score (HSS) are used as the evaluation metrics of model performance. With a given threshold τ , the real and predicted radar echo maps are converted into a 0-1 matrix, and

Table 2

Results of the comparative experiment using CSI and HSS as evaluation metrics on the Guangzhou Station dataset.

Model	CSI					HSS				
	10	20	30	40	avg	10	20	30	40	avg
ConvLSTM	0.6792	0.5997	0.4329	0.1715	0.4708	0.7373	0.7022	0.5728	0.2822	0.5736
PredRNN	0.6601	0.6147	0.4899	0.2104	0.4938	0.7227	0.7156	0.6265	0.3345	0.5998
PredRNN++	0.6453	0.5861	0.4345	0.1764	0.4606	0.7067	0.6878	0.5696	0.2851	0.5623
MIM	0.6520	0.6034	0.4805	0.1805	0.4791	0.7158	0.7068	0.6186	0.2948	0.5840
PhyDNet	0.6050	0.5222	0.3735	0.1744	0.4188	0.6690	0.6328	0.5119	0.2859	0.5249
SA-ConvLSTM	0.6948	0.6305	0.5105	0.2200	0.5140	0.7493	0.7276	0.6462	0.3490	0.6180
SAST-Net	0.6835	0.6335	0.5120	0.2558	0.5212	0.7415	0.7300	0.6451	0.3918	0.6271

when the value in the map is greater than τ , the value at the corresponding position in the matrix is set to 1, otherwise the value is set to 0. In this paper, the value of τ is set to 10, 20, 30 and 40 dBZ. After conversion, the values of TP (real map = 1, predicted map = 1), TN (real map = 0, predicted map = 0), FP (real map = 0, predicted map = 1), FN (real map = 1, predicted map = 0) are obtained. The CSI and HSS values under the current threshold τ can be further calculated according to the following formulas:

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

$$\text{HSS} = \frac{2 \times (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP} + \text{FN}) \times (\text{FN} + \text{TN}) + (\text{TP} + \text{FP}) \times (\text{FP} + \text{TN})} \quad (4)$$

4.3. Experiment results and Analysis

The evaluation results for the extrapolation task on the two real-world radar echo datasets for all the models involved in the comparison are given in Table. 2 and Table. 3 respectively. It can be seen that although the performance of the SAST-Net proposed in this paper under certain thresholds is slightly inferior to the previous model, it is worth noting that when $\tau = 40$, the CSI value of SAST-Net on Guangzhou Station dataset is up to 16.27% higher than the best-performing SA-ConvLSTM in previous works, HSS is improved by 12.26%, and similar results can be seen on CIKM 2017. This shows that SAST-Net can more accurately predict heavy precipitation areas.

Comparing with the results of PredRNN, which composed of the original ST-LSTM stack, it can be seen that

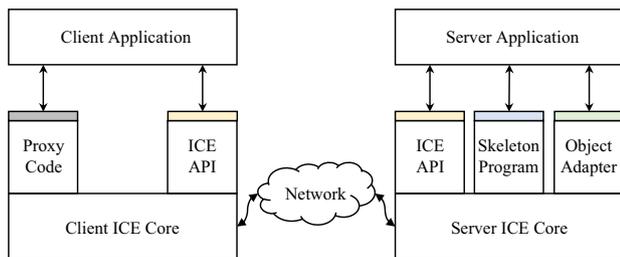


Figure 5: The communication process between server and client based ICE framework.

SAM has played an important role in improving the performance of the model. It can capture more global features and learn the radar echo movement more accurately. On Guangzhou Station dataset, the average CSI increased by 5.55%, the average HSS increased by 4.55%, and both increased by 7.53% and 10.12% on CIKM 2017. Compared with SA-ConvLSTM, which contains SAM but has no spatial temporal memory, SAST-Net captures high-order features through the additional unit, which improves the overall predictive ability of the model. The average CSI increased by 1.40%, the average HSS increased by 1.47% on Guangzhou Station dataset, and the two increased by 2.27% and 0.67% respectively on CIKM 2017. It can be seen that the spatiotemporal memory and SAM together contribute to the excellent prediction performance of the model proposed in this paper.

