Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.088306

ARTICLE

Radar Echo Extrapolation Method Based on Spatio-Temporal Dual-Stream Attention Mechanism

Yuelei Niu[1], **Chuang Liu**[1], **Xiaodong Liu**[2,*] **and Qi Liu**[3]

[1]Guodian Nanjing Automation Co., Ltd., Nanjing, China
[2]School of Computing, Edinburgh Napier University, Edinburgh, UK
[3]NUIST Waterford Institute, Nanjing University of Information Science and Technology, Nanjing, China

*Corresponding Author: Xiaodong Liu. Email: x.liu@napier.ac.uk

Received: 01 July 2026; Accepted: 28 August 2026; Published: 22 October 2026

ABSTRACT: In the scenario where extreme precipitation occurs frequently, we addressed the issues of insufficient modeling of spatial and temporal characteristics and long-term prediction cumulative errors in the existing methods, and proposed a long-term radar echo extrapolation method STDA-Net based on the dual-stream spatio-temporal attention mechanism. With the stacked STDA-LSTM as the core module, we introduced the spatial self-attention mechanism (SSAM) in the image and the time attention mechanism (TAM) between images to capture the global spatial dependence and time evolution characteristics of radar images. At the same time, we designed an adaptive entropy regularization loss function (AER Loss) to balance the pixel-level errors and time-related constraints, effectively suppressing prediction ambiguity and error accumulation. The experiments used the Guangzhou Stationand CIKM 2017 datasets, inputting 10 frames and 5 frames, respectively, and outputting 10 frames. The results show that STDA-Net outperforms the mainstream models in key indicators, has a slower error growth in long-term prediction, and can better retain the echo structure, thereby improving the accuracy and stability of long-term radar echo extrapolation.

KEYWORDS: Radar echo extrapolation; long-term effect; spatio-temporal dual-stream attention

1 Introduction

Due to global warming, as the temperature rises, the water-holding capacity of the atmosphere significantly increases, which in turn leads to an increase in the frequency of extreme precipitation events. Short-term heavy precipitation, as a type of extreme weather, has characteristics such as strong suddenness, high disaster potential, and short forecast time, greatly affecting human production and life [1–3]. Short-term precipitation forecasting is the prediction of the amount of precipitation in a certain area within a short period (such as 0 to 2 h), and it is a high-resolution forecast [4]. Currently, short-term precipitation forecasting is mostly achieved through extrapolation methods based on radar echo data.

The meteorological radar echo extrapolation models can generally be classified into two categories. One type is the traditional numerical calculation method for meteorological radar echo extrapolation, including cross-correlation tracking echo method, centroid tracking method, and optical flow method, etc. [5]. This traditional extrapolation method is difficult to achieve accurate prediction for the occurrence and dissipation process of heavy rain events. The other type is a radar echo extrapolation model based on deep learning. The deep learning method has achieved remarkable results in computer vision [6,7] and information security fields.

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We have proposed a long-range radar echo extrapolation method called STDA-Net, which is constructed based on a spatio-temporal dual-stream attention mechanism. By integrating intra-image spatial self-attention and inter-image temporal attention, the method fully exploits long-range spatial dependencies in radar images and effectively captures the evolutionary characteristics of echoes along the time axis. With the synergy of these two attention modules, STDA-Net suppresses error accumulation in long-term prediction, strengthens the modeling of echo motion trends, and this has thereby improved the accuracy and stability of long-term radar echo predictions. The main contributions of this work are summarized as follows:

(1) We propose a spatial-temporal dual-stream attention network STDA-Net. Built on stacked STDA-LSTM units, it transmits information via spatial and temporal dual paths to alleviate long-term dependency loss and rapid echo trend decay in conventional models, enabling accurate long-term radar echo extrapolation.
(2) We design a spatial-temporal dual-stream attention mechanism. Intra-image spatial self-attention (SSAM) adopts dilated convolution and global self-attention to capture long-range spatial dependencies and enhance modeling of complex echo structures. Inter-image temporal attention (TAM) uses SE-based channel attention to adaptively learn temporal weights and capture echo evolution. Together, they mitigate long-term prediction error accumulation and strengthen spatiotemporal feature representation.
(3) We present an adaptive entropy regularization loss (AER Loss), which combines Mean Squared Error (MSE) pixel loss and entropy distribution constraints. It reduces pixel-level errors while preserving echo spatiotemporal dynamics, alleviating prediction ambiguity and detail blurring caused by conventional loss functions.

2 Related Work

In the extrapolation methods based on recurrent neural networks, Shi et al. [8] proposed the Convolutional Long Short-Term Memory Network (ConvLSTM) in 2015. Wang et al. [9] constructed the PredRNN network by stacking spatio-temporal Long Short-Term Memory (ST-LSTM) units to improve the modeling ability for complex objects. To address the problem of gradient disappearance that occurs when stacking multiple loop unit networks, the variant PredRNN++ [10] was proposed in the following year. PredRNN++ introduced causal Long Short-Term units based on the bipolar connection mechanism, added nonlinear layers of recurrent units, and introduced the Gradient Highway Unit (GHU) to effectively improve the gradient transmission in deeper networks. Wang et al. [11] proposed the Memory In Memory (MIM) method, using difference signals derived from neighboring cyclic states to simulate the stable and unstable behaviors of spatiotemporal dynamic processes. Models such as IDA-LSTM [12] incorporate the attention mechanism, which enhances the focus on high-echo areas and key temporal information, but these methods mostly focus on local spatio-temporal feature modeling and are difficult to explicitly model global spatial dependencies and long-term temporal evolution trends. We embedded the inter-node temporal attention mechanism into the recurrent units to construct the STDA-LSTM new recurrent structure based on the core design ideas of temporal dependency modeling, long-term memory transmission, and spatio-temporal LSTM recurrent units, aiming to address the problems of difficult long-term sequence changes and information attenuation in existing recurrent structures.

An encoder-decoder architecture is predominantly adopted in the extrapolation method built upon convolutional neural networks. Ronneberger et al. [13] first proposed the U-Net network, extracting deep semantic features of images through the convolution and pooling operations of the encoder, and then fusing pixel semantic features and position features through the upsampling and skip connections of the decoder.

Trebing et al. [14] proposed the SmaAt-UNet method, which integrates a small attention mechanism module CBAM into the U-Net architecture and replaces the traditional convolution operations with depthwise separable convolutions. The simple convolution structure is limited by the local receptive field and is prone to problems such as blurred echo morphology and intensity attenuation in long-term time series prediction. We adopt the encoder-decoder structure of the U-Net series, multi-scale convolution, and non-local network. We continue to use its mainstream architectures for spatial feature extraction, downsampling-upsampling, and skip connections. Moreover, we introduce an intra-node spatial self-attention mechanism after the convolutional layers of the encoder to fuse local convolutional features with global spatial-related features, thereby compensating for the deficiency of traditional convolutional methods that can only capture local spatial information.

