GMA: An Improved Framework of Radar Extrapolation Based on Spatiotemporal Sequence Neural Network

Long Huang1, Shengchun Wang1, Kefei Jiang1, and Jingui Huang1

1Hunan Normal University, Changsha, China

Abstract Most previous current spatiotemporal sequence neural networks used for radar extrapolation have difficulties in learning long-term spatiotemporal memories. This is because the spatiotemporal sequential neural networks only use the information from the previous time step node to update the prediction state, and the networks tend to rely on the convolution layers to capture the spatiotemporal features, which are local and inefficient. In order to capture the long-term temporal characteristics and local abrupt spatial characteristics of radar echo sequence, we propose a new framework, Global Memory Attention (GMA), which has two contributions. The first is that we establish a global information flow between calculation units, extract the key historical memory from the global information flow, calculate the correlation between the historical memory and the current time frame prediction state, and determine how much historical memory participates in the prediction state update. It alleviates the current network's difficulty in learning the long-term spatial and temporal characteristics of radar echo sequences and the problem of short-term dependence, and reduces the interference of image noise in the process of radar extrapolation prediction. The second is that the GMA module is a flexible additional module that can be applied to most radar extrapolation algorithms based on timing networks. GMA has reached the highest level on radar extrapolation tasks, and we have provided ablation studies to verify the effectiveness of GMA.

Plain Language Summary Spatiotemporal sequence neural networks are widely used in radar echo extrapolation tasks. However, in the past, most spatiotemporal sequence networks have short-term dependence and long-term dilemma. In this study, we propose a new framework, which can be applied to any radar extrapolation algorithm based on spatiotemporal sequence network. It establishes the global information flow and updates the prediction state by extracting the key historical memory strongly associated with the current prediction state through the attention module. It improves the utilization rate of long-term information and reduces the interference of radar echo image noise, and effectively improves the problem that the extrapolation effect decreases with the increase of extrapolation time.

1. Introduction

The forecast of severe convective weather has always been an important research topic in meteorological work. The approach forecast is defined as the weather forecast within 0–2 hr, of which 0–1 hr is the short-term forecast and 0–2 hr is the long-term forecast. The core content of the approach forecast is to forecast the precipitation intensity in local areas in a relatively short time. Compared with other forecasts, near-range forecasting has higher requirements for timeliness and accuracy of forecasting. Near-range forecasting is a very important and challenging content in meteorological work, and it has always been a hot research topic in meteorology.

Weather radar is a remote sensing detection tool, which sends electromagnetic waves to the atmosphere and receives the reflected echo to get the radar echo reflectivity. There is a strong correlation between radar echo reflectivity and strong convective precipitation weather. Therefore, the corresponding future precipitation can be estimated by predicting the future echo of weather radar. This process of predicting the future echo of weather radar is the extrapolation of radar echo. The accuracy of radar echo extrapolation is the key factor to determine the accuracy of proximity prediction.

The traditional radar echo methods based on morphology mainly include single centroid method (Han et al., 2007), cross-correlation method (Rinehart & Garvey, 1978), and optical flow method (Gibson, 1950). The single centroid method determines the motion vector by calculating the continuous echo single centroid, and then predicts the future echo position. However, when the radar echoes are scattered or merge or split, the prediction effect will decline rapidly with the increase of extrapolation time (Li & Han, 2008). The cross-correlation method...
divides the echo into different tracking regions, calculates the optimal spatial optimization correlation coefficient of adjacent times, and then determines the fitting relationship between the areas and realizes the prediction. However, it is difficult to ensure the accuracy of extrapolation tracking for weather processes with relatively rapid echo changes and strong convection (Liang et al., 2010). Optical flow method was proposed by Gibson in 1950, it obtains the motion vector field by calculating the optical flow field of the continuous echo, and then realizes the extrapolation based on the motion vector. However, the convective weather system is a dynamic and complex system. It has the characteristics of nonlinear motion (such as rotation), moderate deformation (e.g., expansion and contraction), and rapid generation and extinction. The traditional radar echo extrapolation method only assumes the simple linear evolution of the echo, and does not make full use of the long-term historical radar echo data in the time domain, and cannot effectively predict the rotation change of the echo and the process of generation and death. It has low accuracy. Shortcomings such as short forecast time.