Fig. 6 shows the frame-by-frame CSI and HSS variation curves of different models on the two datasets when threshold $\tau = 40$, to compare the long-term prediction capabilities of the models. It can be seen from the figure that in the first and second frames, the previous models have a better performance, but the values of CSI and HSS of these models drop rapidly over time. In contrast, the prediction performance of SAST-Net was not as good as some models in the first two frames, the curves of CSI and HSS decreased much less rapidly than other models, and are always higher than other models from the third frame until the last frame, indicating that SAST-Net can be better applied to long-term prediction tasks and the accuracy of the prediction results is relatively stable.

Fig. 7 and Fig. 8 respectively shows the visualization of the model predictions on the two datasets. In Fig. 7, it can be seen that PhyDNet performs the worst, and its prediction results have been distorted. ConvLSTM and SA-ConvLSTM have the problem of underestimating the high-intensity echo area due to lack of spatiotemporal information. PredRNN++ predicts the high-intensity echo area correctly, but the low-intensity areas are not. PredRNN has the problem of overestimation of echo intensity in a large area. MIM and SAST-Net predict similar results, but SAST-Net predicts more accurately in terms of echo appearance and area boundaries. Fig. 8 is the prediction result on CIKM 2017. It can be clearly seen that ConvLSTM, PhyDNet and

Table 3

Results of the comparative experiment using CSI and HSS as evaluation metrics on the CIKM 2017 dataset.

Model	CSI					HSS				
	10	20	30	40	avg	10	20	30	40	avg
ConvLSTM	0.6283	0.4021	0.2116	0.0809	0.3307	0.6235	0.4913	0.3367	0.1485	0.4000
PredRNN	0.6743	0.3809	0.1819	0.0848	0.3305	0.6582	0.4382	0.2808	0.1528	0.3825
PredRNN++	0.6777	0.3860	0.2073	0.0968	0.3420	0.6569	0.4427	0.3201	0.1731	0.3982
MIM	0.6642	0.3776	0.2092	0.0970	0.3370	0.6539	0.4422	0.3259	0.1735	0.3989
PhyDNet	0.5811	0.3276	0.1576	0.0649	0.2828	0.5677	0.4091	0.2628	0.1212	0.3402
SA-ConvLSTM	0.6642	0.4006	0.2155	0.1096	0.3475	0.6613	0.4739	0.3423	0.1962	0.4184
SAST-Net	0.6856	0.3924	0.2295	0.1140	0.3554	0.6721	0.4553	0.3551	0.2025	0.4212

