Research Article

Radar Echo Maps Prediction Using an Improved MIM with Self-Attention Memory Module

Wei Zhou, Zhongwei Qu, Lianen Qu, Xupeng Wang, Yilin Shi, De Zhang, and Zhenlin Hui

1School of Information and Control Engineering, Qingdao University of Technology, Qingdao, Shandong, China
2International College, Qingdao University of Science and Technology, Qingdao, Shandong, China

Correspondence should be addressed to Zhongwei Qu; 961545038@qq.com

Received 20 January 2023; Revised 14 June 2023; Accepted 19 June 2023; Published 18 July 2023

Academic Editor: Nageswara Lalam

Copyright © 2023 Wei Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A radar echo sequence which includes N frames plays a crucial role in monitoring precipitation clouds and serves as the foundation for accurate precipitation forecasting. To predict future frames, spatiotemporal models are used to leverage historical radar echo sequences. The spatiotemporal information combining both temporal and spatial information is derived from radar echo sequence. The spatiotemporal information reveals the changing trend of intensity in the echo region over time. Dynamic variation information extracted in radar echo maps mainly consists of nonstationary information and spatiotemporal information. However, the changing trends at different locations within the precipitation cloud are different, so the significance of the spatiotemporal information should be different. The current precipitation forecasting model, Memory In Memory (MIM), has the capability to store the nonstationary information derived from radar echo maps. However, the MIM falls short in discerning the significance of the spatiotemporal information extracted from these maps. To address this limitation, we propose a novel model, SAMM-MIM (self-attention memory module-MIM), which regulates the generation of hidden states and spatiotemporal memory states using a SAMM. The proposed SAMM uses the self-attention mechanism and a series of gate mechanisms to concentrate on significant spatiotemporal information, learn changing trends in the echo region, and output predictive information. The predictive information which is stored in hidden states contains predictions of the changing trends of dynamic variation information. Experimental evaluation on a dataset of radar data from Qingdao, China, demonstrates that SAMM-MIM achieves superior prediction performance compared with other spatiotemporal sequence models, as indicated by improved scores on mean squared error, critical success index, and missing alarm rate metrics.

1. Introduction

Severe weather may cause damage to society [1, 2] and even cause natural disasters such as waterlogging and torrential floods, resulting in unnecessary casualties. If we can predict the future rainfall intensity in a local region, we can take timely actions to avoid the harm caused by heavy rainfall. Obtaining accurate weather prediction at least 1 hr in advance could help responsible agencies mobilize and may contribute to saving lives [3]. As global temperatures continue to rise, it becomes more difficult to accurately predict rainfall [4].

Radar maps produced by weather radar can directly reflect the distribution and intensity of rainfall. Deep learning has recently developed rapidly in various fields and made significant contributions [5]. In 2015, Shi et al. [6] proposed that precipitation nowcasting is a spatiotemporal sequence forecasting problem, with the sequence of past radar echo maps as input and the sequence of future radar echo maps as output.

Spatiotemporal information is derived from radar echo sequences that combine temporal and spatial information. The ConvLSTM (convolution long short-term memory) [6] based on LSTM uses convolution operations instead of linear operations to capture spatiotemporal information. The ConvLSTM network stacks ConvLSTM units to predict the sequence of radar maps. Input frames are input into the bottom layer, and the future frame is generated on the top layer.
The prediction information generated by the current layer is conveyed upward to the next layer as input information.

Spatiotemporal motion processes such as precipitation clouds and traffic flow [7] have complex nonstationary behavior. In 2019, Wang et al. [8] proposed MIM (Memory In Memory) to remember higher order nonstationarity in spatiotemporal processes using the concept of difference. It is not efficient to capture long-range spatial dependencies through convolution. In 2020, the SA-ConvLSTM with the self-attention module (SA) was proposed to solve the long-range spatiotemporal dependencies in the ConvLSTM [9].