In recent years, with the great progress of deep learning technology in computer vision and natural language processing, researchers began to apply it to radar extrapolation problem processing. Agrawal et al. (2019) tried to deal with radar echo based on two-dimensional convolutional neural network. However, due to the limitation of its receptive field, two-dimensional convolution calculation can only extract the spatial features of a single image, and lacks the modeling of the temporal relationship of image sequences. Kim et al. (2021) started to try to apply the three-dimensional convolutional neural network to the extrapolation task. LSTM (Hochreiter & Schmidhuber, 1997) proposed the RNN structure and uses gated update memory states to complete predictions. Ranzato et al. (2014) used clustering algorithm to interpolate image sequences and predict future image sequences in the cyclic neural network model for the first time. However, convolutional neural network and cyclic neural network lack reasonable spatial modeling means, and it is difficult to capture the spatial relationship in radar echo extrapolation. Therefore, it is difficult to construct the feature space of spatiotemporal sequence. To solve this problem, Shi et al. (2017) proposed ConvlSTM, it used convolution calculation to replace the Hadamard product of the original cyclic neural network calculation unit, and stacked multiple calculation units to build a convolution long-term and short-term memory network (ConvlSTM) with encoder predictor structure. It integrates the convolution layer that effectively extracts spatial features into the cyclic neural network structure that simulates the state of time evolution. The network model can model time and space at the same time. Therefore, the spatiotemporal sequence network, which can model spatiotemporal characteristics at the same time, has become the research focus of researchers in the direction of radar extrapolation. At present, advanced radar extrapolation neural networks such as Wang et al. (2017, 2018, 2019) are based on this spatiotemporal network structure. Compared with traditional meteorological methods, the radar echo extrapolation method based on deep learning has advantages in the existing performance and development space. Among them, the radar echo extrapolation method based on spatiotemporal sequence network has become the main method of deep learning methods with mature, stable, and effective modeling methods.

The radar extrapolation algorithm based on spatiotemporal sequence network is limited by the network structure. They all have inherent limitations, there are short-term dependency problems and long-term learning dilemmas. The short-term dependency problem is because the prediction process of the spatiotemporal sequence network only accepts the information of the previous time frame, it is difficult to effectively utilize the long-term information in the input sequence. Therefore, the network is susceptible to the interference of anomalous noise in a certain frame (Kelly et al., 1960), and the interference information expands with the extrapolation time. During long-term radar echo extrapolation, the false prediction results gradually accumulate with the extension of the extrapolation time, and the prediction results gradually deviate from the actual situation. Also, there is a problem of gradient disappearance in the process of back-propagation of the spatial and temporal information in the network recursive layer. The gradient information transfer gradually weakens during long-distance time steps, which makes the network not ideally able to use the complete sequence information. When conducting long-term radar extrapolation, it is often difficult to obtain long-term information. For example, when predicting simple linear motion problems, accurate predictions can be done only by relying on short-term information. When predicting the ideal pendulum movement, the future motion trajectory and motion state can be predicted according to its periodicity. However, in reality, radar extrapolation cannot rely solely on short-term information or long-term memory due to the movement, generation or dissipation of radar echoes caused by the interaction of complex atmospheric physical processes.

The above is the current dilemma of radar extrapolation. To solve this problem, we observed the process of human recalling past memories. Then we found that in the process of human recall, global information guides
the whole process of human recall. When people recall an event in the past, especially a long time ago, people first vaguely describe the approximate content, and then recall the details of the event along time. But sometimes, people usually lose information about historical events because of memory confusion at a certain point in time. However, if you vaguely remember the approximate content, even if some small faulty memory interferes, it will not affect the continued recall. Inspired by this, we integrate a method guided by global information into a spatiotemporal sequence network-based radar extrapolation algorithm to propose a new Global Memory Attention (GMA) framework. It contains three parts: base calculation units, global information flow between base units, GMA calculation module.

GMA has the following characteristics:

It establishes a global flow of information between calculation units. Because each radar echo image contains more or less historical information related to the current prediction state, the global information flow contains the main content of the historical memory state.

It extracts the key historical memory from the global information flow through a series of attention mechanisms, calculates the correlation between the historical memory and the current prediction state, and determines how much historical memory participates in the prediction state update of the current time frame. It improves the utilization rate of strong correlation historical memory and reduces the interference of weak correlation historical memory.

It can be flexibly applied to any radar extrapolation algorithm based on timing network. Because a separate module is used between the recurrent units to extract global information, it does not interfere with the internal calculations of the recurrent unit.

We have proved that this unit can help the network learn long-term spatiotemporal memory and reduce image noise interference. It improves the long-term learning dilemma and short-term dependence dilemma of radar extrapolation, and proves the effectiveness of this module in practice.