The hybrid neural network method, defined by Zeng et al. [15], combines a new radar echo extrapolation model T-UNet, which is a combination of CNN and RNN, and can effectively retain the spatio-temporal features of radar images. Models such as 3D-UNet-LSTM [16] extract spatial features through convolution and the recurrent unit models temporal dependencies to achieve joint learning of spatial and temporal features. The researchers have proposed a novel recursive evolutionary memory-aware network called REMNet [17], which incorporates a brand-new long-term evolutionary rule memory (LERM) module. This module is capable of remembering long-term echo evolution rules during the training process. Geng et al. [18] proposed models: residual convolutional LSTM (rcLSTM) and generative adversarial network residual convolutional LSTM (GAN-rcLSTM). These two networks can reduce the degradation phenomenon of the long short-term memory network. RM-ConVLSTM [19] designs a module called Recurrent Memory Perception (RMP) to enhance the model's ability to capture and retain the long-term motion features of echo sequences. We adopt the mature framework of "convolution for extracting spatial features + recurrent network for modeling temporal dependencies", further integrate the memory recurrent perception mechanism and the dual-stream attention mechanism to construct an end-to-end STDA-Net network, addressing the issues of insufficient coupling of spatial features and decreased long-term extrapolation accuracy in existing hybrid models.

In terms of loss functions and model optimization, existing methods mostly use MSE as the core loss, focusing only on pixel-level error constraints and ignoring the temporal changes and intensity distribution characteristics of the radar echo sequence. In the meteorological field, indicators such as Critical Success Index (CSI), Heidke Skill Score (HSS), False Alarm Ratio (FAR) and Probability of Detection (POD) are commonly used to evaluate forecast performance. The forecast accuracy of existing models significantly decreases under high reflectivity thresholds, and the problem of cumulative long-term prediction errors is prominent. At the same time, most methods have not designed specific constraints for the complex dynamic characteristics of radar echo generation, dissipation, and deformation, resulting in poor effect.

3 Methodology

3.1 STDA-Net Radar Echo Extrapolation Network Design

To address the issue that the model is unable to capture the long-term movement and change trends of radar echoes, this paper introduces the spatial self-attention mechanism (SSAM) and the temporal attention mechanism (TAM) within the extrapolation network of MRP-Net, and constructs the STDA-Net (Spatio-Temporal Dual-Attention Net) model based on the dual-stream attention mechanism of space and time. This paper presents STDA-Net, which is composed of 6 layers of spatio-temporal STDA-LSTM modules stacked together. The hidden feature channels are uniformly set to 128. Each layer's encoder contains an SSAM composed of dilated convolutions, and the decoding part of the decoder embeds TAM. Both SSAM and TAM are enabled simultaneously in each layer. The convolution kernel size is uniformly 3×3. The expansion

rate of the dilated convolutions within SSAM is 1, 2, and 3, respectively. Cross-layer skip connections are configured between layers.

This enhances the model's ability to simulate the movement characteristics of radar echoes, and enables the model to accurately predict the long-term motion features of radar echoes. The proposed STDA-Net network is composed of multiple STDA-LSTM recurrent units stacked together, and its network architecture is shown in Fig. 1. This network also consists of four layers, with memory units C_{tl}, hidden states H_{tl}, and memory units M_{tl} transmitting information in both horizontal and vertical directions to obtain spatio-temporal features and avoid the attenuation of trends during long-term motion.

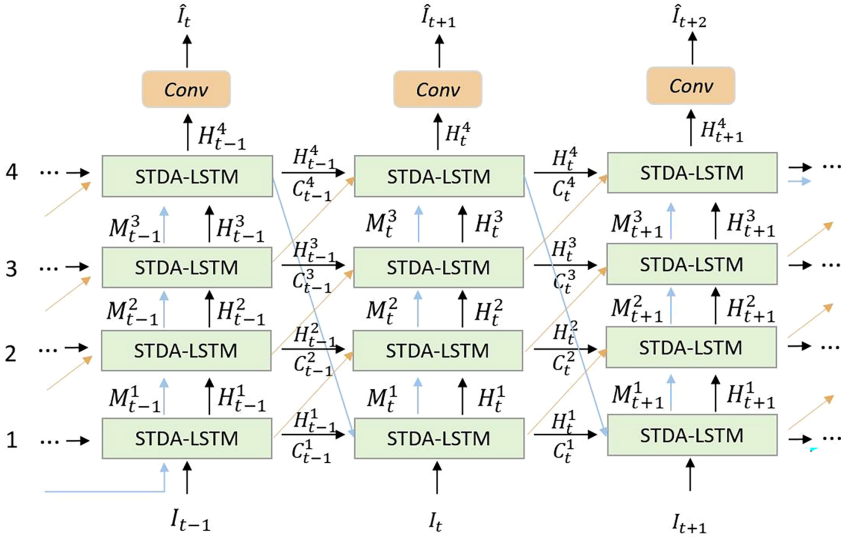


Figure 1: The overall structure of the STDA-Net weather radar echo extrapolation network.

According to the above network architecture, in order to fully demonstrate the roles of the SSAM and TAM that we proposed in the time domain and spatial domain, Fig. 2 shows the unit parameter information at three different time points. In this figure, the light yellow rectangular blocks represent the first layer units in the spatial dimension at a single time point, the dark yellow rectangular blocks represent the second layer units in the spatial dimension at a single time point, the red dashed arrows represent the transmission direction of the temporal dimension memory unit M_{tl} and the hidden state H_{tl}, the blue arrow indicates the transmission direction of the dimension storage unit M_{tl} and the hidden state H_{tl}, and the grey arrows indicate the direction of data transmission between the unit's input and output.

This structure establishes a time transmission path between adjacent time units at the same level and shares parameters to effectively analyze the movement trend of radar echoes. It also builds a spatial transmission path between different hierarchical units at the same time to enable them to obtain the states of adjacent time periods and capture spatial features. The two attention mechanisms help the model better capture complex changes and improve prediction accuracy.

3.2 STDA-LSTM Recurrent Unit Design

The proposed spatiotemporal dual-stream attention mechanism combines intra-graph spatial self-attention and inter-image temporal attention. The former captures long-range dependencies within radar images by expanding the receptive field, while the latter learns channel-wise attention weights to model

temporal evolution. Their structure is illustrated in Fig. 3. This section details the two attention mechanisms and the integrated STDA-LSTM recurrent unit.

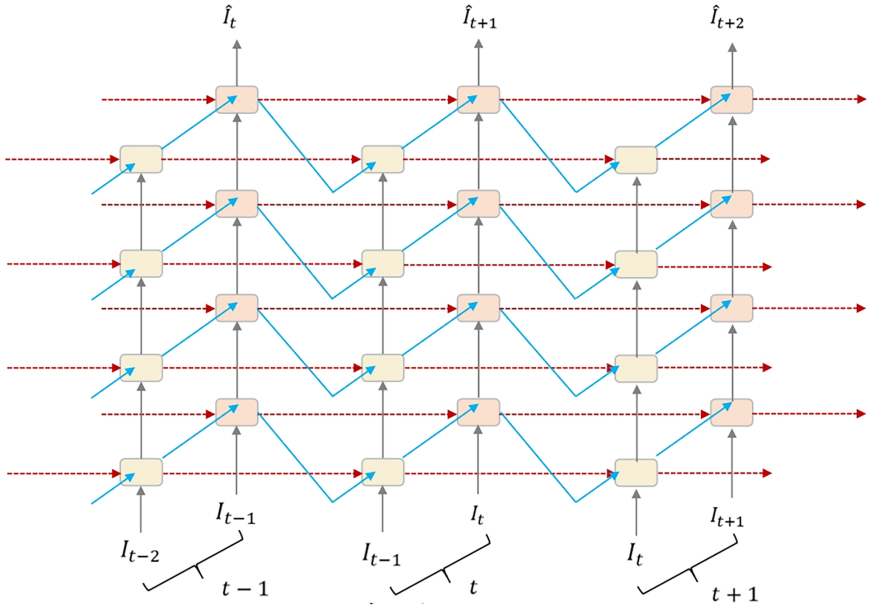


Figure 2: The working principle of the spatio-temporal dual-stream attention mechanism.