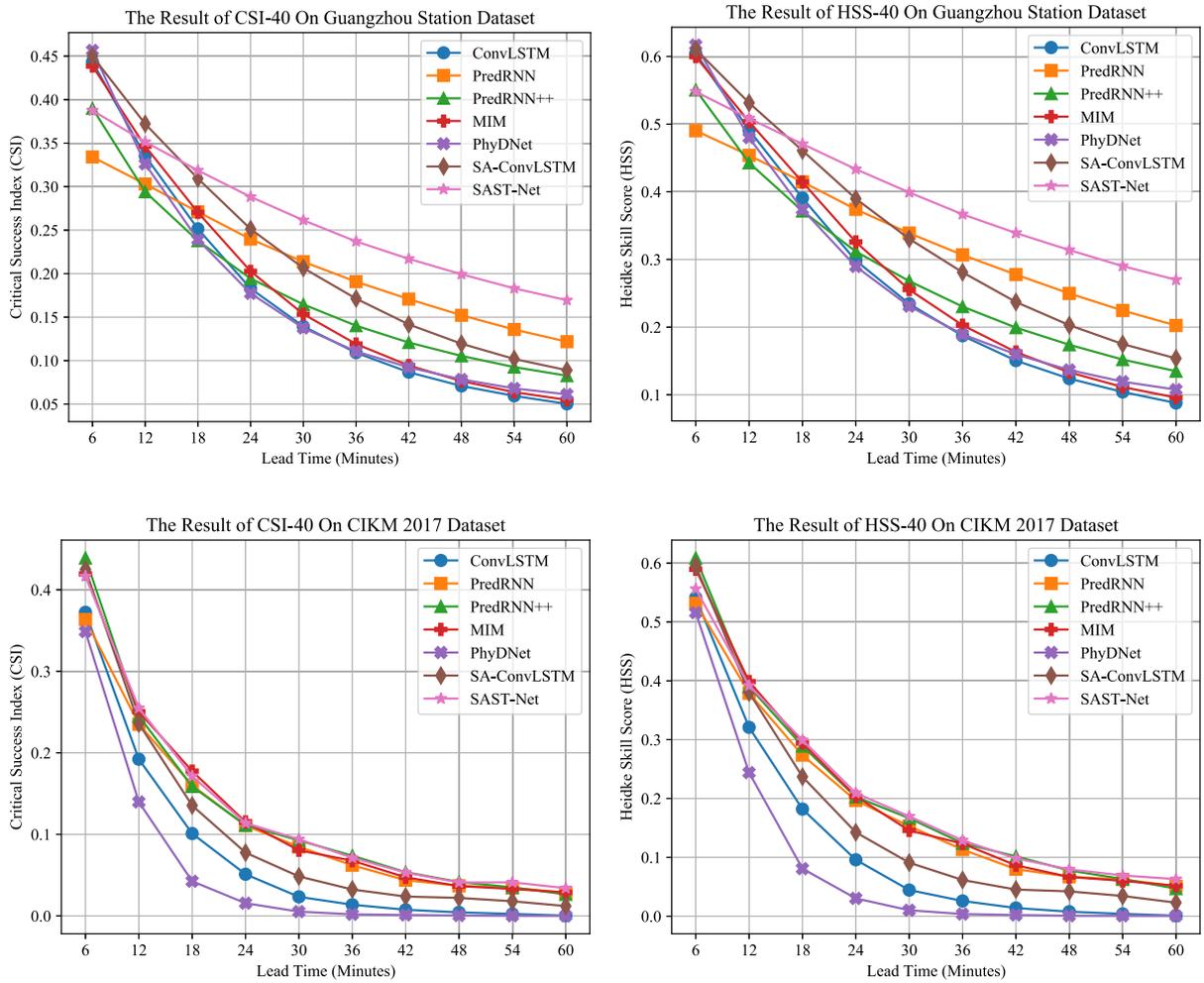


Figure 6: The frame-by-frame variation curves of different models in terms of CSI and HSS scores on **Guangzhou Station** and **CIKM 2017** datasets when threshold $\tau = 40$.

SA-ConvLSTM can't predict the yellow area. The yellow areas predicted by PredRNN and PredRNN++ are not accurate enough. Although MIM predicts the yellow area in the upper right corner correctly, but it can't predict the yellow area in the middle of echo map. In contrast, only SAST-Net predicts the most accurate results.

By conducting comparative experiments on two differ-

ent real-world radar echo data sets, it can be seen that the SAST-LSTM cell and the SAST-Net proposed in this paper has a strong ability to predict radar echo motion in different geographic areas, especially for Guangzhou Station, a kind of data with a wider observation range, reflecting the powerful effect of the self-attention mechanism on global feature capture.