2. Related Work

2.1. Neural Network

Neural networks are widely used in radar extrapolation tasks. Agrawal et al. (2019) applies convolution neural network to radar echo extrapolation, which is limited by two-dimensional convolution receptive field and lacks the establishment of feature space for the time relationship of radar echo. Srivastava et al. (2015) attempted to predict the radar echo sequence with the cyclic neural network. Because the conventional cyclic neural network adopts linear calculation, it is difficult to extract the spatial characteristics of the radar echo better. Shi et al. (2017) replaced Hadamard product with convolution operation, extracted the spatial characteristics of radar echo in convolution layer, and simulated the time evolution of radar echo in recursion layer. ConvLSTM is an important work in the development of radar extrapolation method. Since then, many RNN networks based on ConvLSTM have emerged. For example, Shi et al. (2017) simulated the movement of information area by introducing learnable convolution, and added up sampling and down sampling layers to the prediction structure to alleviate the problem of excessive computation of spatiotemporal sequence network. Lotter et al. (2016) proposed that the extrapolate algorithm PredNet, which extracts the temporal and spatial features of radar echoes for implicit modeling and characterization, and predicts future image sequences through feature learning. In the spatiotemporal sequence network structure, historical memory only propagates in the time step of the recursive layer, and spatial information only propagates in the convolution layer. Therefore, Wang et al. (2017) introduced new memory states and related gating to generate spatiotemporal long-term and short-term memory units (ST-LSTM). This memory state forms a Z-shaped memory flow in the network. It can effectively transfer the spatiotemporal semantic information of the high-level network to the low-level network at the next moment, so as to realize the exchange of spatiotemporal semantic information at different levels. Wang et al. (2018) proposed the concept of gradient highway and CausalLSTM. It makes the jump propagation of spatiotemporal information alleviate the problem of gradient disappearance in back-propagation. At the same time, the depth of convolution layer is increased, and the network extracts multilevel spatial information of radar echo through convolution layer. Wang et al. (2015) integrates three-dimensional convolution into ST-LSTM unit, adds convolution calculation to the time dimension of picture sequence, and strengthens the utilization of upstream semantic information by
downstream tasks. Wu et al. (2021) decomposes the motion of the physical world into instantaneous changes and cumulative motion trends, and designs a Motion-Unit that can unify homeopathic changes and motion trends. As mentioned above, the researcher network improves the radar extrapolation method by simulating the physical motion trajectory of radar echo, effectively establishing the feature space of spatiotemporal information and the propagation route of spatiotemporal memory. However, in essence, the network still only uses the information of the previous time frame node to update the current frame prediction state, and the long-term spatiotemporal memory gradually disappears in the recursive layer propagation process, and falls into the dilemma of being disturbed by the image noise of a certain frame in the whole downstream prediction task. Wu et al. (2021) used continuous historical memory information to participate in prediction status update. Although the current frame can obtain the complete historical state in this way, the participation of all historical information in updating will bring a lot of interference, because not all historical memories are related to the current frame. Inspired by this, we integrate all historical memory states to establish a global information flow. In contrast, we use the attention mechanism to extract historical key memory from the global information flow, reducing the interference of irrelevant information on the current frame.

2.2. Vision Transformer

Since Transformer (Vaswani et al., 2017) was effectively applied to the field of natural language processing, researchers try to transfer transformer to the field of radar extrapolation. For example, Lin et al. (2020) strengthened the global spatial representation in ConvLSTM by introducing spatial self-attention mechanism. Luo (2021) proposed the interactive double attention mechanism in order to make full use of the short-term context of radar echo. Zhong (2020) use attention to model long-term spatiotemporal features to obtain enhanced spatiotemporal representation of radar echoes. Inspired by this, in our work, we use the attention mechanism to calculate the correlation between the key historical memory extracted from the global information flow and the current prediction process, and determine how much historical memory updates the prediction state. It improves the utilization rate of strong correlation historical memory and reduces the interference of weak correlation historical memory.

3. Method

In this section, we will introduce the overall structure of GMA and the specific structures of GMA module in detail. First, it uses all historical information and is not disturbed by long-term redundant information. Second, extract long-term key memory without losing short-term memory. Finally, get the memory state strongly associated with the current prediction state.