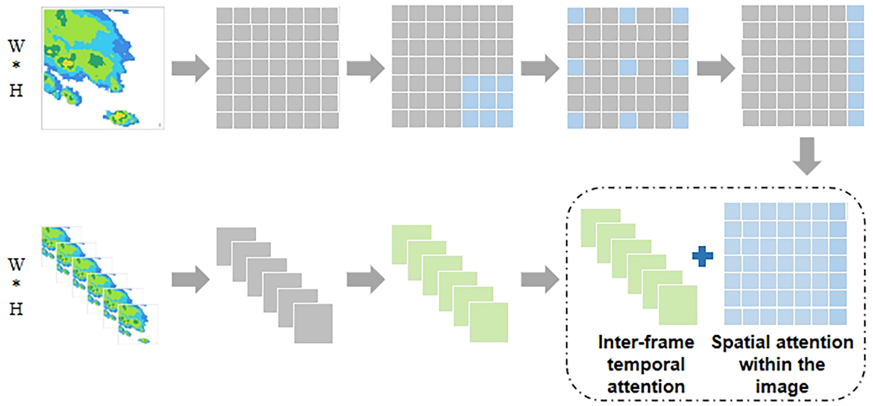


Figure 3: Schematic diagram of the spatio-temporal dual-stream attention mechanism.

3.2.1 Spatial Self-Attention Mechanism within the Image

In radar echo extrapolation, spatial structure information is critical. Traditional convolutional neural networks can extract local features, but they have difficulty simulating long-distance spatial relationships. To address, we introduce SSAM after the encoder's convolutional layers to capture long-range dependencies within the image, as illustrated in Fig. 4.

The spatial self-attention mechanism illustrated in the figure takes the current time-step hidden state H_t as its input and the memory unit M_{t-1} from the previous time-step. The operation process of the entire mechanism can be divided into five stages: feature extraction, attention weight calculation, weighted feature aggregation, feature fusion, and memory unit update. In the feature extraction stage, three dilated convolutions are used to extract and learn radar echo features of different scales from the current time-step's

hidden state Ht and the memory unit M_{t-1}. The convolution kernel size is all 3 × 3, and the dilation rates are successively set as 1, 2, and 3. With the help of the dilated convolution technique, the receptive field size can be expanded to 3 × 3, 5 × 5, and 7 × 7, respectively, which effectively enhances the model's ability in spatial feature extraction. After the convolution transformation, the query vector Q_h, key vector k_h and k_m, value vector V_h and V_m are obtained. The memory unit shares the query vector with the hidden state. The process is shown in Formulas (1) to (5), where W represents the convolution filter.

$$Q_h = W_{hq} \times H_t \tag{1}$$

$$K_h = W_{hk} \times H_t \tag{2}$$

$$V_h = W_{hv} \times H_t \tag{3}$$

$$K_m = W_{MK} \times M_{t-1} \tag{4}$$

$$v_M = W_{MV} \times M_{t-1} \tag{5}$$

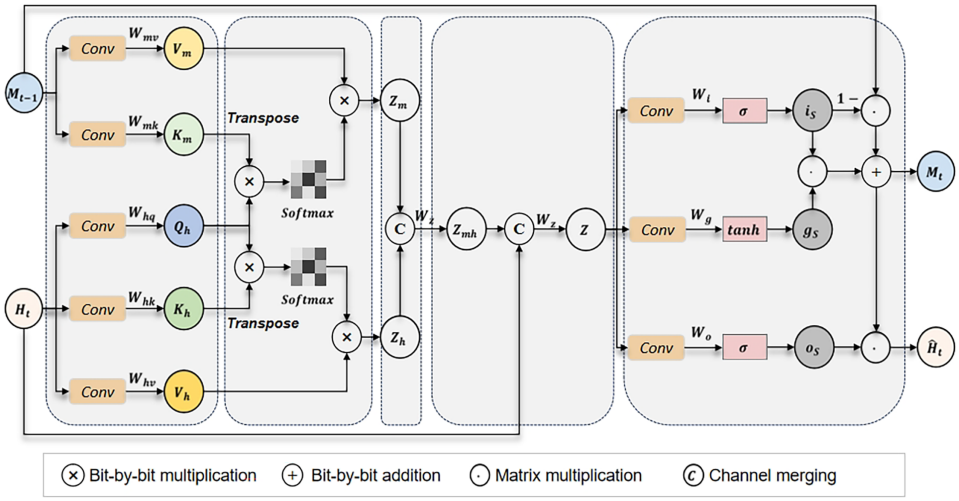


Figure 4: The structure diagram of the intra-image spatial self-attention mechanism (SSAM).

During the calculation of attention weights, the Softmax function is used for result normalization. The calculation process is shown by Formulas (6) and (7), where d represents the channel dimension and is used to prevent gradient explosion. A_h and A_m respectively denote the spatial attention matrices of the hidden state and the memory state.

$$A_h = \text{Softmax}\left(\frac{Q_h \cdot K_h^{\top}}{\sqrt{d}}\right) \tag{6}$$

$$A_m = \text{Softmax}\left(\frac{Q_h \cdot K_m^{\top}}{\sqrt{d}}\right) \tag{7}$$

After obtaining the attention weights through the above process, the value vectors are weighted and summed to obtain the feature maps Z_h and Z_m after extracting the global feature information. The calculation process is shown in Formulas (8) and (9).

$$Z_h = A_h \cdot V_h \tag{8}$$

$$Z_m = A_m \cdot V_m \tag{9}$$

Then, Z_h and Z_m are fused along the 1×1 convolution channels, as shown in Formulas (10) and (11). W_z is the convolution filter and ∘ represents the channel fusion operation.

$$Z_{mh} = W_z \times (Z_h \circ Z_m) \tag{10}$$

$$Z = W_z \cdot Z_{mh} \tag{11}$$

The feature information in the memory unit is updated using i_s and g_s, and it is integrated with the initial memory state M_{t-1} to obtain the final output H_t. The calculation process is shown in Formulas (12) to (16), W_i, W_g, W_o represent the convolution filters, and b is the bias term.

$$i_S = \sigma\left(W_i \cdot Z + W_i \cdot H_t + b_i\right) \tag{12}$$

$$g_S = \tanh\left(W_g \cdot Z + W_g \cdot H_t + b_g\right) \tag{13}$$

$$o_S = \sigma\left(W_o * Z + W_o * H_t + b_o\right) \tag{14}$$

$$M_t = (1 - i_S) \odot M_{t-1} + i_S \odot g_S \tag{15}$$

$$\hat{H}_t = o_S \odot M_t \tag{16}$$

3.2.2 Inter-Image Temporal Attention Mechanism

MRP-Net struggles to effectively capture long-term motion trends in radar echo images. To overcome this limitation, we introduced a feature update module based on inter-image temporal attention (TAM) in the skip connection layer of the STDA-LSTM decoder. TAM models long-term temporal dependencies by learning dynamic relationships across time steps. It computes channel-wise temporal attention weights using squeeze-excitation, average pooling, and fully connected layers, and refines feature representations via a gating mechanism, as illustrated in Fig. 5.