The variation in precipitation clouds shown on the radar map tends to differ in different areas. For example, some areas change rapidly while others tend to be stable. Therefore, the spatiotemporal model should pay different attention to the spatiotemporal information extracted from the radar maps so that it can focus on important spatiotemporal information. For tasks involving the prediction of radar maps, MIM can effectively remember stationary and nonstationary information, but it does not focus on important spatiotemporal information. As shown in Figure 1(a)–1(c) are the 10th, 15th, and 20th frames of the truth picture, respectively, while Figure 1(d)–1(f) correspond to the predicted picture’s 10th, 15th, and 20th frames. As the timestep increases, the accuracy of the prediction decreases gradually.

Hidden states and spatiotemporal memory states are feature maps with a width of W, a height of H, and a channel of C. In the MIM, the hidden state contains the dynamic variation information and predictive information. The spatiotemporal memory state contains spatiotemporal information learned by the MIM. The hidden state conveys the dynamic variation information and the predictive information. The spatiotemporal memory state also conveys spatiotemporal information to next layer. In a MIM network, each layer of SAMM-MIM learns spatiotemporal information and dynamic variation information by the hidden state and spatiotemporal memory state conveyed from the previous layer. The MIM uses convolutional operations to learn spatiotemporal information from the spatiotemporal memory state and hidden state and uses a series of gates to memorize spatiotemporal information. However, convolutional operations can extract information and do not assist the model in further discerning the importance of the information. The model may forget some important spatiotemporal information.

Thus, when the MIM network predicts radar echo maps, two issues emerge: (1) each layer of MIM fails to distinguish the significance of the information in the hidden states and...
spatiotemporal memory state generated by the previous layer of MIM, (2) the quantity of crucial spatiotemporal information in the new hidden state and spatiotemporal memory state generated by the MIM diminishes.

When a MIM network predicts a sequence of radar maps, the redundant spatiotemporal information contained in the spatiotemporal memory state and the hidden state may not be conducive to being learned and memorized by the model, leading to difficulty in achieving accurate predictions.

Inspired by the SA in the SA-ConvLSTM, which utilizes the self-attention mechanism to capture and the gate mechanism to memorize the global information, we propose the SAMM for MIM or SAMM in short. The SAMM is embedded into MIM to construct the SAMM-MIM. The main contribution of this paper can be summarized as follows:

(1) We propose the SAMM, which helps the SAMM-MIM memorize important information in the spatiotemporal memory state and the hidden state and generates new spatiotemporal memory and a new hidden state. The proposed SAMM helps alleviate the problem of gradient vanishing in the SAMM-MIM, allowing the SAMM-MIM to achieve better prediction results.

(2) We propose a variant of MIM called SAMM-MIM. The SAMM-MIM network can perform precipitation nowcasting by predicting radar maps and has better results than other spatiotemporal models in predicting the sequence of radar maps.

(3) The proposed SAMM-MIM network can also be used for other multi-frame spatiotemporal prediction tasks, including a real-world dataset called TaxiBJ and a widely used dataset called Moving MNIST.

The Section 2 in this paper introduces the related method and the architecture of the model. The Section 3 introduces the methods proposed and the architecture of SAMM-MIM in this paper. The Section 4 introduces the experimental results and the discussion of the results.

2. Related Work

LSTM [10–12] alleviates the gradient vanishing problem that may occur during the training process of traditional recurrent neural networks. LSTM is widely used in speech recognition [13], machine translation [14], and video captioning [15]. It can remember one-dimensional time series through various gates, which are generated after the input information is activated. The gate mechanism uses these gates to control the amount of information to be remembered. The activation function in LSTM mainly consists of the sigmoid function and the tanh function.

When LSTM is used to predict radar echo maps, the maps must be processed into one-dimensional sequences, which can result in the loss of some spatial information. This is because LSTM is designed to process one-dimensional time series data and is not well-suited to handling two-dimensional spatial data.

In 2015, Shi et al. [6] proposed ConvLSTM by replacing the fully connected operation in LSTM with a convolution operation. The ConvLSTM network composed of stacked ConvLSTM units can achieve better prediction results than traditional methods on radar map task.