3.1. GMA Framework

It is a conventional stacked structure (Figure 1a). Block is a basic calculation unit (such as ConvLSTM, ST-LSTM). Take ConvLSTM as an example. The update equations of ConvLSTM for the $i^{th}$ time step of the $l^{th}$ layer are as follows:

\[
\begin{align*}
    i_t &= \sigma \left( W_{xi} * X_{t-1} + W_{hi} * H_{t-1}^{l} + W_{ci} \odot C_{t-1}^{l} + b_i \right) \\
    f_t &= \sigma \left( W_{xf} * X_{t-1} + W_{hf} * H_{t-1}^{l} + W_{cf} \odot C_{t-1}^{l} + b_f \right) \\
    g_t &= \tanh \left( W_{xg} * X_{t-1} + W_{hg} * H_{t-1}^{l} + b_g \right) \\
    C_t^{l} &= f_t \odot C_{t-1}^{l} + i_t \odot g_t \\
    o_t &= \sigma \left( W_{xo} * X_{t-1} + W_{ho} * H_{t-1}^{l} + W_{co} \odot C_{t}^{l} + b_o \right) \\
    H_t^{l} &= o_t \odot \tanh \left( C_t^{l} \right)
\end{align*}
\]

(1)

\( \sigma \) is the sigmoid function, \( W \) denotes the convolution kernel, \( \ast \) is the convolution operation, \( \odot \) is the Hadamard product, \( X_{t-1}, H_{t-1}^{l}, C_{t-1}^{l} \) is input state, hidden state, and memory state, respectively, and \( i_t, f_t, g_t, o_t \) are input gate, input modulation gate, forget gate, and output gate, respectively, the subscript \( t \) correspond to the \( t^{th} \) time step.
For the spatiotemporal sequence network of the conventional stacked structure, several layers of computing units are stacked in each frame time step, and several frame time steps are established to simulate the time evolution. The radar image sequence is input to the first layer, and the multilevel spatial features of the radar echo are extracted through the convolution layer of different calculation units with the same time step. Finally, the top-level radar echo space-time memory is mapped back to the original pixel space to predict the next radar image, and the future radar image sequence is generated at the top level. The prediction process of each time step is calculated by convolution layer, and the spatial information is extracted and transferred upward across layers. Then simulate the forward transmission of time evolution and model the time information. This can realize the flow of space-time information. In this case, the spatiotemporal information flow is updated in the calculation unit with the time step. The calculation unit of the current time frame mainly depends on the spatial characteristics of the calculation unit of the previous time step and time memory. This makes each frame of radar extrapolation more inclined to learn short-term historical memory, and the long-term memory decays with the increase of time step.

We try to establish a global information flow to learn long-term spatiotemporal memory (Figure 1b). The orange dotted line represents the global memory flow. The orange arrow indicates the flow of hidden state and memory state.

For the spatiotemporal sequence network of the conventional stacked structure, several layers of computing units are stacked in each frame time step, and several frame time steps are established to simulate the time evolution. The radar image sequence is input to the first layer, and the multilevel spatial features of the radar echo are extracted through the convolution layer of different calculation units with the same time step. Finally, the top-level radar echo space-time memory is mapped back to the original pixel space to predict the next radar image, and the future radar image sequence is generated at the top level. The prediction process of each time step is calculated by convolution layer, and the spatial information is extracted and transferred upward across layers. Then simulate the forward transmission of time evolution and model the time information. This can realize the flow of space-time information. In this case, the spatiotemporal information flow is updated in the calculation unit with the time step. The calculation unit of the current time frame mainly depends on the spatial characteristics of the calculation unit of the previous time step and time memory. This makes each frame of radar extrapolation more inclined to learn short-term historical memory, and the long-term memory decays with the increase of time step.

We try to establish a global information flow to learn long-term spatiotemporal memory (Figure 1b). The orange dotted line represents the global memory flow. The orange arrow indicates the flow of hidden state and memory state.

\[
G_{t+1}^{l+1} = GMA(C_t^{l}, H_t^{l+1}) \\
H_t^{l}, C_t^{l} = \text{Block}(H_t^{l+1}, H_{t-1}^{l}, G_t^{l+1})
\]

Unlike the conventional stacked structure that only uses the hidden state of the previous time step, we use the memory state of all historical time steps to update the prediction state of the current frame at each time step. First, we extract the key historical memory from the global information flow, filter the unimportant historical memory, and reduce the interference to the prediction process. Then we calculate its correlation with the current time step prediction state, so that the strong correlation spatiotemporal memory update calculation unit. The global memory flow established by GMA supports the simultaneous flow of long-term spatiotemporal memory and short-term spatiotemporal memory. The propagation and update of memory state is no longer limited by step size, and the most relevant historical memory can be used in each frame prediction process.
3.2. GMA Module

We consider that the prediction process of the current time step benefits from the temporal and spatial memory of historical frames, we build a global information flow from which to obtain a list of historical memory states. But not every frame of historical memory contributes to the prediction of the current time step. Therefore, we use the GMA module to calculate the correlation between the list of historical memory states and the predicted state of the current frame, and filter the weak correlation information, so that the historical memory with strong correlation guides the update calculation unit prediction process.