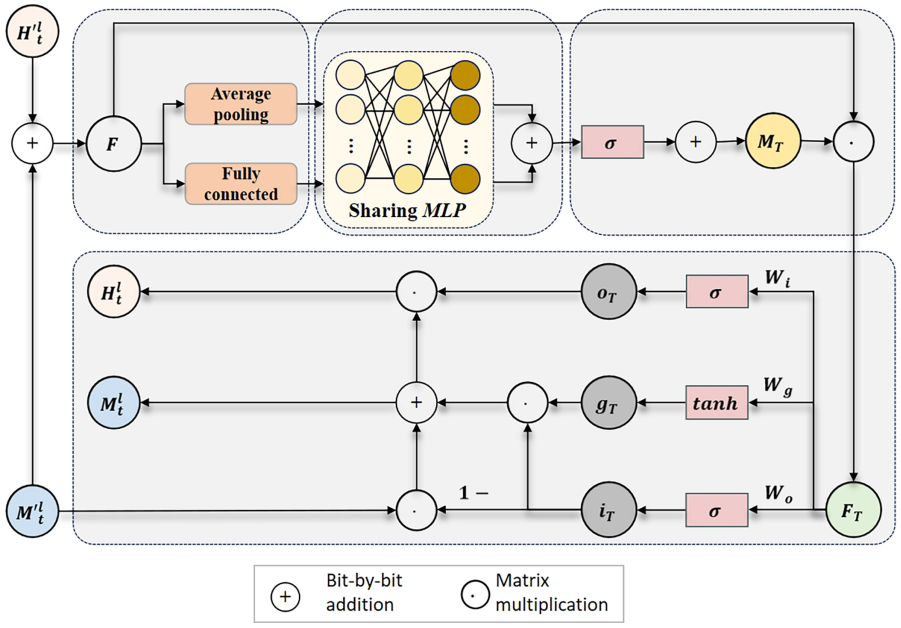


Figure 5: The structure diagram of the inter-image temporal self-attention mechanism (TAM).

The input of the inter-image temporal attention mechanism is the hidden state H'^{l}_{t} of the current time step after the context feature fusion operation, as well as the memory unit M'^{l}_{t}. The operation process of the entire mechanism is similar to that of SSAM, divided into four stages: feature extraction, attention weight calculation, feature fusion, and memory unit update. Firstly, in the feature extraction stage, the hidden state H'^{l}_{t} and the memory unit M'^{l}_{t} are added together to calculate the global feature F, as shown in Eq. (17).

$$F = H'^{l}_{t} + M'^{l}_{t} \tag{17}$$

During the attention weight calculation stage, the time attention is calculated using the squeeze-excitation mechanism. Firstly, the global feature F is averaged pooled. Then, the channel dimension is reduced using a fully connected layer, and adaptive attention weights are calculated using a shared MLP. Finally, element-wise addition is performed and activation operation is carried out using the Sigmoid function to obtain the attention weight matrix M_T. The calculation process is shown in Formulas (18) to (20), where W_f is the dimension reduction mapping matrix and b_f is the bias term.

$$F' = \text{AvgPool}(F) \tag{18}$$

$$F'' = W_f \cdot F' + b_f \tag{19}$$

$$M_T = \sigma\left(\text{MLP}(F') + \text{MLP}(F'')\right) \tag{20}$$

Then, the feature fusion operation is carried out. The attention weight matrix M_T is multiplied with the original input feature F through matrix multiplication to obtain the input feature f_T of the temporal attention. The calculation process is shown in Formula (21).

$$F_T = M_T \odot F \tag{21}$$

Finally, the memory unit is updated. During this process, the calculation processes of the input gate i_t, the input modulation gate g_t and the output gate o_t are as shown in Formulas (22) to (24). The updated memory unit M^{l}_{t} is determined by i_T, g_T and the initial memory state M'^{l}_{t}. The final hidden state H^{l}_{t} is obtained by performing matrix multiplication with o_t and the updated memory unit M^{l}_{t}, as shown in Formulas (25) and (26). Here, W_i, W_g, W_o represent convolution filters, and b is the bias term.

$$i_T = \sigma\left(W_i \cdot F_T + W_i \cdot M'^{l}_{t} + b_i\right) \tag{22}$$

$$g_T = \tanh\left(W_g \cdot F_T + W_g \cdot M'^{l}_{t} + b_g\right) \tag{23}$$

$$o_T = \sigma\left(W_o \cdot F_T + W_o \cdot M'^{l}_{t} + b_o\right) \tag{24}$$

$$M^{l}_{t} = (1 - i_T) \odot M'^{l}_{t} + i_T \odot g_T \tag{25}$$

$$H^{l}_{t} = o_T \odot M^{l}_{t} \tag{26}$$

3.2.3 The Operation Process of the STDA-LSTM Recurrent Unit

In the model MRP-LSTM, SSAM and TAM are introduced to construct STDA-LSTM. This method aims to enhance the spatio-temporal modeling capability, enabling the model to capture more accurate spatial dependencies and temporal dynamic changes and improving the long-term prediction performance of radar echo extrapolation. Its structure is shown in Figs. 4–6.

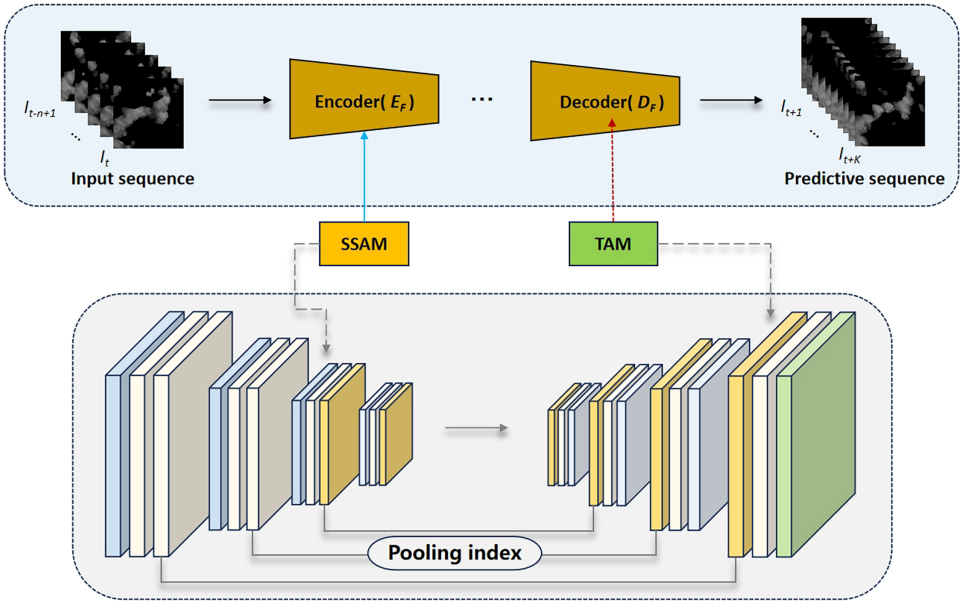


Figure 6: The operation process of the STDA-LSTM recurrent unit.