Each layer of ConvLSTM only relies on the hidden state to convey information, so that some spatiotemporal information may be lost. In 2017, to address this issue, Wang et al. [16] proposed ST-LSTM based on ConvLSTM. The ST-LSTM incorporates a ConvLSTM-like unit that is tasked with encoding spatiotemporal information and producing a spatiotemporal memory state to retain the information. The ST-LSTM conveys information upward through the hidden and the spatiotemporal memory state.

Radar map sequences contain stationary and nonstationary information. Nonstationary information, such as the accumulation and dissipation of clouds in radar maps, is usually difficult to predict. In 2019, Wang et al. [8] proposed using the difference to remember the nonstationarity in radar echo maps and constructed the MIM. The MIM memorizes stationary and nonstationary information by changing the forget gate into two cascading units. The two units are similar to the ConvLSTM. These two units are called MIM-S and MIM-N. The MIM-N can memorize nonstationary information by subtracting the hidden state generated at the previous time from that generated at the current time. The MIM-S is responsible for integrating nonstationary information and stationary information.

The emergence of a self-attention mechanism further alleviates the problem of long-term dependencies. Compared with traditional RNN models, the self-attention mechanism has fewer parameters. The self-attention mechanism is good at capturing the correlation between features. The self-attention mechanism calculates the correlation degree between a feature and other features by weighting and summing them. Therefore, the self-attention mechanism is good at capturing long-term dependencies [17, 18]. As for CNN [19–21], many attention mechanism modules [22–26] can capture the correlation between feature maps.

To address the issue of inefficiency in capturing long-range spatial dependencies through convolution, Lin et al. [9] proposed SA-ConvLSTM in 2020. The SA-ConvLSTM updates the spatiotemporal memory state through the SA. The hidden state $H$ and the spatiotemporal memory state $M$ are mapped to different feature spaces by convolution in the SA. The self-attention mechanism uses a weighted summation operation on the mapped features to obtain the information that can help the model remember long-range spatial dependencies. By memorizing the information, the model alleviates the problem of long-term dependence.

As shown in Figure 2, the self-attention scores $Z$ of the hidden state $H$ and the spatiotemporal memory state $M$ are updated by the SA.

The state changes are shown in Equation (1). $*$ represents convolution operation and $\odot$ represents the Hadamard product operation as:

$$
\bar{f}_t = \sigma (W_{m2} * Z + W_{m1} * H_t + b_{m2}),
$$

(1)
\[ g'_t = \tanh \left(W_{mzg} \ast Z + W_{mhg} \ast H_t + b_{mg} \right), \]  
\[ M_t = (1 - i'_t) \odot M_{t-1} + i'_t \odot g'_t, \]  
\[ o'_t = \sigma \left(W_{mzo} \ast Z + W_{mho} \ast H_t + b_{mo} \right), \]  
\[ \tilde{H}_t = o'_t \odot M_t. \]

3. Method

For the radar maps prediction task, the SAMM-MIM memorizes nonstationary information and focuses on important spatiotemporal information. Some spatiotemporal information that is not conducive to predicting radar maps should be ignored, while others should be paid attention to by the model. In order to improve prediction accuracy, this paper introduces a SAMM-MIM detailed in Section 3.2. The SAMM, detailed in Section 3.3, helps the SAMM-MIM focus on important information and selectively memorize information that has a strong correlation with predicting radar maps.

3.1. The SAMM-MIM Network. A SAMM-MIM network consists of three SAMM-MIM blocks and one ST-LSTM block. Temporal differencing in the SAMM-MIM is achieved by subtracting the hidden state \( H_{t-1} \) from the hidden state \( H_t \).