SAST-LSTM

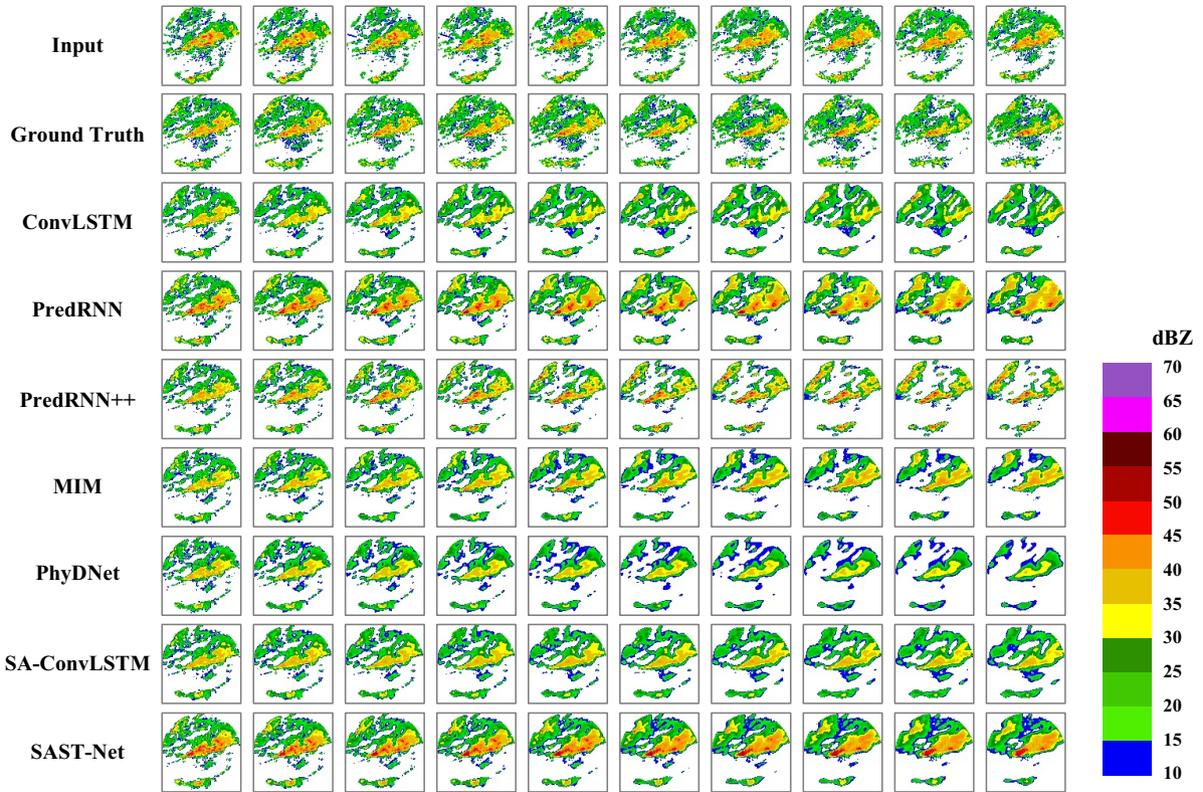


Figure 7: The prediction results of all models on **Guangzhou Station** dataset. The 10 images in the first row are the input, and the 10 images in the second row are the real values. The other rows are the predictions of various models.