Firstly, H_{t-1}, along with the input state I_t at the current time step are input into the encoder that incorporates SSAM for encoding representation, in order to capture the long-range dependencies within the image. The output then yields $\hat{H}'_{t-1}$ and $\hat{I}'_t$. Subsequently, the obtained $\hat{H}'_{t-1}$ and $\hat{I}'_t$ undergo gated operations to obtain the states of the forget gate f_t, input gate i_t, input modulation gate g_t and output gate o_t, after which the cell state C_t and hidden state H_t are updated. The specific calculation process is as shown in Eqs. (27)–(34).

$$\hat{I}'_t = \text{SSAM}\left(I_t\right) \tag{27}$$

$$\hat{H}'_{t-1} = \text{SSAM}\left(H_{t-1}\right) \tag{28}$$

$$f_t = \sigma\left(W_f \cdot \hat{I}'_t + W_f \cdot \hat{H}'_{t-1} + b_f\right) \tag{29}$$

$$i_t = \sigma\left(W_i \cdot \hat{I}'_t + W_i \cdot \hat{H}'_{t-1} + b_i\right) \tag{30}$$

$$g_t = \tanh\left(W_g \cdot \hat{I}'_t + W_g \cdot \hat{H}'_{t-1} + b_g\right) \tag{31}$$

$$o_t = \sigma\left(W_o \cdot \hat{I}'_t + W_o \cdot \hat{H}'_{t-1} + b_o\right) \tag{32}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \tag{33}$$

$$H_t = o_t \odot \tanh\left(C_t\right) \tag{34}$$

The hidden state H_t and memory unit M_t are updated after passing through the context feature fusion module to obtain the updated hidden state H'^l_t and memory unit M'^l_t. These H'^l_t and M'^l_t are then passed into the TAM in the decoder skip connection layer for decoding representation to obtain the final H^l_t and M^l_t. The calculation process is as shown in Formulas (35) and (36).

$$M^l_t = \text{TAM}\left(M'^l_t\right) \tag{35}$$

$$H^l_t = \text{TAM}\left(H'^l_t\right) \tag{36}$$

3.3 Loss Function Optimization

The traditional mean squared error (MSE) mainly focuses on pixel-level errors, only measuring the pixel intensity difference between the predicted image and the real image, and is unable to fully capture the time series characteristics and complex spatial structure relationships of radar echo data. Therefore, we proposed an adaptive entropy regularization loss function (AER Loss). This loss function not only considers pixel-level errors but also combines the spatio-temporal correlation between images, further improving the prediction accuracy of the model.

The adaptive entropy regularization loss function $\mathcal{L}_{\text{AER}}$ consists of MSE and entropy regularization components, as shown in Eq. (37), where λ_1 and λ_2 are weight coefficients used to balance the contributions of MSE and entropy regularization in the total loss.

$$\mathcal{L}_{\text{AER}} = \lambda_1 \mathcal{L}_{\text{MSE}} + \lambda_2 \mathcal{L}_{\text{E}} \tag{37}$$

In this formula, $\mathcal{L}_{\text{MSE}}$ is used to calculate the difference between the predicted image $\hat{I}_t$ and the true image I_t. The calculation process is as shown in Formula (38) and m represents the total number of samples.

$$\mathcal{L}_{\text{MSE}} = \frac{1}{m} \sum_{i=1}^{m} \left(I_i - \hat{I}_i \right)^2 \tag{38}$$

The entropy regularization loss $\mathcal{L}_{\text{E}}$ calculates the probability distribution of the predicted image to make the prediction smoother. Its calculation process is shown in Formula (39). Here, $p\left(I_i\right)$ represents the probability of the pixel value, which is estimated using a Gaussian kernel. The calculation formula of $p\left(I_i\right)$ is as shown in (40), where σ is the bandwidth used to control the smoothness of the distribution.

$$\mathcal{L}_{\text{E}} = -\sum_{i} p(I_i) \log p(I_i) \tag{39}$$

$$p(I_i) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(I_i - I_j\right)^2}{2\sigma^2} \right) \tag{40}$$

The probability $p(I_i)$ in Eq. (40) is estimated via Gaussian-kernel density estimation, which approximates the intensity distribution of predicted radar echoes from sample pixels. The bandwidth parameter σ controls the smoothness of the estimated distribution: larger σ produces smoother probability distributions, while smaller σ retains more local pixel-intensity fluctuations.

The adaptive entropy regularization loss function (AER Loss) we proposed incorporates deep learning on the basis of the traditional mean squared error (MSE) loss. It not only takes into account the pixel-level errors in the image but also integrates the temporal correlation between images, thereby improving the prediction accuracy of the radar echo extrapolation task.

4 Experimental Results

4.1 Experimental Environment

The CIKM 2017 radar echo dataset [20] contains 14,000 real radar echo sequences. The training set, validation set, and test set each consist of 8000, 2000, and 4000 sequences, respectively. The Guangzhou station time span is from May 2018 to September 2022, and the training set, validation set, and test set are 22,914, 2864, and 2864 sequences, respectively. The radar echo dataset covers an area of 1 km × 1 km and has a time resolution of 6 min. The experimental environment uses the Ubuntu 18.04.1 system, combined with

PyTorch 1.12.0, and uses the NVIDIA GeForce RTX 3090 as the GPU device. The model optimizers all use the Adam algorithm for training, with an initial learning rate of 0.001 and a batch size of 4. The validation set is tested after every 200 training iterations. The model inputs 5 frames and 10 frames of historical radar echo data for CIKM and Guangzhou station, respectively, and outputs the predicted results of the next 10 frames of radar echoes, thereby achieving short-term extrapolation prediction of radar echoes. We focus on the long-term extrapolation prediction accuracy of radar echoes. The evaluation of model complexity such as FLOPs and inference speed is not within the scope of this study.

4.2 Comparison of Experimental Results and Analysis

We used ConvLSTM, PredRNN, PredRNN++, MIM, PhyDNet, SA-ConvLSTM and MRP-Net as the comparison models. CSI, HSS, FAR and POD are first computed at 20, 30, 40 dBZ by pixel-wise binarization. CSI-AVG, HSS-AVG, FAR-AVG, POD-AVG denote arithmetic averages over the three thresholds. Each metric is marked with an arrow: an upward arrow indicates higher scores mean better performance, while a downward arrow indicates lower scores mean better performance.

We provide a quantitative comparison of STDA-Net with other models on the Guangzhou radar echo dataset in Table 1. Results across key metrics show that STDA-Net surpasses all baseline models in every indicator, with MRP-Net and PhyDNet ranking second. Notably, the substantial gains in CSI-AVG, HSS-AVG, and POD-AVG demonstrate that STDA-Net predicts precipitation regions more accurately. Meanwhile, the lower FAR-AVG value confirms its effectiveness in reducing false alarms.

Table 1: Quantitative results comparison of each model in the radar echo dataset of Guangzhou station.

Model	CSI-AVG ↑	HSS-AVG ↑	FAR-AVG ↓	POD-AVG ↑	MSE ↓	SSIM ↑
ConvLSTM	0.4268	0.5388	0.5856	0.2703	169.4865	0.5846
PredRNN	0.3243	0.3714	0.4908	0.2259	161.5621	0.4960
PredRNN++	0.4177	0.5297	0.4654	0.3312	135.2987	0.5639
MIM	0.4524	0.5811	0.4553	0.3451	124.9893	0.5978
PhyDNet	0.3845	0.3643	0.4001	0.2604	144.7144	0.4915
SA-ConvLSTM	0.4482	0.5672	0.4304	0.3709	115.3312	0.5924
MRP-Net	0.4648	0.5945	0.4157	0.3832	107.0139	0.6220
STDA-Net	**0.4819**	**0.6088**	**0.3913**	**0.4127**	**100.0012**	**0.6396**

Note: Explanation: If the arrow points upwards, the higher the value, the better. If the arrow points downwards, the lower the value, the better. The bolded content indicates the best effect.