The \( H_{t-1} \) is generated by the previous layer of the SAMM-MIM at the previous timestamp. The \( H_{t-1} \) is generated by the previous layer of the SAMM-MIM at this timestamp. The superscript in each state denotes the layer responsible for generating that state in the network, while the subscript indicates the corresponding time step at which the state is generated. The SAMM allows the hidden state and spatiotemporal memory state to more effectively convey upward spatiotemporal information. By stacking SAMM-MIM blocks, the SAMM-MIM network has the ability to learn higher orders of spatiotemporal information. The SAMM-MIM network can remember the changing trend of precipitation clouds in the radar maps and produce more accurate prediction results. The network generates one prediction frame at one timestamp. As shown in Figure 3, the input can be a truth frame or the generated frame at the previous timestamp, and the output is generated at each timestamp. The blue arrow represents the transmission of the long-term state and the hidden state. The black arrow represents the transmission of the spatiotemporal memory state and the hidden state. The red arrow represents the transmission of differential information.
3.2. The Architecture of SAMM-MIM. The SAMM-MIM accurately predicts variations in radar echo intensity by learning both nonstationary and crucial spatiotemporal information. In the SAMM-MIM, $C_f^t$ contains nonstationary and stationary information. The SAMM-MIM uses the $C_f^t$ to memorize nonstationary information twice. The SAMM-MIM uses $H_{l-1}^t$, $H_{l-1}^t$, MIM-N, and MIM-S to facilitate the storage of nonstationary information in $C_f^{l-1}$ at the current time step, thereby generating $C_f^t$. The specific formula is shown in (6)–(10) as:

$$g_t = \tanh \left( W_{ix} \ast H_{l-1}^t + W_{ih} \ast H_{l-1}^t + b_i \right), \quad (6)$$

$$i_t = \sigma \left( W_{ix} \ast H_{l-1}^t + W_{ih} \ast H_{l-1}^t + b_i \right), \quad (7)$$

$$D_l^t = \text{MIM} - N \left( H_{l-1}^t, H_{l-1}^t, N_{t-1}^t \right), \quad (8)$$

$$T_l^t = \text{MIM} - S \left( D_l^t, C_{l-1}^t, S_{t-1}^t \right), \quad (9)$$

$$C_f^t = T_l^t + i_t \odot g_t, \quad (10)$$

The SAMM-MIM first memorizes the nonstationary information in $C_f^t$. The SAMM-MIM uses the output gate $o_t$ to memorize nonstationary information in $C_f^t$, thereby generating the $H_l^t$. The specific formula is shown in (11) and (12) as:

$$o_t = \sigma \left( W_{xo} \ast H_{l-1}^t + W_{mo} \ast M_{l-1}^t + W_{ho} \ast H_{l-1}^t + b_o \right), \quad (11)$$

$$H_l^t = o_t \odot \tanh \left( C_f^t \right), \quad (12)$$

The $H_l^t$ contains the dynamically variation information learned by the model. Then, the model uses the self-attention mechanism in the SAMM to differentiate the significance of spatiotemporal information within both the $H_l^t$ and the $M_{l-1}^t$. It selectively memorizes spatiotemporal information through a series of gate mechanisms, generating a new spatiotemporal memory state $M_l^t$. The $M_l^t$ contains a substantial amount of important spatiotemporal information. Moreover, the model uses a series of gate mechanisms to simultaneously learn the spatiotemporal information and nonstationary information, resulting in the generation of the new hidden state. The specific formula is shown in (13) as:

$$\tilde{H}_l^t, M_l^t = \text{SAMM} \left( H_l^t, M_{l-1}^t, C_f^t \right). \quad (13)$$

Compared with the predicted picture generated by the MIM, the predicted picture generated by SAMM-MIM has better accuracy of prediction. As shown in Figure 4, the architecture of SAMM-MIM is depicted.

3.3. Self-Attention Memory Module. The SAMM receives three input states: the hidden state $H_l^t$, the long-term memory state $C_f^t$, and the spatiotemporal memory state $M_{l-1}^t$. The SAMM uses the self-attention mechanism and a series of gate mechanisms to memorize important spatiotemporal information in the $M_{l-1}^t$. Through a set of gate mechanisms, it learns both nonstationary and spatiotemporal information, resulting in the generation of a new hidden state $\tilde{H}_l^t$. The $\tilde{H}_l^t$ contains dynamic variation information and predictive information and is the output of the SAMM-MIM. The structure of the SAMM is shown in Figure 5.

The $H_l^t$ assists the $M_{l-1}^t$ in memorizing significant spatiotemporal information through the mechanism of self-attention. As for the dynamic changing information contained in the $H_l^t$, it initially searches for significant spatiotemporal information within itself.