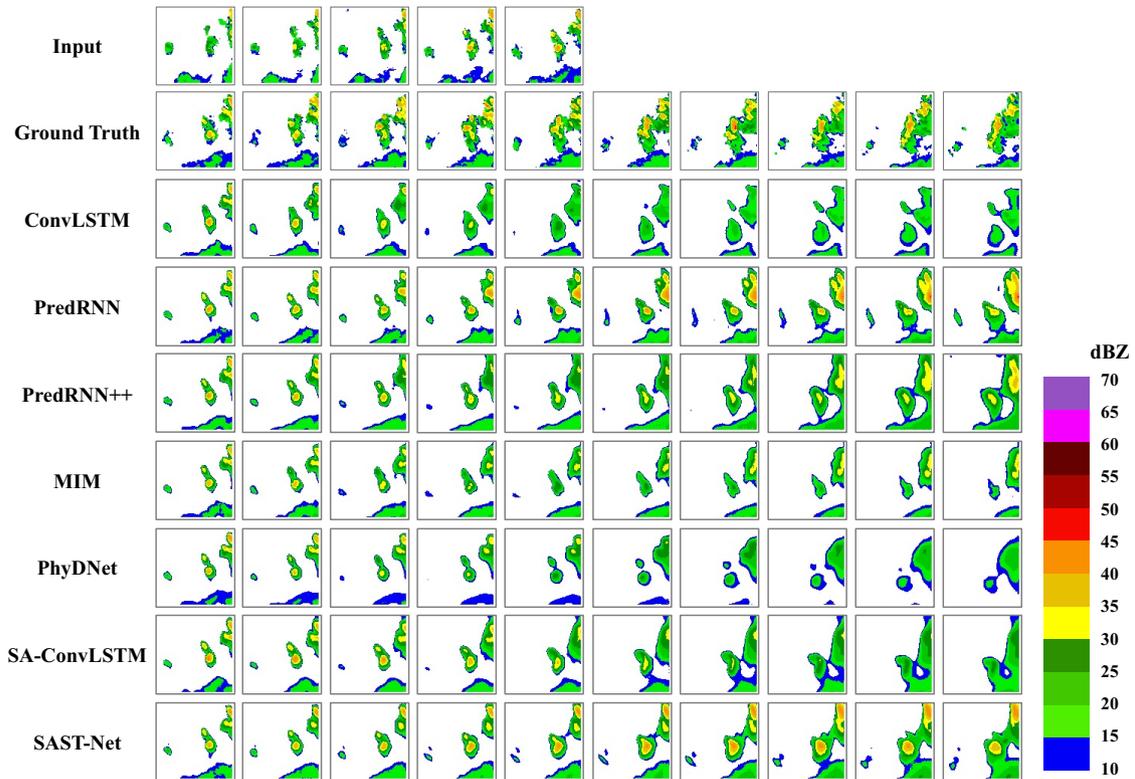


Figure 8: The prediction results of all models on **CIKM 2017** dataset. The 5 images in the first line are the input, and the second line is the real value. The other rows are the predictions of various models.

5. Conclusion

This paper proposes the SAST-LSTM, a spatiotemporal recurrent cell integrates the self-attentive mechanism, which enables the capture of global features by using a self-attentive memory module (SAM) and an additional global feature memory G , extending the perceptual field of previous models and enabling the cell to more accurately model the dynamics of complex moving objects such as radar echoes. In the experiment, the SAST-LSTM was stacked to build a four-layer SAST-Net structure, together with the models in previous works was used to perform radar echo extrapolation tasks on two different real-world radar echo datasets. The results show that the model proposed in this paper obtains the best results on both datasets. In the future, the author will seek solutions to improve the accuracy of the first two predicted frames in radar echo extrapolation tasks.