We compare the MSE and Structural Similarity Index (SSIM) trends of eight models for radar echo extrapolation across different prediction lead times in Fig. 7. As shown in the MSE curve (a), all models exhibit a clear increase in MSE as lead time rises from 2 to 10, while STDA-Net maintains the lowest error throughout. In long-term prediction (Lead Time > 5), its MSE grows much more slowly than other models. According to the SSIM curve (b), structural similarity declines for all models with increasing lead time, yet STDA-Net achieves the highest SSIM values and shows the mildest drop in long-term forecasting (Lead Time = 8–10).

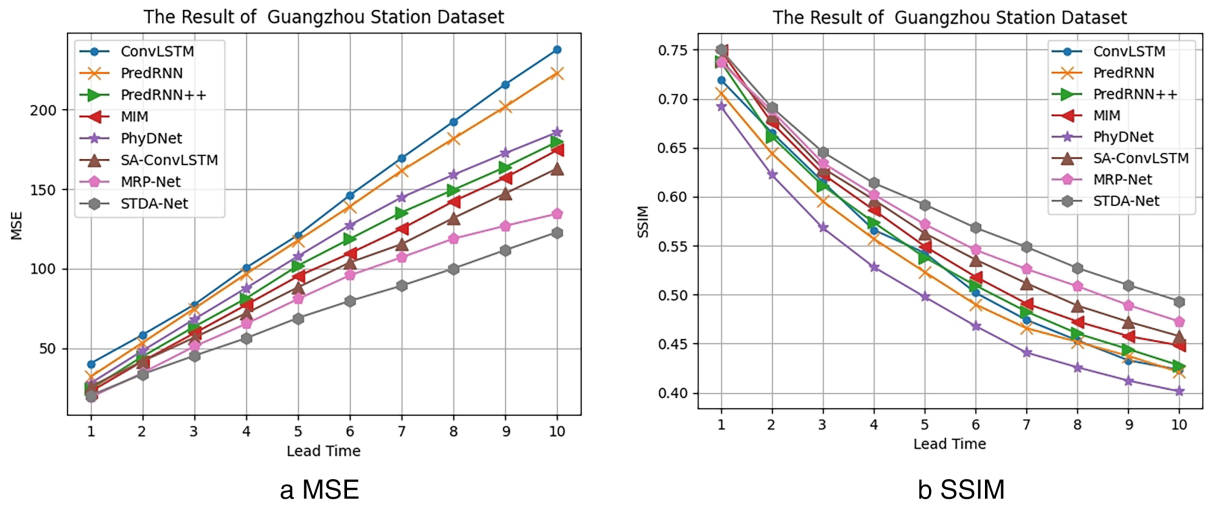


Figure 7: Comparison of the scores at each time step based on the radar echo data set of Guangzhou Station under the MSE and SSIM evaluation indicators.

In Fig. 8, we show the visualization results of the model on a typical sequence of radar echo data from Guangzhou Station, along with comparisons to the ground truth. The color scale indicates radar echo intensities in 5 dBZ intervals. The first row displays the real radar echo images, while the region below the dashed line shows predictions from eight comparison models and STDA-Net. PredRNN, MRP-Net, and SA-ConvLSTM produce increasingly distorted predictions at longer time steps, and ConvLSTM, MIM, PhyDNet, and PredRNN++ suffer from severe shape distortion in long-term extrapolation. In contrast, the proposed STDA-Net maintains excellent echo structure integrity and intensity distribution across the entire time series.

In Table 2, on the CIKM 2017 radar echo dataset, STDA-Net performed exceptionally well in indicators.

As shown in Fig. 9, the growth rate of MSE of STDA-Net was significantly lower than that of other models, while SSIM remained at the highest level as the prediction step size increased.

In Fig. 10, we present the visualized results of a representative sequence of the above model in the 2017 CIKM radar echo dataset. For short-term predictions (Lead Time 2–4), PredRNN, MRP-Net, and SA-ConvLSTM will exhibit edge blurring in high-intensity precipitation areas. ConvLSTM and MIM will also suffer from detail loss in the early prediction stage. As the prediction time interval increases, the prediction quality of all models significantly declines.

4.3 The Results and Analysis of the Ablation Experiment

In Table 3, we conduct ablation studies by comparing the proposed model with stacked-structured models including PredRNN and MRP-Net.

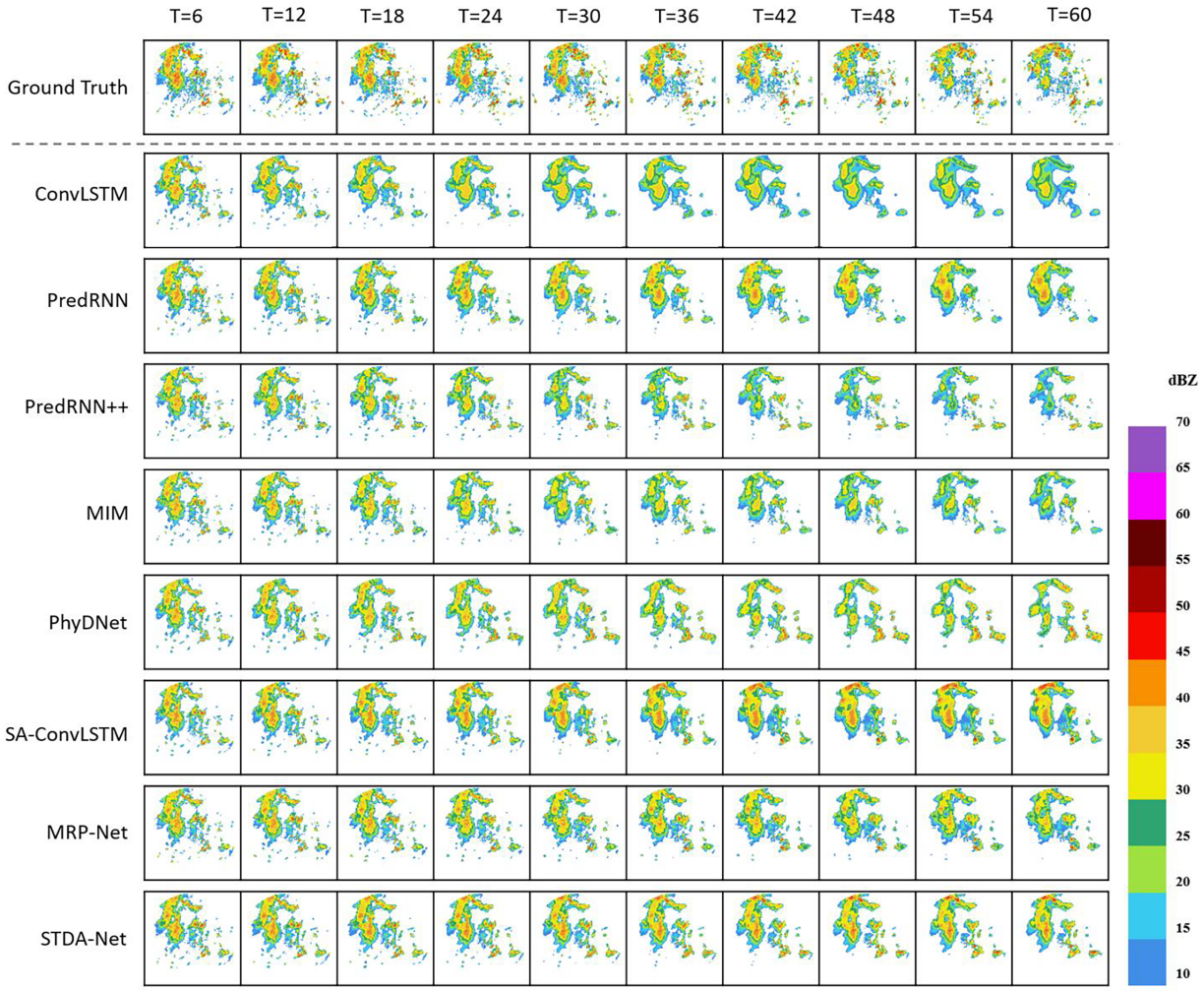


Figure 8: Based on the visualization results of a typical sequence from the radar echo data set of Guangzhou Station.