The $H_l^t \in \mathbb{R}^{W \times H \times C}$ is mapped into different feature spaces as the query: $h_q = W_{hq} \ast H_l^t \in \mathbb{R}^C \times N$, the key: $h_k = W_{hk} \ast H_l^t \in \mathbb{R}^{C \times N}$, and the value: $h_v = W_{hv} \ast H_l^t \in \mathbb{R}^{C \times N}$. $N$ is obtained by multiplying $W$ and $H$. The $\{C, \tilde{C}, W_{hq}, W_{hk}, W_{hv}\}$ is a convolution of size of $1 \times 1$. The $\{C, \tilde{C}\}$ is the number of channels. As shown in (14), the similarity scores of each pair of feature maps are obtained by matrix multiplication. The value of the similarity score represents the degree of importance as:
The shape of feature map can be expressed as shown in (15), where the dimension transformations are shown in Figure 6.

\[
S_h = \mathbb{R}^{N \times N} = (W_{bh} \ast H^l_i) (W_{hk} \ast H^l_i) = h^T_h h_k.
\]

The similarity score of the \(i\)-th feature map to the \(j\)-th feature map can be expressed as shown in (15), where the shape of \(\{H^l_i, H^l_j\}\) is \(N \times 1\). The \(S_{h_{i,j}}\) represents the importance of \(j\)-th to \(i\)-th as:

\[
S_{h_{i,j}} = \left( H^l_{i,j} \ast W_{bh} \right) \left( W_{hk} \ast H^l_{i,j} \right) = h^T_{ij} h_{kj}.
\]

As shown in (16), the similarity scores are normalized as:

\[
\alpha_{h_{i,j}} = \frac{e^{S_{h_{i,j}}}}{\sum_{n=1}^{N} e^{S_{h_{i,n}}}} \quad i, j \in \{1, 2, \ldots, N\}. \tag{16}
\]

As shown in (17), the self-attention scores of the \(i\)-th feature map to other feature maps are obtained by a weighted sum as:

\[
h_{sat} = \sum_{j=1}^{N} \alpha_{h_{i,j}} \left( W_{hv} \ast H^l_{i,j} \right) = \sum_{j=1}^{N} a_{h_{i,j}} h_{v_{ij}}. \tag{17}
\]

As shown in (18), the self-attention score of the hidden state to itself is obtained as:

\[
h_{sat} = [h_{sat,1}; h_{sat,2}; h_{sat,3}; \ldots; h_{sat,N}] \in \mathbb{R}^{C \times N}. \tag{18}
\]

The \(S_{A_h} \in \mathbb{R}^{C \times W \times H}\) is generated by reshaping the \(h_{sat}\). All dimension transformations are shown in Figure 6.

Likewise, the \(H^l_t\) uses the self-attention mechanism to search for crucial spatiotemporal information within the \(M^{l-1}_t\). The \(M^{l-1}_t \in \mathbb{R}^{W \times H \times C}\) is also mapped as the Key: \(m_k = W_{mk} \ast M^{l-1}_t \in \mathbb{R}^{C \times N}\), and the value: \(m_v = W_{mv} \ast M^{l-1}_t \in \mathbb{R}^{C \times N}\). As shown in (19), the similarity scores between the \(H^l_t\) and the \(M^{l-1}_t\) are calculated by matrix multiplication as:

\[
S_m = (W_{bh} \ast H^l_t) (W_{hk} \ast M^{l-1}_t) = h^T_t m_k. \tag{19}
\]

As shown in (20), the self-attention scores of the \(i\)-th feature map in the hidden state to all feature maps in spatiotemporal memory state obtained by a weighted sum as:

\[
m_{sat} = \sum_{j=1}^{N} \frac{e^{S_{m_{i,j}}}}{\sum_{j=1}^{N} e^{S_{m_{i,j}}}} \left( W_{mv} \ast M^{l-1}_{t_{i,j}} \right) = \sum_{j=1}^{N} \alpha_{m_{i,j}} m_{v_{ij}}. \tag{20}
\]

The \(m_m\) is generated by concatenating all \(m_{sat}\) as:

\[
m_m = [m_{sat,1}; m_{sat,2}; m_{sat,3}; \ldots; m_{sat,N}] \in \mathbb{R}^{C \times N}. \tag{21}
\]

The \(S_{A_m} \in \mathbb{R}^{C \times W \times H}\) can be obtained by reshaping the \(m_m\). The \(H^l_t \in \mathbb{R}^{C \times W \times H}\) is generated by \(W_{sat} \ast [S_{A_h}; S_{A_m}; H^l_t]\). Dimension changes of the \(H^l_t\) are shown in Figure 7.

