*remote sensing*

MDPI

*Article*

# An Long Short-Term Memory Model with Multi-Scale Context Fusion and Attention for Radar Echo Extrapolation

**Guangxin He [1,2,3], Haifeng Qu [1,2,3], Jingjia Luo [1], Yong Cheng [1,2,*], Jun Wang [1] and Ping Zhang [4]**

[1] Key Laboratory of Meteorological Disaster, Ministry of Education (KLME), and International Joint Research Laboratory on Climate and Environment Change (ILCEC), and Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China
[2] Guangzhou Institute of Tropical Marine Meteorology, China Meteorological Administration, Guangzhou 510080, China
[3] Key Laboratory of Meteorology and Ecological Environment of Hebei Province, Hebei Provincial Institute of Meteorological Sciences, Shijiazhuang 050021, China
[4] School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China; zhangp@niit.edu.cn
* Correspondence: yongcheng@nuist.edu.cn

**Abstract:** Precipitation nowcasting is critical for areas such as agriculture, water resource management, urban drainage systems, transport and disaster preparedness. In recent years, methods such as convolutional recurrent neural networks (ConvRNN) in deep learning techniques have been used to solve this task. Despite the effective improvement in forecasting quality, there are still problems with blurred and distorted prediction images, as well as difficulties in effectively forecasting high echo regions. To solve the above problems, this article presents a spatio-temporal long–short-term memory network model in view of multi-scale context fusion and attention mechanisms. This method fully extracts the short-term context information of different scales of radar image through the multi-scale context fusion module. The attention module broadens the time perception domain of the prediction unit so that the model perceives more historical time dynamics. Using the Hong Kong region weather radar data as a sample, the results of the experimental comparative analysis show that the spatio-temporal long and short-term memory network in view of multi-scale context fusion and attention mechanism achieves better prediction performance. Our model is effective in improving both image quality and meteorological assessment metrics with higher accuracy and more details.

**Keywords:** radar echo extrapolation; precipitation forecast; multi-scale; long–short-term memory

check for updates

**Citation:** He, G.; Qu, H.; Luo, J.; Cheng, Y.; Wang, J.; Zhang, P. An Long Short-Term Memory Model with Multi-Scale Context Fusion and Attention for Radar Echo Extrapolation. *Remote Sens.* **2024**, *16*, 376. https://doi.org/10.3390/rs16020376

Academic Editor: Steven Dewitte

Received: 12 November 2023
Revised: 9 January 2024
Accepted: 11 January 2024
Published: 17 January 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Precipitation nowcasting has always been an important task in meteorological forecasting, which usually refers to the prediction of short-term (usually 0–2 h) rainfall in a certain area in the future [1]. Usually, accurate precipitation nowcasting predication can provide preventive operations (e.g., weather guidance for agriculture, navigation, etc.), especially in severe weather such as heavy rainfall and thunderstorms, thereby reducing human casualties and property damage. Therefore, how to utilize the technology of radar echo extrapolation to obtain accurate and fast weather short-impeding predication has become a hot issue in meteorological research.

Precipitation forecasting can be seen as a spatial and temporal series prediction problem. The forecast radar map is converted into rainfall intensity through the Z–R relationship [2] as the nowcasting forecast. The main methods of traditional radar echo extrapolation are cross-correlation [3,4], monomeric center-of-mass [5,6] and optical flow [7,8]. The cross-correlation method divides the entire data area into several small regions, calculates the correlation coefficient between adjacent small regions of the radar echo image,

determines the corresponding relationship between the regions in the adjacent time images through the maximum correlation coefficient, and then determines the average motion of the echo region. However, tracking failures significantly increase in severe convective weather. The monomeric center-of-mass method involves identifying, analyzing and tracking thunderstorms as three-dimensional monomers, as well as fitting extrapolations of the thunderstorms to make proximity forecasts. The accuracy of this method is greatly reduced when the radar echoes are more fragmented or appear to merge or split. The optical flow method essentially obtains the motion vector field of the radar echo by calculating the optical flow field of the echo and extrapolates the radar echo based on this motion vector field. However, the optical flow method accumulates errors in the two steps of calculating the optical flow vector and extrapolating. Therefore, the traditional method often cannot model the spatio-temporal relationship well or obtain better forecast results.

In recent times, deep learning has become the most rapidly developing technique in machine learning, and in response to the problems of traditional methods, more and more people are trying to use deep learning methods to solve video prediction problems [9–11], traffic flow prediction [12–14] and precipitation nowcasting [15–22], as well as other spatio-temporal sequence prediction problems. Deep learning methods can handle complex spatio-temporal relationships in order to adaptively learn the patterns of rainfall variability from a large number of previous radar echo sequences. For medium-term weather forecasting on a global scale, the Google DeepMind team [23] designed a deep generative model, DGMR, based on conditional generative adversarial networks, to accurately predict echo motions and precipitation while generating clear forecasts. For medium-term weather forecasting on a global scale. The Nvidia team [24] designed a high-resolution weather model, FourCastNet, that uses adaptive Fourier neural operators to generate global data-driven forecasts of key atmospheric variables at a 0.25° resolution. More deep learning models for global weather have since been proposed [25–28]. For the purpose of localized precipitation forecasting, deep-learning-based extrapolation of radar echoes is more advantageous. For example, Shi et al. [29] suggested a convolutional LSTM (ConvLSTM) model that combines convolutional neural networks (CNN) and short-term memory (LSTM) networks for precipitation prediction in Hong Kong. LSTM is used to extract temporal dynamic information and storing it in temporal memory units, while CNN is responsible for extracting spatial information. Therefore, the network can learn and model spatio-temporal information better. Considering that ConvLSTM only focuses on temporal information and ignores the spatial information from between layers, Wang et al. [30] suggested the ST-LSTM unit (Spatiotemporal LSTM) to preserve the spatial features of each layer and apply them to the new end-to-end model PredRNN by adding a new parallel spatial memory unit to ConvLSTM. Wang et al. [31] further constructed the Causal-LSTM unit by cascading the dual memory units and adding the Gradient Highway Unit (GHU), which is used to alleviate the problem of gradient vanishing, and formed a new end-to-end model, PredRNN++. Wang et al. [32] suggested a new model Eidetic 3D LSTM (E3D-LSTM) that integrates 3D convolution into the RNN, enabling the storage unit to store better short-term features. For long-term relationships, the current memory state is made to interact with its historical record by a gate-controlled self-attentive mechanism. However, the integrated 3D convolution makes the computational load of E3D-LSTM very high. Wang et al. [33] suggested a Memory In Memory network (MIM) that can capture the non-smooth and near-smooth features in radar echo images. Additionally, several variant structures based on ConvLSTM and PredRNN have emerged, such as PredRANN [34], SAST-LSTM [35] and PrecipLSTM [36], among others.