Table 2: Quantitative results comparison of each model in the radar echo dataset of CIKM 2017.

Model	CSI-AVG ↑	HSS-AVG ↑	FAR-AVG ↓	POD-AVG ↑	MSE ↓	SSIM ↑
ConvLSTM	0.2247	0.2879	0.3914	0.6124	146.0147	0.5013
PredRNN	0.2831	0.3542	0.3647	0.5915	140.2514	0.5142
PredRNN++	0.3006	0.3691	0.4215	0.4353	104.6356	0.5531
MIM	0.3375	0.4122	0.4351	0.5737	100.4924	0.5993
PhyDNet	0.2337	0.2853	0.3235	0.6686	137.6692	0.4979
SA-ConvLSTM	0.3341	0.4277	0.4972	0.5649	115.3728	0.5925
MRP-Net	0.3513	0.4410	0.3803	0.6492	94.7241	0.6094
STDA-Net	**0.3654**	**0.4487**	**0.3054**	**0.7681**	**90.3535**	**0.6217**

Note: Explanation: If the arrow points upwards, the higher the value, the better. If the arrow points downwards, the lower the value, the better. The bolded content indicates the best effect.

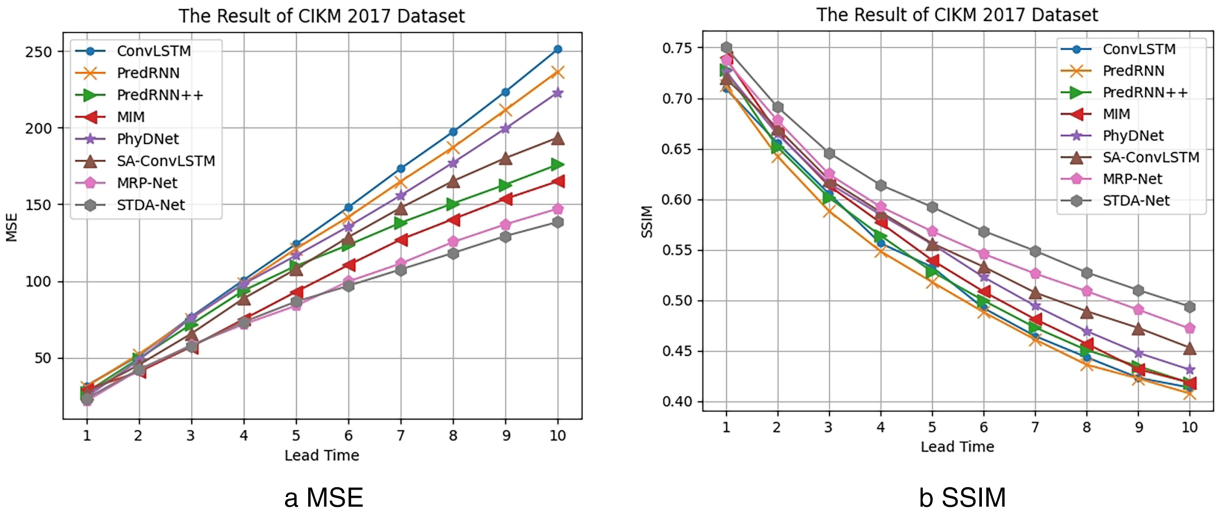


Figure 9: Based on the radar echo dataset of CIKM 2017 in terms of MSE and SSIM evaluation metrics.

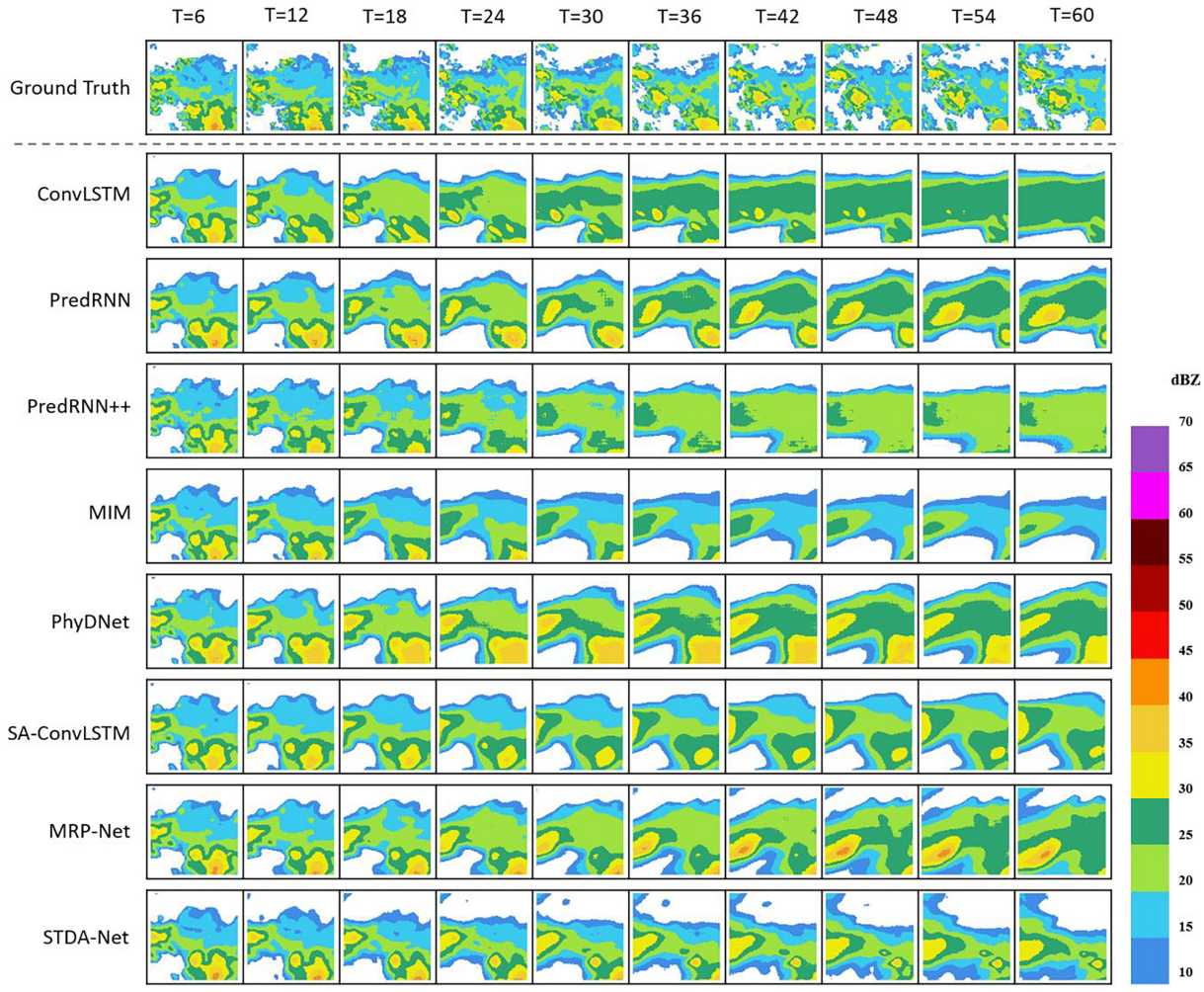


Figure 10: Visual comparison of the results for a typical sequence based on the radar echo dataset from CIKM 2017.