The \(H^l_{t_{a}}\) contains a significant amount of critical spatiotemporal information. Through a series of gate mechanisms, the SAMM helps the SAMM-MIM memorize important information in \(H^l_{t_{a}}\) and \(M^{l-1}_t\), and then generates a new state \(M^{l}_t\). The specific formula is shown in (22)–(25) as:

![Figure 5: The architecture of self-attention memory module.](image-url)
The SAMM generates the $\tilde{H}_l^t$ by learning important spatiotemporal information and nonstationary information. The hidden state $\tilde{H}_l^t$ is the output of the SAMM-MIM. The specific formula is shown in (26) and (27) as:

$$g'_t = \tanh \left( W'_{h_g} \ast H_{sa}^l + W'_{m_q} \ast M_{i}^{t-1} + b'_g \right),$$  

$$i'_t = \sigma \left( W'_{h_i} \ast H_{sa}^l + W'_{m_i} \ast M_{i}^{t-1} + b'_i \right),$$  

$$f'_t = \sigma \left( W'_{h_f} \ast H_{sa}^l + W'_{m_f} \ast M_{i}^{t-1} + b'_f \right),$$  

$$M_{i}^{t} = f'_t \odot M_{i}^{t-1} + i'_t \odot g'_t.$$  

(24)

4. Experiment

To ensure a fair comparison of predictive capabilities for radar echo maps, all models were configured with identical hyperparameter and optimizer settings. As FRNN [27] is a network architecture and not a module, the parameter settings for the FRNN rely on the authors’ default configuration in their code. The specific hyperparameter settings are shown in Table 1.

The training process involved minimizing the mean squared error (MSE) metric. This paper evaluates the validity of models on three datasets. Experiments on the radar echo maps of Qingdao, China, demonstrate that the SAMM-MIM can predict precipitation clouds.

Experiments on the TaxiBJ dataset and the Moving MNIST dataset demonstrate that the SAMM-MIM can predict other spatiotemporal prediction tasks.

4.1. Radar Echo Dataset. Radar echo maps reflect the rainfall intensity in an area. Predicting radar echo maps is a challenging task and the basis of precipitation prediction [16]. The radar map data from July 17, 2020, to September 1, 2020 is collected as experimental data. Blank pictures are removed to eliminate the interference of unnecessary data on the model’s prediction ability. Every radar map is resized to a $64 \times 64 \times 1$ grid of image, and each radar map is processed into a grayscale image to reduce training time. The training dataset contains 9,624 pictures and the test dataset contains...
2,020 pictures. This paper uses a 20-frame-wide window to slice consecutive images, resulting in each sequence consisting of 20 consecutive images, with the first 10 as inputs and the last 10 as outputs. When evaluating the model’s prediction ability, the first predicted picture is used as the input to iteratively predict nine frames. The experimental results of the models are obtained by several experiments.

In this paper, MSE and MAE are used as metrics to evaluate models. The results are shown in Table 2.

To better compare the predicted results of the models, this paper visualizes 11th and 17th frames of a sequence. The results are shown in Figure 8.

As shown in Figure 8, models can not accurately predict certain regions with the increase of frames, resulting in increasingly inaccurate prediction results.

For the radar echo maps, the direction of gradient descent is different for models when they learn in high-dimensional space. The MIM network may find a better solution to MAE along gradient descent in higher dimensional space, which makes the MIM network slightly better than the SAMM-MIM network on MAE.

To compare the prediction ability of the SAMM-MIM and the MIM, MSE and MAE difference are calculated by subtracting the map generated by MIM from the map generated by SAMM-MIM. The final result is shown in Figure 9.