Despite the significant improvements in the above methods, these networks still have two shortcomings. Firstly, the prediction unit does not fully consider the contextual correlation between the previous output and current input, resulting in the input and hidden states not being able to assist each other in identifying and preserving important information. Therefore, as the depth of the model increases, the correlation between the contexts gradually decreases and short-term correlation information is lost. Second, as

the prediction time increases, the problem of gradual decay of the stored information in the memory unit occurs, i.e., it is difficult for the memory unit at the current moment to effectively recall the stored memory at the previous moment. These problems lead to the gradual blurring of the radar echo prediction image with the increase in prediction time in the radar echo extrapolation task, and the trend in the disappearance of radar echo areas with high reflectivity greatly affects the prediction accuracy.

With the aim of enhancing the level of detail in prediction images, improving the capability to forecast high echo regions and achieving accurate precipitation forecasts over extended time periods, this paper proposes a spatio-temporal LSTM model with multi-scale context fusion and attention mechanism (MCA-LSTM) to address the above problems. First, this paper proposes a multi-scale context fusion module to effectively extract multi-scale spatio-temporal information of images and improve contextual relevance. Then, an attention module is proposed to make the model perceive more temporal information by widening the temporal perceptual domain of the network model. By integrating these two modules into the network unit, the performance is significantly improved, especially in areas with heavy rainfall.

Our approach can be summarized as follows:

1. We propose a multi-scale contextual information fusion module for efficient multi-scale feature extraction, as well as for the improvement in the correlation between contexts. It effectively improves the blurring problem of predicted images and enhances the details.
2. We propose an attention module that effectively improves the forgetting problem of the prediction unit during information transmission. A better establishment of long-term time dependence improves the prediction of high echo regions.
3. Combining the above two methods, CAST-LSTM is constructed. Experiments show that CAST-LSTM achieves state-of-the-art results on long-term prediction tasks.

The rest of the article is organized as follows: Section 2 shows the dataset of standard Moving MNIST and the real radar echo used for experiments. Section 3 presents the details of the method proposed in this article. Section 4 carries out the experimental analysis to compare and analyze the results of the dataset tests. Section 5 summarizes the paper to draw conclusions and provides an outlook for future research work.

## 2. Data

### *2.1. Moving MNIST Dataset*

The Moving MNIST dataset is the most widely used dataset in spatio-temporal sequence prediction, where several shapes randomly move within a limited range and have several motion modes, including rotation, rescaling, illumination changes and so on. Every 20 consecutive frames are divided into a sequence. There are 10 frames for input, 10 frames for prediction, and each frame is 64 $\times$ 64. The training dataset consists of 10,000 sequences, the validation dataset consists of 2000 sequences, and the testing dataset consists of 3000 sequences.

### *2.2. Radar Dataset*

This paper uses the well-known public radar dataset HKO-7 provided by the Hong Kong Observatory (HKO) to evaluate the model performance. The dataset is the Hong Kong weather radar data from 2009–2015, which is produced by the Hong Kong Observatory (HKO) and stored in the form of gray-scale maps. The time interval is 6 min and the grid size of the single-time data (i.e., single image) is 480 $\times$ 480 pixels, taken at an altitude of 2 km and covering a 512 km $\times$ 512 km area centered on Hong Kong. Data are recorded every 6 min, resulting in 240 frames per day. In order to be more suitable for model training and testing, our experimental radar images were scaled to 256 $\times$ 256 by bilinear interpolation.

As precipitation does not occur every day, the radar echo images when there is no precipitation are not meaningful for the development of the network, so the part of the

data without precipitation needs to be filtered before dividing the training and test sets. Over 20,000 radar images were filtered as the training set and 5000+ radar images were used as the test set. In this paper, 30 radar images with an interval of 6 min are used as a sequence sample. The first 10 radar echo images are used as input and the last 20 radar echo images are used as predicted output in each sample. Therefore, a two-hour extrapolation is predicted based on the observations of the past hour.

## 3. Algorithm Description

This section describes the detail of the MCA-LSTM model. First, the multi-scale context fusion module is introduced, then the attention module is elaborated, and how to embed the multi-scale context fusion module and the attention module into the ST-LSTM unit is described. Finally, the overall extrapolation structure of the proposed MCA-LSTM model is presented.

### *3.1. Context Fusion Module*

In LSTM-based models (e.g., ConvLSTM, PredRNN, etc.) with a gating structure consisting of input gates, forgetting gates, input modulation gates and output gates, which learn new input features and previous features in the current input and previous hidden states, respectively, there is not only a sequential relationship between and in time, but also a low-level and high-level relationship in space. This close connection between contexts is therefore crucial to the accuracy of the prediction results. However, the previous hidden state and current input of existing networks can only interact individually through convolutional layers and additional operations. As the depth of the model increases, the contextual relationship between the current input and the previous hidden state gradually weakens, which leads to the loss of short-term relevance information of the model and makes the prediction results inaccurate. Therefore, this paper proposes a multi-scale context fusion module for extracting multi-scale features and improving contextual relevance, as shown in Figure 1.

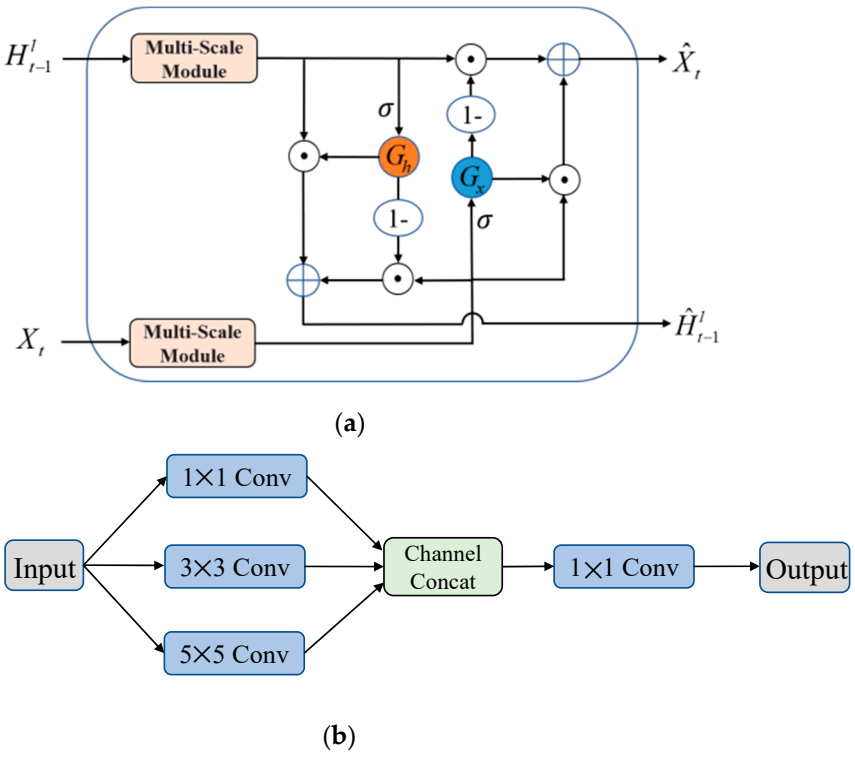



**Figure 1.** (**a**) Multi-Scale Context Fusion Module. (**b**) Multi-Scale Module.

First, spatio-temporal information at different scales of the context is extracted by means of a multi-scale module, as shown in Equation (1):

$$\begin{aligned} X_t' &= Concat(W_x^{k\times k} * X_t), k = 1,3,5 \\ H_{t-1}^l{}' &= Concat(W_h^{k\times k} * H_{t-1}^l), k = 1,3,5 \end{aligned} \tag{1}$$

where "$*$" denotes two-dimensional convolution, "$W$" denotes the weight matrix and "Concat" denotes channel stitching. From this equation, the inputs in the MCA-LSTM model are $X_t$ and $H_{t-1}^l$. Convolution operations are performed on the inputs $X_t$ and $H_{t-1}^l$ using convolution kernels of $1 \times 1$, $3 \times 3$ and $5 \times 5$ sizes. This is used to help the contextual information extract detailed features at different scales. Channel concatenation is performed separately, followed by convolutional operations to restore the channels, resulting in the acquisition of the current input $X_t'$ and the previous hidden state $H_{t-1}^l{}'$, both enriched with multi-scale feature information.