Table 3: Ablation results comparison of the STDA-Net model on the radar echo dataset of the Guangzhou station.

Model	CSI-AVG ↑	HSS-AVG ↑	FAR-AVG ↓	POD-AVG ↑	MSE ↓	SSIM ↑
PredRNN	0.3872	0.3497	0.4821	0.2556	166.4815	0.4588
MRP-Net	0.4768	0.5983	0.4047	0.3849	106.7541	0.6119
w/o.SSAM	0.4732	0.5849	0.3985	0.4090	100.4852	0.5971
w/o.TAM	0.4701	0.5814	0.3959	0.4144	99.8974	0.5842
w/o.AER	0.4821	0.5908	0.4009	0.4179	99.2316	0.6154
STDA-Net	**0.4906**	**0.6263**	**0.3877**	**0.4240**	**97.4852**	**0.6389**

Note: Explanation: If the arrow points upwards, the higher the value, the better. If the arrow points downwards, the lower the value, the better. The bolded content indicates the best effect.

5 Conclusion

STDA-Net is a long-term weather radar echo extrapolation network constructed based on the spatio-temporal dual-stream attention mechanism, which improves the accuracy of modeling and predicting the long-term movement trend of radar echoes. This study, by combining the required indicators, demonstrates that STDA-Net outperforms existing methods in long-term trend modeling and prediction. It can more effectively capture long-term movement characteristics and achieve higher forecast accuracy. The STDA-Net is mainly trained and evaluated based on the local radar observation dataset. The model's transferability in different regions and different observation devices has not been fully verified.

Acknowledgement: None.

Funding Statement: This work has received funding from National Natural Science Foundation of China (No. U23B20141).

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Yuelei Niu; methodology, Yuelei Niu and Chuang Liu; software, Yuelei Niu and Chuang Liu; validation, Yuelei Niu; formal analysis, Yuelei Niu and Chuang Liu; investigation, Yuelei Niu and Chuang Liu; resources, Qi Liu; data curation, Yuelei Niu; writing—original draft preparation, Yuelei Niu; writing—review and editing, Yuelei Niu; visualization, Xiaodong Liu and Yuelei Niu; supervision, Yuelei Niu; project administration, Yuelei Niu; funding acquisition, Yuelei Niu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The CIKM 2017 radar echo dataset is publicly available. The radar echo dataset of the Guangzhou station is internal meteorological observation data and cannot be publicly released.

Ethics Approval: Not applicable.

Conflicts of Interest: Given his role as Editorial Board Member of this journal, Xiaodong Liu had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. The authors declare no other conflicts of interest.

References

1. Shen X, Meng K, Zhang L, Zuo X. A method of radar echo extrapolation based on dilated convolution and attention convolution. Sci Rep. 2022;12(1):10572. doi:10.1038/s41598-022-13969-6.

2. Prein AF, Mearns LO. US extreme precipitation weather types increased in frequency during the 20th century. J Geophys Res Atmos. 2021;126(7):e2020JD034287. doi:10.1029/2020jd034287.
3. Liu Q, Sun J, Zhang Y, Liu X. DenMerD: a feature enhanced approach to radar beam blockage correction with edge-cloud computing. J Cloud Comput. 2024;13(1):32.
4. Wu D, Wu L, Zhang T, Zhang W, Huang J, Wang X. Short-term rainfall prediction based on radar echo using an improved self-attention PredRNN deep learning model. Atmosphere. 2022;13(12):1963. doi:10.3390/atmos13121963.
5. Wang C, Wang P, Wang P, Xue B, Wang D. Using conditional generative adversarial 3-D convolutional neural network for precise radar extrapolation. IEEE J Sel Top Appl Earth Obs Remote Sens. 2021;14:5735–49. doi:10.1109/jstars.2021.3083647.
6. Sun M, Wang W, Zhu X, Liu J. Moso: decomposing motion, scene and object for video prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023 Jun 17–24; Vancouver, BC, Canada. p. 18727–37.
7. Shen X, Li X, Elhoseiny M. Mostgan-v: video generation with temporal motion styles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023 Jun 17–24; Vancouver, BC, Canada. p. 5652–61.
8. Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst. 2015;28:1–9.
9. Wang Y, Long M, Wang J, Gao Z, Yu PS. Predrnn: recurrent neural networks for predictive learning using spatiotemporal LSTMs. Adv Neural Inf Process Syst. 2017;30:1–10.
10. Wang Y, Gao Z, Long M, Wang J, Yu PS. Predrnn++: towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. Int Conf Mach Learn. 2018;80:5123–32. doi:10.48550/arxiv.1804.06300.
11. 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: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15–20; Long Beach, CA, USA. p. 9154–62.
12. 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. doi:10.3390/rs13020164.
13. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference; 2015 Oct 5–9; Munich, Germany. p. 234–41.
14. Trebing K, Stanczyk T, Mehrkanoon S. Smaat-unet: precipitation nowcasting using a small attention-unet architecture. Pattern Recognit Lett. 2021;145:178–86. doi:10.1016/j.patrec.2021.01.036.
15. Zeng Q, Li H, Zhang T, He J, Zhang F, Wang H. Prediction of radar echo space-time sequence based on improving TrajGRU deep-learning model. Remote Sens. 2022;14(19):5042. doi:10.3390/rs14195042.
16. Guo S, Sun N, Pei Y, Li Q. 3D-UNET-LSTM: a deep learning-based radar echo extrapolation model for convective nowcasting. Remote Sens. 2023;15(6):1529.
17. Jing J, Li Q, Ma L, Chen L, Ding L. REMNet: recurrent evolution memory-aware network for accurate long-term weather radar echo extrapolation. IEEE Trans Geosci Remote Sens. 2022;60:1–13.
18. Geng H, Wang T, Zhuang X, Xi D, Hu Z, Geng L. GAN-rcLSTM: a deep learning model for radar echo extrapolation. Atmosphere. 2022;13(5):684.
19. Ji R, Liu Q, Zhang Y, Liu X. A new improved method of recurrent memory perception for radar echo extrapolation. In: Proceedings of the 2024 IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C); 2024 Jul 1–5; Cambridge, UK. p. 779–85.
20. Shenzhen Meteorological Bureau, Alibaba Group. CIKM AnalytiCup 2017: short-term quantitative precipitation forecasting challenge. 2017 [cited 2026 Jan 1]. Available from: https://tianchi.aliyun.com/competition/entrance/231596/introduction.