

Received 7 May 2023, accepted 24 May 2023, date of publication 29 May 2023, date of current version 6 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3280932

RESEARCH ARTICLE

A Forecast-Refinement Neural Network Based on DyConvGRU and U-Net for Radar Echo Extrapolation

JINLIANG YAO^{1,2}, FEIFAN XU¹, ZHENG QIAN³, AND ZHIPENG CAI¹

¹School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

²Zhejiang Key Laboratory of Brain-Machine Collaborative Intelligence, Hangzhou 310018, China

³Ningbo Meteorological Service Center, Ningbo 315010, China

Corresponding author: Feifan Xu (212050251@hdu.edu.cn)

This work was supported by the Zhejiang Provincial Basic Public Welfare Research Project under Grant LGG20F020012, and the General Scientific Research Project of the Zhejiang Provincial Department of Education (ZPDE) under Grant Y202146867.


ABSTRACT Precipitation nowcasting is very important for the sectors which critically depend on timely and accurate weather information. One of the challenges of precipitation nowcasting is radar echo extrapolation which predicts the radar echo images accurately. Nowadays, the methods of radar echo extrapolation are mostly based on ConvRNNs. Unfortunately, as lead time increases, these methods unavoidably suffer from the problem that high reflectivity values are underestimated. Therefore, we propose a forecast-refinement neural network based on DyConvGRU and U-Net to improve the predicting ability for high reflectivity during radar echo extrapolation. To improve the model's ability to predict high reflectivities, dynamic convolution, and the forecast-refinement architecture are applied. And to obtain more realistic results, the WGAN's training strategy is adopted to train the forecast module and the refinement module. Through experiments on a radar dataset from Shanghai, China, the results show that our proposed method obtains higher Probability of Detection (POD), Critical Success Index (CSI), Heidke Skill Score (HSS), and lower False Alarm Rate (FAR).

INDEX TERMS Spatiotemporal sequence prediction, radar echo extrapolation, DyConvGRU, U-Net, dynamic convolution, WGAN.

I. INTRODUCTION

Precipitation nowcasting is to predict the precipitation intensity in a specific area two hours ahead. Short-term heavy precipitation frequently causes floods, mudslides, landslides, and other serious natural disasters that pose risks to public health and safety. Therefore, it is important to carry out precipitation nowcasting and early warning for disaster prevention.

Precipitation forecasting methods are broadly divided into two major categories which are numerical weather prediction (NWP) and radar echo extrapolation based methods [1]. NWP is a technique for forecasting the weather by using mathematical computer models. It uses datasets (including satellite data and observations from weather stations) to create a

The associate editor coordinating the review of this manuscript and approving it for publication was Shovan Barma .

mathematical representation of the current state of the atmosphere. This mathematical model is then used to calculate the likely weather conditions at different points in the future. NWP is widely used to improve the accuracy of weather forecasts. However, for precipitation at 0-2 hours lead time, NWP has a poor ability to provide an accurate prediction as this is less than the time needed for model spin-up and due to difficulties in non-Gaussian data assimilation [2], [3], [4].

Radar echo based methods are divided into two categories. One is to predict the precipitation intensity directly when the inputs are the radar echo images. And the other one is to predict future radar echoes which are then converted into precipitation intensity. Radar echo extrapolation can track and forecast Mesoscale Convective Systems (MCS) quickly so it can be used to predict precipitation. Previous radar echo extrapolation is mainly based on an optical flow algorithm, which achieved promising performance [5], [6], [7]. The

optical flow based methods are based on two basic assumptions: the pixel intensity is constant between adjacent frames, and the adjacent pixels move in the same directions. However, these two assumptions are violated in radar echo extrapolation in some cases because the intensity may vary over time and the motion may be highly dynamic and nonlinear. In addition, the optical flow based methods only consider a given radar echo image sequence for prediction, which does not make full use of a large number of radar echo records in the dataset.

Currently, as deep learning technology develops, it is used to solve the problem of radar echo extrapolation and precipitation prediction. Many works have shown that deep learning based methods significantly outperform optical flow based methods. Radar echo extrapolation can be considered as a problem of image sequence prediction. There are many sequence modeling methods such as RNN [8], LSTM [9], and GRU [10] to solve this problem. Chiang et al. [11] adopted RNN to estimate and forecast precipitation based on radar observations. However, RNN, LSTM, and GRU focus on temporal features, and can not catch spatial features of the image. Therefore, ConvLSTM was proposed by Shi et al. [12] which combines CNN and RNN to catch spatial and temporal features simultaneously. And Tran and Song [13], Zhong et al. [14], and Jing et al. [15] used ConvLSTM to extrapolate the radar echo. However, the images which are predicted by ConvRNNs tend to be blurry and unrealistic, and the motion of the region of high radar echo is not accurately predicted [16]. Tian et al. [17], Xie et al. [16], and Zheng et al. [18] proposed three GAN based methods including the ConvRNN based generators and the CNN based discriminators. These models were trained with the GAN's training strategy. The GAN based methods obtain more clear and realistic results compared with the ConvRNN based methods. However, as lead time increases, these methods always underestimate the radar reflectivity, especially in the high reflectivity regions which often indicate heavy precipitation regions, because of the low percentage of the high reflectivity in the real world.

To alleviate this problem, a forecast-refinement natural network was proposed in this study. The forecast module is based on DyConvGRU which predicts the motion of the clouds. DyConvGRU incorporates dynamic convolution into ConvGRU. With different convolution strategies applied to different inputs, DyConvGRU can better focus on the high reflectivity regions than with the static convolution. The refinement module is based on U-Net which recovers more fine texture details of high reflectivity regions. The WGAN's training strategy which introduces Wasserstein distance is used to train DyConvGRU and U-Net to retain more shape details of the predicted results. The experiments show that the proposed method obtains higher POD, CSI, HSS, and lower FAR. And our proposed method gives a new idea to prevent natural disasters caused by strong convective weather. The main contributions of our work are summarized as follows:

- This work proposed DyConvGRU which incorporates dynamic convolution into the ConvGRU network. With different convolution strategies responding to different inputs, the model with dynamic convolution can better focus on the different reflectivity regions than with static convolution.
- A forecast-refinement architecture is used in the proposed method. A forecast module is used to predict the future radar echo images of the previous observations and a refinement module is used to reconstruct the predictions which are generated by the forecast module to make them more similar to the real radar echo images.
- Dynamic convolution based discriminators are used to train DyConvGRU and U-Net. The WGAN's training strategy which introduces Wasserstein distance is used to train DyConvGRU and U-Net to retain more shape details of the predicted radar echo images.

II. RELATED WORK

The goal of radar echo extrapolation is to use the previously observed radar echo sequence to extrapolate a fixed length of the future radar echo extrapolation, which means that the radar echo extrapolation problem can be regarded as a spatiotemporal sequence prediction problem and we can solve this problem using a spatio-temporal modeling approach. In this section, we will review the main methods for precipitation nowcasting and radar echo extrapolation. Firstly, we introduce the deep learning based methods including the ConvRNNs based methods and the GAN based methods. Then, the dynamic convolution is presented for a better understanding of our method.

A. ConvRNNs BASED METHODS

Shi et al. [12] defined the precipitation nowcasting problem as a spatiotemporal sequence prediction problem and proposed a new model named convolutional LSTM (ConvLSTM) by converting input-to-state and state-to-state fully connected operations into convolutional operations in fully connected LSTM (FC-LSTM) cells. ConvLSTM can capture spatial and temporal features simultaneously and outperforms both the optical flow based method and FC-LSTM. However, the convolutional recursive architecture in ConvLSTM is location-invariant, but natural motions and transformations (e.g., rotations) are usually location-variant. To overcome the shortcoming of the ConvLSTM, Shi et al. [19] proposed the TrajGRU based on ConvGRU, which can actively learn the location-variant architecture of recurrent connections. Zeng et al. [20] proposed a deep learning model combining U-Net architecture and TrajGRU named T-UNet. This model uses an efficient convolutional neural network of UNet architecture with a residual network and a TrajGRU recurrent neural network is added at each layer. Wang et al. [21] proposed a novel model PredRNN based on ConvLSTM. In PredRNN, the core module is spatiotemporal LSTM which can extract and store both spatial and temporal