The GMA module is mainly composed of two modules (Figure 2). First, Temporal-SE extracts key historical memory from the list of historical memory states. Since each radar picture contains more or less the radar echo trajectory and echo intensity change trend, the extracted information includes the main content of all historical memory states. Then, through the attention module, the degree of relevance to the predicted state at the current time step is calculated, making the strong associative memory state more involved in the updating of calculation units. The overall equations of the GMA module for the $A_{th}$ time step of the $A_{th}$ layer are as follows:

$$ S'_l = T - SE(C_{0,l}) $$

$$ G'_l = Attention(H_{l-1}, S'_l) $$

(3)

The Temporal – SE module consists of three layers (Figure 3). We consider that the prediction process of each frame does not need to use all the historical spatial information, and there is a lot of repetition and redundancy in the spatial memory between historical frames, and we pay more attention to the changes in the spatial position of radar echoes over time. The spatiotemporal sequence network extracts the spatial information of each dimension of the radar echo through the convolution layer and stores it in the memory state, we establish Channel – SELayer to extract the key spatial historical memory of the radar echo. First, each frame of historical memory state input Channel – SELayer to learn the importance of different channel features. It performs a squeeze operation on the memory state to learn the global characteristics of the channel level. Then perform an excitation operation on the global feature to get the weight of each channel multiplied by the historical memory to extract key spatial history memory to radar echoes. In the same way that the prediction process does not benefit from each historical frame, we use Step – SELayer to extract the key time historical memory of the radar echo sequence. Similarly, it performs squeeze and excitation on the state of historical memory. Capture the global characteristics in the time dimension and capture the historical memory state with important information. This makes the model filter irrelevant information and obtain the main content from all historical memory states. Finally, we use 3D-convolution to integrate all global historical memory state lists to obtain key historical memories. $S'_{l-1}$ represents key historical memory from at $t-1$ th time step of the $l$ th layer.

Taking into account the diversity and uniqueness of the prediction process at each time step, the prediction process at each frame time step is not necessarily applicable to the key historical memory. We use the attention...
mechanism to calculate the correlation between key historical memory and the predicted state at the current time step, the strongly correlated key historical memory guides the current frame time step prediction state update, avoiding the interference of the weakly correlated key historical memory on the prediction process (Figure 4). This helps the model to overcome the step size limit and learn the most relevant historical information.

Figure 3. The Temporal-SE module overall structure.

Figure 4. The attention module overall structure.
At the $i$th time step of the $l$th layer, we map the current prediction state to query vector $Q_c$, Temporal – SE module extracted key historical memory $S^l_l$ is mapped into key vector $K_c$ and value vector $V_c$, flattening and linear full connection processing by the Map method in Figure 4. Calculate the correlation of $K$ and $Q$ through the dot product, and normalized by softmax to get the correlation score $A_n$. It is regarded as the degree of correlation between the predicted state of the current time step and the key historical memory, and determines the degree of participation in the update of the predicted state of key historical memory. Multiply the correlation score by $V_c$ to get the key historical memory state $\tilde{S}^l_l$ that best matches the current prediction process.

4. Experiment

In order to test the extrapolation accuracy of this method, this paper uses Hunan Meteorological Bureau (2018–2020) radar echo image data set to verify. This data set is the radar echo reflectivity of Changsha (105.00°–117.09°E, 21.99°–33.01°N) collected by Doppler weather radar of Hunan meteorological station. Collect a 256 × 256 × 1 radar image every 6 min. Forty-one continuous radar echo images are taken as a set of data sequences. The first 21 frames are input images and the last 20 frames are extrapolated images. The data set has the same geographical location, image size, time cycle spanning 2 years, and has both day and night radar images. It ensures the diversity, particularity, and consistency of data. In many complex situations, semantic information guidance based on long-term memory is very useful for radar extrapolation.