Then, the current input $X_t'$ and the previously hidden state $H_{t-1}^l{}'$ are fused, and in order to control the fusion rate of the information, the two fusion gates are shown in Equation (2):

$$\begin{aligned} G_x &= \sigma(W_{xu} * X_t') \\ G_h &= \sigma(W_{hu} * H_{t-1}^l{}') \end{aligned} \tag{2}$$

where $G_x$ denotes the current moment fusion gate, $G_h$ denotes the previous moment fusion gate and "$\sigma$" denotes the sigmoid function. The fusion is performed by two gates as shown in Equation (3):

$$\begin{aligned} \hat{X}_t &= G_x \odot (W_{xx} * X_t') + (1 - G_x) \odot (W_{hx} * H_{t-1}^l{}') \\ \hat{H}_{t-1}^l &= G_h \odot (W_{hh} * H_{t-1}^l{}') + (1 - G_h) \odot (W_{xh} * X_t') \end{aligned} \tag{3}$$

where "$\odot$" means the Hadamard product.

As seen by the above equation, finer multi-scale spatio-temporal features are extracted by convolving the contextual information with different sizes. The use of fusion gates to control the context fusion process improves the contextual relevance of the current input and previously hidden states. Therefore, this module can effectively solve the problem of weakening contextual relevance with increasing prediction time. Meanwhile, the multi-scale feature extraction method adopted by this module can effectively improve the details and clarity of the prediction results.

### *3.2. Attention Module*

In this paper, an attention module is proposed as shown in Figure 2 to further improve the model's long-term dependability and the reduction in information loss. This module obtains the corresponding attention scores based on the correlation between the current spatial state $M_t^{l-1}$ and the historical spatial state $M_{t-\tau:t-1}^{l-1}$. Different degrees of attention are given to historical temporal states $C_{t-\tau:t-1}^l$ based on the attention scores, and the attended historical temporal states are aggregated into long-term memory units $C_{att}$. As a result, the prediction unit can perceive more temporal information from a wider sensory domain. Then, the long-term memory unit $C_{att}$ and the short-term memory unit $C_{t-1}^l$ are further fused into the final enhanced memory unit $C_{ATT}$.

To achieve this important process, the correlation attention fraction between the spatial state of current time and the spatial state of historical time is first calculated, as shown in Equation (4):

$$\begin{aligned} \beta_i &= M_t^{l-1} \cdot M_{t-i}^{l-1}, i = 1,2\cdots,5 \\ Score_i &= Softmax(\beta_i) \end{aligned} \tag{4}$$

This is true particularly when $l = 1$, $M_t^{l-1} = X_t$, $M_{t-\tau:t-1}^{l-1} = X_{t-\tau:t-1}$, $(\tau = 5)$, where "$\cdot$" denotes the matrix dot product operation and "$\beta_i$" denotes the correlation coefficient. The dot product of the current spatial state $M_t^{l-1}$ and the spatial memory $M_{t-\tau:t-1}^{l-1}$ is calculated

at multiple time steps in history, respectively. Then, it is further normalized using the Softmax activation function to attentional fraction $Score_i$.

In order to aggregate the multi-step historical temporal information in the time domain, the attention scores $Score_i$ is applied to the corresponding temporal memory units and then fused by summation, as shown in Equation (5):

$$C_{att} = \sum_{i=1}^{\tau} Score_i \cdot C_{t-i}^{l}, \tau = 5 \tag{5}$$

where "$C$" denotes the temporal memory unit in the prediction unit. The attention score is obtained by the correlation between the current spatial state and the historical spatial state and can better and selectively retain the information of the historical temporal memory unit; $C_{att}$ can be represented as temporal attention information, which represents a long-term motion trend.

In order to effectively aggregate long-term motion trend information $C_{att}$ and short-term motion information $C_{t-1}^{l}$, the fusion rate between the two is controlled by setting a temporal fusion gate $G_f$, as shown in Equation (6):

$$\begin{aligned} G_f &= \sigma(W_f * C_{t-1}^{l}) \\ C_{ATT} &= G_f \odot C_{t-1}^{l} + (1 - G_f) \odot C_{att} \end{aligned} \tag{6}$$

The final enhanced motion information $C_{ATT}$ is obtained by using $G_f$ to control the percentage of short-term motion state information retained and the percentage of long-term motion trend information retained by $(1 - G_f)$.

The above process widens the time-receptive domain of the prediction unit so that it can capture more historical information. The problem of irreversible information loss in the prediction process is effectively improved, and the prediction capability for high echo regions is enhanced.

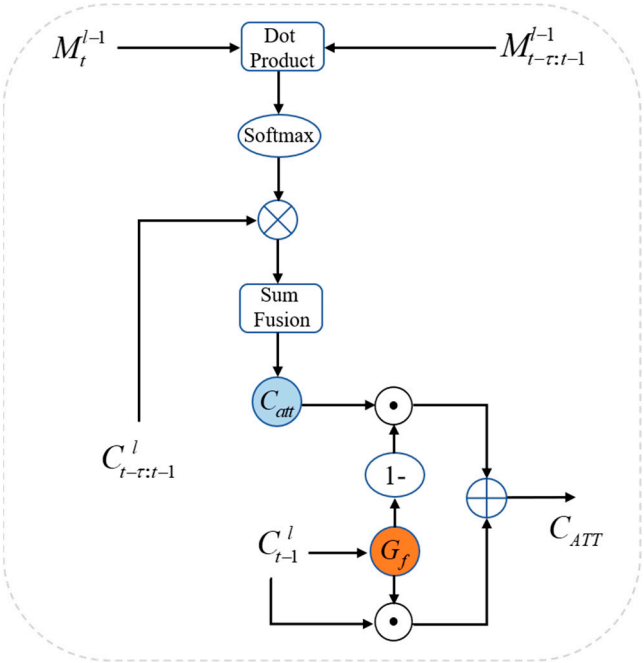



**Figure 2.** Attention module embedded in model.

### 3.3. MCA-LSTM Cell

In this subsection, the internal structure of the MCA-LSTM cell is introduced. As shown in Figure 3, the input of MCA-LSTM cell includes current input $X_t$, spatial memory $M_t^{l-1}$, temporal memory $C_{t-1}^{l}$, historical temporal memory set $C_{t-\tau:t-1}^{l}$, historical spatial

memory set $M_{t-\tau:t-1}^{l-1}$ and hidden state $H_t^{l-1}$. The current input $X_t$ and hidden states $H_t^{l-1}$ are firstly fused by extracting detailed spatio-temporal features at different scales through the context fusion block, and then the new input $\hat{X}_t$ and hidden states $\hat{H}_t^{l-1}$ are obtained. The current spatial memory $M_t^{l-1}$, historical spatial memory set $M_{t-\tau:t-1}^{l}$, temporal memory $C_{t-1}^{l}$ and historical temporal memory set $C_{t-\tau:t-1}^{l}$ are used as the input of the attention module to obtain the enhanced memory unit $C_{ATT}$. The MCA-LSTM unit is calculated as shown in Equation (7):

$$\begin{aligned}
&\hat{X}_t, \hat{H}_{t-1}^{l} = MSCF(X_t, H_{t-1}^{l}) \\
&i_t = \sigma(W_{xi} * \hat{X}_t + W_{hi} * \hat{H}_{t-1}^{l} + b_i) \\
&g_t = \tanh(W_{xg} * \hat{X}_t + W_{hg} * \hat{H}_{t-1}^{l} + b_g) \\
&f_t = \sigma(W_{xf} * \hat{X}_t + W_{hf} * \hat{H}_{t-1}^{l} + b_f) \\
&C_t^{l} = i_t \odot g_t + f_t \odot ATT(C_{t-1}^{l}, C_{t-\tau:t-1}^{l}, M_t^{l-1}, M_{t-\tau:t-1}^{l-1}) \\
&i_t' = \sigma(W_{xi}' * \hat{X}_t + W_{mi} * M_t^{l-1} + b_i') \\
&g_t' = \tanh(W_{xg}' * \hat{X}_t + W_{mg} * M_t^{l-1} + b_g') \\
&f_t' = \sigma(W_{xf}' * \hat{X}_t + W_{mf} * M_t^{l-1} + b_f') \\
&M_t^{l} = i_t' \odot g_t' + f_t' \odot M_t^{l-1} \\
&o_t = \sigma(W_{xo} * \hat{X}_t + W_{ho} * \hat{H}_{t-1}^{l} + W_{co} * C_t^{l} + W_{mo} * M_t^{l} + b_o) \\
&H_t^{l} = o_t \odot \tanh(W_{1\times1} * [C_t^{l}, M_t^{l}])
\end{aligned} \tag{7}$$

where "MSCF" denotes the multi-scale context fusion module; "ATT" denotes the attention module; $i_t$ is the first input gate; $g_t$ is the first input modulation gate; $f_t$ is the first forgetting gate; $i_t'$ is the second input gate; $g_t'$ is the second input modulation gate; $f_t'$ is the second forgetting gate; $o_t$ is the output gate; $C_t^{l}$ denotes the updated temporal memory unit; $M_t^{l}$ denotes the updated spatial memory unit; $W$ denotes the corresponding convolution kernel; and $b$ denotes the corresponding deviation value. "*" denotes the 2D convolution operation, "$\odot$" denotes the Hadamard product and $\tau$ is the historical time step. In particular, in the ATT equation, when l = 1, $M_t^{l-1} = X_t$, $M_{t-j}^{l-1} = X_{t-j}$.

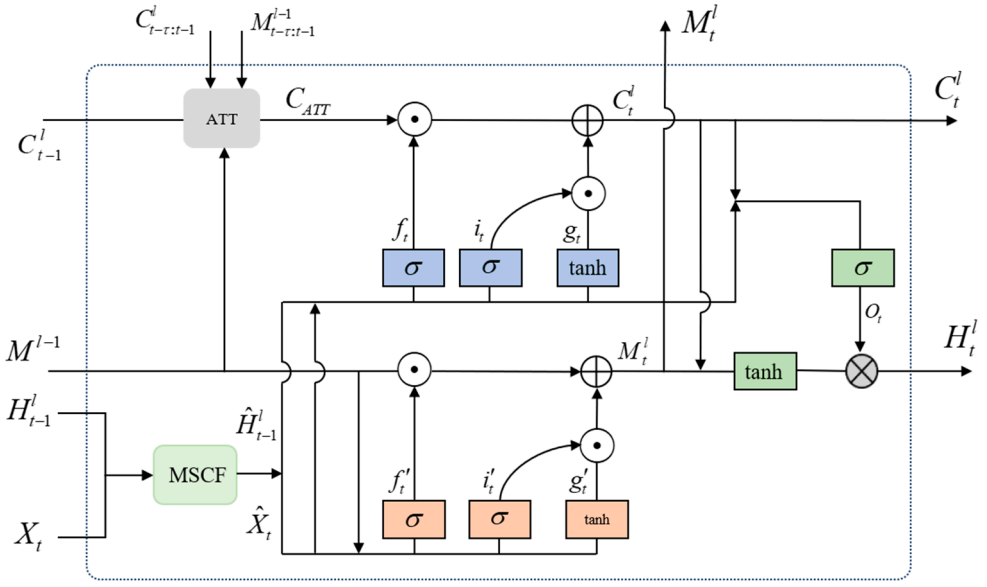



**Figure 3.** Internal structure diagram of context fusion attention long–short-term memory unit.

### *3.4. MCA-LSTM Network Structure*

The network structure of the MCA-LSTM model is presented in Figure 4. This network is constructed by stacking four layers of MCA-LSTM cells, in which the spatial storage cell M (shown by the black dashed line) is updated in the zigzag direction and the temporal

storage cell C (shown by the black solid line) is updated in the horizontal direction, and the top layer outputs the prediction results $\hat{I}_t$.

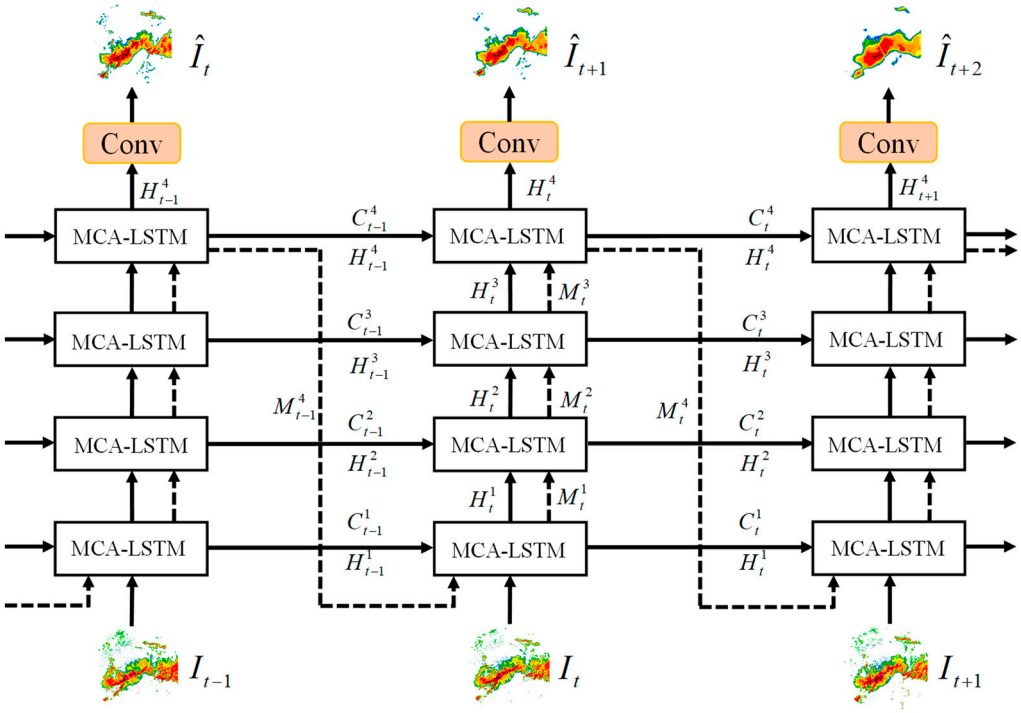



**Figure 4.** MCA-LSTM network model structure.

### *3.5. Evaluation Metrics*

For evaluation, the critical success index (CSI), Heidke skill score (HSS) and probability of detection (POD) metrics are used in this paper to assess the results. For this purpose, the following transformations are applied to convert the pixel values pixel of ground truth and predicted echo maps to reflectance dBZ as shown in Equation (8):

$$dBZ = p \times 70/255 - 10 \tag{8}$$

The predicted echo maps and ground truth maps are converted to binary matrices by setting thresholds. If the radar echo value is greater than the given threshold, the corresponding value is set to 1; otherwise, it is set to 0. Analogously to the meteorology as shown in the confusion matrix in Table 1 the true positive prediction TP (prediction = 1, true value = 1), false positive prediction FP (prediction = 1, true value = 0), true negative prediction TN (prediction = 0, true value = 0) and false negative prediction FN (prediction = 0, true value = 1) are calculated.

**Table 1.** Confusion matrix.

| | **Forecast** | **Forecast** |
|---|---|---|
| True | TN (True negative) | FP (False positive) |
| True | FN (False negative) | TP (True positive) |

The specific formulas for CSI, HSS and POD are shown in Equation (9):

$$\begin{aligned} CSI &= \frac{TP}{TP+FN+FP} \\ HSS &= \frac{2(TP \times TN - FN \times FP)}{(TP+FN)(FN+TN)+(TP+FP)(FP+TN)} \\ POD &= \frac{TP}{TP+FN} \end{aligned} \tag{9}$$

Specifically, 20, 35 and 45 dBZ are chosen as thresholds. CSI and HSS are composite measures that take into account the detection probability and false alarm rate can directly reflect the advantages of the model. The better performance with the larger CSI and HSS. Again, the larger the POD, the better the forecasting performance of the model.

## 4. Experiments and Analysis

The experiments are conducted on the Moving MNIST dataset and the Hong Kong Radar dataset, respectively, and are analyzed in comparison with the existing models in this section. Four layers of MCA-LSTM cells are applied according to Figure 4, with the number of channels per cell set to 64 and the convolutional kernel size set to $5 \times 5$. The comparison models are all used in the same way as in Figure 4. All models are trained and tested on the Pytorch-based framework, and the experiments are implemented on an NVIDIA A10 GPU (NVIDIA Santa Clara, CA, USA). The Adam optimizer is selected for optimization, with a learning rate of 0.0001 and a batch size of 4. To stabilize the training process, LeakyReLU activation function is used after each convolutional layer in MCA-LSTM.

### *4.1. Moving MNIST Experiments*

Results and Analysis

This article uses two commonly used metrics to evaluate performance, including mean square error (MSE) and structural similarity index (SSIM). As shown in Table 2, lower MSE and higher SSIM indicate better predictive performance.

**Table 2.** Results of different methods on moving MNIST datasets (10 frames $\rightarrow$ 10 frames).

| Method | MSE/Frame↓ | SSIM/Frame↑ |
|---|---|---|
| ConvGRU | 103.4 | 0.713 |
| ConvLSTM | 102.3 | 0.725 |
| PredRNN | 55.8 | 0.866 |
| PredRNN++ | 45.6 | 0.895 |
| MIM | 44.1 | 0.905 |
| IDA-LSTM | 38.4 | 0.916 |
| MCA-LSTM | 29.7 | 0.938 |

As shown in Figure 5, the MCA-LSTM proposed in this paper clearly outperforms other methods, especially in the prediction of the last two time steps. The MCA-LSTM network retains the details of digit variations well, especially when dealing with overlapping trajectories, and maintains clarity over time. This is due to the fact that MCA-LSTM introduces a multi-scale context fusion module for extracting the detail information of moving digits and, at the same time, increases the interactivity between the contexts to recognize important information from each other. In addition, MCA-LSTM introduces an attention mechanism module, which effectively ameliorates the forgetfulness problem in information transfer by obtaining more multi-step history information from a wider temporal sensory domain. In comparison, the prediction results of ConvGRU and ConvLSTM networks become blurred very quickly and gradually lose the detailed information; this is because ConvGRU and ConvLSTM only focus on the temporal information of lateral propagation and ignore the spatial information between different cell layers. Although other methods can also achieve some prediction effect, such as PredRNN and PredRNN++ models, which have made some improvements for temporal and spatial information, the effect is still unsatisfactory. As the prediction time goes by, only MCA-LSTM can retain more detailed information in the prediction result of the last time step, which is more advantageous in terms of localization accuracy and spatial appearance.

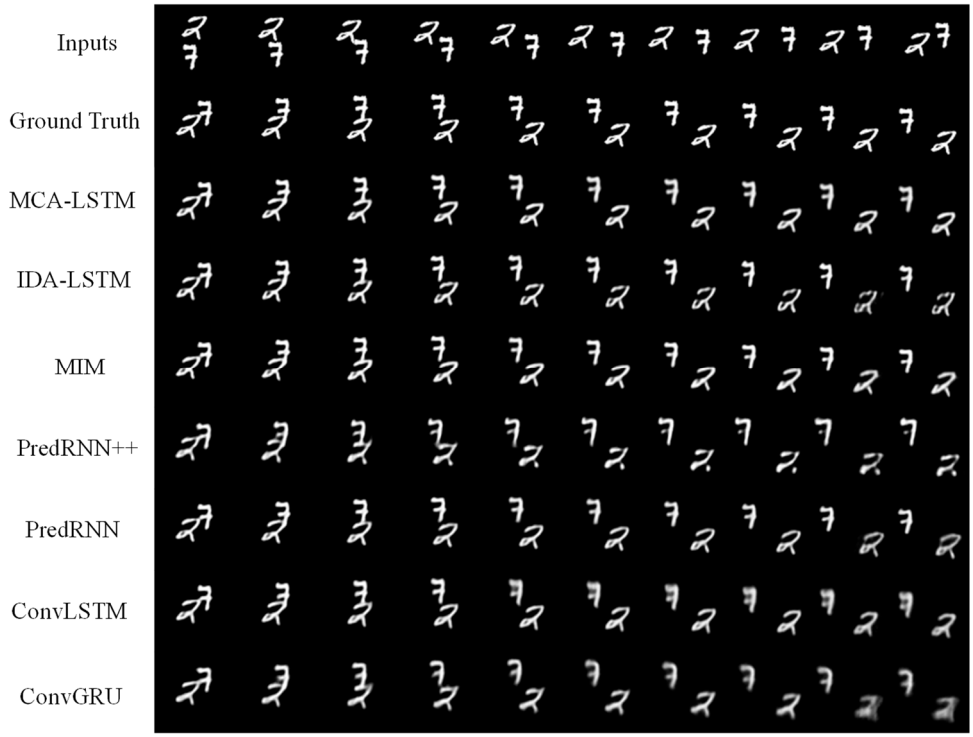



**Figure 5.** Results of different methods on mobile MNIST dataset.

### *4.2. Radar Dataset Experiments*

Results and Analysis

In the experiments, our model was compared with state-of-the-art models such as ConvGRU, ConvLSTM and PredRNN on the predictive evaluation metrics CSI, HSS and POD.

In order to provide a comprehensive assessment of the algorithm's prediction accuracy performance, we also provide prediction evaluation scores for multiple thresholds (20 dBZ, 35 dBZ and 45 dBZ) corresponding to different rainfall levels. Tables 3–5 show the comparison results of different methods. The best results are also marked in bold. It can be seen that the MCA-LSTM model proposed in this paper has the better performance at all thresholds, and the advantage of the model becomes increasingly evident as the threshold increases. In particular, the evaluated CSI, HSS and POD metrics reach 0.1852, 0.2725 and 0.2239, respectively, when the threshold is 45dBZ, where CSI, HSS and POD are 11.6%, 11.3% and 14.8% better, respectively, than the PredRNN algorithm, and 38.8%, 35.0% and 47.7% higher, respectively, than the IDA-LSTM algorithm. This implies that the developed multi-scale context fusion module and attention module help to improve the prediction of high rainfall areas.

**Table 3.** Scores of CSI (↑) at echo thresholds = 20, 35 and 45 dBZ.

| Method | 20 dBZ | 35 dBZ | 45 dBZ | Avg |
|---|---|---|---|---|
| ConvGRU | 0.5805 | 0.4461 | 0.1605 | 0.3957 |
| ConvLSTM | 0.5829 | 0.4588 | 0.1647 | 0.4021 |
| PredRNN | 0.5894 | 0.4323 | 0.1660 | 0.3959 |
| PredRNN++ | 0.5753 | 0.4606 | 0.1491 | 0.3950 |
| MIM | 0.5808 | 0.4488 | 0.1636 | 0.3977 |
| IDA-LSTM | 0.5721 | 0.4267 | 0.1334 | 0.3774 |
| MCA-LSTM | 0.5803 | 0.4631 | 0.1852 | 0.4095 |

**Table 4.** Scores of HSS (↑) at echo thresholds = 20, 35 and 45 dBZ.

| Method | 20 dBZ | 35 dBZ | 45 dBZ | Avg |
|---|---|---|---|---|
| ConvGRU | 0.6541 | 0.5497 | 0.2396 | 0.4811 |
| ConvLSTM | 0.6558 | 0.5629 | 0.2445 | 0.4877 |
| PredRNN | 0.6609 | 0.5646 | 0.2449 | 0.4901 |
| PredRNN++ | 0.6491 | 0.5435 | 0.2216 | 0.4714 |
| MIM | 0.6529 | 0.5521 | 0.2413 | 0.4821 |
| IDA-LSTM | 0.6446 | 0.5294 | 0.2019 | 0.4586 |
| MCA-LSTM | 0.6511 | 0.5673 | 0.2725 | 0.4970 |

**Table 5.** Scores of POD (↑) at echo thresholds = 20, 35 and 45 dBZ.

| Method | 20 dBZ | 35 dBZ | 45 dBZ | Avg |
|---|---|---|---|---|
| ConvGRU | 0.6585 | 0.5177 | 0.1886 | 0.4549 |
| ConvLSTM | 0.6651 | 0.5408 | 0.1937 | 0.4665 |
| PredRNN | 0.6791 | 0.5425 | 0.1951 | 0.4722 |
| PredRNN++ | 0.6448 | 0.5078 | 0.1741 | 0.4422 |
| MIM | 0.6642 | 0.5209 | 0.1908 | 0.4586 |
| IDA-LSTM | 0.6551 | 0.4921 | 0.1516 | 0.4329 |
| MCA-LSTM | 0.6755 | 0.5561 | 0.2239 | 0.4852 |

To better illustrate the results, curves for CSI, HSS and POD at different forecast moments (6–120 min) are shown in Figure 6 to show the performance of the various models at different time steps. It can be seen that within the first hour of the prediction time, the variability between the networks is not significant. Even at the threshold of 20 dBZ, our model has no significant advantage, and our model becomes more and more advantageous as the prediction time goes on, due to the fact that MCA-LSTM incorporates a multi-scale context fusion module and an attention module, which fully extracts spatio-temporal information at different scales to improve contextual relevance; the attention module can perceive more temporal dynamics from a wider sensory domain, reducing information forgetting and better modelling of short-term and long-term dependencies. As a result, MCA-LSTM can better retain the detail of the prediction results and perform better in the strong echo region. In addition, at higher thresholds, the results of PredRNN are not as good as the model proposed in this paper. This is because PredRNN suffers from the problem of not fully extracting contextual relevance information and loss of memory unit information.

In order to better compare and understand the results, we visualized the extrapolation results of the radar echo from 8:00–10:00 a.m. on 2 April 2014 in Hong Kong from the China radar dataset, as shown in Figure 7. As can be seen from the figure, the prediction results within the first hour do not differ much between the methods, and all of them achieve good results. As the prediction time continues to increase, the extrapolation results of the ConvGRU, ConvLSTM and IDA-LSTM models gradually become blurred, the high echo region gradually becomes smaller or even disappears and the whole prediction boundary region is gradually smoothed. This is because ConvGRU and ConvLSTM pay attention to the time information of lateral propagation and lack attention to the spatial information between cell layers, which leads to the inability to model the spatio-temporal information well in long time series prediction. IDA-LSTM, despite the use of the self-attention mechanism, is still not able to aggregate the multi-step history information effectively, which leads to its failure to achieve good results in long-term prediction. Although PredRNN and PredRNN++ networks take into account the extraction and preservation of temporal and spatial information, MIM also considers non-smooth and nearly smooth characteristics. However, the forgetting problem in the information transfer process has not been effectively improved. Our proposed MCA-LSTM adopts multi-scale contextual information fusion to effectively extract multi-scale detailed spatio-temporal features, which improves the

interaction capability between contexts. The attention module preserves more historical information by widening the time-receptive domain, which effectively improves the information decay problem. It effectively improves the prediction ability for high echo regions, and the details of the prediction results are also higher.

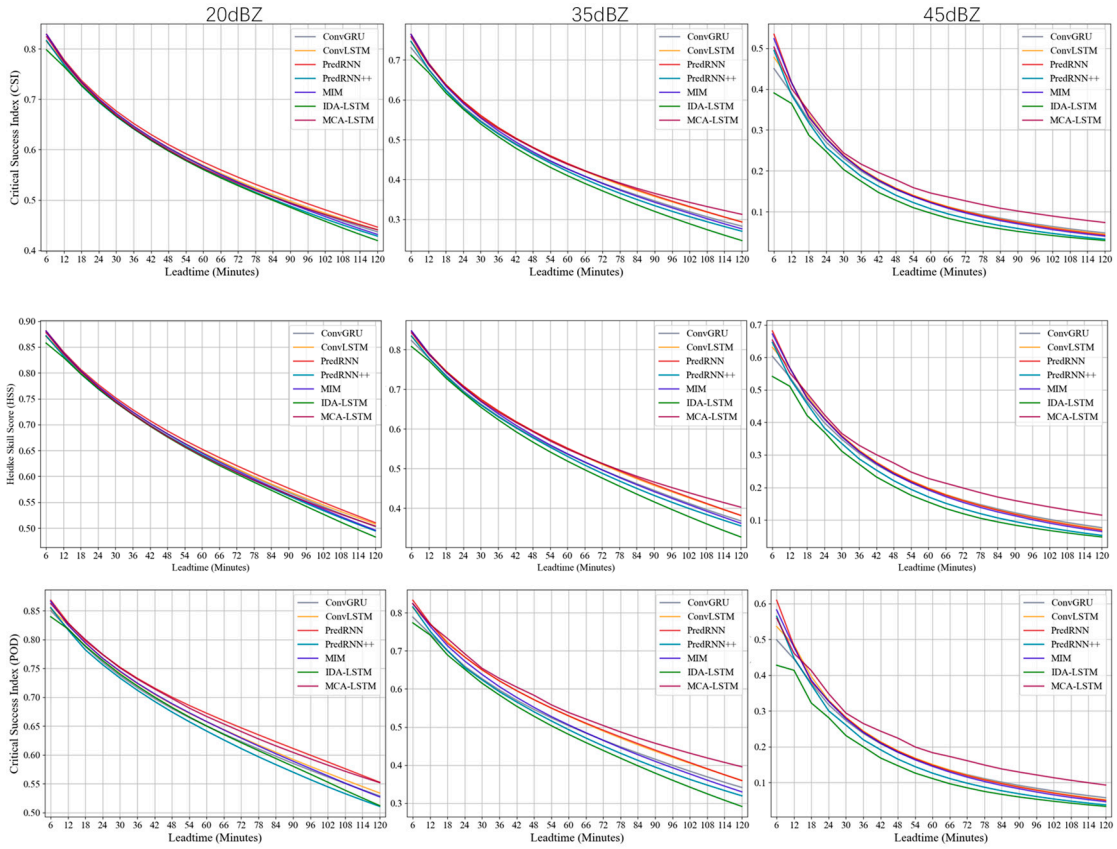



**Figure 6.** CSI, HSS and POD scores of echo forecasts at 20, 35 and 45 dBZ thresholds for each algorithm.

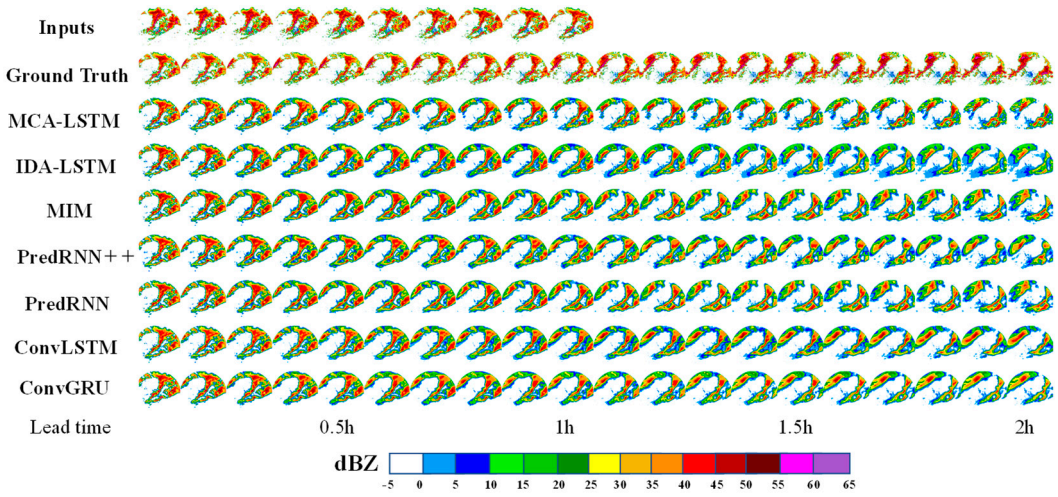



**Figure 7.** The prediction results of all methods are from an example of the radar dataset. The first line is input, and the second line is ground-truth output. The other lines are predictions under different models.

In addition, as seen from the ground truth sequence, the intensity of the high echo value region becomes higher and the location changes with time. the MCA-LSTM model can predict the trend well and the prediction results are more detailed. For other deep

learning models, they cannot predict the high echo region, which gradually blurs or even disappears as the prediction time increases.

## 5. Conclusions

In this paper, we address the issue of blurry distortions in radar echo extrapolation results caused by the ConvRNN-based method, particularly the problem of underestimation in high echo regions. We propose a novel deep-learning-based radar echo image extrapolation model called MCA-LSTM for short-term precipitation forecasting based on weather radar data within the 0-2 h range. Comparative experiments are conducted using the Moving MNIST dataset and Hong Kong meteorological radar data. Through a comparative analysis with existing algorithms, the following conclusions are drawn:

1. The proposed multi-scale context information fusion module effectively enhances the contextual relevance of network units and improves the detail of the predicted images by extracting multi-scale feature information.
2. The proposed attention module captures more historical temporal dynamics from a broader perception field, reducing information loss and enhancing the prediction capability for strong echo regions.

By incorporating these two modules into the ST-LSTM network units, a four-layer radar echo extrapolation network (MCA-LSTM) is constructed. Experimental results on the Moving MNIST dataset and Hong Kong weather radar dataset demonstrate that, compared to recent alternative methods, this approach achieves higher prediction detail and stronger forecasting capability for high echo regions, meeting the requirements for fine-grained predictions in long-term forecasting tasks. The current deep learning algorithm has led to a significant improvement in the extrapolation of radar echoes, but it is still some way from real-life conditions. In subsequent studies, we will investigate how to take more meteorological factors into account in the radar echo extrapolation task and explore more effective algorithms to further improve the prediction capability of short-range precipitation forecasts.

**Author Contributions:** Conceptualization, G.H. and H.Q.; methodology, G.H. and H.Q.; data curation, H.Q. and P.Z.; software, Y.C.; validation, J.W.; formal analysis, J.L.; writing—review and editing, G.H. and H.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research supported by the "The Pearl River Talent Recruitment Program of Guangdong" (2019ZT08G669); in part by the National Natural Science Foundation of China under Grant Nos. 41975183 and 41875184; in part by the funding of Fengyun Application Pioneering Project (FY-APP); in part by the China Meteorological Administration Youth Innovation Team (CMA2023QN10); in part by the S&T Program of Hebei (21567624H).

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was obtained from the Hong Kong Observatory (HKO) and are available from https://github.com/sxjscience/HKO-7 with the permission of the Hong Kong Observatory (HKO).

**Acknowledgments:** The authors would like to thank the Hong Kong Observatory (HKO) for providing public radar dataset HKO-7. We also thank sincerely the editor and anonymous reviewers for their constructive suggestions and improvements to our work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Singh, S.; Sarkar, S.; Mitra, P. A deep learning based approach with adversarial regularization for Doppler weather radar ECHO prediction. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 5205–5208.
2. Marshall, J. The distribution of raindrops with size. *J. Meteor.* **1948**, *5*, 165–166. [CrossRef]
3. Rinehart, R.; Garvey, E. Three-dimensional storm motion detection by conventional weather radar. *Nature* **1978**, *273*, 287–289. [CrossRef]
4. Zou, H.; Wu, S.; Shan, J. A method of radar echo extrapolation based on TREC and Barnes filter. *J. Atmos. Ocean. Technol.* **2019**, *36*, 1713–1727. [CrossRef]

5. Lakshmanan, V.; Hondl, K.; Rabin, R. An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Atmos. Ocean. Technol.* **2009**, *26*, 523–537. [CrossRef]
6. Chung, K.; Yao, I. Improving radar echo Lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases. *Mon. Weather. Rev.* **2020**, *148*, 1099–1120. [CrossRef]
7. Woo, W.; Wong, W. Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere* **2017**, *8*, 48. [CrossRef]
8. Ayzel, G.; Heistermann, M.; Winterrath, T. Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1). *Geosci. Model Dev.* **2019**, *12*, 1387–1402. [CrossRef]
9. Chang, Z.; Zhang, Y.; Wang, S. A Motion-Aware Unit for Video Prediction and Beyond. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 26950–26962.
10. Tamaru, R.; Siritanawan, P.; Kotani, K. Interaction Aware Relational Representations for Video Prediction. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Melbourne, Australia, 17–20 October 2021; pp. 2089–2094.
11. Bei, X.; Yang, Y.; Soatto, S. Learning semantic-aware dynamics for video prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 902–912.
12. Tian, C.; Chan, W. Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dependencies. *IET Intell. Transp. Syst.* **2021**, *15*, 549–561. [CrossRef]
13. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Deep learning on traffic prediction: Methods, analysis and future directions. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 4927–4943. [CrossRef]
14. Zhao, J.; Liu, Z.; Sun, Q.; Li, Q.; Jia, X.; Zhang, R. Attention-based dynamic spatial-temporal graph convolutional networks for traffic speed forecasting. *Expert Syst. Appl.* **2022**, *204*, 117511. [CrossRef]
15. Guo, S.; Xiao, D.; Yuan, X. Short-term rainfall prediction method based on neural network and model ensemble. *Adv. Meteor. Sci. Technol.* **2017**, *7*, 107–113.
16. Huang, J.; Cao, R.; Yao, R. Application of deep learning network in precipitation phase identification and prediction. *Meteor. Mon.* **2021**, *47*, 317–326.
17. Guo, H.; Chen, M.; Han, L.; Zhang, W.; Qing, R.; Song, L. Correlation analysis between vegetation coverage and climate drought conditions in North China during 2001–2013. *J. Geogr. Sci.* **2017**, *27*, 143–160.
18. Chen, J.; Feng, Y.; Meng, W. Research on hourly precipitation forecast correction method based on convolutional neural network. *Meteor. Mon.* **2021**, *47*, 60–70.
19. Li, Y.; Li, Q.; Wei, J. Meteorological radar echo extrapolation based on ConvLSTM. *J. Qinghai Univ.* **2021**, *39*, 93–100.
20. Yin, Q.; Gan, J.; Qi, H.; Hu, W.; Zhang, Y.; Li, R.; Tang, W. An improved recurrent neural network radar image extrapolation algorithm. *Meteor. Sci. Technol.* **2021**, *49*, 18–24.
21. Huang, X.; Ma, Y.; Hu, S. Extrapolation and effect analysis of weather radar echo sequence based on deep learning. *Acta Meteor. Sin.* **2021**, *27*, 817–827.
22. Luo, C.; Li, X.; Ye, Y. A spatiotemporal LSTM model with pseudo flow prediction for precipitation nowcasting. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 843–857. [CrossRef]
23. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature* **2021**, *597*, 672–677. [CrossRef]
24. Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv* **2022**, arXiv:2202.11214.
25. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F. GraphCast: Learning skillful medium-range global weather forecasting. *arXiv* **2022**, arXiv:2212.12794.
26. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast. *arXiv* **2022**, arXiv:2211.02556.
27. Andrychowicz, M.; Espeholt, L.; Li, D.; Merchant, S.; Merose, A.; Zyda, F.; Agrawal, S.; Kalchbrenner, N. Deep Learning for Day Forecasts from Sparse Observations. *arXiv* **2023**, arXiv:2306.06079.
28. Chen, L.; Du, F.; Hu, Y.; Wang, Z.; Wang, F. SwinRDM: Integrate SwinRNN with diffusion model towards high-resolution and high-quality weather forecasting. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 322–330. [CrossRef]
29. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.; Wong, W.; Woo, W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 28–39.
30. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P. Predrnn: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *45*, 2208–2225. [CrossRef] [PubMed]
31. Wang, Y.; Gao, Z.; Long, M. Predrnn++: Towards a resolution of thedeep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132.
32. Wang, Y.; Jiang, L.; Yang, M. Eidetic 3D LSTM: A model for video prediction and beyond. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.

33. Wang, Y.; Zhang, J.; Zhu, H. 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, Long Beach, CA, USA, 15–20 June 2019; pp. 9154–9162.
34. Luo, C.; Zhao, X.; Sun, Y.; Li, X.; Ye, Y. Predrann: The spatiotemporal attention convolution recurrent neural network for precipitation nowcasting. *Knowl.-Based Syst.* **2022**, *239*, 107900. [CrossRef]
35. Yang, Z.; Wu, H.; Liu, Q.; Liu, X.; Zhang, Y.; Cao, X. A self-attention integrated spatiotemporal LSTM approach to edge-radar echo extrapolation in the Internet of Radars. *ISA Trans.* **2023**, *132*, 155–166. [CrossRef]
36. Ma, Z.; Zhang, H.; Liu, J. Preciplstm: A meteorological spatiotemporal lstm for precipitation nowcasting. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4109108. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.