features. Then Wang et al. [22] improved PredRNN and proposed PredRNN++ for Spatiotemporal prediction learning. PredRNN++ has a new LSTM cell named Casual LSTM and a Gradient Highway Unit (GHU). GHU provides a fast passing route for gradients from the outputs to the previous long term inputs to alleviate the difficulty of gradient propagation in deep models. GHU and Casual LSTM work simultaneously to allow the model to capture both short term and long term temporal features adaptively. The model performs well on the Moving MNIST Dataset and KTH Action Dataset. Bonnet et al. [23] used PredRNN++ to predict the future sequence of radar echo images for up to one hour lead time in São Paulo, Brazil. Chen et al. [24] proposed ConvLSTM layers with a star-shape bridge to transfer features across time steps. In this method, they also proposed a raining-oriented loss function and used normalization techniques to improve the convergence performance of the model. Liu et al. [25] proposed a spatiotemporal prediction model, namely the Spatial-Temporal Long Short-Term Memory based on the self-attentive mechanism (ST-LSTM-SA). The 3D convolution is developed to exploit the short-term spatiotemporal information, and the channel correlation is modeled by the self-attention mechanism to improve representations in long-term interaction. Zhang et al. [26] introduced the self-attention mechanism and an extra memory that saves global spatiotemporal features into the original Spatiotemporal LSTM (ST-LSTM) to capture both spatial and temporal global features of radar echo motion. And these units are stacked to build the radar echo extrapolation network SAST-Net. Lin et al. [27] extended ConvSLTM with a self-attention memory module to generate new spatiotemporal features by aggregating features from all inputs and memory features with pairwise similarity scores. Luo et al. [28] proposed a new pseudo flow spatiotemporal LSTM (PFST-LSTM) cell in which a spatial storage unit and a position calibration module were embedded into. And a new sequence-to-sequence precipitation nowcasting architecture based on the PFST-LSTM cells was proposed. The experiment results demonstrated that the PFST-LSTM is superior to the previous models. Huang et al. [29] proposed the Location-Refining (LR) network. The LR network consists of two networks: a location network and a refining network. The former predicts rainfall locations and motion trends, and the latter predicted rainfall intensity and distribution. To make the model focus more on the high reflectivity regions, they also proposed a new loss function which is combined with the F-MSE Loss, Modified Dice Loss, and Structural Loss. Yu et al. [30] proposed an axial attention memory module and embedded it into a standard ConvGRU to develop an axial attention memory cascaded ConvGRU (ATMConvGRU). The axial attention memory module can aggregate and embed spatiotemporal features into a standard ConvGRU. The experiments demonstrated that ATMConvGRU is effective for forecasting four types of weather data, including temperature, relative humidity, wind, and radar echoes.

ConvRNN based methods convert fully connected operations into convolutional operations to catch the temporal features and spatial features simultaneously. The previous works have shown that the ConvRNN based models outperform the optical flow based models. However, these ConvRNN based models are almost optimized by the mean square error loss. The mean square error loss is good at modeling unimodal distribution, while the intensity of real radar images is multi-modal and skewed, which makes the extrapolated radar echo images blurring and unrealistic. Moreover, the ConvRNNs based methods adopt static convolution to catch the spatial features, which treat the different reflectivities equally. But, treating high and low echoes equally will make the model more likely to predict low echoes, because high echo samples are very rare in real life.

B. GAN BASED METHODS

Generative Adversarial Nets (GAN) train two models at the same time: a generator model G to generate the data distribution and a discriminator model D to calculate the probability that the samples come from the training data instead of G [31]. However, during the training phase, GAN often suffers from pattern collapse and failure to converge. Arjovsky et al. [32] proposed WGAN, a new algorithm that introduces the Wasserstein distance in the training process. The Wasserstein distance can theoretically alleviate the gradient vanishing due to its superior smoothness compared to KL divergence and JS divergence. In the training process, Wasserstein distance is transformed into an approximate and solvable form, which is minimized by a discriminator. When training the generators and discriminators, WGAN differs from the traditional GAN in four respects. Firstly, the sigmoid function is removed from the last layer of the discriminators. Secondly, the loss of the generator and discriminator does not adopt the log function. Thirdly, the absolute values of the discriminator parameters are truncated to less than a fixed constant c after each update of their parameters. Fourthly, momentum based optimization algorithms are not used during the training process. WGAN not only alleviates the problem of training instability but also provides meaningful learning curves for debugging and hyperparameter searches.

Currently, many researchers use GAN to solve precipitation prediction and radar echo extrapolation problems. Jing et al. [33] proposed an adversarial extrapolation neural network named AENN, which is a generative adversarial model with a conditional generator and two discriminators including a radar echo frame discriminator and a radar echo sequence discriminator. The experimental results at 30, 60, and 90 minutes lead time show that AENN achieves the expected results and outperforms the contemporaneous model significantly. Tian et al. [17] proposed a generative adversarial ConvGRU (GA-ConvGRU) which consists of a ConvGRU based generator and a CNN based discriminator. The model generated more realistic and accurate radar echo extrapolations, and the results are better

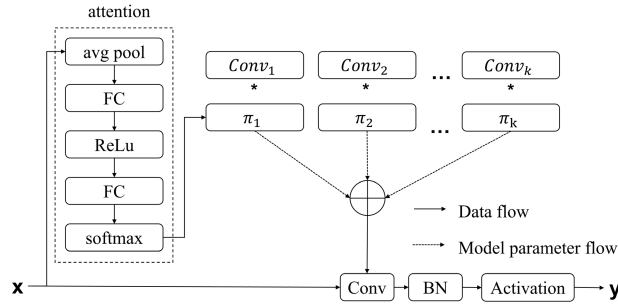


FIGURE 1. A dynamic convolution layer.

than the images generated by the optical flow method and ConvGRU. Ravuri et al. [34] proposed a deep generative model (DGM) for long-term precipitation forecasting, which was expected to solve the problem of long-term blurring and poor performance when meeting rare heavy rainfall. Zheng et al. [18] proposed GAN-argcPreNet v1.0 which used GAN to overcome the problem of blurred and unrealistic radar echo images. In the generator, a gate that regulates the memory and output was designed in the argcLSTM to reduce the loss of spatiotemporal information. The discriminator has two input channels, which can score more strictly according to the real echo distribution. Then Zheng et al. [35] designed GAN-argcPredNet v2.0 based on GAN-argcPredNet v1.0. A Spatiotemporal Information Changes Prediction (STIC-Prediction) network is designed as a generator and the discriminator is a Channel-Spatial Convolution (CS-Convolution) network. The discriminator intensifies the discrimination of echo information by strengthening the spatial information of a single image, which reduces the loss of important evolutionary information.

Compared with the ConvRNN based methods, GAN based methods have a ConRNN based generator and a CNN based discriminator. The discriminator is trained to discriminate between the generated images and real images and provide the adversarial loss to help the generator produce more realistic images. However, the discriminators of these methods are trained to give the real image the value of true and give the generated images the value of false, which can not evaluate the gap between the real images and the generated images. This will cause the problem that the generator and discriminator are not easy to be coordinately trained.

C. DYNAMIC CONVOLUTION

For traditional static convolution, once the training is completed, the parameters of all convolution kernels are fixed, which means that all inputs are treated equally by the convolution kernels. Yang et al. [36] argued that most strategies for increasing the capacity of a CNN based model are increasing the depth of the model or the number of channels of convolution layers, which can improve the model performance but at the cost of high computational complexity. To decrease the computational complexity, they proposed conditionally

parameterized convolution (CondConv) which learns specialized convolutional kernels for each example. Replacing traditional static convolutions for each example. Replacing traditional static convolutions with CondConv increases the size and capacity of a network while maintaining efficient inference. Inspired by the idea of CondConv, Chen et al. [37] proposed dynamic convolution. Compared with traditional convolutions, dynamic convolutions aggregate multiple parallel convolution kernels dynamically based on their attentions, which are input dependent. The dynamic convolution layer is shown in Fig.1. The specific form of Dynamic Convolution is shown as follows:

$$y = f(\tilde{W}^T(x)x + \tilde{b}) \tag{1}$$

$$\tilde{W}^T(x) = \sum_{k=1}^K \pi_k(x)\tilde{W}_k \tag{2}$$

$$\tilde{b}(x) = \sum_{k=1}^K \pi_k(x)\tilde{b}_k \tag{3}$$

$$\pi_k(x) = \frac{\exp(z_j/\tau)}{\sum_j \exp(z_j/\tau)} \tag{4}$$

$$s.t. 0 \leq \pi_k(x) \leq 1, \sum_{k=1}^K \pi_k(x) = 1 \tag{5}$$

where f is the activation function; \tilde{W}_k is the weight of the k^{th} perceptron. \tilde{b}_k is the bias of the k^{th} perceptron. $\pi_k(x)$ is the weight of the k^{th} linear function $\tilde{W}^T(x)x + \tilde{b}$ which varies with the inputs. z_j is the output of the second FC layer of the attention branch in Fig.1; τ is the temperature, and the output is less sparse as τ increases. K is the number of kernels. By merging multiple convolutional kernels, dynamic convolution with the kernel attention mechanism improves the model's ability to capture spatial features. Compared with standard convolution, Dynamic convolution is input-dependent.