4.1. Data Set and Setup

In order to prevent data leakage and generalization of the network model for special data samples, we unevenly segment the data set to ensure that the training set and the test set do not repeat, and the test set contains special sample sequences. We set to 16,378 training sequences and 2,447 test sequences. Among them, about 280 special test sequences are radar echo data of different time or different areas in the training set data, which are used to test the generalization ability of the model.

4.2. Main Result

We compare GMA with five benchmark networks: ConvLSTM, ConvGRU, PredRNN, PredRNN++, MotionRNN. Especially MotionRNN is the latest CVPR network dedicated to radar echo extrapolation missions in 2021. We use the general image quality evaluation index of deep learning: mean squared error (MSE) to estimate radar image absolute pixel error, peak signal-to-noise ratio (PSNR) to evaluate the degree of image distortion, and radar echo image sharpness (Sharp). As shown in Table 1, after using the GMA framework, the prediction performance of all baseline networks is improved. Among them, PredRNN++ obviously outperforms other models after using the GMA framework, and achieves the best results.

At the same time, we use several meteorological precipitations near prediction indicators to evaluate critical success index (CSI), false alarm rate (FAR), and probability of detection (POD). The above calculation indicators comprehensively consider hits, false alarms, and misses events. They are defined as

$$\begin{align*}
\text{CSI} &= \frac{\text{Hits}}{\text{Hits} + \text{Misses} + \text{Falsealarms}} \\
\text{FAR} &= \frac{\text{Falsealarms}}{\text{Hits} + \text{Falsealarms}} \\
\text{POD} &= \frac{\text{Hits}}{\text{Hits} + \text{Misses}}
\end{align*}$$

(4)

Hits correspond to the true positive, misses correspond to the false positive, and false alarms correspond to the false negative. Higher CSI and POD indicate higher forecasting accuracy, lower FAR indicates lower false alarm rate. For the radar echo intensity, we convert the pixel value to dBZ, and calculate the CSI, FAR, and POD with

| Model       | MSE (per) | PSNR (per) | Sharp (per) |
|-------------|-----------|------------|-------------|
| ConvLSTM    | 46.37     | 30.21      | 14.46       |
| ConvGRU     | 46.23     | 30.25      | 15.52       |
| PredRNN     | 46.01     | 30.16      | 15.56       |
| PredRNN++   | 46.52     | 30.21      | 17.17       |
| MotionRNN   | 45.43     | 30.38      | 17.52       |
| GMA + ConvLSTM | 44.70     | 30.39      | 15.47       |
| GMA + ConvGRU | 44.96     | 30.38      | 15.53       |
| GMA + PredRNN | 44.51     | 30.39      | 18.83       |
| GMA + PredRNN++ | 42.44     | 30.60      | 18.68       |
| GMA + MotionRNN | 44.39     | 30.41      | 16.93       |

Note. Bold represents the best results.
20 and 30 dBZ as thresholds, respectively. The forecasting results are shown in Tables 1 and 2. After using the GMA framework, benchmark networks achieved the better results in the CSI, POD, FAR of two thresholds, which shows that benchmark networks after using the GMA framework can predict precipitation with different precipitation intensities.

As shown in Figure 5, we show five benchmark networks and their radar echo images after using GMA framework. In the first few frames of radar extrapolation, the radar echo predicted by the reference network is basically consistent with the real results. With the increase of extrapolation time, the prediction accuracy gradually decreases. ConvLSTM and ConvGRU can only predict the future echo trajectory, cannot predict the echo shape, gradually lose the radar echo information, and it is difficult to predict the future echo shape until it is completely blurred. PredRNN can still predict echo shape, but it loses low-intensity echo information and cannot predict low-intensity radar echo dissipation. After using GMA framework, they can better capture the approximate range of low-intensity radar echo in the future. In contrast, PredRNN++, MotionRNN predicts the large-scale echo shape details in more detail after using the GMA framework, while accurately predicting the dissipation of small-scale radar echoes.

As shown in Figure 5, ConvLSTM, ConvGRU predicted radar echo map gradually blurred with the extrapolation time. Although the radar echo trajectory predicted by PredRNN and MotionRNN is basically consistent with the actual situation, as the extrapolation time increases, it becomes difficult to predict the small-scale echo trajectory. For low-intensity radar echoes, it can only predict the approximate range of future echoes, and the details and clarity of the echo shape are not enough. PredRNN++ can predict the trajectory of low-intensity radar echoes, but cannot predict the dissipation of small-range radar echoes. In contrast, PredRNN++ after using GMA due to the guiding role of global memory in the motion trajectory and echo shape for the prediction process, they can predict the shape of future low-intensity radar echoes in more detail, and can more accurately predict small areas dissipation of internal radar echo.