In this paper, structure similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) are used to evaluate the image quality of the generated images. The range of SSIM is between $[-1, 1]$. The minimum value of PSNR is 0. The higher the value of SSIM and PSNR, the higher the prediction accuracy. As shown in Table 3, the SAMM-MPR performs better than others in terms of SSIM and PSNR.

The SSIM and the PSNR difference are calculated by subtracting the map generated by MIM from the map generated by SAMM-MIM. The final result is shown in Figure 10.
At the same time, the critical success index (CSI), equitable threat score (ETS), and the missing alarm rate (MAR) are introduced as evaluation indicators to evaluate the prediction ability of the models. CSI, ETS, and MAR are three commonly used meteorological indicators. CSI is defined as
\[
CSI = \frac{\text{hits}}{\text{hits} + \text{misses} + \text{false alarms}}.
\]
MAR is defined as
\[
MAR = \frac{\text{misses}}{\text{misses} + \text{hits}}.
\]
ETS is an improvement on CSI. The hits correspond to true positives, the misses correspond to false positives, and the false alarms correspond to false negatives. Radar echoes can be extracted by removing background information from the radar echo maps. Because the radar echo maps are grayscale images, the radar maps value is 0–255.

This paper chose respectively 10 and 60, respectively, as thresholds to calculate CSI. The final results are shown in Table 4.

This paper uses 10, 30, and 60 as the threshold to calculate MAR. The final results are shown in Table 5.
As shown in Table 6, by embedding the SAMM, the MIM network can better learn the important information contained in the radar echo maps, thereby reducing the MAR. Similarly, this paper uses 10, 30, and 60 as thresholds to calculate ETS.

The Predrnn network contains four ST-LSTM units and also relies on the hidden state and the spatiotemporal memory state to convey information upwards. To verify the validity of SAMM, SAMM is embedded in the ST-LSTM to change the hidden and spatiotemporal memory states. The final results are shown in Table 7.

As shown in Table 7, by embedding the SAMM, the MIM network can better learn the important information contained in the radar echo maps, thereby reducing the MAR. Similarly, this paper uses 10, 30, and 60 as thresholds to calculate ETS.

The Predrnn network contains four ST-LSTM units and also relies on the hidden state and the spatiotemporal memory state to convey information upwards. To verify the validity of SAMM, SAMM is embedded in the ST-LSTM to change the hidden and spatiotemporal memory states. The final results are shown in Table 7.

As shown in Table 7, the modified ST-LSTM performs better than the ST-LSTM. Experiments show that SAMM can help ST-LSTM get better results.

4.2. TaxiBJ Dataset. The TaxiBJ is a dataset about crowd flow collected in a natural environment. Each frame is a 32 × 32 × 2 grid of image. The two channels represent the traffic flow entering and leaving the same area simultaneously. Because every frame is the real traffic flow, there is a strong temporal correlation between the pictures of neighboring timesteps. The TaxiBJ dataset contains 19,788 training sequences and 1,344 test sequences. Each sequence includes eight consecutive frames, with four for input and four for prediction. As shown in Figure 11(a)–11(c) are the visualization results of the second frame, the fourth frame, and the eighth frame of a sequence. The white area in the graph represents traffic. The more obvious the white area is, the higher the traffic density.

For the TaxiBJ dataset, this paper uses RMSE as the metric. The result is shown in Table 8. Compared with other spatiotemporal models, the proposed SAMM-MIM network achieves good results.

### Table 4: Comparison of average scores on CSI of different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>CSI-10</th>
<th>CSI-60</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConvLSTM</td>
<td>0.7154</td>
<td>0.6809</td>
</tr>
<tr>
<td>ST-LSTM</td>
<td>0.7392</td>
<td>0.6864</td>
</tr>
<tr>
<td>Causal LSTM</td>
<td>0.7620</td>
<td>0.6910</td>
</tr>
<tr>
<td>MIM</td>
<td>0.7899</td>
<td>0.7107</td>
</tr>
<tr>
<td>SAMM-MIM</td>
<td>0.7912</td>
<td>0.7111</td>
</tr>
</tbody>
</table>