References

- [1] L. Li, S. Chen, X. Mai, Sub-pixel precipitation nowcasting over guangdong province using optical flow algorithm, in: 2017 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2017, pp. 4638–4641.
- [2] S. Mecklenburg, J. Joss, W. Schmid, Improving the nowcasting of precipitation in an alpine region with an enhanced radar echo tracking algorithm, *Journal of Hydrology* 239 (1–4) (2000) 46–68.
- [3] Q. Chu, Z. Xu, Y. Chen, D. Han, 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, *Hydrology and Earth System Sciences* 22 (6) (2018) 3391–3407.
- [4] O. B. Akan, M. Arik, Internet of radars: Sensing versus sending with joint radar-communications, *IEEE Communications Magazine* 58 (9) (2020) 13–19.
- [5] Q. Liang, Y. Feng, W. Deng, S. Hu, Y. Huang, Q. Zeng, Z. Chen, A composite approach of radar echo extrapolation based on trec vectors in combination with model-predicted winds, *Advances in Atmospheric Sciences* 27 (5) (2010) 1119–1130.
- [6] R. Rinehart, E. Garvey, Three-dimensional storm motion detection by conventional weather radar, *Nature* 273 (5660) (1978) 287–289.
- [7] J. Handwerker, Cell tracking with trace3d—a new algorithm, *Atmospheric Research* 61 (1) (2002) 15–34.
- [8] J. Johnson, P. L. MacKeen, A. Witt, E. D. W. Mitchell, G. J. Stumpf, M. D. Eilts, K. W. Thomas, The storm cell identification and tracking algorithm: An enhanced wsr-88d algorithm, *Weather and forecasting* 13 (2) (1998) 263–276.
- [9] W.-c. Woo, W.-k. Wong, Operational application of optical flow techniques to radar-based rainfall nowcasting, *Atmosphere* 8 (3) (2017) 48.
- [10] H. Sakaino, Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow, *IEEE Trans. Geosci. Remote. Sens.* 51 (5–2) (2013) 3023–3036.
- [11] P. Li, E. S. Lai, Applications of radar-based nowcasting techniques for mesoscale weather forecasting in hong kong, *Meteorological Applications* 11 (3) (2004) 253–264.
- [12] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, Deep learning for computer vision: A brief review, *Computational intelligence and neuroscience* 2018.
- [13] X. Xu, X. Zhang, X. Liu, J. Jiang, L. Qi, M. Z. A. Bhuiyan, Adaptive computation offloading with edge for 5g-envisioned internet of connected vehicles, *IEEE Trans. Intell. Transp. Syst.* 22 (8) (2021) 5213–5222.
- [14] X. Xu, Z. Fang, J. Zhang, Q. He, D. Yu, L. Qi, W. Dou, Edge content caching with deep spatiotemporal residual network for iov in smart city, *ACM Trans. Sens. Networks* 17 (3) (2021) 29:1–29:33.
- [15] X. Xu, Z. Fang, L. Qi, X. Zhang, Q. He, X. Zhou, Tripres: Traffic flow prediction driven resource reservation for multimedia iov with edge computing 17 (2).
- [16] X. Xu, H. Li, W. Xu, Z. Liu, L. Yao, F. Dai, Artificial intelligence for edge service optimization in internet of vehicles: A survey, *Tsinghua Science and Technology* 27 (2) (2021) 270–287.
- [17] Y. N. Malek, M. Najib, M. Bakhouya, M. Essaaidi, Multivariate deep learning approach for electric vehicle speed forecasting, *Big Data Mining and Analytics* 4 (1) (2021) 56–64.
- [18] R. Bi, Q. Liu, J. Ren, G. Tan, Utility aware offloading for mobile-edge computing, *Tsinghua Science and Technology* 26 (2) (2020) 239–250.
- [19] W. Wang, Z. Wang, Z. Zhou, H. Deng, W. Zhao, C. Wang, Y. Guo, Anomaly detection of industrial control systems based on transfer learning, *Tsinghua Science and Technology* 26 (6) (2021) 821–832.
- [20] S. C. K. Tekouabou, S. Hartini, Z. Rustam, H. Silkan, S. Agoujil, et al., Improvement in automated diagnosis of soft tissues tumors using machine learning, *Big Data Mining and Analytics* 4 (1) (2021) 33–46.
- [21] Y. Djenouri, A. Belhadi, G. Srivastava, U. Ghosh, P. Chatterjee, J. C.-W. Lin, Fast and accurate deep learning framework for secure fault diagnosis in the industrial internet of things, *IEEE Internet of Things Journal*.
- [22] Y. Liu, Z. Song, X. Xu, W. Rafique, X. Zhang, J. Shen, M. R. Khosravi, L. Qi, Bidirectional gru networks-based next poi category prediction for healthcare, *International Journal of Intelligent Systems*.
- [23] L. Qi, H. Song, X. Zhang, G. Srivastava, X. Xu, S. Yu, Compatibility-aware web api recommendation for mashup creation via textual description mining, *ACM Transactions on Multimedia Computing Communications and Applications* 17 (1s) (2021) 1–19.
- [24] U. Ahmed, G. Srivastava, J. C.-W. Lin, Reliable customer analysis using federated learning and exploring deep-attention edge intelligence, *Future Generation Computer Systems* 127 (2022) 70–79.
- [25] H. Li, J. Liu, R. W. Liu, N. Xiong, K. Wu, T.-h. Kim, A dimensionality reduction-based multi-step clustering method for robust vessel trajectory analysis, *Sensors* 17 (8) (2017) 1792.
- [26] Z. Ren, Y. Liu, T. Shi, L. Xie, Y. Zhou, J. Zhai, Y. Zhang, Y. Zhang, W. Chen, Aiperf: Automated machine learning as an ai-hpc benchmark, *Big Data Mining and Analytics* 4 (3) (2021) 208–220.
- [27] Y. Shao, J. C.-W. Lin, G. Srivastava, D. Guo, H. Zhang, H. Yi, A. Jolfaei, Multi-objective neural evolutionary algorithm for combinatorial optimization problems, *IEEE Transactions on Neural Networks and Learning Systems* (2021) 1–11.
- [28] Z. Tong, F. Ye, M. Yan, H. Liu, S. Basodi, A survey on algorithms for intelligent computing and smart city applications, *Big Data Mining and Analytics* 4 (3) (2021) 155–172.
- [29] W. Fang, X. Yao, X. Zhao, J. Yin, N. Xiong, A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48 (4) (2016) 522–534.
- [30] N. Bhardwaj, P. Sharma, An advanced uncertainty measure using fuzzy soft sets: Application to decision-making problems, *Big Data Mining and Analytics* 4 (2) (2021) 94–103.
- [31] D. Wei, H. Ning, F. Shi, Y. Wan, J. Xu, S. Yang, L. Zhu, Dataflow management in the internet of things: Sensing, control, and security, *Tsinghua Science and Technology* 26 (6) (2021) 918–930.
- [32] Y. Liu, D. Li, S. Wan, F. Wang, W. Dou, X. Xu, S. Li, R. Ma, L. Qi, A long short-term memory-based model for greenhouse climate prediction, *International Journal of Intelligent Systems*.
- [33] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: *Advances in neural information processing systems*, 2015, pp. 802–810.
- [34] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, W.-c. Woo, Deep learning for precipitation nowcasting: A benchmark and a new model, *arXiv preprint arXiv:1706.03458*.
- [35] Y. Wang, M. Long, J. Wang, Z. Gao, P. S. Yu, Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms, in: *Proceedings of the 31st International Conference on Neural Infor-*