As shown in Figure 6, the radar echo image is a gray-scale image, so the visual differentiation of different echo reflectivity regions is not intuitive enough. Therefore, this paper carries out pseudo color processing on the image sequence here. The image sequence processed by pseudo color is only used as the display in this paper, and the actual model still uses the gray-scale image without pseudo color processing.

As shown in Figure 6, we show five benchmark networks and their radar echo images of the first 20 frames after using GMA framework. Although this image sequence does not belong to the scope of radar extrapolation prediction, this paper discusses the image sequence predicted by each network when each frame is input with a real image, and compares it better with the radar extrapolation prediction image sequence.

First, for the heavy precipitation area in the upper right corner, Although the five benchmarks predict the approximate range of heavy rainfall, it is impossible to accurately distinguish the areas around heavy rainfall from nonheavy rainfall areas. However, PredRNN and MotionRNN after using the GMA, under the guidance of global

| Model                      | CSI20 | CSI30 | POD20 | POD30 | FAR20 | FAR30 |
|----------------------------|-------|-------|-------|-------|-------|-------|
| ConvLSTM                   | 0.234 | 0.124 | 0.345 | 0.256 | 0.219 | 0.343 |
| ConvGRU                    | 0.226 | 0.092 | 0.344 | 0.236 | 0.222 | 0.391 |
| PredRNN                    | 0.230 | 0.120 | 0.337 | 0.240 | 0.219 | 0.349 |
| PredRNN++                  | 0.214 | 0.097 | 0.360 | 0.258 | 0.248 | 0.384 |
| MotionRNN                  | 0.207 | 0.089 | 0.372 | 0.239 | 0.262 | 0.396 |
| GMA + ConvLSTM             | 0.237 | 0.120 | 0.371 | 0.277 | 0.221 | 0.356 |
| GMA + ConvGRU              | 0.227 | 0.106 | 0.357 | 0.263 | 0.226 | 0.372 |
| GMA + PredRNN              | 0.232 | 0.124 | 0.355 | 0.267 | 0.218 | 0.344 |
| GMA + PredRNN++            | 0.237 | 0.120 | 0.371 | 0.277 | 0.221 | 0.356 |
| GMA + MotionRNN            | 0.226 | 0.109 | 0.361 | 0.261 | 0.230 | 0.368 |

Note. Bold represents the best results.
information, the particle size of heavy precipitation area is smaller, and the boundary of heavy precipitation and nonheavy precipitation is finely divided.

As shown in Figure 7, we show five benchmark networks and their radar echo images of the last 20 frames after using GMA framework. Different from the time step of the first 20 frames, the prediction time frame input of the last 20 frames comes from the network prediction image of the previous frame. Therefore, each frame of the predicted image is crucial to the radar extrapolation prediction process. Some benchmark networks, such as ConvGRU, it predicts that the picture has been mostly blurred. Compare the upper right corner and center area of the images. The strong echo gradually dissipates, and there is no precipitation in some areas in the middle of the original strong precipitation area. The ConvLSTM and ConvGRU lost their long-term memories. Although no precipitation areas were predicted, but it is wrong to predict a large area of heavy precipitation. After using the GMA, they revised the heavy rainfall forecast results under the guidance of global information. Contrast the lower left corner of the images, there are a small number of strong precipitation areas in the large area of no precipitation and low-intensity precipitation areas. After PredRNN++ using GMA, it finely divides the boundaries of heavy precipitation regions. At the same time, compared with other networks, it is better to predict the shape of precipitation region for the middle weak precipitation region.
5. Conclusions

We propose an improved framework GMA based on timing network radar extrapolation algorithm, which can be applied to any timing network. It establishes a global information flow between network calculation units, and extracts key historical memories through a series of attention mechanisms to guide the prediction and update process. It enhances the ability of network learning and long-term historical memory, and effectively alleviates the dilemma of network interference by image noise. The experimental results on the real radar data set of Hunan meteorological station show that we propose that GMA can effectively optimize the prediction results of radar extrapolation tasks.

**Figure 6.** The first 20 frames prediction example on the radar echo data set.
Conflict of Interest
The authors declare no conflicts of interest relevant to this study.