### Table 5: Comparison of average scores on MAR of different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAR-10</th>
<th>MAR-30</th>
<th>MAR-60</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConvLSTM</td>
<td>0.2095</td>
<td>0.2776</td>
<td>0.2281</td>
</tr>
<tr>
<td>ST-LSTM</td>
<td>0.1924</td>
<td>0.2617</td>
<td>0.2318</td>
</tr>
<tr>
<td>Causal LSTM</td>
<td>0.1887</td>
<td>0.292</td>
<td>0.2313</td>
</tr>
<tr>
<td>MIM</td>
<td>0.1448</td>
<td>0.2051</td>
<td>0.1740</td>
</tr>
<tr>
<td>SAMM-MIM</td>
<td>0.1394</td>
<td>0.1990</td>
<td>0.1574</td>
</tr>
</tbody>
</table>

### Table 6: Comparison of average scores on ETS of different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>ETS-10</th>
<th>ETS-30</th>
<th>ETS-60</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConvLSTM</td>
<td>0.3980</td>
<td>0.5630</td>
<td>0.5980</td>
</tr>
<tr>
<td>ST-LSTM</td>
<td>0.3798</td>
<td>0.5628</td>
<td>0.6080</td>
</tr>
<tr>
<td>Causal LSTM</td>
<td>0.4394</td>
<td>0.5727</td>
<td>0.6079</td>
</tr>
<tr>
<td>MIM</td>
<td>0.5001</td>
<td>0.6132</td>
<td>0.6408</td>
</tr>
<tr>
<td>SAMM-MIM</td>
<td>0.5120</td>
<td>0.6198</td>
<td>0.6425</td>
</tr>
</tbody>
</table>

### Table 7: Compare the standard ST-LSTM with the modified ST-LSTM.

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>MAE</th>
<th>SSIM</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-LSTM</td>
<td>43.09</td>
<td>221.41</td>
<td>0.7182</td>
<td>20.24</td>
</tr>
<tr>
<td>ST-LSTM (modified)</td>
<td>42.98</td>
<td>219.20</td>
<td>0.7204</td>
<td>20.26</td>
</tr>
</tbody>
</table>

4.3. Moving MNIST Dataset. Each sequence contains 20 consecutive frames, with 10 for input and 10 for prediction. Each frame is a 64 × 64 × 1 grid of image. Each frame contains two handwritten digits. Each sequence depicts two handwritten digits with constant velocity moving within the image and bouncing off the edges. The two handwritten digits may overlap as they move. We visualize a sequence of truth frames in Figure 12.

We use SSIM, MSE, and MAE to evaluate models. The results are shown in Table 9.
The Moving MNIST is a commonly used dataset for testing the learning and prediction abilities of spatiotemporal models. As shown in Table 9, the proposed SAMM-MIM performs better than the standard MIM and other models. The proposed SAMM-MIM network also has strong prediction ability on the Moving MNIST dataset.

5. Conclusion

The degree of change in precipitation clouds in radar echo maps varies, so using the same level of attention for all spatiotemporal information is not helpful in predicting radar echo maps. The MIM cannot selectively memorize the information conveyed by the hidden and the spatiotemporal memory state from the previous layer. This paper constructs a SAMM-MIM. The spatiotemporal memory state in the SAMM-MIM can focus on more important spatiotemporal information and selectively memorize important information. At the same time, the hidden state is adjusted by the SAMM, allowing it to memorize more spatiotemporal and nonstationary information. The SAMM-MIM unit conveys more important spatiotemporal and predictive information compared with the MIM unit. In comparison to other models, the proposed SAMM-MIM network has more accurate prediction results on the radar echo dataset. The proposed SAMM-MIM network also has strong learning and prediction capabilities for other datasets.

Data Availability

The radar echo map dataset generated during the current study is available from the corresponding author upon reasonable request.
Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Z.Q. conducted the research, and W.Z. and L.Q. provided guidance. Z.Q. and X.W. wrote the main manuscript text. W.Z. and L.Q. revised the manuscript. Y.S., Z.H., and D.Z. analyzed data. All authors reviewed the manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (NSFC), grant number 61502262.

References