III. METHOD

The aim of radar echo extrapolation is to predict future radar echo images based on historical radar echo images. Given a historical radar echo image sequence $X_{1:T} = \{X_1, X_2, \dots, X_T\}$, we predict the future radar echo image sequence $X_{T+1:T+L} = \{X_{T+1}, X_{T+2}, \dots, X_{T+L}\}$. The radar echo extrapolation can be defined as follows:

$$\hat{X}_{T+1}, \dots, \hat{X}_{T+L} = \arg \max_{X_{T+1}, \dots, X_{T+L}} (X_{T+1}, \dots, X_{T+L} | X_1, \dots, X_T) \tag{6}$$

where T represents the length of the input sequence and L represents the length of the predicted image sequence.

Radar echo extrapolation is essentially a spatiotemporal sequence prediction problem, and shares similarities with video frame prediction [14]. However, there are some differences between the predicting video frames and extrapolating radar echo images. There are multiple frames in each second in video prediction, and the variations between neighboring

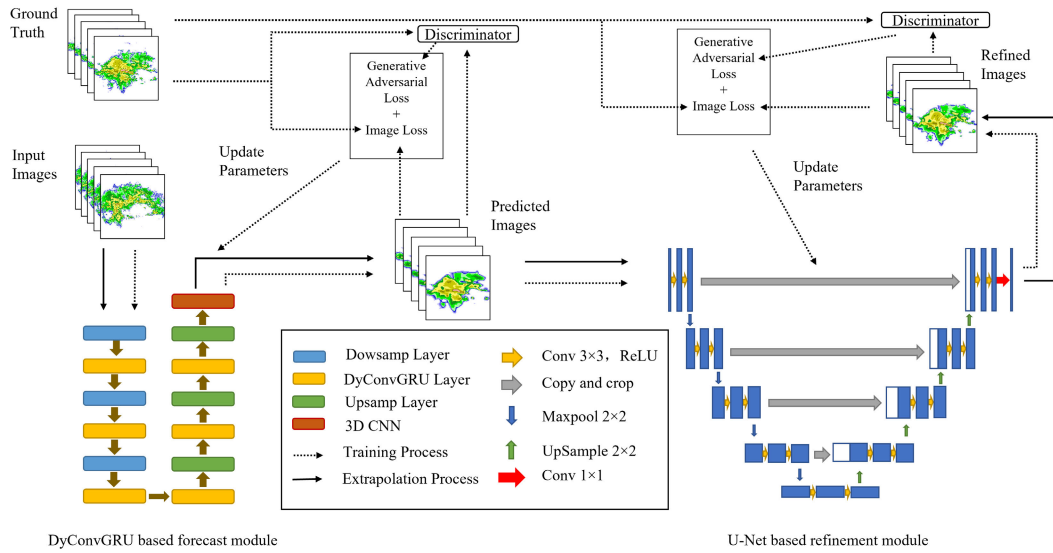


FIGURE 2. The process of our proposed methods. The proposed model is divided into two parts: a DyConvGRU based forecast module and a U-Net based refinement module. The forecast module predicts the variation of radar echo and it is trained with WGAN’s training strategy. The refinement module corrects the reflectivities that are predicted by the forecast module and it is trained with WGAN’s training strategy.

frames are limited. But there are several minutes between consecutive radar echo images. When confronting strong convective weather, the variations in adjacent radar echo images are noticeable. In addition, radar echo extrapolation has to capture the changes of non-rigid radar echoes and learn the radar echo generations, accumulations, and dissipations, which makes it more difficult to obtain accurate radar reflectivities. As mentioned above, the ConvRNNs based methods adopt the static convolution to catch the spatial features. When the training is completed, the parameters of the models are fixed, and these models treat high reflectivity regions and low reflectivity regions equally. However, high reflectivities account for a small part of the total data, which causes the values of high regions to be underestimated. To alleviate the underestimation of the radar reflectivities, a forecast-refinement neural network is proposed. The model consists of two parts: DyConvGRU based forecast module and U-Net based refinement module. The dynamic convolution in DyConvGRU calculates weights based on the varied radar echo reflectivities in the radar images and applies the weights to the K convolution kernels, causing the K kernels to adopt different convolution strategies depending on the different inputs. The refinement module reconstructs the images generated by DyConvGRU and corrects the reflectivities. The complete method is shown in Fig.2.

A. DyConvGRU BASED FORECAST MODULE

As mentioned previously, radar echo extrapolation can be formalized as a spatiotemporal sequence prediction. ConvRNNs are suitable for this problem. ConvGRU proposed by Ballas et al. [38] is one of the variants of ConvRNNs and its performance has been proved in spatiotemporal sequence

prediction. ConvGRU has two trainable gate units: the update gate z_t and the reset gate r_t . The reset gate controls how much of the past knowledge to forget. The update gate controls how much of the past knowledge needs to be passed along into the future. These two gates can save information from long term sequences without being cleared or removed as time passes. The specific expression form of ConvGRU is shown as follows:

$$z_t = \sigma(W_{xz} * x_t + W_{hz} * h_{t-1}) \quad (7)$$

$$r_t = \sigma(W_{xr} * x_t + W_{hr} * h_{t-1}) \quad (8)$$

$$\hat{h}_t = f(W_{xh} * x_t + r_t \odot (W_{hj} * h_{t-1})) \quad (9)$$

$$h_t = (1 - z_t) \odot \hat{h}_t + z_t \odot h_{t-1} \quad (10)$$

where σ is the Sigmoid function; $*$ and \odot are convolution operation and Hadamard product; z_t and r_t are the update gate and reset gate; h_{t-1} and h_t are the hidden state at the previous moment and the current moment; f is an activation function.

As shown in the above equation (7) to (10), ConvGRU adopts the static convolution operations by inputs and hidden states. The original ConvGRU suffers from the same shortcoming as ConvRNNs which has a poor ability to predict high reflectivities. DyConvGRU we proposed replaces the static convolution operations with dynamic convolutional operations that apply the attention mechanism to the convolutional kernels. The K convolutional kernels make up each dynamic convolutional unit. An attention module first extracts k weights $\{\pi_1, \dots, \pi_k\}$ from the input image x , the K weights are then applied to the k convolutional kernels $\{Conv_1, \dots, Conv_k\}$, and finally the K convolutional kernels are aggregated based on the weights to provide the results of each dynamic convolutional unit. This enables

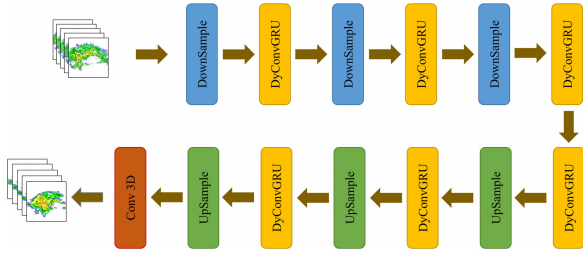


FIGURE 3. The encoder-decoder architecture of the forecast module. The blue blocks are the DownSample layers. The yellow blocks are the DyConvGRU layers. The green blocks are the UpSample layers. The red block is the 3D convolution layer.

DyConvGRU to employ different convolution algorithms in response to reflectivities of varying intensities. The formula of the DyConvGRU is shown as (11) to (14).

$$z_t = \sigma(Dy(W_{xz}, x_t) + Dy(W_{hz}, h_{t-1})) \quad (11)$$

$$r_t = \sigma(Dy(W_{xr}, x_t) + Dy(W_{hr}, h_{t-1})) \quad (12)$$

$$\hat{h}_t = f(Dy(W_{xh}, x_t) + r_t \odot Dy(W_{hj}, h_{t-1})) \quad (13)$$

$$h_t = (1 - z_t) \odot \hat{h}_t + z_t \odot h_{t-1} \quad (14)$$

where σ is the Sigmoid function; $Dy(\cdot)$ and \odot are the dynamic convolution operation and Hadamard product; z_t and r_t are the update gate and reset gate; h_{t-1} and h_t is the hidden state at the previous moment and the current moment; f is the activation function which is a tanh activation function in this study.

The forecast module adopts an Encoder-Decoder architecture as shown in Fig.3. The encoder comprises three downsampling layers and three DyConvGRU layers and it extracts hidden states from previous radar echo images. The downsampling layer reduces the size of the feature maps and captures the spatial features by convolutional operation. DyConvGRU layers learn the spatiotemporal feature of the radar echo sequences. The decoder uses the hidden states to forecast future radar echo images and consists of three upsampling layers, three DyConvGRU layers, and a 3D Convolution layer. The upsampling layers increase the size of feature maps by deconvolution operations.

B. U-NET BASED REFINEMENT MODULE

The refinement module is the other major part of the proposed model, which aims to correct the value of reflectivities that are predicted by the forecast module. The refinement module is based on U-Net [39] which is widely used in medical image segmentation. And some researchers applied it to predict short-term precipitation [40], [41], [42], [43]. They feed the previous observations $\{X_1, X_2, \dots, X_{t-1}\}$ into the model to predict the radar echo image \hat{X}_t , and then the radar echo images $\{X_2, X_3, \dots, X_t\}$ are fed into the model to predict the radar echo images X_{t+1} . They repeat these steps to achieve extended lead time results. However, there are errors in the \hat{X}_t generated from $\{X_1, X_2, \dots, X_{t-1}\}$ compared with the real obivation X_t , and then \hat{X}_t is utilized as input to predict the

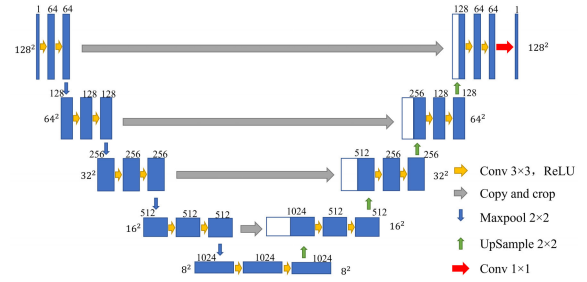


FIGURE 4. The Encoder-Decoder architecture of the U-Net. It has four layers. The left side of the model is the encoder which reduces the size of the image and extracts the low-level image features. The right side of the U-Net is the decoder which extracts high-level features and restores the output to its original size.

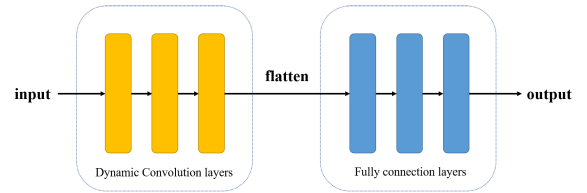


FIGURE 5. The architecture of the discriminator consists of three dynamic convolution layers and three fully connected layers. The yellow blocks are the dynamic convolutional layers and the blue blocks are the fully connected layers.

following image X_{t+1} , which makes the predictions suffer from accumulative error as lead time increases. To avoid introducing the accumulative error caused by the U-Net based methods and inspired by the image transformation methods based on GAN [44], U-Net is adopted to correct the value of high reflectivities instead of predicting radar echo images directly in this study. When using the U-Net based methods to predict the future radar echo images, the prediction error will grow and accumulate exponentially as the extrapolation goes deeper, and the future frames will be extrapolated by referring to increasingly unrealistic past frames [15]. U-Net accepted the images generated by the forecast module one by one, and then reconstruct these images to corrected images sequentially. Using this method, we create a one-to-one mapping between the images generated by the prediction module and the images generated by the refinement module, avoiding the inaccuracies associated with the direct recycling of U-Net prediction results. The detailed architecture of the U-Net in the refinement module is shown in Fig.4. The left side of U-Net is the encoder which reduces the size of the image and extracts the image features. The right side of U-Net is the decoder which increases the size of the feature maps.

C. DISCRIMINATORS

In the proposed method, DyConvGRU and U-Net are trained independently with WGAN's training strategy. The discriminators in this work are divided into a temporal discriminator and a spatial discriminator. These discriminators both consist of three dynamic convolutional layers and three fully connected layers, as shown in Fig.5. The dynamic convolutional

layers extract the spatial features and the fully connected layers decrease the dimensions of the features.

The spatial discriminator has three 2D dynamic convolution layers and three fully connected layers. In the training phase of the forecast module, it is trained to distinguish each generated radar echo image $\hat{X}^{T+6}, \hat{X}^{T+12}, \hat{X}^{T+18}, \dots, \hat{X}^{T+90}$ from the real radar echo images $X^{T+6}, X^{T+12}, X^{T+18}, \dots, X^{T+90}$, trying to judge the radar echo images generated by DyConvGRU as fake when the ground truth is judged as real. In the refinement module training phase, the spatial discriminator is trained to judge the radar echo images generated by U-Net as fake when the ground truth is judged as real.

The temporal discriminator has three 3D dynamic convolution layers and three fully connected layers. In the forecast module training phase, it is trained to judge the reality of the radar echo image sequence $\hat{X}^{T+6:T+90}$ generated by DyConvGRU, trying to determine whether it is fake while the ground truth is real. It could assure the temporal consistency of the whole generated radar echo images in a sequence, which is a complement of the spatial discriminator when training the forecast module.

The DyConvGRU is trained with a temporal and spatial discriminator. Following the completion of DyConvGRU training, the radar echo image is fed into the trained DyConvGRU, and the DyConvGRU prediction result is utilized to train U-Net with a spatial discriminator. These discriminators produce the adversarial loss and the DyConvGRU and U-Net are optimized respectively by combining it with the image loss. The training process is shown in Fig.2.

D. LOSS FUNCTION

In radar echo images, the high reflectivity indicates heavy precipitation, so it is important to predict the high reflectivity regions accurately. However, there is a low percentage of high reflectivities in the real world. In order to focus more on the high reflectivity during the training phase, we extended the weighted loss function which Shi et al. [19] proposed to make their model pay more attention to heavy precipitation. Our layering of radar echo reflectivity is more precise, and high echo locations are given more weight. And we introduced adversarial loss in weighted loss to help DyConvGRU and U-Net generate more realistic images. The weighted function is defined as follows:

$$weight_{i,j} = \begin{cases} 0, & 0 \leq x_{i,j} \leq 15 \\ 1, & 15 < x_{i,j} \leq 30 \\ 2, & 30 < x_{i,j} \leq 40 \\ 5, & 40 < x_{i,j} \leq 50 \\ 10, & 50 < x_{i,j} \leq 55 \\ 20, & 55 < x_{i,j} \leq 60 \\ 50, & 60 < x_{i,j} \leq 65 \\ 90, & 65 < x_{i,j} \leq 70 \end{cases} \quad (15)$$

where $weight_{i,j}$ and $x_{i,j}$ are the weight and the radar reflectivity at coordinate (i, j), respectively.

The forecast module includes a DyConvGRU, a temporal discriminator, and a spatial discriminator, and they are trained with WGAN’s training strategy. We define the weighted loss function of each batch of the samples between the real radar echo images and coarse predicted images as follows:

$$loss_{image} = WMAE_{image} + WMSE_{image} \quad (16)$$

$$WMAE_{image} = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{j=1}^H \sum_{i=1}^W weight_{n,t,i,j} * (|target_{n,t,i,j} - output_{n,t,i,j}|) \quad (17)$$

$$WMSE_{image} = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^T \sum_{j=1}^H \sum_{i=1}^W weight_{n,t,i,j} * (target_{n,t,i,j} - output_{n,t,i,j})^2 \quad (18)$$

where $loss_{image}$ is the loss of a batch of predicted radar echo images; N is the batch size; T is the length of the radar echo image sequence which is predicted by DyConvGRU network; H and W are the height and width of the radar echo images, respectively; $target_{n,t,i,j}$ is the value of real radar echo images; $output_{n,t,i,j}$ is the value of the radar echo images predicted by the DyConvGRU; $weight_{n,t,i,j}$ is the weight which is calculated according to ground truth at coordinate (n, t, i, j).

In the training phase of the forecast module, the DyConvGRU network is adopted as a generator and it is trained with a temporal discriminator and a spatial discriminator. The loss function of the DyConvGRU network is defined as follows:

$$loss_{G_{forecast\ module}} = -D_T(output) - D_S(output) + loss_{image} \quad (19)$$

where $loss_{image}$ is the weighted loss of the radar echo images predicted by the DyConvGRU network; D_T and D_S are the temporal discriminator and the spatial discriminator, respectively; $output$ is the radar echo images predicted by the DyConvGRU network.

The loss function of the discriminators in the forecast module is defined as follows:

$$loss_{D_{forecast\ module}} = D_T(output) + D_S(output) - D_T(target) - D_S(target) \quad (20)$$

where D_T and D_S are the temporal discriminator and spatial discriminator, respectively; $output$ is the radar echo images generated by the DyConvGRU network; $target$ is the real radar echo images.

In the refinement module, a U-Net is adopted as the generator, and a spatial discriminator is used to help to train the U-Net. The U-Net and the spatial discriminator are trained with WGAN’s training strategy. The loss function of U-Net is specified as follows:

$$loss_{G_{refinement\ module}} = -D_S(output) + loss_{image} \quad (21)$$

where D_S is the spatial discriminator, $output$ is the image generated by the U-Net; $loss_{image}$ is the radar echo image weighted loss.

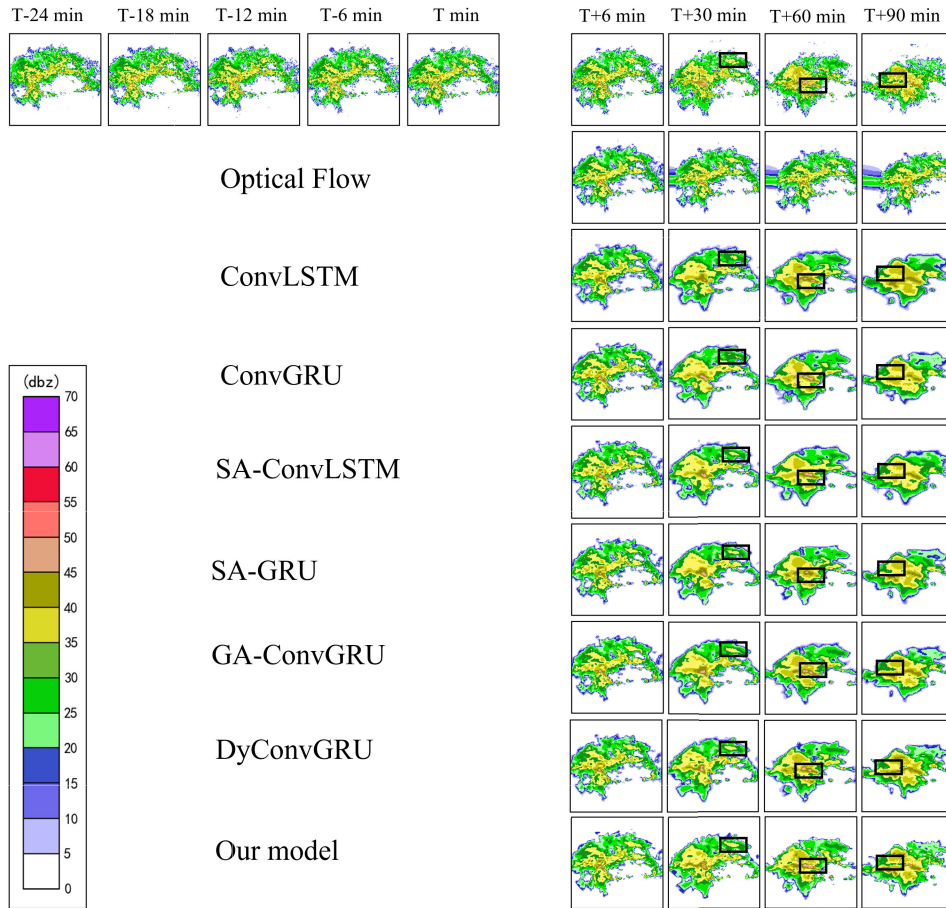


FIGURE 6. One of the results of the comparative experiments. The radar echo images in the first line are the input and the ground truth at T+6 min, T+30 min, T+60 min, and T+90 min. The images in the following lines are the results of Optical flow, ConvLSTM, ConvGRU, SA-ConvLSTM, SA-GRU, GA-ConvGRU, DyConvGRU, and our proposed model, respectively.

The loss function of the spatial discriminator in the refinement module is

$$loss_{D_{refinement\ module}} = D_S(output) - D_S(target) \quad (22)$$

where D_S denotes the spatial discriminator, $target$ is the real radar echo image, and $output$ is the radar echo image generated by U-Net.

IV. EXPERIMENT

A. DATA

We evaluated our proposed method with a public radar echo dataset proposed by Chen et al. [24]. The dataset includes composite reflectivity data collected from the Dual Polarization Weather Surveillance Radar-1988 Doppler Radar (WSR-88D) located in Pudong, Shanghai, China. The dataset contains 170,000 radar echo images generated by volume scans at intervals of approximately 6 minutes from October 2015 to July 2018. To reduce the impact of a large number of rainless data, the radar echo images whose maximum reflectivity is less than 15dbz were removed. We also filtered out nonconsecutive frames and apply a stride-1 sliding window to

generate the sequences of continuous radar echo images with the length of 20. For better training, we scaled the reflectivity R to the interval $[0, 1]$ by setting $\hat{R} = \frac{R}{70}$.

B. IMPLEMENTATION DETAILS

The optical flow based method, ConvGRU, ConvLSTM, GA-ConvGRU, SA-GRU, SA-ConvLSTM, and DyConvGRU were employed as comparative models in this work. To make fair comparisons with previous works, we applied almost the same experimental setting. We trained all models on the radar echo image dataset for 100 epochs. During the training process, the mini-batch was set to 16. The parameters of comparative models were optimized by the Adam algorithm and the learning rates are set to 0.0002. For our proposed model, the parameters of the DyConvGRU network and the U-Net were optimized by the Adam algorithm and the learning rates were set to 0.0002 and 0.00005, respectively. The parameters of the discriminators including the spatial discriminator and the temporal discriminator were optimized by the RMSprop algorithm and the learning rates were set to 0.00005. The

TABLE 1. The CSI, FAR, POD, and HSS of the comparative experiments under the three different thresholds (20dbz, 30dbz, and 40dbz) and at T+30 min, T+60 min, and T+90 min, respectively. Bold values are the best results, underlined values are the second best results.

	T+30 min			T+60 min			T+90 min		
	20dbz	30dbz	40dbz	20dbz	30dbz	40dbz	20dbz	30dbz	40dbz
CSI↑									
Optical Flow	0.6662	0.4189	0.1510	0.5352	0.2830	0.0455	0.4421	0.2000	0.0160
ConvLSTM	0.7710	0.5499	0.2401	0.7658	0.5407	0.2129	0.7553	0.5101	0.1646
ConvGRU	0.7662	0.5510	0.2620	0.7601	0.5336	0.2120	0.7440	0.4988	0.1161
SA-ConvLSTM	0.7641	0.5493	0.2020	0.7547	0.5112	0.1408	0.7304	0.4602	0.0768
SA-GRU	0.7757	0.5635	0.2839	0.76760	0.5481	0.2175	0.7477	0.5004	0.1450
GA-ConvGRU	0.7757	0.5625	0.3020	0.7730	0.5519	0.2262	0.7496	0.5000	0.1286
DyConvGR	0.7875	0.5788	0.3333	0.7896	0.5774	0.2825	0.7710	0.5387	0.2173
Our model	0.8108	0.6288	0.4251	0.8165	0.6351	0.4485	0.7970	0.5884	0.3310
POD↑									
Optical Flow	0.7754	0.5608	0.2320	0.6634	0.4120	0.0784	0.5752	0.3085	0.0276
ConvLSTM	0.8608	0.6965	0.2923	0.8534	0.6900	0.2682	0.8716	0.7124	0.2293
ConvGRU	0.8516	0.7151	0.3284	0.8423	0.6875	0.2698	0.8365	0.6586	0.1478
SA-ConvLSTM	0.8709	0.7225	0.2481	0.8557	0.6534	0.1709	0.8390	0.6009	0.0846
SA-GRU	0.8743	0.7526	0.3799	0.8597	0.7102	0.2798	0.8572	0.6608	0.1919
GA-ConvGRU	0.8852	0.7915	0.4085	0.8680	0.7405	0.2915	0.8389	0.6499	0.1595
DyConvGRU	0.8737	0.7417	0.4493	0.8739	0.7389	0.3619	0.8647	0.7162	0.3014
Our model	0.8941	0.8013	0.5713	0.9029	0.8162	0.6112	0.8823	0.7564	0.4330
FAR↓									
Optical Flow	0.1804	0.3846	0.6784	0.2738	0.5314	0.8876	0.3507	0.6419	0.9518
ConvLSTM	0.1229	0.2839	0.3790	0.1224	0.2928	0.3661	0.1553	0.3666	0.4233
ConvGRU	0.1194	0.2995	0.4380	0.1178	0.3039	0.4161	0.1352	0.3383	0.3961
SA-ConvLSTM	0.1419	0.3067	0.3303	0.1396	0.3035	0.2794	0.1562	0.3370	0.2412
SA-GRU	0.1301	0.3135	0.4427	0.1262	0.3015	0.4288	0.1510	0.3387	0.4038
GA-ConvGRU	0.1407	0.3429	0.4346	0.1280	0.3205	0.5000	0.1297	0.3278	0.4187
DyConvGRU	0.1143	0.2811	0.4213	0.1122	0.2813	0.3833	0.1276	0.3234	0.4465
Our model	0.1054	0.2592	0.3687	0.1076	0.2636	0.3570	0.1116	0.2810	0.3712
HSS↑									
Optical Flow	0.2975	0.1046	-0.1585	0.2003	-0.0339	-0.3369	0.1234	-0.1378	-0.4125
ConvLSTM	0.3635	0.209	0.0611	0.3620	0.2013	0.0446	0.3393	0.1383	-0.0260
ConvGRU	0.3640	0.1985	0.0579	0.3628	0.1916	0.0292	0.3482	0.1586	-0.0002
SA-ConvLSTM	0.3503	0.1927	0.0497	0.3487	0.1867	0.0423	0.3315	0.1477	0.0278
SA-GRU	0.3608	0.1908	0.03457	0.3603	0.1963	0.0289	0.3394	0.1592	-0.0054
GA-ConvGRU	0.3544	0.1668	0.0406	0.3611	0.1827	0.0139	0.3529	0.1659	-0.0098
DyConvGRU	0.3734	0.2187	0.0606	0.3751	0.2178	0.0659	0.3608	0.1773	-0.0026
Our model	0.3857	0.2463	0.1189	0.3862	0.2446	0.1284	0.3780	0.2202	0.0844

parameters of the discriminators were truncated to the range from -0.01 to 0.01 after each update of their parameters. All experiments were implemented by PyTorch and ran on an NVIDIA TITAN RTX GPU.

C. EVALUATION METRICS

In the experiment, we used five radar echo images as input to predict the next fifteen radar echo images. To evaluate the performance of the models, we adopted four widely utilized metrics, namely Critical Success Index (CSI), Probability of Detection (POD), False Alarm Rate(FAR), and Heidke Skill Score (HSS). The higher POD, CSI, HSS, and lower FAR indicate better prediction accuracy. To calculate the metrics, the radar echo images were changed to binary maps according to a threshold τ . If the reflectivity is larger than the threshold, the value is set to true, otherwise, the value is set to false. By doing so, the predicted radar echo images and the ground truth are both transformed into a binary matrix, respectively. With the two matrices, we can calculate the

true positive (TP), false negative (FN), true negative (TN), and false positive (FP). Since it is important to show the model performance for different levels of radar reflectivities, we considered three thresholds which are the reflectivity 20dbz, 30dbz, and 40dbz. The definition of POD, FAR, CSI, and HSS are as follows:

$$CSI = \frac{TP}{TP + FP + FN} \tag{23}$$

$$POD = \frac{TP}{TP + FN} \tag{24}$$

$$FAR = \frac{FP}{TP + FP} \tag{25}$$

$$HSS = \frac{FN \times TN - FN \times FP}{(TP + FN)^2 + (TP + FP) \times (FP + FN)} \tag{26}$$

where TP denotes the number of true positive predictions (prediction is true and ground truth is true), FP indicates the number of false positive predictions (prediction is true and ground truth is false), TN is the number of true negative

TABLE 2. The CSI, FAR, POD, and HSS of the ablation experiments under the three thresholds (20dbz, 30dbz, and 40dbz) and at the three moments (T+30 min, T+60 min, and T+90 min). Bold values are the best results, underlined values are the second best results. In this table, W represents that the model is trained with WGAN’s training strategy, and R represents the model with the refinement module.

	T+30 min			T+60 min			T+90 min		
	20dbz	30dbz	40dbz	20dbz	30dbz	40dbz	20dbz	30dbz	40dbz
CSI↑									
ConvGRU	0.7662	0.5510	0.2620	0.7601	0.5336	0.2120	0.7440	0.4988	0.1161
DyConvGRU	0.7875	0.5788	0.3333	0.7896	0.5774	0.2825	0.7710	0.5387	0.2173
DyConvGRU+W	0.7910	0.5859	0.3199	0.7953	0.5845	0.2946	0.7732	0.5419	0.1886
DyConvGRU+R	<u>0.8054</u>	<u>0.6176</u>	0.4396	<u>0.8068</u>	<u>0.6163</u>	<u>0.4000</u>	<u>0.7881</u>	<u>0.5673</u>	<u>0.3076</u>
DyConvGRU+W+R	0.8108	0.6288	<u>0.4251</u>	0.8165	0.6351	0.4485	0.7970	0.5884	0.3310
POD↑									
ConvGRU	0.8516	0.7151	0.3284	0.8423	0.6875	0.2698	0.8365	0.6586	0.1478
DyConvGRU	0.8737	0.7417	0.4493	0.8739	0.7389	0.3619	0.8647	0.7162	0.3014
DyConvGRU+W	0.8816	0.7703	0.4274	<u>0.8872</u>	0.7715	0.4090	0.8605	0.7050	0.2445
DyConvGRU+R	0.8861	0.7719	0.5850	0.8868	<u>0.7752</u>	<u>0.5204</u>	<u>0.8815</u>	<u>0.7508</u>	0.4351
DyConvGRU+W+R	0.8941	0.8013	<u>0.5713</u>	0.9029	0.8162	0.6112	0.8823	0.7564	<u>0.4330</u>
FAR↓									
ConvGRU	0.1194	0.2995	0.4380	0.1178	0.3039	0.4161	0.1352	0.3383	0.3961
DyConvGRU	0.1143	0.2811	0.4213	0.1122	0.2813	0.3833	0.1276	0.3234	0.4465
DyConvGRU+W	0.1180	0.2955	0.4264	0.1182	0.2993	0.4111	0.1202	0.3072	<u>0.4220</u>
DyConvGRU+R	0.1039	0.2487	0.3516	0.1034	0.2546	0.3339	<u>0.1222</u>	<u>0.3088</u>	<u>0.4222</u>
DyConvGRU+W+R	<u>0.1054</u>	<u>0.2592</u>	<u>0.3687</u>	<u>0.1076</u>	<u>0.2636</u>	<u>0.3570</u>	0.1116	0.2810	0.3712
HSS↑									
ConvGRU	0.3640	0.1985	0.0579	0.3628	0.1916	0.0292	0.3482	0.1586	-0.0002
DyConvGRU	0.3734	0.2187	0.0606	0.3751	0.2178	0.0659	0.3608	0.1773	-0.0026
DyConvGRU+W	0.3725	0.2101	0.0532	0.3736	0.2066	0.0337	0.3658	0.1901	0.0122
DyConvGRU+R	<u>0.3849</u>	0.2505	0.1319	<u>0.3854</u>	0.2459	<u>0.1280</u>	<u>0.3691</u>	<u>0.1946</u>	<u>0.0315</u>
DyConvGRU+W+R	0.3857	<u>0.2463</u>	<u>0.1189</u>	0.3862	<u>0.2446</u>	0.1284	0.3780	0.2202	0.0844

predictions (prediction is false and ground truth is false), and *FN* represents the number of false negative predictions (prediction is false and ground truth is true).

D. RESULTS AND ANALYSIS

1) COMPARATIVE EXPERIMENTS

We randomly chose 40 sequences of radar echo images to evaluate the models and calculated the evaluation metrics. Table 1 shows the experimental results of our proposed model and other models. Besides, Fig.6 presents a relatively simple and intuitive example, which compared our proposed method with other approaches. From this experiment, it can be concluded as follows:

The main advantage of the proposed model can be shown in the prediction of high reflectivity regions. Compared with other methods, the proposed model obtains the highest CSI, HSS, and POD. And CSI and POD are improved significantly when the threshold is 40dbz. For FAR, our proposed model obtains the lowest score when the threshold is 20dbz and 30dbz and obtains the second-lowest score when the threshold is 40dbz. According to Fig.6, the optical flow based method can capture the trend of cloud motion in the radar echo images but has a poor ability to predict the variation such as the dissipation. At T+30 min and T+60 min, ConvLSTM, ConvGRU, SA-ConvLSTM, SA-GRU, and GA-ConvGRU underestimate the value of the regions which are chosen

by rectangles. At T+90 min, compare with other models, our proposed model can preserve the higher spatial similarity with ground truth and improve the prediction furthermore. In conclusion, compared with other models, our model can not only predict the motion of the cloud and the high reflectivity regions with a long lead time but also retains the details of the radar echo images.

2) ABLATION STUDY

To further illustrate the influence of the dynamic convolution, the refinement module, and the WGAN’s training strategy, the ablation experiments were conducted to validate the performance of the model with the dynamic convolution, the refinement module, and the WGAN’s training strategy, respectively. Table 2 shows the experimental results of these models. Here, DyConvGRU represents the ConvGRU with dynamic convolutions, W represents that the model is trained with WGAN’s training strategy, and R represents the model with the refinement module. For the convenience of reading, DyConvGRU which is trained with WGAN’s training strategy is denoted as WGA-DyConvGRU. Similarly, we plotted the visualization results of a sample in Fig.7. We can obtain the following conclusions:

Firstly, dynamic convolution is helpful to improve the prediction of those high echo reflectivity regions. By comparing the evaluation metric on the ConvGRU and DyConvGRU,

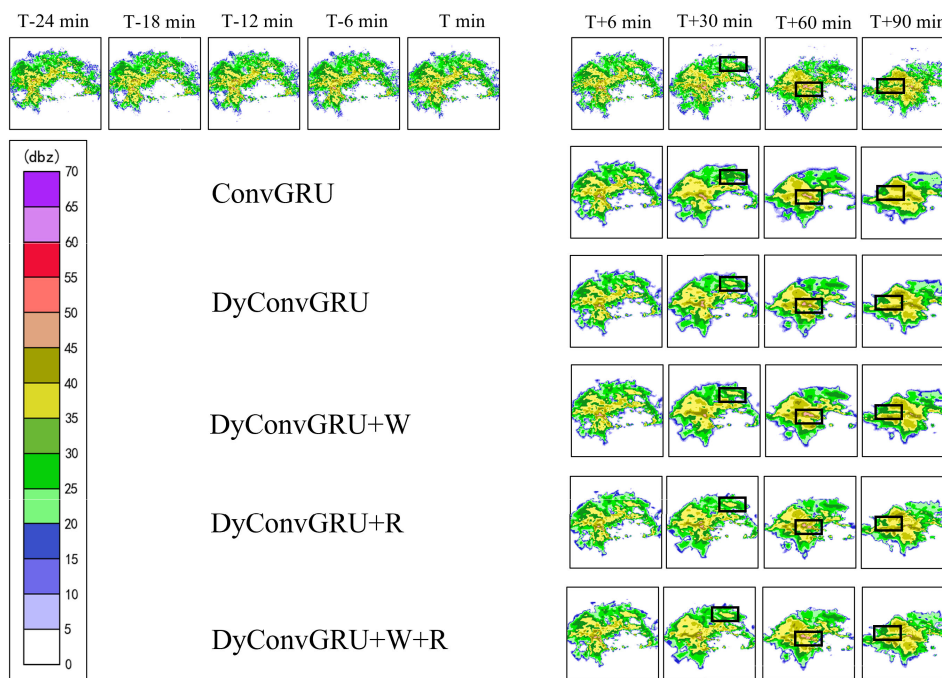


FIGURE 7. One of the results for ablation experiments. The radar echo images in the first line denote the input and the ground truth radar echo images at T+6 min, T+30min, T+60 min, and T+90 min, respectively. The images in the following lines are the corresponding prediction results of ConvGRU, DyConvGRU, DyCnovGRU+W, DyConvGRU+R, and DyConvGRU+W+R, respectively. In this figure, W represents that the model is trained with WGAN’s training strategy, and R represents the model with the refinement module.

we find that DyConvGRU has a better CSI and POD. For FAR and HSS, DyConvGRU obtains better scores except when the threshold is 40dbz at T+90 min. It means that the introduction of dynamic convolution contributes to improving the performance of prediction. From Fig.7, we observe that DyConvGRU generates a wider region with a higher value (yellow and red parts), which implies that dynamic Convolution can enhance the performance for the high region prediction.

Secondly, WGAN’s training strategy is helpful to improve the model’s ability to produce the shape details of radar echo images. According to Fig.7, we notice that the radar echo images generated by the models trained with WGAN’s training strategy can produce more details that are marked by a rectangle. It means that WGAN’s training strategy contributes to improving the details of the radar echo images. In Table 2, Compared with DyConvGRU, the evaluation of WGA-DyConvGRU had little improvements and was even worse in some cases. However, compared with DyConvGRU with a refinement module, WGA-DyConvGRU with a refinement module has a higher CSI and POD in most cases, which means that adopting the WGAN’s training strategy to train the DyConvGRU can help the refinement module correct the value of the rada echo images.

Thirdly, the refinement module is helpful to correct the values of the radar reflectivities. As Table 2 shows, DyConvGRU with a refinement module and WGA-DyConvGRU with a refinement module are better than the model without the

TABLE 3. The CSI and POD of the loss function experiments under the thresholds of 40dbz.

CSI↑	T+30 min	T+60 min	T+90 min
MSE+MAE	0.1274	0.0732	0.0294
Original loss	0.3186	0.2692	0.1847
Improved loss	0.3333	0.2825	0.2173
POD↑	T+30 min	T+60 min	T+90 min
MSE+MAE	0.1367	0.0784	0.0311
Original loss	0.4390	0.3421	0.2307
Improved loss	0.4493	0.3619	0.3014

refinement models among all the measure indexes. It means that the introduction of the refinement module contributes to improving the performance of prediction, especially in the reflectivity regions.

Fourthly, by integrating dynamic convolution, refinement module, and WGAN’s training strategy, the proposed method has further improved especially for the prediction in the position with a high reflectivity value. According to Fig.7, the proposed model produces wider yellow and red regions which means that it has a high probability of heavy rainfall. Table 2 shows the advantage of our model. As Table 2 shows, the proposed method which integrates dynamic convolution, refinement module, and WGAN’s training strategy is better than ConvGRU among all the measure indexes and thresholds.

3) LOSS FUNCTION STUDY

We trained DyConvGRU using MSE+MAE, the original loss function, and our suggested loss function with weights to better understand the role of our proposed loss function with weights. Meanwhile, at a threshold of 40 dbz, we calculated CSI and POD, and the results are shown in Table 3. According to Table 3, our proposed loss functions produce the greatest outcomes in both experiments. The defect that the high echo reflectivity accounts for a small percentage of the sample is compensated by stratifying the radar echo reflectivity more accurately and giving more weight to the high echo regions, allowing the model to better capture the characteristics of the high echo region in the sample.

V. CONCLUSION

In this paper, a forecast-refinement neural network based on DyConvGRU and U-Net is proposed to extrapolate radar echo images. It alleviates the problem that the high reflectivities fade as the lead time increase by introducing dynamic convolution, refinement module, and WGAN's training strategy. The proposed DyConvGRU which integrates dynamic convolution and ConvGRU uses multiple convolution strategies for different inputs and captures more information on high reflectivity regions in radar echo images than traditional convolution. But we found that the details of the results are lost. To predict the details of the radar echo images, the WGAN's training strategy is adopted in the training process of the forecast module. Then a forecast-refinement architecture is used in our model. The forecast module predicts future radar echo images based on the given observations and the refinement module is used to reconstruct the forecast results to correct the values of high reflectivities. The combination of these can enhance the accuracy of radar echo extrapolation, especially in high reflectivity regions. This work contributes to the prevention of natural disasters caused by convective weather. However, instead of end-to-end prediction, the method proposed in this paper runs prediction and refinement in two steps. Our loss function gives higher weights to the high echo region, which makes our model less effective for the low echo region. Although our method obtains higher CSI and POD but also has a higher FAR. In future work, we will combine prediction and refining into one procedure and use radar echo images to estimate precipitation distribution directly.

REFERENCES

- [1] C. Bai, F. Sun, J. Zhang, Y. Song, and S. Chen, "Rainformer: Features extraction balanced network for radar-based precipitation nowcasting," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [2] J. Bech and J. L. Chau, *Doppler Radar Observations: Weather Radar, Wind Profiler, Ionospheric Radar, and Other Advanced Applications*. Rijeka, Croatia: IntechOpen, Apr. 2012.
- [3] J. Sun, "Convective-scale assimilation of radar data: Progress and challenges," *Quart. J. Roy. Meteorol. Soc.*, vol. 131, no. 613, pp. 3439–3463, Oct. 2005.
- [4] M. Buehner and D. Jacques, "Non-Gaussian deterministic assimilation of radar-derived precipitation accumulations," *Monthly Weather Rev.*, vol. 148, no. 2, pp. 783–808, Feb. 2020.
- [5] N. E. H. Bowler, C. E. Pierce, and A. Seed, "Development of a precipitation nowcasting algorithm based upon optical flow techniques," *J. Hydrol.*, vol. 288, nos. 1–2, pp. 74–91, Mar. 2004.
- [6] W.-C. Woo and W.-K. Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 3, p. 48, Feb. 2017.
- [7] G. Ayzel, M. Heistermann, and T. Winterrath, "Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainmotion v0. 1)," *Geoscientific Model Develop.*, vol. 12, no. 4, pp. 1387–1402, 2019.
- [8] C. L. Giles, G. M. Kuhn, and R. J. Williams, "Dynamic recurrent neural networks: Theory and applications," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 153–156, Mar. 1994.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*.
- [11] Y.-M. Chiang, F.-J. Chang, B. J.-D. Jou, and P.-F. Lin, "Dynamic ANN for precipitation estimation and forecasting from radar observations," *J. Hydrol.*, vol. 334, nos. 1–2, pp. 250–261, Feb. 2007.
- [12] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
- [13] Q. K. Tran and S. K. Song, "Multi-channel weather radar echo extrapolation with convolutional recurrent neural networks," *Remote Sens.*, vol. 11, no. 19, p. 2303, Oct. 2019.
- [14] S. Zhong, X. Zeng, Q. Ling, Q. Wen, W. Meng, and Y. Feng, "Spatiotemporal convolutional LSTM for radar echo extrapolation," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 58–62.
- [15] J. Jing, Q. Li, X. Peng, Q. Ma, and S. Tang, "HPRNN: A hierarchical sequence prediction model for long-term weather radar echo extrapolation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4142–4146.
- [16] P. Xie, X. Li, X. Ji, X. Chen, Y. Chen, J. Liu, and Y. Ye, "An energy-based generative adversarial forecaster for radar echo map extrapolation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [17] L. Tian, X. Li, Y. Ye, P. Xie, and Y. Li, "A generative adversarial gated recurrent unit model for precipitation nowcasting," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 601–605, Apr. 2020.
- [18] K. Zheng, Y. Liu, J. Zhang, C. Luo, S. Tang, H. Ruan, Q. Tan, Y. Yi, and X. Ran, "GAN-argcPredNet v1.0: A generative adversarial model for radar echo extrapolation based on convolutional recurrent units," *Geoscientific Model Develop.*, vol. 15, no. 4, pp. 1467–1475, Feb. 2022.
- [19] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [20] Q. Zeng, H. Li, T. Zhang, J. He, F. Zhang, H. Wang, Z. Qing, Q. Yu, and B. Shen, "Prediction of radar echo space-time sequence based on improving TrajGRU deep-learning model," *Remote Sens.*, vol. 14, no. 19, p. 5042, Oct. 2022.
- [21] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.
- [22] Y. Wang, Z. Gao, M. Long, J. Wang, and S. Y. Philip, "PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5123–5132.
- [23] S. M. Bonnet, A. Evsukoff, and C. A. M. Rodriguez, "Precipitation nowcasting with weather radar images and deep learning in São Paulo, Brasil," *Atmosphere*, vol. 11, no. 11, p. 1157, Oct. 2020.
- [24] L. Chen, Y. Cao, L. Ma, and J. Zhang, "A deep learning-based methodology for precipitation nowcasting with radar," *Earth Space Sci.*, vol. 7, no. 2, Feb. 2020, Art. no. e2019EA000812.
- [25] J. Liu, L. Xu, and N. Chen, "A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images," *J. Hydrol.*, vol. 609, Jun. 2022, Art. no. 127748.
- [26] Z. Yang, H. Wu, Q. Liu, X. Liu, Y. Zhang, and X. Cao, "A self-attention integrated spatiotemporal LSTM approach to edge-radar echo extrapolation in the internet of radars," *ISA Trans.*, vol. 132, pp. 155–166, Jan. 2023.
- [27] Z. Lin, M. Li, Z. Zheng, Y. Cheng, and C. Yuan, "Self-attention ConvLSTM for spatiotemporal prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11531–11538.

- [28] C. Luo, X. Li, and Y. Ye, "PFST-LSTM: A SpatioTemporal LSTM model with pseudoflow prediction for precipitation nowcasting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 843–857, 2021.
- [29] X. Huang, C. Luo, Y. Ye, X. Li, and B. Zhang, "Location-refining neural network: A new deep learning-based framework for heavy rainfall forecast," *Comput. Geosci.*, vol. 166, Sep. 2022, Art. no. 105152.
- [30] T. Yu, Q. Kuang, and R. Yang, "ATMConvGRU for weather forecasting," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.
- [33] J. R. Jing, Q. Li, X. Y. Ding, N. L. Sun, R. Tang, and Y. L. Cai, "AENN: A generative adversarial neural network for weather radar echo extrapolation," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 89–94, Oct. 2019.
- [34] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas, and S. Mohamed, "Skillful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, Sep. 2021.
- [35] K. Zheng, Q. Tan, H. Ruan, J. Zhang, C. Luo, S. Tang, Y. Yi, Y. Tian, and J. Cheng, "GAN-argPredNet v2.0: A radar echo extrapolation model based on spatiotemporal process intensification," *Geosci. Model Develop. Discuss.*, vol. 2023, pp. 1–25, 2023.
- [36] B. Yang, G. Bender, Q. V. Le, and J. Ngiam, "CondConv: Conditionally parameterized convolutions for efficient inference," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.
- [37] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11027–11036.
- [38] N. Ballas, L. Yao, C. Pal, and A. Courville, "Delving deeper into convolutional networks for learning video representations," 2015, *arXiv:1511.06432*.
- [39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Munich, Germany: Springer*, 2015, pp. 234–241.
- [40] K. Trebing, T. Stanczyk, and S. Mehrkanoon, "SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture," *Pattern Recognit. Lett.*, vol. 145, pp. 178–186, May 2021.
- [41] A. Punjabi and P. I. Ayala, "Efficient spatio-temporal weather forecasting using U-Net," 2021, *arXiv:2112.06543*.
- [42] L. Han, H. Liang, H. Chen, W. Zhang, and Y. Ge, "Convective precipitation nowcasting using U-Net model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4103508.
- [43] J. Ko, K. Lee, H. Hwang, S.-G. Oh, S.-W. Son, and K. Shin, "Effective training strategies for deep-learning-based precipitation nowcasting and estimation," *Comput. Geosci.*, vol. 161, Apr. 2022, Art. no. 105072.
- [44] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.



JINLIANG YAO received the Ph.D. degree in computer application technology from the Institute of Automation, Chinese Academy of Sciences, in 2009. In 2009, he went to Hangzhou Dianzi University to conduct teaching and research work. During the study period, he participated in a number of national and provincial-level projects, published 20 papers at domestic and international journals and conferences, and obtained two national patents granted by the first inventor. His current research interests include pattern recognition and image processing. He received the Third Prize of the Zhejiang Science and Technology Progress Award.



FEIFAN XU was born in Zhejiang, China, in 1998. He received the bachelor's degree from China Jiliang University, in 2017. He is currently pursuing the master's degree with Hangzhou Dianzi University.



ZHENG QIAN was born in 1979. He is currently a Senior Engineer in meteorological information technology with the Ningbo Meteorological Service Center. His current research interests include the application of cutting-edge information technology in meteorology and dedicated to improve the degree of automation of meteorological services.



ZHIPENG CAI received the bachelor's degree from Hangzhou Dianzi University, in 2016, where he is currently pursuing the master's degree.

• • •