Data Availability Statement
The data sets in the manuscript were collected by the local meteorological department (Hunan Meteorological Bureau). The authors collaborate with the institution for scientific research, they grant permission for the data set to be used for research purposes. Due to laws and regulations, the disclosure of government depart-
Supplementary data is prohibited. The authors are sorry that we can only provide partial data, we randomly selected 100 radar echo sequences from the data set as submissions. Radar echo data set URL: https://zenodo.org/badge/latestdoi/531896507.

References

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., & Hickey, J. (2019). Machine learning for precipitation nowcasting from radar images. arXiv:1912.12132.

Gibson, J. (1950). The ecological approach visual perception. Boston: Houghton Mifflin.

Han, L., Wang, H. Q., Tan, X. G., & Lin, Y. J. (2007). Research progress of storm body identification, tracking and early warning based on radar data. Meteorology, 33(1), 3–10.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kelly, E. J., Reed, I. S., & Root, W. L. (1960). The detection of radar echoes in noise. I. Journal of the Society for Industrial and Applied Mathematics, 8(2), 309–341. https://doi.org/10.1137/0108020

Kim, D. K., Suezawa, T., Mega, T., Kikuchi, H., Yoshikawa, E., Baron, P., & Ushio, T. (2021). Improving precipitation nowcasting using a three-dimensional convolutional neural network model from multi parameter phased array weather radar observations. Atmospheric Research, 262, 105774. https://doi.org/10.1016/j.atmosres.2021.105774

Li, Y. J., & Han, L. (2008). Research on storm body tracking algorithm based on trunk 3D radar image data. Computer Application, 28(4), 1078–1080.

Liang, Q. Q., Feng, Y., Deng, W., Hu, S., Huang, Y., Zeng, Q., & Chen, Z. (2010). A composite approach of radar echo extrapolation based on TREC vectors in combination with model-predicted winds. Advances in Atmospheric Sciences, 27(5), 1119–1130. https://doi.org/10.1007/s00376-009-9093-4

Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C. (2020). Self-attention ConvLSTM for spatiotemporal prediction. In Proc. 34st AAAI (pp. 11531–11538).

Lotter, W., Kreiman, G., & Cox, D. (2016). Deep predictive coding networks for video prediction and unsupervised learning. arXiv:1605.08104.

Luo, C. Y. (2021). A novel LSTM model with interaction dual attention for radar echo extrapolation. Remote Sensing, 13(2), 164.

Ranzato, M., Szlam, A., Bruna, J., Mathieu, M., Collobert, R., & Chopra, S. (2014). Video (language) modeling: A baseline for generative models of natural videos. arXiv:1412.6604.

Rinehart, R. E., & Garvey, E. T. (1978). Three-dimensional storm motion detection by conventional weather radar. Nature, 273(5660), 287–289. https://doi.org/10.1038/273287a0

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In Proc. 31st Conf. Neural Inf. Process (NIPS) (pp. 5622–5632).

Srivastava, N., Mansimov, E., & Salahudinov, R. (2015). Unsupervised learning of video representations using LSTMs. In International Conference on Machine Learning (PMLR) (pp. 843–852).

Warsini, A., Shazer, N., Parman, N., Uszkoreit, J., Jones, L., & Gomez, A. N. (2017). Attention is all you need. In Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS) (pp. 6000–6010).

Wang, Y., Gao, Z., Long, M., Wang, J., & Philip, S. Y. (2018). PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In ICML (pp. 5110–5119).

Wang, Y., Jiang, L., Yang, M. H., Li, L. J., Long, M., & Li, F.-F. (2015). Eidetic 3D LSTM: A model for video prediction and beyond. In Proc. 7th Int. Conf. Learn. Represent. (ICLR).

Wang, Y., Long, M., Wang, J., Gao, Z., & Philip, S. Y. (2017). PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In Proc. Conf. Neural Inf. Process. Syst. (NIPS) (pp. 879–888).

Wang, Y., Zhang, J., Zhu, H., Long, M., Wang, J., & Yu, P. S. (2019). Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In Proc. Comput. Vis. Pattern Recognit. (CVPR) (pp. 9154–9162). https://doi.org/10.1109/cvpr.2019.00937

Wu, H., Yao, Z., Wang, J., & Long, M. (2021). MotionRNN—A flexible model for video prediction with spacetime-varying motions. In Proc. Comput. Vis. Pattern Recognit. (CVPR) (pp. 15435–15444). https://doi.org/10.1109/cvpr46437.2021.01518

Zhong, S. X. (2020). Spatiotemporal convolutional LSTM for radar echo extrapolation. In 2020 54th Asilomar Conference on Signals, Systems, and Computers (pp. 58–62). IEEE.