- mation Processing Systems, 2017, pp. 879–888.
- [36] Y. Wang, Z. Gao, M. Long, J. Wang, S. Y. Philip, Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning, in: International Conference on Machine Learning, PMLR, 2018, pp. 5123–5132.
- [37] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, J. Hickey, Machine learning for precipitation nowcasting from radar images, arXiv preprint arXiv:1912.12132.
- [38] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 4905–4913.
- [39] Z. Lin, M. Li, Z. Zheng, Y. Cheng, C. Yuan, Self-attention convlstm for spatiotemporal prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 11531–11538.
- [40] J. Kim, M. El-Khamy, J. Lee, 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), IEEE, 2020, pp. 6649–6653.
- [41] K. J. Han, R. Prieto, T. Ma, 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), IEEE, 2019, pp. 54–61.
- [42] Y. Wu, Y. Ma, J. Liu, J. Du, L. Xing, Self-attention convolutional neural network for improved mr image reconstruction, Information sciences 490 (2019) 317–328.
- [43] H. Zhao, J. Jia, V. Koltun, Exploring self-attention for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10076–10085.
- [44] B. Klein, L. Wolf, Y. Afek, A dynamic convolutional layer for short range weather prediction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4840–4848.
- [45] W. Zhuang, W. Ding, Long-lead prediction of extreme precipitation cluster via a spatiotemporal convolutional neural network, in: Proceedings of the 6th International Workshop on Climate Informatics: CI, 2016.
- [46] G. Ayzel, T. Scheffer, M. Heistermann, Rainnet v1.0: a convolutional neural network for radar-based precipitation nowcasting, Geoscientific Model Development 13 (6) (2020) 2631–2644.
- [47] J. G. Fernandez, S. Mehrkanoon, Broad-unet: Multi-scale feature learning for nowcasting tasks, arXiv preprint arXiv:2102.06442.
- [48] L. Han, H. Liang, H. Chen, W. Zhang, Y. Ge, Convective precipitation nowcasting using u-net model, IEEE Transactions on Geoscience and Remote Sensing (2021) 1–8.
- [49] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, P. S. Yu, Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9154–9162.
- [50] V. L. Guen, N. Thome, Disentangling physical dynamics from unknown factors for unsupervised video prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11474–11484.
- [51] Z. Chai, C. Yuan, Z. Lin, Y. Bai, Cms-lstm: Context-embedding and multi-scale spatiotemporal-expression lstm for video prediction, arXiv preprint arXiv:2102.03586.
- [52] S. Zhong, X. Zeng, Q. Ling, Q. Wen, W. Meng, Y. Feng, Spatiotemporal convolutional lstm for radar echo extrapolation, in: 2020 54th Asilomar Conference on Signals, Systems, and Computers, IEEE, 2020, pp. 58–62.
- [53] K. Trebing, T. Stanczyk, S. Mehrkanoon, Smaat-unet: Precipitation nowcasting using a small attention-unet architecture, Pattern Recognit. Lett. 145 (2021) 178–186.
- [54] C. Luo, X. Li, Y. Wen, Y. Ye, X. Zhang, A novel lstm model with interaction dual attention for radar echo extrapolation, Remote Sensing 13 (2) (2021) 164.
- [55] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
- [56] S. Bengio, O. Vinyals, N. Jaitly, N. Shazeer, Scheduled sampling for sequence prediction with recurrent neural networks, arXiv preprint arXiv:1506.03099.
- [57] J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450.

Declaration of interests

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: