Strong Spatiotemporal Radar Echo Nowcasting Combining 3DCNN and Bi-Directional Convolutional LSTM

Suting Chen [1,2,*], Song Zhang [1,2], Huantong Geng [1], Yaodeng Chen [1], Chuang Zhang [1] and Jinzhong Min [1]

[1]   Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing University of Information Science & Technology, Nanjing 210044, China; 20181219118@nuist.edu.cn (S.Z.); 001692@nuist.edu.cn (H.G.); 001772@nuist.edu.cn (Y.C.); 001981@nuist.edu.cn (C.Z.); 000648@nuist.edu.cn (J.M.)

[2]   Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China

*   Correspondence: sutingchen@nuist.edu.cn; Tel.: +86-139-1386-4015

check for
updates

Abstract: In order to solve the existing problems of easy spatiotemporal information loss and low forecast accuracy in traditional radar echo nowcasting, this paper proposes an encoding-forecasting model (3DCNN-BCLSTM) combining 3DCNN and bi-directional convolutional long short-term memory. The model first constructs dimensions of input data and gets 3D tensor data with spatiotemporal features, extracts local short-term spatiotemporal features of radar echoes through 3D convolution networks, then utilizes constructed bi-directional convolutional LSTM to learn global long-term spatiotemporal feature dependencies, and finally realizes the forecast of echo image changes by forecasting network. This structure can capture the spatiotemporal correlation of radar echoes in continuous motion fully and realize more accurate forecast of moving trend of short-term radar echoes within a region. The samples of radar echo images recorded by Shenzhen and Hong Kong meteorological stations are used for experiments, the results show that the critical success index (CSI) of this proposed model for eight predicted echoes reaches 0.578 when the echo threshold is 10 dBZ, the false alarm ratio (FAR) is 20% lower than convolutional LSTM network (ConvLSTM), and the mean square error (MSE) is 16% lower than the real-time optical flow by variational method (ROVER), which outperforms the current state-of-the-art radar echo nowcasting methods.

Keywords: radar echo nowcasting; 3DCNN; bi-directional convolutional LSTM; spatiotemporal correlation

1. Introduction

Radar echo nowcasting is a crucial method in the field of atmospheric science. The goal of this task is to carry out prediction timely and accurately for weather conditions of local areas in a relatively short period (such as 0–2 h) in the future [1–3]. Currently, this technology has been widely applied to provide flood prevention information for resident trip, agricultural production, flight safety, and other aspects. It is both convenient for people and conducive to disaster prevention and mitigation, and it has always been a pivotal task in weather forecast field. With climate change and the rapid process of urbanization, atmospheric conditions have become more complex, various meteorological phenomena frequently happen, such as precipitation, hail, high temperature, typhoon, etc. Climate change has brought about many adverse impacts on the life and work of people and increased many dangers of uncertainty. If effective forecast and precaution can be made regarding the aforementioned meteorological phenomena, the losses will be reduced dramatically [4]. However, for nowcasting,

the process of future strong convection formation is affected with the current convection situation in the region and previous change trend. Under the influences of these climate factors, it is difficult to determine the shape and size of convection, the distribution of convection presents a complex change trend, which requires prediction model and data with strong spatiotemporal correlation to solve the problem [5,6]. On the other hand, due to the higher requirements for accuracy and timeliness compared with traditional forecast tasks, the work is very challenging and also gradually becomes a hot study topic in the meteorological community.

At present, the regular methods of nowcasting are mainly based on radar echo extrapolation methods [7,8]. The radar echo extrapolation can be divided into traditional methods and deep learning methods. Specifically, traditional methods generally include the centroid tracking method, cross correlation method, and the optical flow-based methods. Initially, a simple tracking algorithm [9] obtains the moving vector by continuously tracking the position of the centroid in an alpine region, so as to predict the position of the echo, but it is only suitable for strong single echo. The TRACE3D algorithm [10] identifies convective cells and tracks them by exclusively using radar reflectivity data as input. This method shows promising preliminary results for centroid tracking. An enhanced centroid tracking method [11] makes least square fitting for the position of the echo centroid at the adjacent moment, and obtains the atmospheric parameters such as the moving vector, the maximum reflection factor and the centroid coordinates of the single echo, and proposes a dynamic constraint-based combinatorial optimization method to track storms. This method is effective for the echo block with larger intensity, when the echo splits or merges, the accuracy of tracking and prediction is low. A cross correlation method based on TREC vectors [12] and an improved cross correlation method [13] are mainly used to calculate out spatial correlation of two consecutive moments and establish fitting relation for echoes, but for the strong convection echoes that evolve fast, these methods cannot ensure the effect of tracking, the data utilization rate and nowcasting accuracy are obviously low. In addition, the optical-flow-based methods in computer visual technology [14,15] has proved to be effective for radar echo extrapolation prediction with fast evolution, especially the real-time optical flow by variational methods for echoes of radar algorithm (ROVER) that was recently proposed [16]. ROVER completes echo prediction by use of the changes of pixels of image sequence in time threshold and calculation of optical flow correlation between adjacent frames, combing with climate and other factors. However, the optical flow estimation step and the radar echo extrapolation step are separated, so it is difficult to decide model parameters to obtain good prediction, which results in a limitation of the optical-flow-based method for nowcasting of strong convection. Moreover, an excellent review of these methods was given by Keenan et al. [17], it presents an overview and comparison of nine existing nowcasting systems deployed in the forecast demonstration project during the 2000 Olympic Games in Sydney, Australia. This fully reflects the practicability of the nowcasting technique in big events. Traditional radar echo nowcasting extrapolation methods only assume simple linear evolution of echo, and the utilization of massive radar echo image data is low, so there are defects in nowcasting accuracy.

Compared with traditional radar echo prediction methods, deep learning is able to better mining and analyze big data in depth and improve prediction performance of models [18], which has been practically applied in many fields. Therefore, the application of this technology in weather nowcasting field is also a meaningful research task. The dynamic convolutional layer network [19] was first proposed and used for short range weather nowcasting, it applies a dynamically varied convolution kernel to extrapolate radar echoes, but it is limited to extrapolating only one echo at a time. For the characteristics of convolutional neural networks such as local perception of images and feature extraction, the recurrent dynamic convolution neural network model (RDCNN) [20] further was established to learn changing features of echoes by adding a dynamic sub network and a probability prediction layer, improved the accuracy of echo extrapolation prediction. In recent years, the recurrent neural network (RNN) [21] and long short-term memory (LSTM) [22] bring some new solutions for radar echo nowcasting task [23,24]. RNN has excellent effect in dealing with time series problems, and as an emerging technology driven by big data, deep learning can make full use of large amount of

collected radar echo data, this will train the network model more effectively and predict the future echo trend more accurately. The unsupervised video representation learning model based on LSTM structure was proposed [25], and by using this encoding-decoding structure, multi-frame actions in the future can be predicted, which has laid a foundation for the spatiotemporal sequence prediction. Subsequently, in order to capture long-term time features more fully, the bidirectional LSTM network with 1D CNN model [26] was constructed to solve precipitation nowcasting problem. The forecast of radar echoes has comparatively strong spatiotemporal correlation, spatiotemporal information at previous moment can decide the prediction of next moment, but general, LSTM does not consider spatial correlation in temporal dimension. Considering the problems of LSTM structure such as containing too much redundant data and easy spatial information loss, a convolutional LSTM network (ConvLSTM) [27] was proposed on this basis, which can learn the spatial features and temporal features at the same time, and it is more suitable to solve the problem of radar echo prediction. Tan et al. [28] proposed a hierarchical convolutional LSTM network named FORECAST-CLSTM. The model is designed to fuse multi-scale features in the hierarchical network structure to predict the pixel value and the morphological movement of the cloudage simultaneously. Thereafter, a ST-LSTM method [29] with convolution calculation and spatiotemporal memory flow was introduced into a radar nowcasting task, which makes it possible to extract spatiotemporal features of echoes in different time and sizes, but, the computational complexity is increased. Furthermore, a 3D convolution method [30] was proposed to capture motional information between consecutive frames, it made convolutional neural networks be suitable for dealing with the information of spatiotemporal features. Compared with the method [30], a 3DCNN video generation model combining generative adversarial networks was proposed [31], this method used 3D convolution network to extract spatiotemporal features efficiently and generated new dynamic echo sequences.

Therefore, this paper proposes a novel 3DCNN-BCLSTM radar echo nowcasting model with encoding-forecasting structure to tackle the challenging task of low forecast accuracy and easy spatiotemporal information loss. Because inputs and outputs are both multi-frame radar echo sequences, the prediction of radar echo evolution trend can be expressed as a video sequence prediction with spatiotemporal features [32]. In order to achieve a more accurate nowcasting result, it first introduces a 3D convolution network that is usually used for feature extraction of continuous video frames. This can preserve the feature information of motion in the temporal dimension and extract local short-term spatiotemporal features of consecutive images more effectively, which then enters the bi-directional convolutional LSTM networks. Its state to state transitions are all convolutional structures, and the bi-directional structure can learn the global long-term motion trend of the front and back echoes more fully, then completes prediction of future echoes through forecasting network. Finally, we evaluate and compare it with traditional extrapolation algorithms and other deep learning algorithms, the experiment fully proves that the comprehensive evaluation of the improved deep learning model proposed in this paper is always better than other compared models.

2. 3DCNN-BCLSTM Model

In order to further improve nowcasting accuracy and make better use of spatiotemporal correlation between radar echo images, this paper proposes a encoding-forecasting structure combining 3DCNN and bi-directional convolutional LSTM according to the multiple deep learning technologies, this can capture spatiotemporal feature relation of consecutive radar echoes more effectively and enhance transmission ability between spatiotemporal features, the specific model architecture is shown in Figure 1.
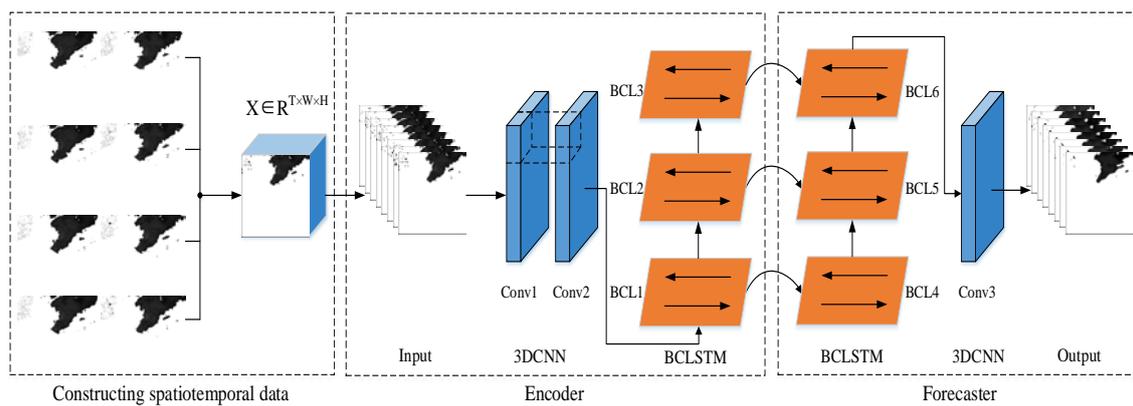
Figure 1. Model structure of 3DCNN-BCLSTM.

First of all, the consecutive radar image sequences are constructed as model input with uniformly spatial and temporal dimensions, for this treatment of data dimension, tensors with complete spatiotemporal features can be obtained. In terms of main structure, a generative model of encoding-forecasting structure is established which is mainly consisted of two networks—one is encoding network and the other one is forecasting network. Second, this paper extracts local short-term spatiotemporal features of consecutive multi-frame images through 3DCNN, then learns dependencies of global long-term bi-directional spatiotemporal features through three-layer bi-directional convolutional LSTM networks, and compresses captured and learned echo motion features into hidden state tensors (the former part is the encoding network of model). After that, the forecasting network is composed of three-layer bidirectional convolutional LSTM connected with the internal states of the encoding network and the last layer of 3DCNN, which is used to fuse the multi-frame spatiotemporal states, the spatiotemporal feature information learned by the encoding network is transmitted into the forecasting network, the future echo image sequences are reconstructed according to the current input and feature information. In addition, the batch normalization (BN) method [33] is introduced, and the rectified linear unit (ReLU) as nonlinear activation function is used to replace the traditional Sigmoid to improve network convergence speed and alleviate the over fitting phenomenon. This deep learning structure can obviously enhance learning capability of model, and the model possesses stronger expression capability of spatiotemporal features for multi-frame radar echo images; therefore, the prediction accuracy is improved effectively.

2.1. Construction of 3D Spatiotemporal Data

In terms of radar echo prediction problem, the original input data dimensions can no longer meet the requirements of network model, its main disadvantage is that the convective spatiotemporal feature information cannot be encoded completely. In order to solve this problem, all input, unit output and cell states need to be transformed to 3D tensors $X \in R^{T \times W \times H}$, where R denotes the domain of atmospheric data features. The first dimension T is temporal dimension, the second dimension W, and the third dimension H are spatial dimensions of row and column, respectively. In fact, the 3D spatiotemporal data is different to the use of the volumetric data of the weather radar. As showed in Figure 2, the original single echo image has been transformed to vectors of multi-frame temporal dimension in spatial grid, a 3D spatiotemporal stereostructure is generated by stacking consecutive images in turn, then the neural networks may predict future states of unit in grid through local adjacent information and past states.
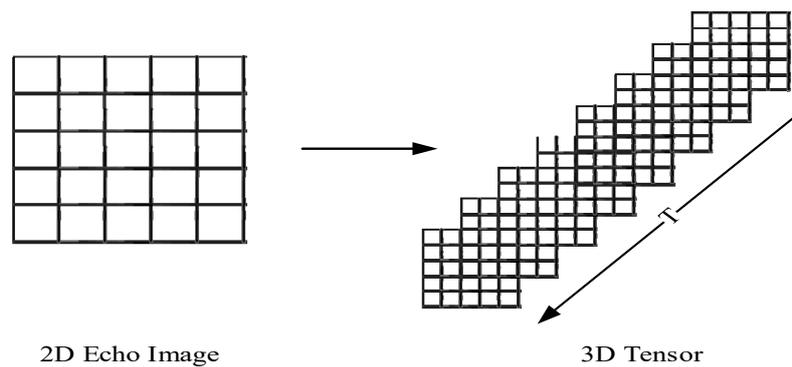
**Figure 2.** Dimension construction of radar echo data.

For the 3DCNN-BCLSTM network structure, input data dimensions of echo images need to be restructured, the temporal dimension and spatial dimension are constructed respectively. During the process of spatiotemporal feature extraction and motional information learning, input and output are both 3D tensors, the transitions between states are also convolution calculation of 3D tensors, which makes the radar echo data have a unified dimension, preserves all spatial and temporal features at the same time, and the radar echo nowcasting in the region is more comprehensive and accurate.

2.2. 3DCNN Module

The convolutional neural networks are very suitable for image data processing due to its local connection, feature mapping and weight sharing. Even though traditional 2DCNN possesses strong feature extraction capability of image data, when it deals with consecutive echo image tasks, it fails to consider the impact of relation between multi-frame images on prediction, and is easy to lose motion trend information of target features, thus cannot solve the problem of motional echo prediction effectively. We utilize constructed 3DCNN instead of traditional 2DCNN for more accurate results. The calculation formula for 3DCNN is showed as follows:

$$
v_{ij}^{WHT} = f\left(b_{ij} + \sum_{m} \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(W+p)(H+q)(T+r)}\right) \tag{1}
$$

There are multiple convolution kernels in the convolution layers of the neural networks, each convolution kernel corresponds one echo feature, the more convolution kernels, the more feature maps are generated. In the formula, the value at position (W, H, T) on the jth feature map in the ith layer is given by v_{ij}^{WHT}, R_i is the size of the 3D kernel along the temporal dimension. w_{ijm}^{pqr} is the (p, q, r)th value of the kernel connected to the mth feature map in the previous layer. b_{ij} is the bias for this feature map, f is a nonlinear activation function introduced to improve the expression capability of neural networks. This 3DCNN structure can preserve more information of continuous multi-frame images and can be used for meteorological nowcasting tasks effectively. In the process of dimension reconstruction of the input radar echo images, several consecutive frames of uniform spatial size are stacked in time order to form 3D data with spatiotemporal features. Then, as shown in Figure 3, the 3D convolution kernel is used for operation in this continuous 3D data, the 3D convolution kernel in the figure contains three frames of temporal dimension, that is, the convolution operation for three consecutive maps are required. The feature data extracted by 3DCNN in the last layer of the encoding network will be transmitted to the next network as input [30]. In this structure, every feature map in convolution layer will be connected to several consecutive frames in the previous layer, and the specific value of each position of feature map is obtained through local feeling of successive multiple same positions in the previous layer, thereby captures spatiotemporal motional information of echo images.

Temporal Dimension

Temporal Dimension

Layer i Feature Map   Layer i+1 Feature Map
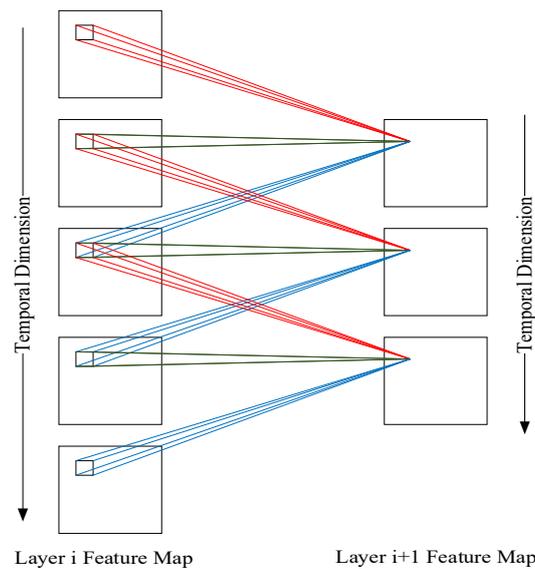
Figure 3. Illustration of 3D convolution to extract spatiotemporal features.

In the encoding network part of the radar echo extrapolation model, we improve the problems of multi-frame images that are difficult to deal with, and the spatiotemporal information is easily lost. The input of the network is composed of consecutive image sequences, and then successively enters to Conv1 and Conv2 for short-term feature extraction. This part is mainly composed of two Conv3D layers, each Conv3D layer is followed by a batch normalization (BN) and a ReLU nonlinear activation function layer. The convolution kernels of two-layer Conv3D are small size $3 \times 3 \times 3$, the number of filters is 16 and 32, respectively, and each 3D convolution kernel has the same weight coefficient. In order to keep the size of feature maps constant, the padding operation is carried out before convolution operation. In order to accelerate the deep learning network training and effectively avoid the related gradient problems, we increase the BN after each 3D convolution layer [33] and normalize the data distribution of each batch in the network calculation process. The derivative range of the traditional activation function is less than 1, and the gradient will be continuously attenuated when passing through each layer, with the deepening of the network structure, the gradient may disappear. Thus, the ReLU activation function is selected to replace the traditional Sigmoid activation function. The formula is as follows:

$$\text{ReLU}(x) = \max(0, x) \tag{2}$$

When the input x is less than 0, the mandatory output will be 0; when input x is larger than 0, it is constant. ReLU increases sparsity of networks and makes convergence rate grow, then the generalization capability of the feature extraction is stronger, the over fitting phenomenon is alleviated, and the accuracy is improved to a certain extent. 3DCNN module uses two shallow layers here, this is to capture spatiotemporal features of images more effectively by combining bi-directional convolutional LSTM layers afterward; this reduces feature loss and accelerates convergence speed of neural networks.

A 3DCNN network is also used in the forecasting network part, followed by a ReLU nonlinear activation function layer. The number of filters is set to 1 here, so that the model can finally generate the gray images with the same channel number as the original input and outputs the visualization results.

2.3. Bi-Directional Convolutional LSTM Module

Recurrent neural network (RNN) can handle the time series problem of meteorological forecast, long short-term memory (LSTM) is a special structure based on RNN, this network structure is used for learning the changes with temporal sequence factor. In recent years, LSTM is frequently used in

fields such as natural language processing (NLP), and in this paper, we try to learn spatiotemporal dependencies of consecutive echo images through improved LSTM structure.

As a special variant of a recurrent neural network, the innovation of LSTM is the memory units whose essence is the place for continuous update and interaction of information. However, the traditional recurrent update structure cannot either realize update and filter of information or meet long distance dependency of information; therefore, the three gates structure is introduced to fulfill those requirements. LSTM relies on memory units to update continuously state information of current moment uses forget gate, input gate, and output gate to decide what information to forget, what information to input, and what information to output. LSTM network solves long-term dependency problem of RNN and extends extrapolation timeliness, which makes input sequences effectively map to hidden nodes, and can learn the relation between the front and the back of the long time series through training. The LSTM structure possesses strong capability of solving time series problems; however, for the processing of spatial data, it contains too much redundant information. Spatiotemporal information cannot be encoded; if it is directly applied to radar echo nowcasting, the loss of spatiotemporal information will be inevitable. A convolutional LSTM [27] was proposed whose structure is still LSTM in essence, but the transitions between states is changed from multiplication to convolution. It establishes a time series relation like LSTM does and also depicts spatial features like CNN does, effectively overcoming the problem of spatial information loss in sequence transmission process. Based on this structure, this paper constructs bi-directional convolutional LSTM, the structure is showed in Figure 4.
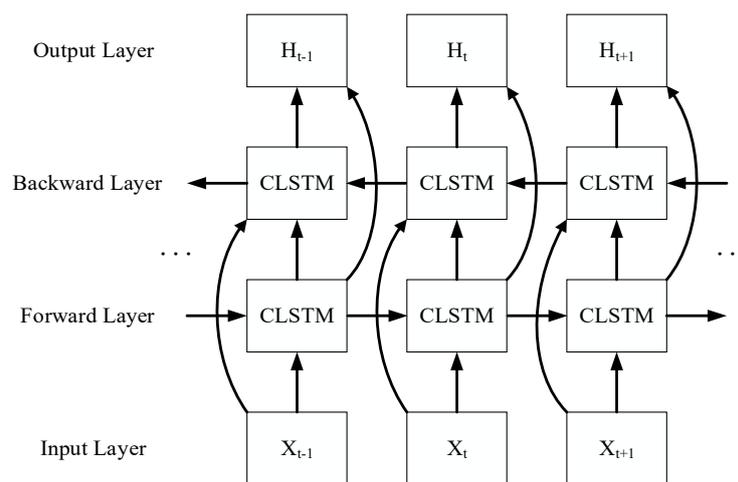
Figure 4. Structure diagram of bi-directional convolutional long short-term memory (LSTM).

A bi-directional convolutional LSTM network is composed of one forward transmission and one backward transmission. This network comprehensively combines the forward and backward information, outputs the radar echo results, and solves the problem that single directional transmission cannot handle the information from the back to the front. In the network, each bi-directional convolutional LSTM memory unit contains the spatial and temporal output from 3D convolution network, the calculation process of each part in the structure is as follows:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{4}$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \tag{5}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \tag{6}$$

$$H_t = o_t \circ \tanh(C_t) \tag{7}$$

where X_t is the input of current moment; H_{t-1} is the output of t-1 moment; f_t, i_t, and O_t denote forget gate, input gate, and output gate in CLSTM respectively; and W and b are connection weight and bias of gate structure. Let the convolution operation $*$ replace original multiplication of LSTM, and let $\circ$ denote the Hadamard product, which is the multiplication of corresponding elements of matrix. The nonlinear activation function σ used here is Sigmoid, with the formula $S(x) = [1 + \exp(-x)]^{-1}$, and the value range of three gates is controlled to [0,1]. C_t is state update unit which is the core part of bi-directional convolutional LSTM.

The 3DCNN-BCLSTM model proposed in this paper, three-layer bi-directional convolutional LSTM is placed at encoding network, three layers are in prediction network, and the number of filters in two parts is both 32, 48, 64 with size of convolution kernel 3x3. In the bidirectional convolutional LSTM, the padding operation is also performed in order to make the size of spatiotemporal features unified, and each layer is followed by a layer of Batch Normalization. The spatiotemporal information of continuous multi-frame image sequences is transmitted by bi-directional convolutional LSTM, which has been effectively fused in the global long-term range. Compared with single direction convolution LSTM, bi-directional convolutional LSTM can learn the global long-term feature dependencies in the forward and reverse directions and completes the nowcasting task more efficiently.

2.4. EncodingFforecasting Network Structure

For the radar spatiotemporal sequence nowcasting, when there is a set of 3D tensor sequence data $\{X_1, \ldots, X_t\}$, given the previously fixed length of L observation sequence data, the radar echo image sequences with the future length of K $\{\tilde{y}_{t+1}, \ldots, \tilde{y}_{t+K}\}$ can be generated through the encoding-forecasting network structure, where the t denotes current moment, and $\tilde{y}$ represents the prediction output, as shown in Equation (8), taking the past prediction echoes as the condition, $\underset{X_{t+1}, \ldots, X_{t+K}}{\text{argmax}}$ represents the maximum probability to make the prediction of future moment as close to reality as possible.

$$\tilde{y}_{t+1}, \ldots, \tilde{y}_{t+K} = \underset{X_{t+1}, \ldots, X_{t+K}}{\text{argmax}} P(X_{t+1}, \ldots, X_{t+K} | \tilde{y}_{t-L+1}, \tilde{y}_{t-L+2}, \ldots, \tilde{y}_t) \tag{8}$$

The generative model of encoding-forecasting network structure showed in Figure 5 is mainly used in this paper, it is composed of encoding network and forecasting network [25]. The network combines encoding network of stacked two-layer 3DCNN and three-layer BCLSTM, and the forecasting network of three-layer BCLSTM and one-layer 3DCNN, which receives internal state of encoding network. This structure compresses the captured feature information of motional echoes into hidden tensor format by the encoding network and then the forecasting network will unfold hidden state tensors and generate new radar echo prediction results based on the feature information of last moment. The network is as follows.
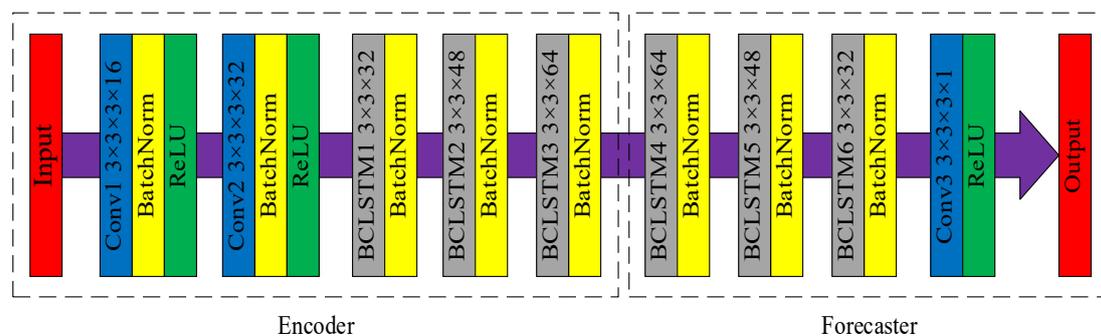

Figure 5. Network structure of encoding-forecasting.

Step 1: Constructing input data with spatiotemporal features

When data is input, radar echo images in the dataset need to be narrowed to single channel gray images with 100x100 pixel spatial dimension, then images are transformed to array format and save it in numpy array to wait for extraction and use. For pre-processing process, temporal dimension of data also needs to be constructed, which constitutes eight consecutive frames of images input and predicts right frames of images in the future. Thus, radar data is transformed to 3D tensors with spatiotemporal features to facilitate model inputting and training.

Step 2: Extracting local short-term spatiotemporal feature information

Consecutive radar echo images with uniformly spatial and temporal dimensions after dimension construction processing as a whole are the network input. Thirty-two echo feature maps are extracted through two layers of 3DCNN, and the ReLU activation function is used to replace the original Sigmoid to alleviate over fitting phenomenon and effectively increase prediction accuracy.

Step 3: Learning global long-term spatiotemporal feature information

Then, the global long-term spatiotemporal correlation from delivered feature information is learned through three-layer BCLSTM, the learned spatiotemporal features are compressed into hidden state tensors. Up to this point, it is the encoding network of first half of whole generative network, the forward and backward structure in this part can learn bi-directional spatiotemporal feature dependency fully.

Step 4: Reconstructing and generating predicted radar images

Finally, a forecasting network composed of three-layer BCLSTM and the last layer 3DCNN is constructed, the atmospheric spatiotemporal feature information learned by the encoding network is transmitted into the forecasting network, and the future prediction images are generated according to the current input and hidden states.

3. Experiment and Evaluation Analysis

3.1. Dataset

The radar echo nowcasting task is the prediction of the evolution of multi-frame radar echo images in the future using the multi-frame radar echo images in the previous moment. In order to verify the effectiveness of this method, we use Standardized Radar Dataset 2018 (SRAD2018) [34], which was established by the Shenzhen Meteorological Bureau and the Hong Kong Observatory based on radar data from Guangdong, Hong Kong, and Macao in recent years as experimental data. Quality controlled radar echo gray images from March to July every year in which seasonal strong convection happens frequently is selected as the data in dataset. The range control of data in this figure is 0-80dBZ, the resolution of provided radar echo images is 501x501 pixel, the images are collected at 3000 meters above sea level, covering an area of 500 km x 500 km. Data is obtained every six minutes from meteorological radar, and every record is a frame and named after original data sequence to facilitate experimental indexing in the future.

Here, 6283 pieces of radar echo images are adopted as training samples in this experiment—1775 as validation set and 1775 as testing set (no crossing and overlapping in data). In order to accelerate training speed and improve training effect, the original radar echo images whose resolution is 501×501 pixel are not suitable to directly input model to carry out training, it needs to be compressed to 100×100 pixel gray images and inputs constructed spatiotemporal sequences into model.

3.2. Network Training

The whole model structure is showed in Figure 1. We realize the proposed algorithm network with the TensorFlow framework, and a NVIDIA Tesla V100 GPU (NVIDIA, Santa Clara, CA, US) is used to accelerate various experimental models training. In this paper, the radar echo encoding-forecasting model is verified by multiple trainings, the adaptive learning rate Adadelta algorithm [35] is used to

optimize the loss function, the attenuation coefficient is 0.95 in the specific parameters, and the batch size is set to 4. After 5000 iterations on the basis of GPU acceleration, the network has converged well.

3.3. Experimental Quantitative Analysis

In order to verify the effectiveness of this 3DCNN-BCLSTM radar echo nowcasting model, pixel-level mean square error (MSE), the number of network parameters, critical success index (CSI), probability of detection (POD), and false alarm ratio (FAR) are commonly used by the meteorological community [36]. These measures are similar to the commonly used concepts of accuracy, recall, and precision in deep learning. The MSE measures the mean square error of every location pixel between actual radar echoes and prediction. Before calculation, the pixel value of each location needs to be transformed to rainfall intensity value based on Z-R relation [37], and then it is used to calculate error loss. In terms of CSI, POD, and FAR evaluation indexes, the regression problem of radar echo prediction needs to be transformed into a 0/1 classification problem for calculation. In this paper, the echo thresholds of 10dBZ, 20dBZ, and 40dBZ are used to distinguish whether it is a positive case or a negative case (a positive case is greater than the threshold value while a negative case is less than the threshold value). Then, the prediction and the fact are transformed into a 0/1 matrix by combining the threshold value, and calculate, respectively, the number of TP (prediction = 1, truth = 1), FN (prediction = 0, truth = 1), FP (prediction = 1, truth = 0), and TN (prediction = 0, truth = 0). Calculation formulas for mean evaluation indexes are as follows:

$$CSI = \frac{TP}{TP + FN + FP} \tag{9}$$

$$POD = \frac{TP}{TP + FN} \tag{10}$$

$$FAR = \frac{FP}{TP + FP} \tag{11}$$

Considering the fact that the frequencies of different rainfall levels are highly imbalanced, the weight w in (12) is added to the MSE loss function to alleviate this problem, where $R(y)$ denotes rainfall intensity, the higher rain rate are multiplied by a higher weight, and the weight of masked pixels is set to 0 [2].

$$w = \begin{cases} 1 & R(y) < 2 \\ 2 & 2 \le R(y) < 5 \\ 5 & 5 \le R(y) < 10 \\ 10 & 10 \le R(y) < 30 \\ 30 & R(y) \ge 30 \end{cases} \tag{12}$$

As shown in Equation (13), the pixel-level mean square error (MSE) of the radar echo is constructed as the loss function of the model to measure the similarity between the predicted results and the actual results. In the formula, w is the weight, y is the actual output, $\tilde{y}$ is the predicted output, N means the total number of current output frames, and W and H represent respectively horizontal and vertical coordinate of radar echo images.

$$L(y, \tilde{y}) = \frac{1}{N} \sum_{n=1}^{N} \sum_{W=1}^{100} \sum_{H=1}^{100} w_{n,W,H} (y_{n,W,H} - \tilde{y}_{n,W,H})^2 \tag{13}$$

Input is the sample of previous moment while actual output is sample of next moment, when current moment is t, and the eight frames input of data are $\{X_{t-7}, X_{t-6}, \ldots, X_t\}$, then the 8 frames of radar echo image output in the future which can be predicted are $\{\tilde{y}_{t+1}, \tilde{y}_{t+2}, \ldots, \tilde{y}_{t+8}\}$. Radar echo images are continuously input the model for training, the deviation of the predicted results and the actual results is calculated, and the network weights and other parameters are updated constantly by back propagation, the loss function value can be constantly reduced, and repeats iterations until

convergence [18], so that the reconstructed echo image sequences are more and more like the real image sequences. This defined similarity loss function improves the feature expression ability of generated images.

3.3.1. Evaluation Analysis of Convolution Kernel Size

During the experiments, the size of the convolution kernel in the convolution layer is an important factor impacting accuracy of echo prediction. In this paper, we select 3, 5, 7, and 9 specifications of convolution kernel as experimental parameters to study the impact of the convolution kernel size for prediction accuracy. The provided layers of network, ReLU activation function and other parameters remain a constant. The mean square error (MSE) for similarity is used to test the accuracy of radar echo extrapolation, and the number of parameters is used to represent the computational complexity of network space, the evaluation result is showed in Table 1.

Table 1. Performance comparison among different convolution kernel sizes. Each number in bracket represents corresponding size of convolution kernel in each network layer. M denotes million.

Models	Number of Parameters	Mean Square Error
Proposed (3-3-3-3-3-3-3-3-3)	3.08M	1.398
Proposed (5-5-5-5-5-5-5-5-5)	8.58M	1.487
Proposed (7-7-7-7-7-7-7-7-7)	16.87M	1.590
Proposed (9-9-9-9-9-9-9-9-9)	27.97M	1.735
Proposed (3-3-5-7-9-9-7-5-3)	21.27M	1.554
Proposed (9-9-7-5-3-3-5-7-9)	7.23M	1.572

We can see from Table 1 that with the increase of size of convolution kernel, the network parameter value grows constantly, when convolution kernel is 9 in all layers. The network parameters have already reached 27.97 million, which reflects that the computational complexity of the neural network is the largest in this case, and the operation time will be longest. Moreover, MSE loss increases gradually as well, and the loss of last two models is also lager than the convolution kernel in small size 3. The convolution kernel sizes are all 3 for best results, and the mean square error is only 1.398, which means that the smaller the loss, the higher the similarity between the predicted images and the real images. Plenty of network parameters cause dramatic increase of computational complexity while the stacking of multiple small convolution kernels improves accuracy of prediction. In addition, we have not considered the case of 1×1×1 (3DCNN) or 1×1 (BCLSTM) because it cannot effectively enhance the receptive field and capture the spatiotemporal features of radar echo prediction. Combining the size of radar echo images at the same time, we use small convolution kernel in size 3 as experimental parameter in this paper.

3.3.2. Evaluation Analysis of Network Layers

The prediction accuracy of deep learning algorithms also depends on setting of layers in neural networks. In this paper, one layer of BCLSTM is used for encoding network and one layer of BCLSTM is used for forecasting network firstly, and then three different network layers of two layers, three layers and four layers are also respectively used to test the impact on the prediction accuracy. The number of filters is 32, 48, 64, and 128, respectively, which correspond to the number of network layers from small to large. Other network layers remain a constant, the size of convolution kernel is 3, and the test results are shown in Table 2.

Table 2. Performance comparison of different network layers.

Models	Number of Parameters	Mean Square Error
Proposed (3-3-3-3-3)	0.39M	1.545
Proposed (3-3-3-3-3-3)	1.35M	1.440
Proposed (3-3-3-3-3-3-3-3)	3.08M	1.398
Proposed (3-3-3-3-3-3-3-3-3-3)	9.57M	1.376

We can see from Table 2 that the changes between whole performance test results in two evaluation indexes is not very large. With the increase of network layers, the parameters of the network model increase constantly, and the computational complexity also increases constantly, especially slow increase before three layers. MSE reduces gradually, and the experiment shows that network model with deeper layers provides better prediction effect. However, after three layers of bi-directional convolutional LSTM in the encoding and forecasting network, the number of parameters still increases but the changes of error loss slow down a lot, with only 0.022 reductions in MSE. In this algorithm network, extrapolation accuracy between two layers and three layers makes progress, while there is no obvious difference between three layers and four layers. Setting three layers bi-directional convolutional LSTM network respectively is suitable, and it does not consume much memory and simultaneously keeps a decent prediction accuracy.

3.3.3. Evaluation Analysis of Performances of Various Models

In the evaluation analysis of performance of various models, we select six different radar echo nowcasting methods to conduct comparative experiments, including FC-LSTM, the optical flow-based method (ROVER), 3DCNN, ConvLSTM, and the 3DCNN-BCLSTM model (which respectively uses Sigmoid activation function and ReLU activation function) proposed in this paper. During the training process of the meteorological nowcasting models, we select the best experimental effect through multiple adjustments on each parameter for ROVER. In addition, the experiment selects MSE with good performance as a loss function and Adam as an optimization function of deep learning model besides the model newly put forward in the paper. The leaning rate is set to 0.001, and the batch size is 4. ConvLSTM uses encoding-forecasting structure of three-layer encoding network and three-layer prediction network with small convolution kernel, the number of filters in two sub networks are 32, 64, and 128. Better situations are also selected for other experimental parameters and the general evaluation results are shown in Tables 3–5 and Figure 6.

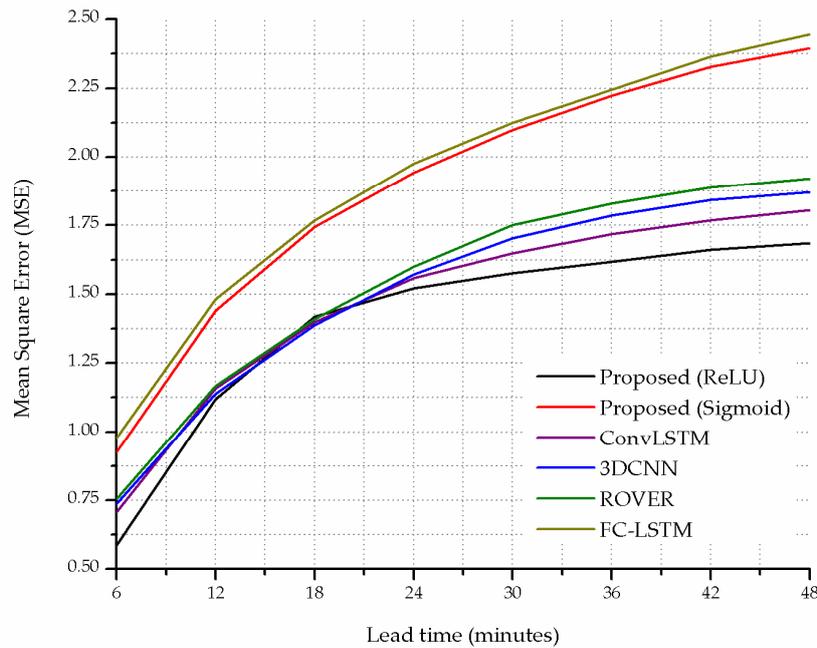**Table 3.** Performance comparison of different models of the 10dBZ threshold.

Models	CSI	POD	FAR
FC-LSTM	0.305	0.376	0.368
ROVER	0.480	0.624	0.312
3DCNN	0.512	0.641	0.280
ConvLSTM	0.547	0.659	0.239
Proposed (Sigmoid)	0.325	0.395	0.379
Proposed (ReLU)	0.578	0.673	0.192

Table 4. Performance comparison of different models of the 20dBZ threshold.

Models	CSI	POD	FAR
FC-LSTM	0.178	0.235	0.556
ROVER	0.302	0.423	0.480
3DCNN	0.337	0.442	0.421
ConvLSTM	0.361	0.476	0.395
Proposed (Sigmoid)	0.173	0.241	0.591
Proposed (ReLU)	0.375	0.489	0.382

**Table 5.** Performance comparison of different models of the 40dBZ threshold.

Models	CSI	POD	FAR
FC-LSTM	0.032	0.042	0.836
ROVER	0.047	0.064	0.805
3DCNN	0.055	0.082	0.812
ConvLSTM	0.069	0.101	0.781
Proposed (Sigmoid)	0.034	0.045	0.828
Proposed (ReLU)	0.084	0.121	0.736

Figure 6. Mean square error comparison of different models.

In the field of atmospheric science, the prediction performance of different rainfall conditions is different, this section studies the accuracy of various radar echo nowcasting methods under different thresholds. Divided into three rainfall threshold conditions of 10dBZ, 20dBZ, and 40dBZ, the following tables compare CSI, POD, and FAR evaluation indexes in each model based on the situation described in Figure 7, in which the second work is traditional extrapolation prediction method based on optical flow and other five methods are deep learning methods. In general, the proposed model (ReLU) is always better than other compared models under various thresholds. Compared with 10dBZ and 20dBZ, the performance of each model is not good enough at 40dBZ. We think that heavy rain data is rare, and these events are considered mistakenly as outliers by the network models. Although advanced extrapolation results for our model have been achieved so far, the indexes need to be further improved in heavy rainfall or severe weather. Under all rainfall thresholds, the overall performance ranking of each model is basically maintained. Specifically, it can be found from experiment results that deep learning models besides of FC-LSTM and our model (Sigmoid) are generally better than the traditional method based on optical flow. The FC-LSTM model seems to not be suitable for echo nowcasting task, as each meteorological evaluation index is very poor. Multi-frame radar echoes possess very strong spatiotemporal correlation, and this fully connected structure is easy to cause large amount of redundant atmospheric information, so this common LSTM structure cannot accurately predict the motion trend of echoes. Besides, the performance of various evaluation indexes in model proposed in this paper (Sigmoid) is also not good—FAR reaches 0.379 at 10dBZ. It reflects that there are many wrong rainfall predictions in this method, during the experiments, gradient disappearance phenomenon is easy to happen with bad convergence, therefore the predicted echo images are usually

distorted and vague. Among deep learning models, 3DCNN-BCLSTM model (ReLU) performs best, its CSI and POD evaluation indexes reach 0.578 and 0.673, respectively at 10dBZ, both higher than other models (higher is better), and FAR is the lowest. This is because 3DCNN can better extract the spatiotemporal features of echoes, and the special bidirectional structure is more stable under the complex atmospheric conditions. The ConvLSTM model is better than the 3DCNN model, which confirms that the convolutional LSTM structure is often more powerful in time series. The general evaluation of traditional ROVER is worse than other three deep learning methods, mainly because ROVER algorithm is difficult to deal with atmospheric boundary conditions and update the future flow fields end-to-end.
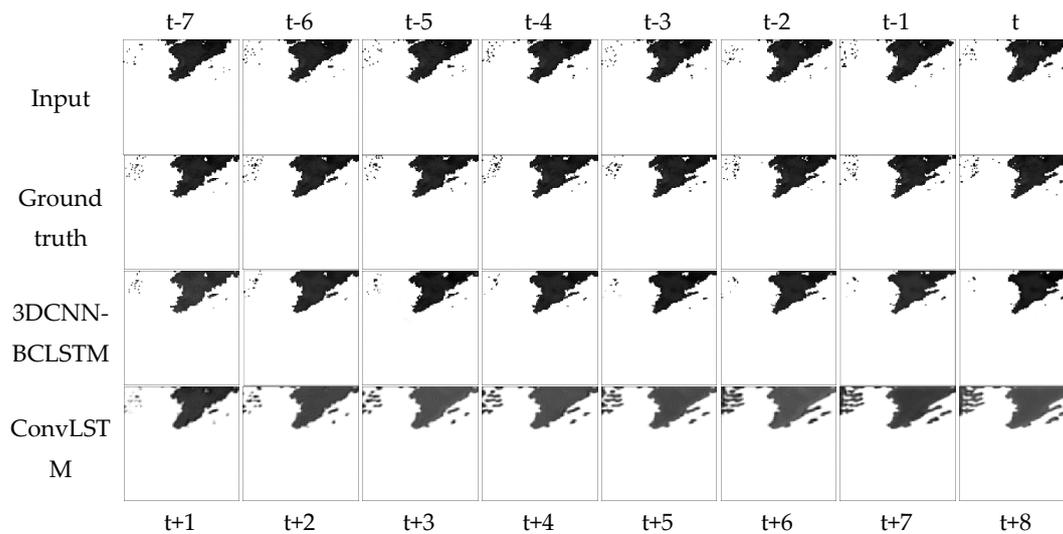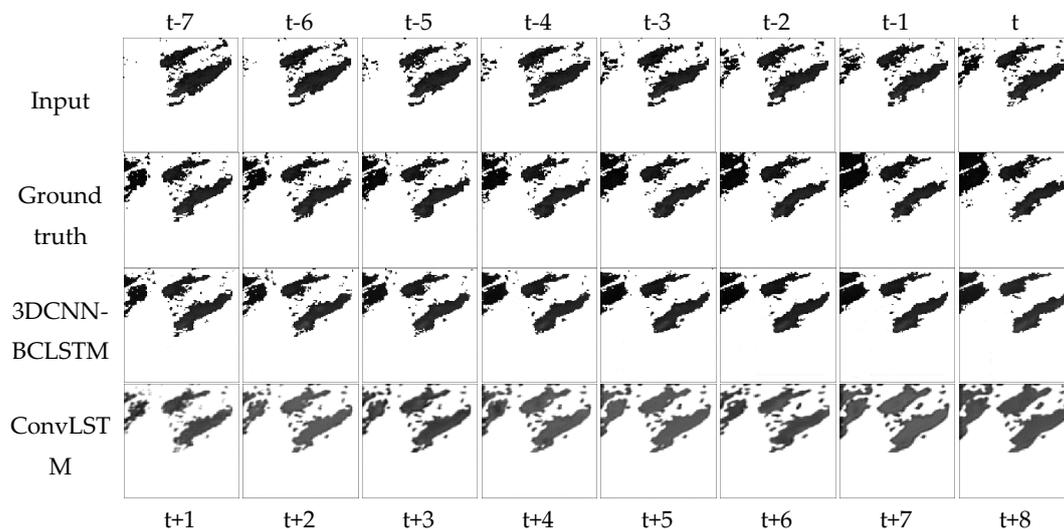

Figure 7. The prediction example of advection motion for radar echo nowcasting at 15:06 UTC on 16 July 2015. From top to bottom: input, ground truth, prediction of 3DCNN-BCLSTM model, and prediction of the ConvLSTM model.

To view the results more clearly, the comparison of the mean squared errors of different models for the radar echo nowcasting task is described in Figure 6. The result shows that the performance of the proposed model (Sigmoid) and FC-LSTM is similar, and their errors are obviously larger than other models. At the initial stage, the mean square errors of other models are similar and alternate. Over time, the performance of various algorithms reduces, but the proposed (ReLU) still significantly outperforms other models—this means the model is more tolerant to high and increasing meteorological uncertainty, and it possesses the higher transmission ability of spatiotemporal features by constructing two spatiotemporal convolutional structures with 3DCNN and bidirectional convolutional LSTM. The test result of ConvLSTM model is close to that of model proposed (ReLU) in this paper, ConvLSTM is the second-best performing model, and the 3DCNN model follows closely. Besides, it can be obviously seen that the error loss of the FC-LSTM model is the largest, which makes the echo prediction accuracy comparatively low. This shows that 3DCNN possesses a stronger capability to extract spatiotemporal features than FC-LSTM. The spatiotemporal convolutional structure substantially improves MSE and extends radar echo extrapolation timeliness.

3.4. Experimental Qualitative Analysis

We compare 3DCNN-BCLSTM model (ReLU) with the second best ConvLSTM model, the extrapolation results are showed in Figures 7 and 8. During nowcasting, the continuous 8 frames of radar echo images (from t-7 to t) are used as the input of the model, and the radar images of the next 8 frames (from t+1 to t+8) are output. On the whole, ConvLSTM in the figures can predict the general motion trend of radar echoes, but compared with actual images, the detailed part is not precise enough, there are more error predictions in echo shape and intensity. The prediction result of the

model proposed in this paper is more accurate, it not only roughly predicts along motion direction, but also effectively predicts boundary conditions and detailed changes of inner part of convection. In Figure 7, it can be seen that the main echo moves slowly towards right direction and merges a generated small echo on the right gradually. Following the moving direction, the rainfall rate has no obvious change. It can also be noticed that the echo predicted by ConvLSTM is smoother and many details are lost, because the transmission capability of spatiotemporal features is poor for this structure, the modeling ability is limited, some echo features are lost in the learning process. In Figure 8, there is almost no echo on the left side in the first few frames of input, then the left echo enhances obviously and constantly. ConvLSTM is vague in every time period, whereas the proposed model can predict the future echoes well under the formation situation of complex direction change, the detailed part is more in line with the actual status and the images are clearer and more stable. We argue that because our model first constructs 3DCNN to extract spatiotemporal features of radar echoes, avoids the confusion of spatial features caused by directly using ConvLSTM for learning, and the bi-directional convolutional LSTM structure is more stable in the learning process of echo features. As explained in Section 3.3.3, over time, the uncertainty increases, the accuracy of the two methods is slightly reduced in the figures, but 3DCNN-BCLSTM (ReLU) is still closer to reality than ConvLSTM.

Figure 8. The same that Figure 7, but for the case of formation for radar echo nowcasting at 21:18 UTC on 10 August 2016. From top to bottom: input, ground truth, prediction of 3DCNN-BCLSTM model, prediction of ConvLSTM model.

The proposed model realizes accurate prediction of future short-term echo images. As expected, 3DCNN possesses strong capability to extract spatiotemporal features of multi-frame echo images and bi-directional convolutional LSTM can learn more comprehensive spatiotemporal correlation compared with single direction network, this improves the problem of vague prediction images. And through the combination of 3DCNN and bi-directional convolutional LSTM, enhances transmission capability of spatiotemporal features, avoids the loss of radar echo information, so it improves predictive skills.

4. Conclusions

Utilizing and mining massive radar echo data is low in traditional radar echo nowcasting, and the meteorological process of future echo formation is affected with the current echo situation in the region and previous change trend, which possesses strong spatiotemporal correlation. In this paper, a novel deep learning model of 3DCNN-BCLSTM encoding-forecasting structure is proposed and applied to radar echo nowcasting task. This model captures and learns spatiotemporal feature dependencies of consecutive radar echo images more effectively by utilizing the constructed 3D spatiotemporal data and the encoding-forecasting network combining two spatiotemporal convolutional structures.

Three-dimensional spatiotemporal data contains spatial and temporal dimensions in the atmospheric change, which is more suitable for radar echo nowcasting tasks with strong spatiotemporal correlation. The constructed 3DCNN is first used to extract the local short-term spatiotemporal features, avoids the confusion of spatial features caused by directly utilizing the convolutional LSTM network for learning, and the bi-directional convolutional LSTM structure can learn the global long-term motion trend of the forward and backward radar echoes more fully. This model improves the situation of vague prediction images and solves the problems of easy spatiotemporal information loss and low forecast accuracy. It is shown in the evaluation result that the performance of this model is obviously better than other models under various rainfall threshold conditions, and the predicted future echo images are more accurate, which fully proves the effectiveness of this method.

In future work, we will try to integrate the generative adversarial network (GAN) into the meteorological deep learning network proposed in this paper. The current encoding-forecasting network model is regarded as a generator, and we plan to add a discriminator network and reconstruct the loss function to force the generation of more accurate echo images through the adversarial training. On the other hand, although our model has some advantages under various rainfall thresholds by setting the weights of different rainfall levels, there is still room for improvement in heavy rainfall or severe weather; therefore, we argue that the model needs to try more heavy rainfall data further or introduce the parameters of rainfall intensity change such as humidity and topography for correction.

References

1. Tran, Q.K.; Song, S.K. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. *Atmosphere* **2019**, *10*, 244. [CrossRef]

2. Franch, G.; Nerini, D.; Pendesini, M.; Coviello, L.; Jurman, G.; Furlanello, C. Precipitation Nowcasting with Orographic Enhanced Stacked Generalization: Improving Deep Learning Predictions on Extreme Events. *Atmosphere* **2020**, *11*, 267. [CrossRef]

3. Heuvelink, D.; Berenguer, M.; Brauer, C.C.; Uijlenhoet, R. Hydrological application of radar rainfall nowcasting in the Netherlands. *Environ. Int.* **2020**, *136*, 105431. [CrossRef] [PubMed]

4. McGovern, A.; Elmore, K.L.; Gagne, D.J.; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Smith, T.; Williams, J.K. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2073–2090. [CrossRef]

5. Zhang, W.; Han, L.; Sun, J.Z.; Guo, H.Y.; Dai, J. Application of multi-channel 3D-cube successive convolution network for convective storm nowcasting. *arXiv* **2017**, arXiv:1702.04517. Available online: https://arxiv.org/abs/1702.04517 (accessed on 15 November 2018).

6. Wang, C.; Hong, Y. Application of Spatiotemporal Predictive Learning in Precipitation Nowcasting. In Proceedings of the American Geophysical Union, Fall Meeting 2018, Washingtong, DC, USA, 10–14 December 2018.

7. Zou, H.B.; Wu, S.S.; Shan, J.S.; Yi, X.T. A Method of Radar Echo Extrapolation Based on TREC and Barnes Filter. *J. Atmos. Ocean. Technol.* **2019**, *36*, 1713–1727. [CrossRef]

8. Otsuka, S.; Tuerhong, G.; Kikuchi, R.; Kitano, Y.; Taniguchi, Y.; Satoh, S.; Ushio, T.; Miyoshi, T. Precipitation nowcasting with three-dimensional space-time extrapolation of dense and frequent phased-array weather radar observations. *Weather. Forecast.* **2016**, *31*, 329–340. [CrossRef]

9.   Mecklenburg, S.; Joss, J.; Schmid, W. Improving the nowcasting of precipitation in an Alpine region with an enhanced radar echo tracking algorithm. *J. Hydrol.* **2000**, *239*, 46–68. [CrossRef]

10. Handwerker, J. Cell tracking with TRACE3D—A new algorithm. *Atmos. Res.* **2002**, *61*, 15–34. [CrossRef]

11. Han, L.; Fu, S.X.; Zhao, L.F.; Zheng, Y.G.; Wang, H.Q.; Lin, Y.J. 3D convective storm identification, tracking, and forecasting—An enhanced TITAN algorithm. *J. Atmos. Ocean. Technol.* **2009**, *26*, 719–732. [CrossRef]

12. Liang, Q.Q.; Feng, Y.R.; Deng, W.J.; Hu, S.; Huang, Y.Y.; Zeng, Q.; Cheng, Z.T. A composite approach of radar echo extrapolation based on TREC vectors in combination with model-predicted winds. *Adv. Atmos. Sci.* **2010**, *27*, 1119–1130. [CrossRef]

13. Fletcher, T.D.; Andrieu, H.; Hamel, P. Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Adv. Water Resour.* **2013**, *51*, 261–279. [CrossRef]

14. Sakaino, H. Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow. *IEEE Trans. Geosci. Remote. Sens.* **2012**, *51*, 3023–3036. [CrossRef]

15. Li, L.; He, Z.W.; Chen, S.; Mai, X.F.; Zhang, A.; Hu, B.Q.; Li, Z.; Tong, X.H. Subpixel-Based Precipitation Nowcasting with the Pyramid Lucas–Kanade Optical Flow Technique. *Atmosphere* **2018**, *9*, 260. [CrossRef]

16. Woo, W.C.; Wong, W.K. Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere* **2017**, *8*, 48. [CrossRef]

17. Keenan, T.; Joe, P. The Sydney 2000 world weather research programme forecast demonstration project: Overview and current status. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 1041–1054. [CrossRef]

18. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

19. Klein, B.; Wolf, L.; Afek, Y. A dynamic convolutional layer for short range weather prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4840–4848. [CrossRef]

20. Shi, E.; Li, Q.; Gu, D.Q.; Zhao, Z.M. A method of weather radar echo extrapolation based on convolutional neural networks. In Proceedings of the International Conference on Multimedia Modeling, Bangkok, Thailand; Springer: Cham, Switzerland, 2018; pp. 16–28. [CrossRef]

21. Cho, K.; Merrienboer, B.V.; Gulcehre, C.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078. Available online: https://arxiv.org/abs/1406.1078 (accessed on 22 May 2019).

22. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

23. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In Proceedings of the NIPS, Montreal, QC, Canada, 8–13 December 2014; pp. 3104–3112.

24. Graves, A. Generating sequences with recurrent neural networks. *arXiv* **2013**, arXiv:1308.0850. Available online: https://arxiv.org/abs/1308.0850 (accessed on 22 May 2019).

25. Srivastava, N.; Mansimov, E.; Salakhutdinov, R. Unsupervised learning of video representations using LSTMs. In ICML. *arXiv* **2015**, arXiv:1502.04681. Available online: https://arxiv.org/abs/1502.04681 (accessed on 22 May 2019).

26. Patel, M.; Patel, A. Precipitation Nowcasting: Leveraging bidirectional LSTM and 1D CNN. *arXiv* **2018**, arXiv:1810.10485. Available online: https://arxiv.org/abs/1810.10485 (accessed on 18 October 2019).

27. Shi, X.J.; Chen, Z.R.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 802–810.

28. Tan, C.; Feng, X.; Long, J.W.; Geng, L. FORECAST-CLSTM: A New Convolutional LSTM Network for Cloudage Nowcasting. In Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2019; pp. 1–4. [CrossRef]

29. Wang, Y.B.; Long, M.S.; Wang, J.M.; Gao, Z.F.; Philip, S.Y. PredRNN: Recurrent Neural Networks for Predictive Learning Using Spatiotemporal LSTMs. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 879–888.

30. Ji, S.W.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 221–231. [CrossRef] [PubMed]

31. Vondrick, C.; Pirsiavash, H.; Torralba, A. Generating videos with scene dynamics. In Proceedings of the 30th Annual Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 613–621. *arXiv* **2016**, arXiv:1609.02612. Available online: https://arxiv.org/abs/1609.02612 (accessed on 22 May 2019).

32. Shi, X.J.; Gao, Z.H.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Deep learning for precipitation nowcasting: A benchmark and a new model. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 5617–5627.

33. Loffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning. *arXiv* **2015**, arXiv:1502.03167. Available online: https://arxiv.org/abs/1502.03167 (accessed on 5 August 2019).

34. Standardized Radar Dataset 2018 (SRAD2018). Available online: https://tianchi.aliyun.com/competition/entrance/231662/information (accessed on 22 May 2019).

35. Zeiler, M.D. ADADELTA: An adaptive learning rate method. *arXiv* **2012**, arXiv:1212.5701. Available online: https://arxiv.org/abs/1212.5701 (accessed on 5 August 2019).

36. Lin, C.; Vasic, S.; Kilambi, A.; Turner, B.; Zawadzki, I. Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.* **2005**, *32*, 1–4. [CrossRef]

37. Uijlenhoet, R. Raindrop size distribution and radar reflectivity-rain rate relationships for radar hydrology. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 615–627. [CrossRef]