GAN-argcPredNet v2.0: A Radar Echo Extrapolation Model based on Spatiotemporal Process Enhancement

Kun Zheng[1*], Qiya Tan[1*], Huihua Ruan[2], Jinbiao Zhang[2], Cong Luo[3], Siyu Tang[3], Yunlei Yi[4], Yugang Tian[1], Jianmei Cheng[5]

[1]School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430074, China
[2]Guangdong Meteorological Observation Data Center, Guangzhou, 510080, China
[3]Guangdong Meteorological Observatory, Guangzhou, 510080, China
[4]Wuhan Zhaotu Technology Co. Ltd., Wuhan, 430074, China
[5]School of Environmental Studies, China University of Geosciences, Wuhan, 430074, China
*These authors have contributed equally to this work

Correspondence to: Kun Zheng (ZhengK@cug.edu.cn); Huihua Ruan (ruanhuihua@163.com)

Abstract. Precipitation nowcasting has important implications for urban operation and flood prevention. Radar echo extrapolation is a common method in precipitation nowcasting. Using deep learning models to extrapolate radar echo data has great potential. The increase of lead time leads to a weaker correlation between the real rainfall evolution and the generated images. The evolution information is easily lost during extrapolation, which is reflected as echo attenuation. Existing models, including Generative Adversarial Network (GAN)-based models, are difficult to curb attenuation, resulting in insufficient accuracy in rainfall prediction. ~~Aim to the problem~~ To solve this issue, a Spatiotemporal Process Enhancement Network (GAN-argcPredNet v2.0) based on GAN-argcPredNet v1.0 is designed. GAN-argcPredNet v2.0 curbs attenuation by avoiding blurring or maintaining the intensity. A Spatiotemporal Information Changes Prediction (STIC-Prediction) network is designed as the generator. By suppressing the blurring effect of rain distribution and reducing the negative bias by STIC Attention, the generator generates more accurate images. Furthermore, the discriminator is a Channel-Spatial Convolution (CS-Convolution) network. The discriminator enhances the discrimination of echo information and provides better guidance to the generator in image generation by CS Attention. The experiments are based on the radar dataset of South China. The results show that GAN-argcPredNet v2.0 performs better than other models. In heavy rainfall prediction, compared with the baseline, the Probability of Detection (POD), the Critical Success Index (CSI), the Heidke Skill Score (HSS) and Bias Score increase by 18.8 %, 17.0 %, 17.2% and 26.3 % respectively. The False Alarm Ratio (FAR) decreases by 3.0 %.

1 Introduction

Accurate precipitation nowcasting, especially heavy precipitation nowcasting, plays a key role in hydrometeorological applications such as urban-operation safety and flash-flood warnings (Liu et al., 2015). It can effectively prevent the hazards and losses caused by heavy precipitation to economy and people (Luo et al., 2020). Radar echo extrapolation is the method most often used to nowcast precipitation (Reyniers, 2008). The essence is tracking areas of reflectivity to derive motion vectors, and then using the motion vectors to determine future location of the reflectivity (Austin and Bellon, 1974).

Traditional radar echo extrapolation methods include cross-correlation, individual radar echo-tracking, and the optical flow method (Bowler et al., 2004). Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) is a classical centroid tracking algorithm (Dixon and Wiener, 1993). This algorithm achieves precipitation nowcasting by real-time tracking and automatic identification of individual storms. However, the tracking performance of TITAN is poor during multi-cell storms. To address this, an enhanced TITAN (ETITAN) was proposed (Han et al., 2009). By combining cross-correlation and individual radar echo-tracking, ETITAN achieves more accurate tracking and prediction. While the cross-correlation method is effective, it has lower prediction accuracy when echoes change rapidly. On the other hand, the optical flow method achieves local prediction by treating echoes motion as fluid (Sakaino, 2013). Additionally, some traditional nowcasting systems combined different information and further improved the ability of nowcasting. A Bayesian precipitation nowcasting system based on the ensemble Kalman filter was formulated. The system correctly captures the flow dependence of both the numerical weather prediction (NWP) forecast and the Lagrangian persistence of radar observations (Nerini et al., 2019). Furthermore, the variational algorithm is used to improve the nowcasting system to achieve three-hour nowcasting (Chung and Yao, 2020). The Lagrangian Integro-Difference equation model with Autoregression (LINDA) also performs better for prediction accuracy and duration (Pulkkinen et al., 2021). As the storm evolution like merging, splitting, growth and decay, traditional methods are difficult to predict accurately. Besides, these traditional methods do not intend to utilize the large amounts of historical images.

Deep learning has powerful nonlinear mapping ability. By analyzing the motion process through a large number of historical radar echo images, deep learning achieves better results (Shi et al., 2015; Pan et al., 2021). Radar echo extrapolation can be regarded as an image sequence prediction problem. Therefore, the problem can be solved by implementing an end-to-end sequence learning method (Sutskever et al., 2014; Shi et al., 2015). ConvGRU learns video features through convolution operation, enabling sparse connection of model units (Ballas et al., 2015). Convolution operation is also used in ConvLSTM. By replacing the step of internal data state transformation in LSTM, ConvLSTM can better extract features (Shi et al., 2015). Convolutional recursive structure is position invariant, which is not consistent with natural motion and transformation. Trajectory GRU (TrajGRU) was further proposed (Shi et al., 2017). Both LSTM and GRU models have long-term memory. However, this capability is limited to historical spatial information. RainNet utilizes a convolutional network architecture in precipitation nowcasting, which avoids the brittleness of LSTM structure (Ayzel et al., 2020).

Attention mechanisms are also frequently employed in sequential networks. By learning the importance of different image parts, attention mechanisms can improve prediction accuracy. For example, the self-attention mechanism combines the spatial relationships of different locations and emphasizes important areas (Wang et al., 2018a). Eidetic 3D LSTM (E3D-LSTM) introduces self-attention to enhance long-term memory in LSTM (Wang et al., 2018b). However, it lacks attention in the channel dimension. Interaction Dual Attention LSTM (IDA-LSTM) expands the spatial and channel attention based on self-attention to improve representation learning (Luo et al., 2021). Due to the high hardware load, self-attention is hard to train for high-resolution images. Convolutional Block Attention Module (CBAM) was developed simultaneously as a less computational attention mechanism. It can be flexibly applied in sequential networks (Woo et al., 2018).

Compared to sequence prediction networks, Generative Adversarial Network (GAN)-based models have significant advantages in generating high-quality echo images. High quality refers to images that are more realistic and structurally similar to real images (Tian et al., 2020; Xie et al., 2022). GANs consist of a generator and a discriminator. The generator is responsible for generating new synthetic data that follows the distribution of the training data. The discriminator is trained to distinguish between samples generated by the generator and real samples from the training set. The generator and discriminator are trained against each other to achieve balance. GANs have powerful data generation capabilities (Goodfellow et al., 2020). This is because the model with anti-loss can better realize multi-modal modeling (Lotter et al., 2016). For instance, Deep Generative Models of Rainfall (DGMR) generates more accurate reflectivity by adversarial training (Ravuri et al., 2021). GANs are also used to generate realistic details for a broader extrapolation range (Chen et al., 2019). GA-ConvGRU uses ConvGRU as the generator. The image quality is far better than ConvGRU by implementing multi-modal data modeling (Tian et al., 2020). A number of studies contribute to improving the stability of GAN training. Energy-Based Generative Adversarial Forecaster (EBGAN-Forecaster) combines a convolution structure with a codec framework to improve stability (Xie et al., 2022). Additionally, our proposed GAN-argcPredNet v1.0 has more advantages in improving the details of predicted echoes and stabilizing GAN training (Zheng et al., 2022).

However, the radar echo images are forecasted for future time periods based on the real echo sequence. In deep learning models, the increase of lead time leads to a weaker correlation between the real images at the front of the sequence and the generated images. The influence of the real echo evolution diminishes rapidly, resulting in the loss of rainfall evolution information. This loss is reflected as the attenuation of echo shape and intensity in the generated images. Due to the smaller percentage of heavy rainfall areas, the attenuation is ~~more severe~~ severer. To the knowledge of authors, existing deep learning models, including GAN-based models, lack a method to curb this attenuation, which leads to low accuracy in predicting heavy rainfall.

In this study, a Spatiotemporal Process Enhancement Network (GAN-argcPredNet v2.0) was proposed based on Generative Adversarial Advanced Reduced-Gate Convolutional Deep Predictive Coding Network (GAN-argcPredNet v1.0) (Zheng et al., 2022), which aims at curbing attenuation. In GAN-argcPredNet v2.0, a Spatiotemporal Information Changes Prediction (STIC-Prediction) network is designed as the generator. The generator focuses on the spatiotemporal variation of the radar echo feature sequence. ~~The purpose of the generator is to curb echo attenuation by suppressing the blurring effect of rain distribution and reducing the negative bias~~ The purpose of the generator is to more accurately forecast future precipitation distributions by curbing echo attenuation. Furthermore, a Channel-Spatial Convolution (CS-Convolution) network is designed as the discriminator. The discriminator aims to guide the generator to better retain echo shape and intensity by enhancing the ability to identify echo information. The generator and discriminator are trained against each other to achieve accurate rainfall prediction.

2 Model

2.1 GAN-argcPredNet v1.0 overview

GAN–argcPredNet v1.0 (Fig. 1) uses Wasserstein GAN with gradient penalty (WGAN-gp) as its predictive framework. The model solves the problem of training instability by utilizing gradient penalty measures (Gulrajani et al., 2017). The generator in GAN-argcPredNet v1.0 is an argcPredNet model, responsible for learning the data features of rainfall and simulating the real echo distribution to generate predictive images. The discriminator is a four-layer CNN model with a dual-channel input method. The predictive and real images are fed into the discriminator, which judges the real images as true and the predictive images as false. Adam is used as the optimizer, which is an extension of stochastic gradient descent (Kingma Diederik and Adam, 2014). The model parameters are updated through adversarial loss optimization. The generator is updated once after every five updates of the discriminator.

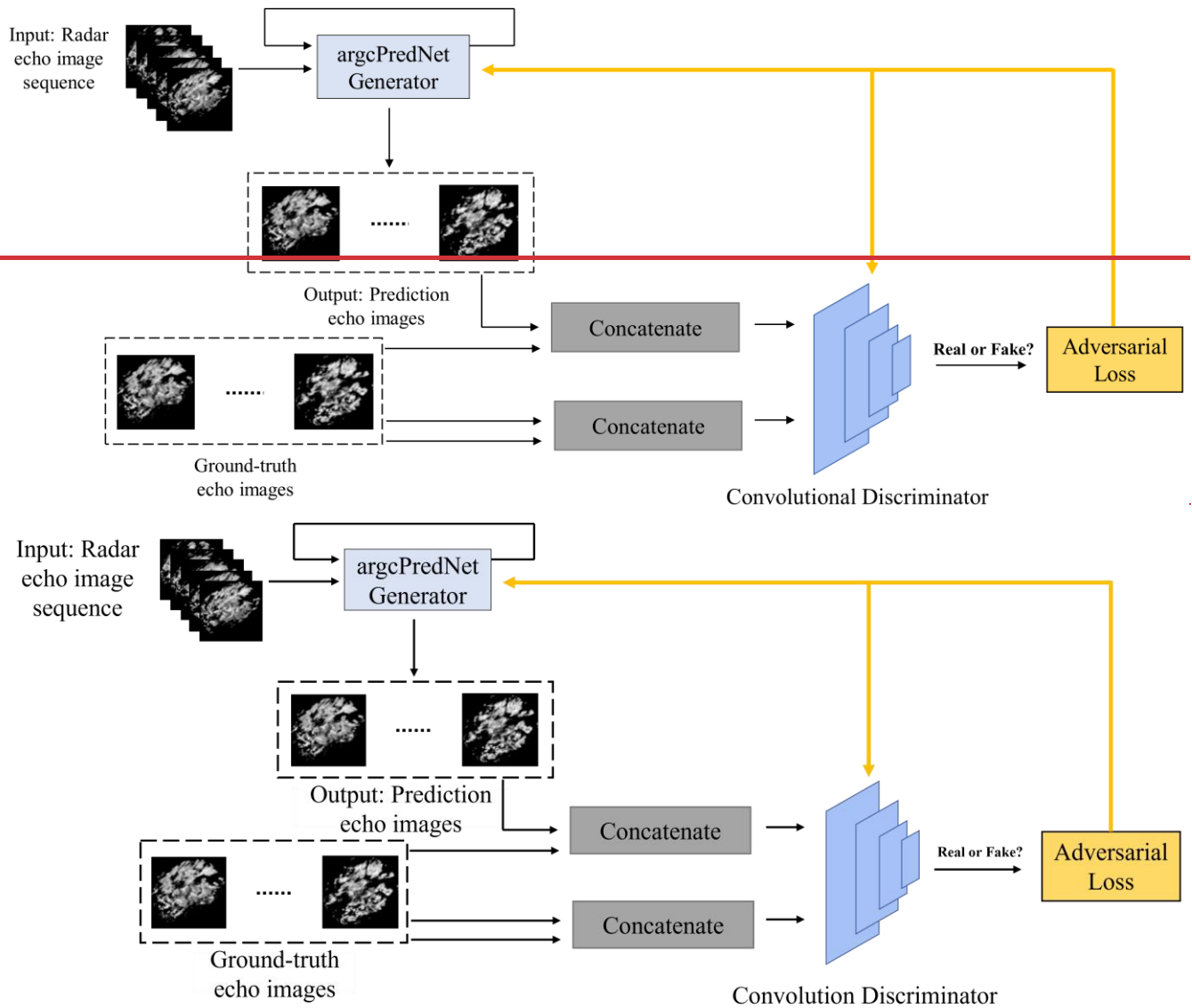



Figure 1: The structure of GAN-argcPredNet v1.0. More information about the model can be found in the paper of Zheng et al. (Zheng et al., 2022).

2.2 GAN-argcPredNet v2.0 overview

GAN-argcPredNet v2.0 model was built based on GAN-argcPredNet v1.0. GAN-argcPredNet v2.0 consists of a STIC-Prediction generator and a CS-Convolution discriminator (Fig. 2). The STIC-Prediction generator is designed to reduce echo attenuation ~~by suppressing the blurring effect and reducing the negative bias. The generator~~ and ~~is composed~~ consists of the argcPredNet and the STIC Attention module (Fig. 3). The argcPredNet is composed of a series of repeatedly stacked modules, with a total of four layers (Zheng et al., 2022). Each layer of the module consists of the input convolutional layer (A_l), the

recurrent representation layer (R_l), the prediction convolutional layer ($\hat{A}_l$) and the error representation layer (E_l). R_l learns image features and generates the feature map $R_l^T \in R^{H \times W \times C}$, where l, T, H, W and C denote layer, current prediction time, map height, map width and feature channel, respectively. The feature map guides the lower layers to generate images. STIC Attention is designed to focus on the importance of different rainfall regions after the third layer. The purpose is to better maintain spatiotemporal features during information transmission within the model. After passing through STIC Attention, the features are fed to the lower layers, aiming to avoid blurring or to maintain the intensity during extrapolation. The calculation method of STIC-Prediction is:

$$A_l^T = \begin{cases} x_T & if\ l = 0 \\ MAXPOOL(\gamma(f(E_{l-1}^T))) & 0 < l < L \end{cases}, \tag{1}$$

$$\hat{A}_l^T = \gamma(f(R_l^T)), \tag{2}$$

$$E_l^T = \left[\gamma(A_l^T - \hat{A}_l^T)\ ;\ \gamma(\hat{A}_l^T - A_l^T) \right], \tag{3}$$

$$R_l^T = \begin{cases} argcLSTM(E_l^{T-1},\ R_l^{T-1}) & if\ l = L \\ argcLSTM(E_l^{T-1},\ R_l^{T-1},\ UPSAMPLE(STIC(R_{l+1}^0 : R_{l+1}^T))) & if\ l = 1 \\ argcLSTM(E_l^{T-1},\ R_l^{T-1},\ UPSAMPLE(R_{l+1}^T)) & l = 0\ and\ 1 < l < L \end{cases} \tag{4}$$

Here, x_T denotes the initial input, $MAXPOOL$ denotes the maximum pooling operation, γ denotes relu activation function, f denotes the convolution operation, $argcLSTM$ denotes Advanced Reduced-Gate Convolutional LSTM (Zheng et al., 2022), and $STIC$ denotes STIC Attention.

The CS-Convolution discriminator is composed of a four-layer convolution structure and a CS Attention module. The convolution structure is responsible for extracting the echo features from the input radar echo images. CS Attention is embedded after the first-layer convolution structure. This module is designed to focus on radar echo features, especially heavy rainfall features. The purpose is to enhance discriminative ability and provide better guidance for the generator. The hyperparameters of the generator and discriminator are provided in the supplement (Table S7 and S8).

Input: Radar echo image sequence

STIC-Prediction Generator

Ground-truth echo images

Output: Prediction echo images

Concatenate

Concatenate

CS Attention module

CS-Convolution Discriminator

Real or Fake?

Adversarial Loss

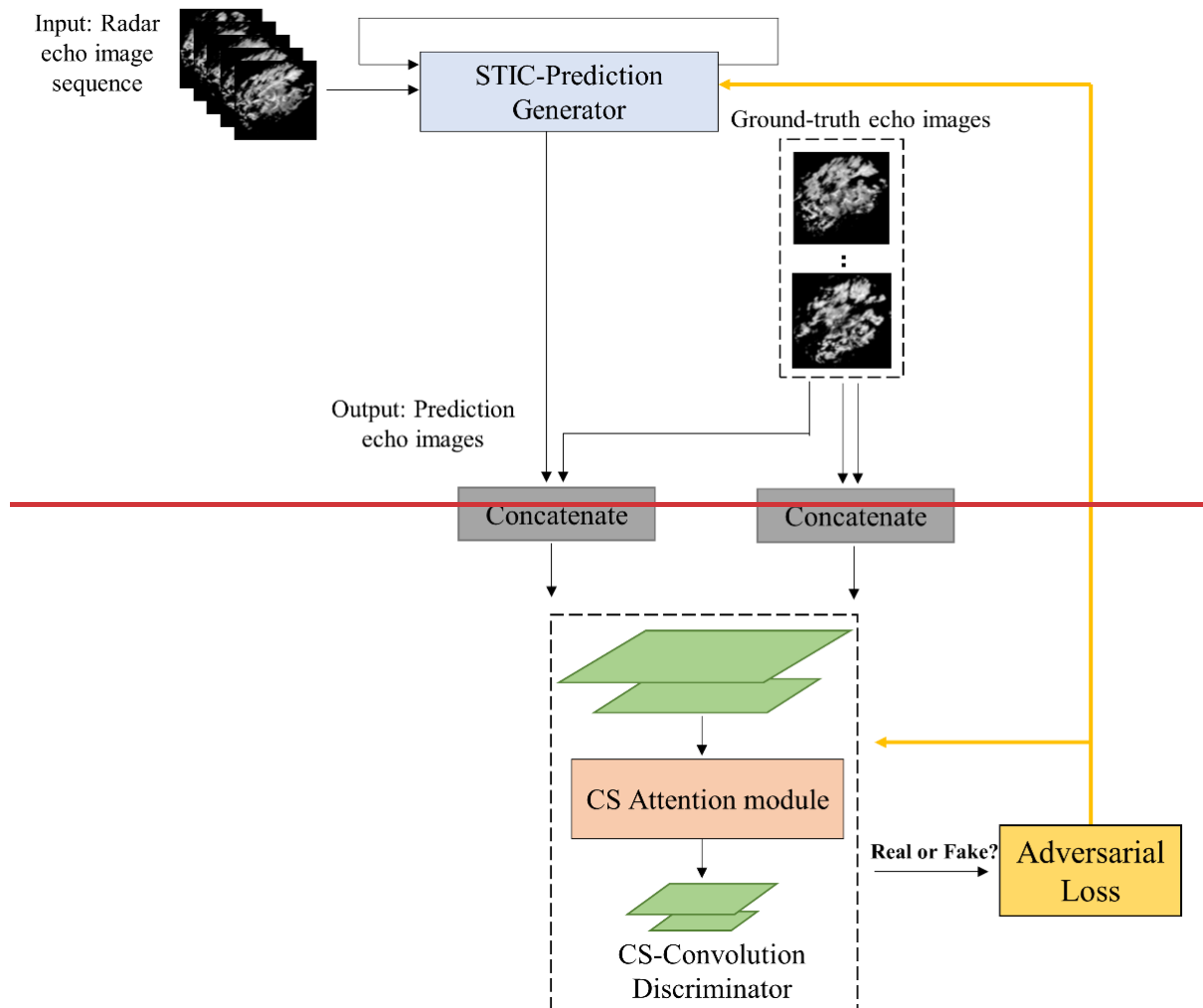
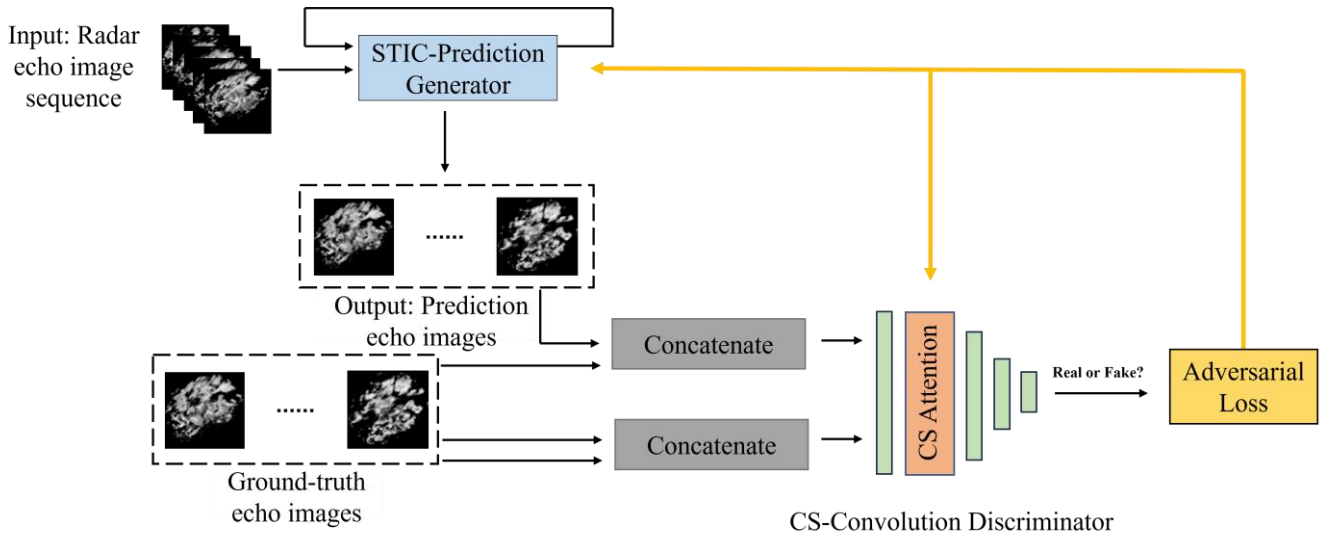



Figure 2: The structure of GAN-argcPredNet v2.0. Fifteen radar echo images are used as a test sequence. Here, the first five images are used as the input sequence, and the last ten images are used as the Ground-truth images. The generator generates images based on input sequences. The discriminator judges both the predictive images and Ground-truth images. The adversarial loss is then obtained to optimize the model.

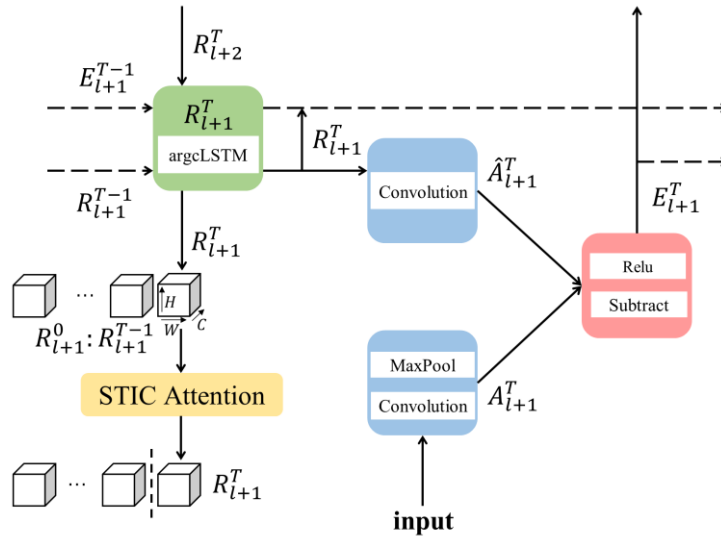



Figure 3: ~~The STIC-Prediction structure at time T with $l = 1$ ($l = 0, 1, 2, 3$). STIC Attention is after the third layer. Before R_{l+1}^T enters the R_l^T module, R_{l+1}^T forms a sequence $R_{l+1}^0 : R_{l+1}^T$ with the previous R_{l+1}. STIC Attention focuses on the importance of different regions of the sequence. Finally, the new R_{l+1}^T is fed to the R_l^T module. See Section 2.2 for further explanation.~~ The STIC-Prediction structure at time T with four layers ($l=0, 1, 2, 3$). The STIC Attention is located between the second layer ($l=1$) and the third layer ($l=2$). During prediction, R_{l+1}^T forms the sequence $R_{l+1}^0 : R_{l+1}^T$ with R_{l+1} prior to the moment T. Then, the sequence fed into the STIC module, which captures the correlation between the sequences and adjusts R_{l+1}^T. Finally, the new R_{l+1}^T is output. See Section 2.2 for further explanation.

2.3 STIC Attention

STIC Attention (Fig. 4) combines MaxPool3D (3D = map height, map width and time) and AvgPool3D to focus on the spatial information in feature sequences from both the maximum and average perspectives. This step focuses on heavy rainfall echoes while considering non-heavy rainfall. The introduction of 3D convolution enables extraction of spatiotemporal changes in the feature sequences. The module calculates the importance of evolutionary information~~, which aims to suppress the blurring effect and reduce the negative bias~~. Following are the detailed steps.

Given the feature sequence $F \in R^{t \times H \times W \times C}$ as input, where t denotes time. Two feature sequences $F_{max}^{ts} \in R^{t \times H \times W \times 1}$ and $F_{avg}^{ts} \in R^{t \times H \times W \times 1}$, are obtained by pooling operations, which denote the maximum and average feature along the channel axis, respectively. The feature sequences are then connected and 3D convoluted. The activation function is hard_sigmoid. By introducing linear behavior, hard_sigmoid allows gradients to flow easily in the unsaturated state, and provides a crisp decision in the saturated state, resulting in far less computational expense (Courbariaux et al., 2015; Gulcehre et al., 2016; Nwankpa et al., 2018). Then, the STIC Attention map sequence $M_{STIC} \in R^{t \times H \times W \times 1}$ is obtained. Finally, the output feature sequence $F_1 \in R^{t \times H \times W \times C}$, is calculated by element-wise multiplication of M_{STIC} and F. In short, the calculation method of STIC Attention is:

$$M_{STIC} = \sigma(f_{7 \times 7 \times 5}((MaxPool3D(F))concat(AvgPool3D(F)))) = \sigma(f_{7 \times 7 \times 5}(F_{max}^{ts} concat F_{avg}^{ts})) , \qquad (5)$$

$$F_1 = M_{STIC} \otimes F , \qquad (6)$$

Here, $concat$ denotes the connection operation, $MaxPool3D$ denotes the 3D maximum pooling operation, $AvgPool3D$ denotes the 3D average pooling operation, $f_{7 \times 7 \times 5}$ denotes the 3D convolution operation with a convolution kernel of 7×7×5, σ denotes the hard_sigmoid activation function, and $\otimes$ denotes the element-wise multiplication.

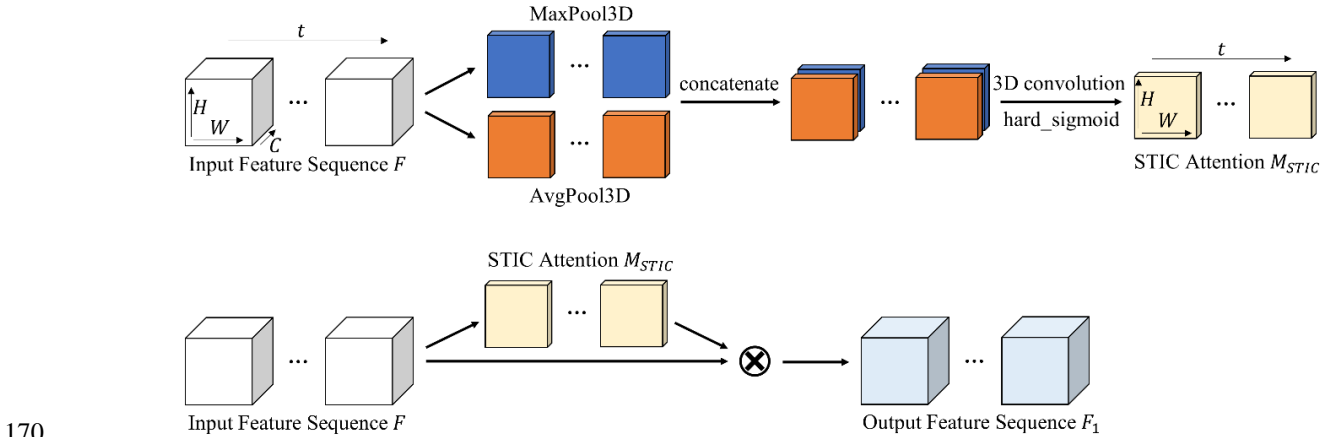



Figure 4: The structure of STIC Attention. The first line represents the calculation process of the STIC Attention map sequence. The second line represents the process of applying the STIC Attention map sequence to the input feature sequence.

2.4 CS Attention

CS Attention consists of Channel Attention and Spatial Attention (Fig. 5). For input feature $F' \in R^{H \times W \times C}$, the channel attention map $M_c \in R^{1 \times 1 \times C}$ is first generated. After element-wise multiplication with initial feature image, the spatial attention map $M_s \in R^{H \times W \times 1}$, is generated by the Spatial Attention module. Finally, the output feature $F'_2 \in R^{H \times W \times C}$ is obtained in the same way. In short, the calculation process is as follows:

$$F'_1 = M_c \otimes F' , \tag{7}$$

$$F'_2 = M_s \otimes F'_1 , \tag{8}$$

Here, $\otimes$ denotes the element-wise multiplication, and F'_1 is the middle feature.

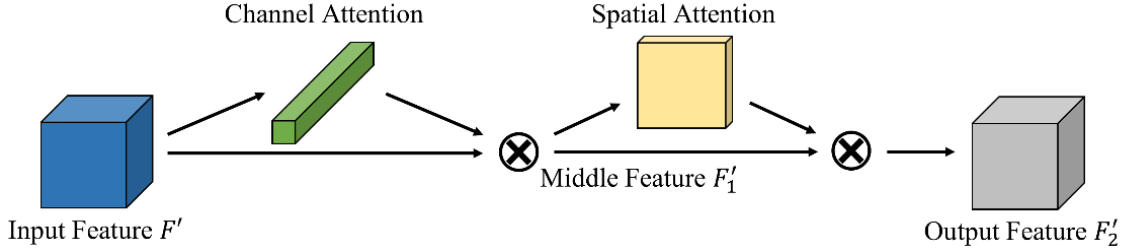



Figure 5: The structure of CS Attention. The calculation order is first through the Channel Attention module and then through the Spatial Attention module.

The Channel Attention module (Fig. 6) studies the relationship among different feature channels. The global maximum and average pooling are used to gather spatial maximum and average information for each channel. The combination of these methods allows for a more comprehensive judgment of the importance of different feature channels. Then, the correlation between feature channels is obtained by learning the respective parameters in a dense layer. The Channel Attention module focuses more on meaningful feature channels. The detailed steps are as follows:

The feature map F' is input into the Channel Attention module. Two 1D feature maps $F^c_{max} \in R^{1 \times 1 \times C}$ and $F^c_{avg} \in R^{1 \times 1 \times C}$ are obtained by applying global pooling, which denote the global maximum and global average pooling features, respectively. Then, the correlation between these features is extracted through dense layers. In order to reduce parameter overhead, the number of neurons in the first dense layer is set to C/r, where r is the compression ratio. Finally, the hard_sigmoid activation function is applied to obtain final channel attention map M_c. In short, the channel attention map is calculated as follows:

$$M_c = \sigma(\varphi_{11}(\gamma(\varphi_{10}(GMP(F'))))+\varphi_{21}(\gamma(\varphi_{20}(GAP(F')))))=\sigma(\varphi_{11}(\gamma(\varphi_{10}(F^c_{max})))+\varphi_{21}(\gamma(\varphi_{20}(F^c_{avg})))), \tag{9}$$

Here, GMP denotes the global maximum pooling, GAP denotes the global average pooling, $\varphi_{10}, \varphi_{11}$ denotes the first and second dense layer of F^c_{max}, $\varphi_{20}, \varphi_{21}$ denotes the first and second dense layer of F^c_{avg}, and γ denotes the relu activation function.

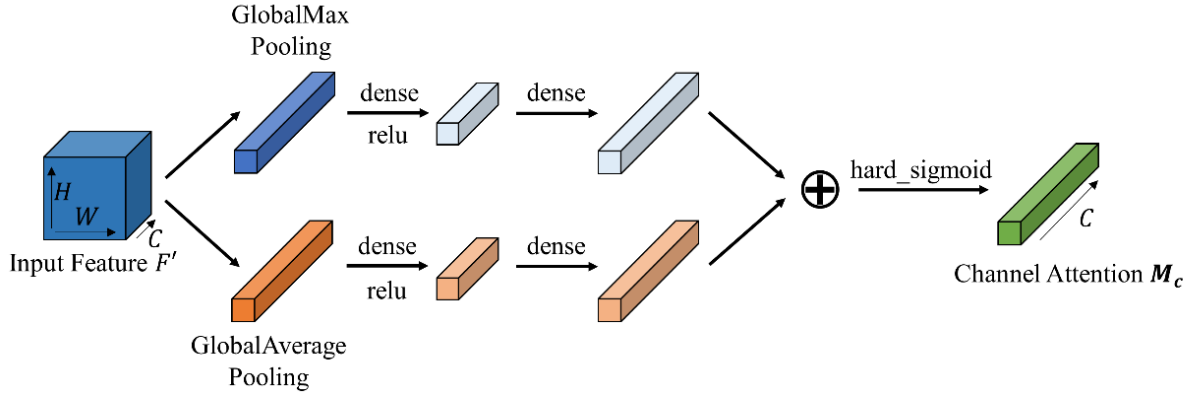



Figure 6: The structure of Channel Attention. Channel Attention utilizes maximum pooling outputs and average pooling outputs with their respective networks.

The Spatial Attention module (Fig. 7) studies the importance of each part in the same channel. The maximum and average pooling are used along the channel axis, which obtains echoes information of the feature image. The 2D convolution operation extracts feature and generates a spatial attention map with the same size as the input image. The detailed steps are as follows: After the Channel Attention module, the feature map $F_1' \in R^{H \times W \times C}$ is input into the Spatial Attention module. Two 2D feature maps $F_{max}^s \in R^{H \times W \times 1}$ and $F_{avg}^s \in R^{H \times W \times 1}$, are obtained by applying pooling operation, which denote the maximum pool feature and average pool feature on the channel respectively. The feature maps are then connected and 2D convoluted, using hard_sigmoid as the activation function to obtain final spatial attention map M_s. In short, the calculation method of spatial attention map is:

$$M_s = \sigma(f_{7\times7}((MaxPool2D(F_1'))concat(AvgPool2D(F_1')))) = \sigma(f_{7\times7}(F_{max}^s concat F_{avg}^s)), \tag{10}$$

Here, $MaxPool2D$ denotes the 2D maximum pooling operation, $AvgPool2D$ denotes the 2D average pooling operation, and $f_{7\times7}$ denotes the 2D convolution operation with a convolution kernel of 7×7.

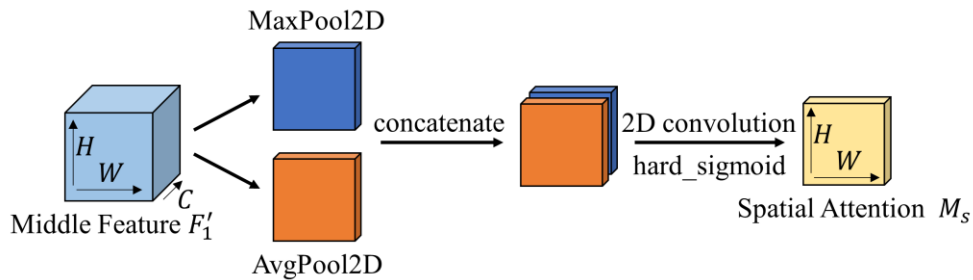



Figure 7: The structure of Spatial Attention. Spatial Attention utilizes the maximum and average pooling outputs along the channel axis and forwards them to the convolutional layer.

3 Data and experimental setup

3.1 Dataset description

This study used the South China radar echo data provided by Guangzhou Meteorological Administration. The radar mosaic comes from 11 weather radars. The median filtering algorithm was used to control radar data quality, which eliminates errors caused by isolated clutter. In addition, the mirror filling and continuity checks were applied to remove traditional radar error sources. After quality control, the data contains only an extremely small amount of strong interference, which has a negligible impact on the training of the model.

From 2015 to 2016, a total of 32,010 consecutive echo images with rainfall were randomly selected as the training set. For the testing phase, 7,995 consecutive images were randomly selected from March to May 2017. The original resolution of each image is 1050 × 880 pixels, covering an area of 1050 km × 880 km. Each pixel denotes a resolution of 1 km × 1 km. The reflectivity values range from 0 to 80 dBZ, with the amplitude limit set between 0 and 255. The data was collected every 6 minutes, which is the Constant Altitude Plan Position Indicator (CAPPI) data at a height of 1 km. To speed up training and reduce hardware load, the central 128 × 128 images were segmented.

Due to the relationship between radar reflectivity and rainfall type, the value on the radar echo image is converted to the corresponding rainfall rate. The calculation formula is as follows:

$$Z = 10 \, log_{10} \, a + 10b \, log_{10} \, R \,, \tag{11}$$

Here, a is set to 58.53 and b is set to 1.56, Z denotes the radar reflectivity intensity, R denotes the rainfall intensity. The correspondence between rainfall rate, radar reflectivity intensity and rainfall level are referred to Table 1 (Shi et al., 2017).

Table 1: The rainfall level table. Rain Rate, Radar Reflectivity Intensity and Rainfall Level correspond to each other.

Rain Rate (mm h^{-1})	Radar Reflectivity Intensity (dBZ)	Rainfall Level
$0 \leq R < 0.5$	$Z < 12.98$	No / Hardly noticeable
$0.5 \leq R < 2$	$12.98 \leq Z < 22.37$	Light
$2 \leq R < 5$	$22.37 \leq Z < 28.58$	Light to moderate
$5 \leq R < 10$	$28.58 \leq Z < 33.27$	Moderate
$10 \leq R < 30$	$33.27 \leq Z < 40.72$	Moderate to heavy
$30 \leq R$	$40.72 \leq Z$	Rainstorm warning

3.2 Evaluation metrics

As for evaluation, the study used four metrics to evaluate the prediction accuracy of all 128 × 128 pixels, which are Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI) and Heidke Skill Score (HSS). POD evaluates hit ability, while FAR is the metric of false alarms. The combination of them can evaluate the model more objectively. CSI and HSS are composite metrics that provide a direct judgment of the model effectiveness. CSI measures the fraction of observed

and/or forecast events that are correctly predicted. HSS measures the fraction of correct forecasts after eliminating those
forecasts which would be correct due purely to random chance. To measure the blurring effect, the study also used the Bias Score, which evaluates the ratio between the frequency of forecast events and the frequency of observed events. The formulas for calculating these five metrics are as follows:

$$POD = \frac{TP}{TP+FN} , \tag{12}$$

$$FAR = \frac{FP}{TP+FP} , \tag{13}$$

$$CSI = \frac{TP}{TP+FN+FP} , \tag{14}$$

$$HSS = \frac{2(TP \times TN - FN \times FP)}{(TP+FN)(FN+TN)+(TP+FP)(FP+TN)} , \tag{15}$$

$$Bias = \frac{TP+FP}{TP+FN} , \tag{16}$$

Here, TP denotes that the real and predicted value reach specified threshold, FN denotes that the real value reaches the specified threshold and the predicted value does not reach, FP denotes that the real value does not reach specified threshold and the predicted value reaches, and TN denotes that the real value and predicted value do not reach specified threshold. In the study, we applied threshold rain rates of 0.5, 2, 5, 10 and 30mm h^{-1} for calculating these metrics.

In order to evaluate the quality of generated images objectively, Mean Square Error (MSE) and Mean Structural Similarity (MSSIM) were also chosen for the experiment (Wang et al., 2004; Inoue and Misumi, 2022).

3.3 Experimental setting

Radar echo extrapolation is the prediction of future radar echo images based on real images. This study set the input sequence M and output sequence N to 5 and 10, respectively.

GAN-argcPredNet v2.0 was first compared with the optical flow, ConvLSTM, ConvGRU, GA-ConvGRU, and GAN-argcPredNet v1.0 in comparison experiments. The first one is a traditional method, and the code comes from the ~~Pysteps~~ pySTEPS library (Pulkkinen et al., 2019), which we performed using a local tracking approach (Lucas-Kanade). The last one is the model we designed before. Other models are commonly used deep learning models in radar echo extrapolation. The hyperparameters for deep learning models were provided in the supplement (Table S1-6). Then the ablation study was designed to verify the effectiveness of STIC Attention and CS Attention.

Before training, each pixel of the radar echo image was normalized to [0, 1]. All experiments were implemented by Python. Model training and testing were performed using the Keras deep learning library with Tensorflow as the backend. The operating environment is a Linux workstation equipped with two NVIDIA RTX 2080 Ti 11G GPUs.

4 Results

4.1 Comparison study

In order to observe the performance of the models and evaluate the blurring effect more easily, the average scores of POD, FAR, CSI, HSS and Bias were calculated for ~~all~~ 30 and 60 minutes lead time in the experiment. Fig. 8 shows the average scores of the metrics used for predicting 30 minutes, and the scores of GAN-argcPredNet v2.0 are better than the other models except Bias Score, especially the thresholds 10 and 30 mm h^{-1}, which indicates that GAN-argcPredNet v2.0 has a significant advantage in forecasting heavy precipitation. In Bias Score, GAN-argcPredNet v2.0 is second only to the optical flow and better than other deep learning models. The optical flow extrapolates the echo motion as linear, which is not realistic and leads to poorer forecasts. ~~Fig. 8 shows that,~~ Fig. 9 shows the average scores of the metrics used for predicting 60 minutes, except for the Bias Score, most deep learning models have better scores than the optical flow. Besides the threshold of 30 mm h^{-1}, the Bias Score of the optical flow is second only to GAN-argcPredNet v2.0. In deep learning models, the FAR Score of GAN-argcPredNet v2.0 is slightly higher than those of ConvGRU and ConvLSTM at thresholds of 0.5 and 2 mm h^{-1}, respectively. However, the other scores of GAN-argcPredNet v2.0 are always the best. The comprehensive score of GAN-argcPredNet v1.0 is second only to GAN-argcPredNet v2.0. GA-ConvGRU performs better than the two non-GAN models in most cases. Between the two non-GAN models, all scores of ConvLSTM are better than ConvGRU when the threshold is set to 5 and 10 mm h^{-1}.

GAN-argcPredNet v2.0 shows excellent performance in heavy rainfall prediction. During 60 minutes of forecasting, ~~Compared~~ compared to the baseline GAN-argcPredNet v1.0, the POD, CSI, HSS and Bias scores of GAN-argcPredNet v2.0 increase by 20.1 %, 16.0 %, 15.4 % and 25.0 % when the threshold is set to 10 mm h^{-1}. The FAR Score also decreases by 2.3 %. When the threshold is set to 30 mm h^{-1}, the POD, CSI, HSS and Bias scores increase by 18.8 %, 17.0 %, 17.2% and 26.3 % respectively. The FAR Score decreases by 3.0 %.

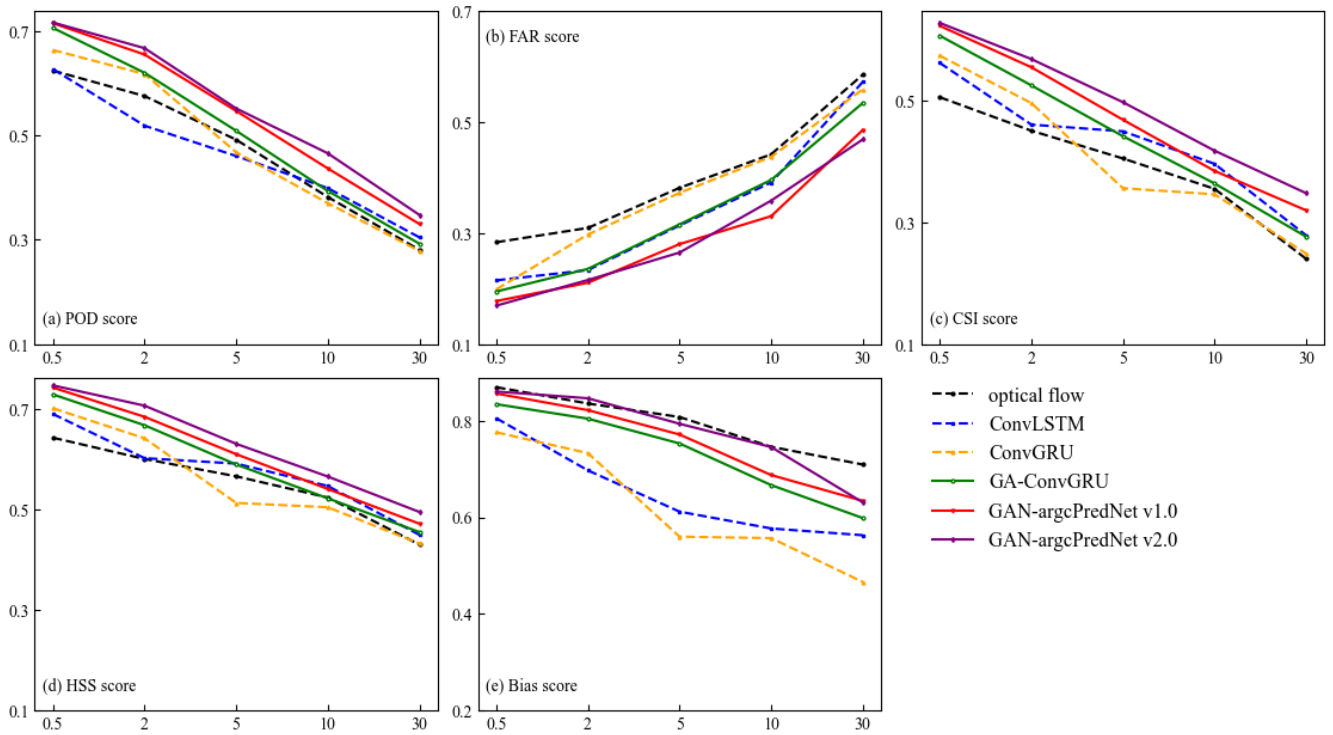



Figure 8: The average scores of (a) POD, (b) FAR, (c) CSI, (d) HSS and (e) Bias under different thresholds for forecasting 30 minutes. All the time steps were used to compute the scores. The horizontal axis represents the threshold, in units of mm h^{-1}. The perfect score of FAR is 0, while the others are 1.

290

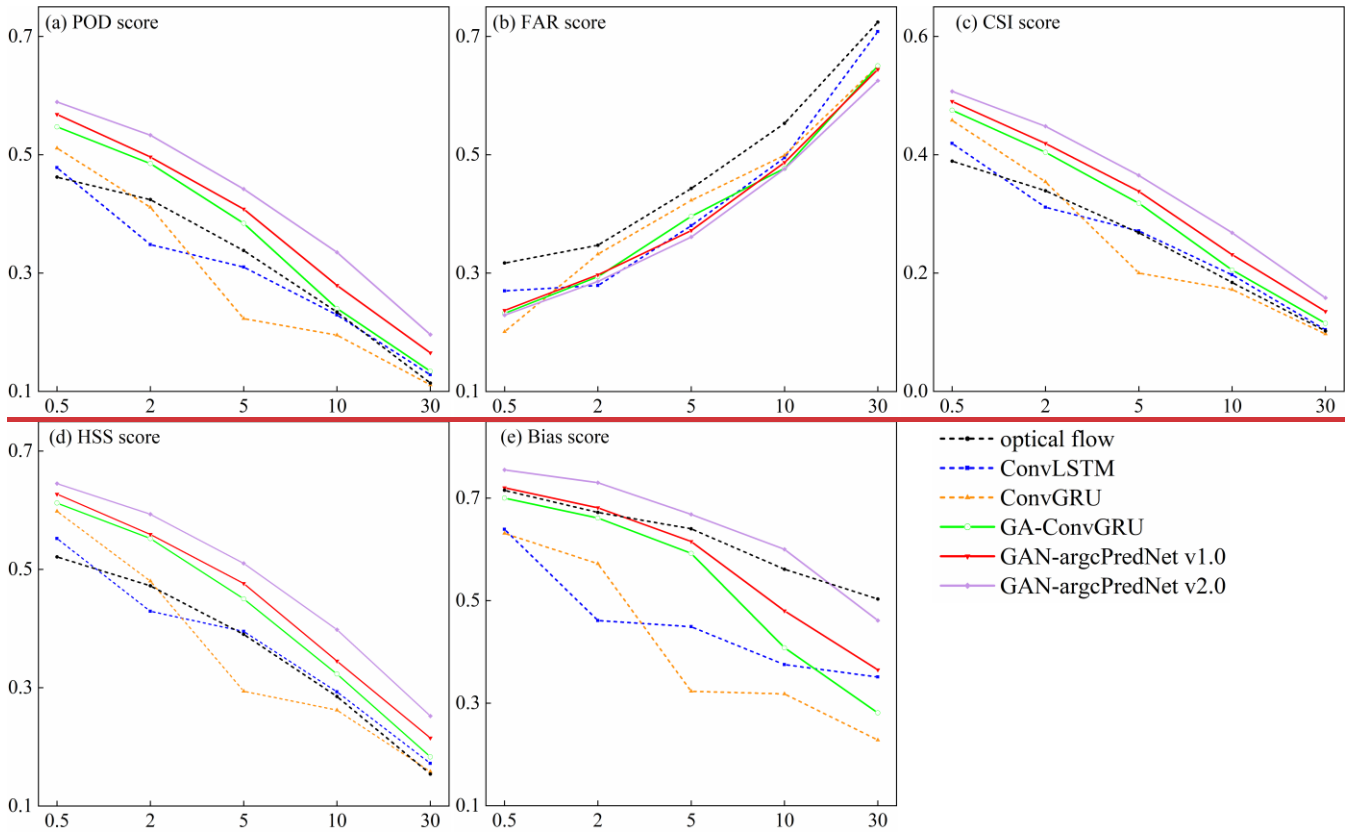
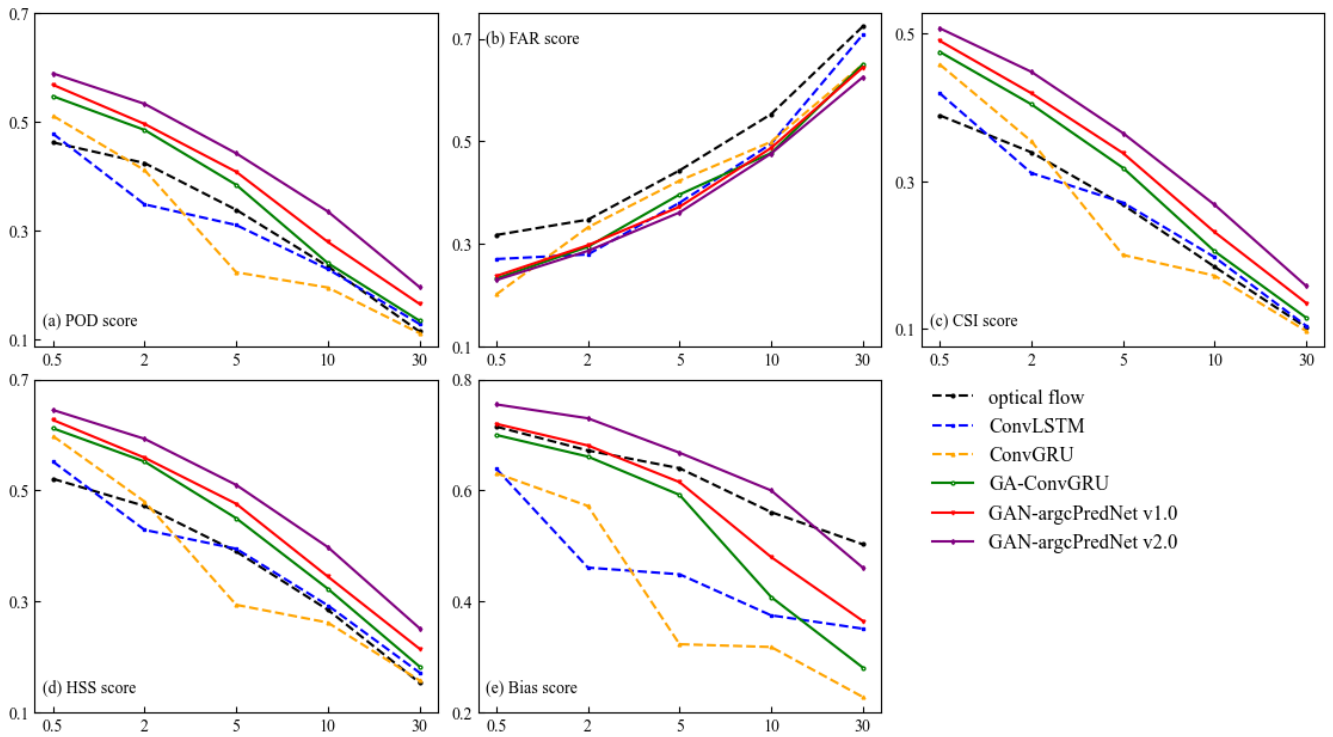



Figure ~~8~~9: The average scores of (a) POD, (b) FAR, (c) CSI, (d) HSS and (e) Bias under different threshold~~s~~ for forecasting 60 minutes. All the time steps were used to compute the scores. The horizontal axis represents the threshold, in units of mm h⁻¹. The perfect score of FAR is 0, while the others are 1.

The experiment selected three prediction examples and drew the extrapolation comparison images (Fig. ~~9-11~~10-12). ~~In Fig. 9, the optical flow method has significant overestimation of the heavy rainfall areas. As the lead time increases, all deep learning models exhibit the phenomenon of echo attenuation in the predicted images. The attenuation of GAN-argcPredNet v2.0 is noticeably slower. In the circular and rectangular regions, the intensity and shape of strong echoes are well retained until 60 minutes. There are some false predictions on the bottom right corner, and the heavy rain area near the top right goes out of the domain. Nonetheless, the other predictions of GAN-argcPredNet v2.0 are more consistent with Ground truth.~~

~~In Fig. 10, as the storm rapidly grows, the optical flow method misestimates the location of strong echoes. In the rectangular region, GAN-argcPredNet v2.0 performs better than other deep learning models in maintaining intensity and provides more accurate predictions compared to the optical flow method.~~

~~Fig. 11 shows that the optical flow method overestimates the intensity and distorts the predicted shape of the echoes. When faced with storm decay, GAN-argcPredNet v2.0 still performs the best among all models, especially in the heavy rainfall areas.~~In Figs. 10, 11, and 12, in the 0-30 minutes predictions, the optical flow has an advantage in terms of less change in echo strength and echo shape preservation. However, in regions where the echoes with rapid change, such as the center region of Fig. 10, and the upper-left region of Figs. 11 and 12, the optical flow overestimates the strength of the echoes, especially

310 the strong echoes. In Fig. 11, the optical flow incorrectly estimates the location of the strong echoes as the storm grows rapidly. Fig. 12 shows that the optical flow overestimates the intensity and distorts the predicted echo shape. In contrast, the deep learning model predicted the changes in the echoes, while GAN-argcPredNet v2.0 better preserved the shape of the echoes. In 30-60 minutes of prediction, all images predicted by the deep learning models showed echo decay. the GAN-argcPredNet v2.0 decayed significantly slower. The intensity and shape of the strong echoes are well maintained in the circular or rectangular

315 labeled areas in the figure, and the prediction results are more consistent with the ground truth. Overall, compared with other methods, GAN-argcPredNet v2.0 can provide more accurate predictions, showing its stronger prediction ability.

The optical flow method consistently exhibits overestimation or misestimation in these examples. In deep learning models, the three GAN-based models also present better echo structure and prediction performance compared to the two non-GAN models. For the echo above 40 dBZ, ConvLSTM exhibits the most severe attenuation. This suggests that echo attenuation and

320 prediction blurring seriously affect the performance of precipitation prediction, especially heavy precipitation.

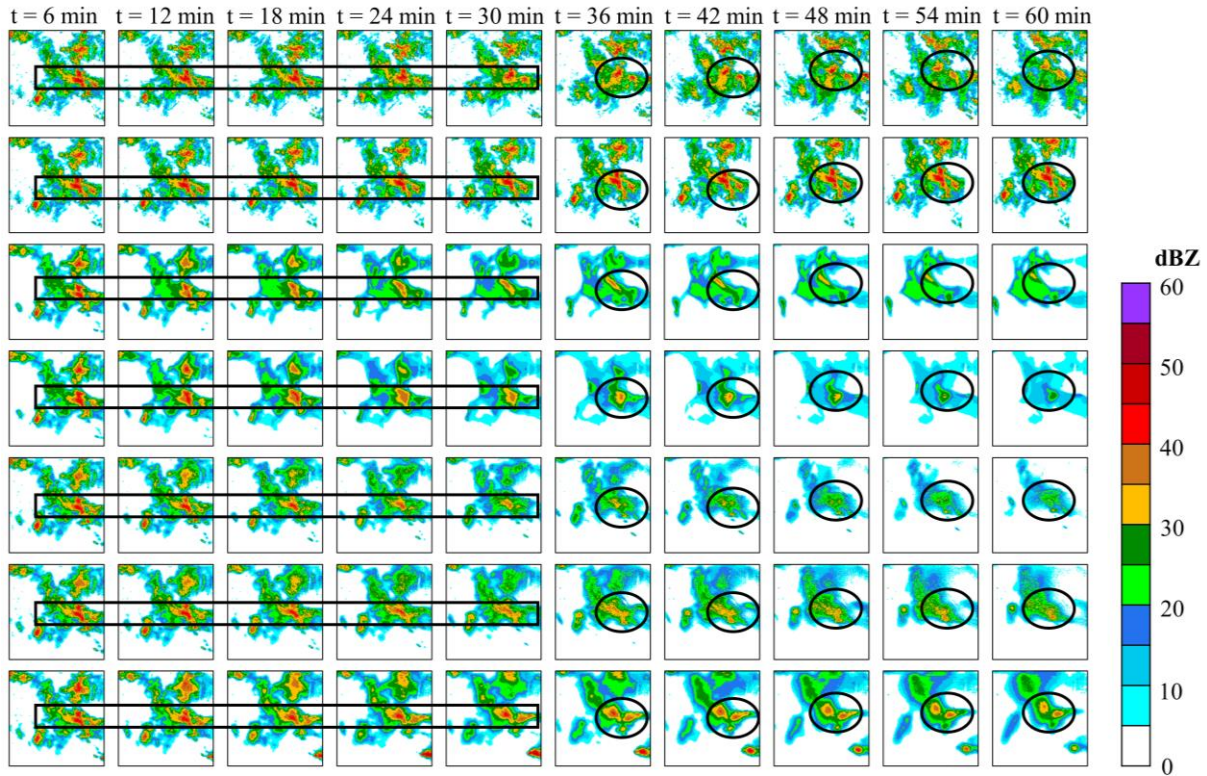



Figure 910: The example of the growth of weather system in radar echo extrapolation. From top to bottom: ground truth frames; prediction by the optical flow; prediction by ConvGRU; prediction by ConvLSTM; prediction by GA-ConvGRU; prediction by GAN-argcPredNet v1.0; prediction by GAN-argcPredNet v2.0. The circular and rectangular regions represent heavy rainfall

325 prediction, where the performance of the models in extrapolating heavy rainfall can be observed.

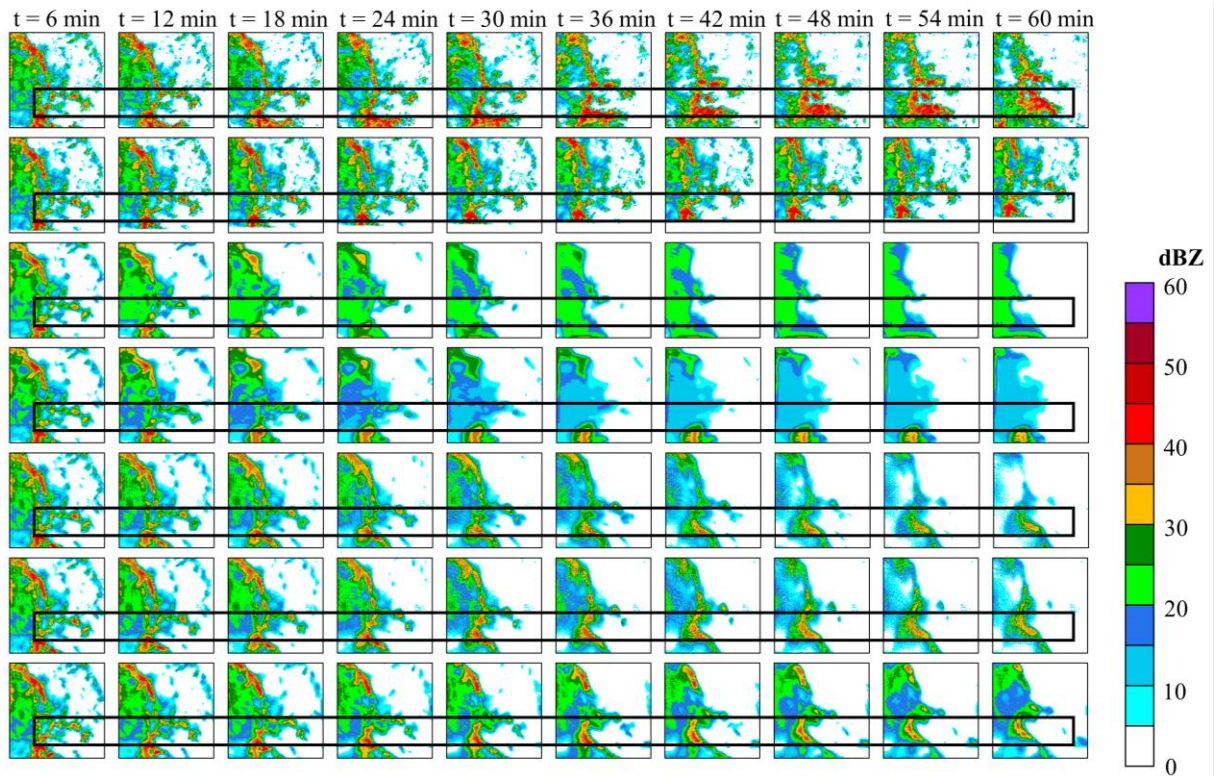



Figure ~~10~~11: The example of rapid storm growth in radar echo extrapolation. From top to bottom: ground truth frames; prediction by the optical flow; prediction by ConvGRU; prediction by ConvLSTM; prediction by GA-ConvGRU; prediction by GAN-argcPredNet v1.0; prediction by GAN-argcPredNet v2.0. The rectangular region represents the prediction for rapid storm growth.

330

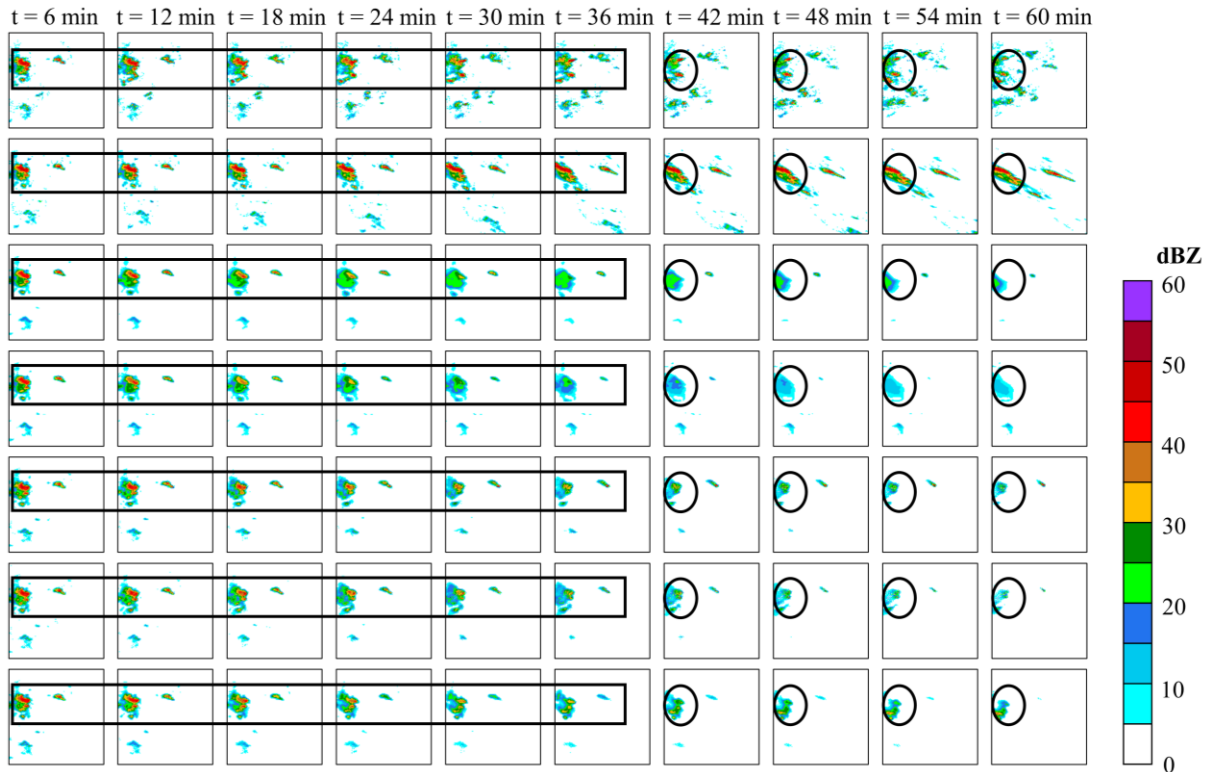



Figure ~~11~~12: The example of storm decay in radar echo extrapolation. From top to bottom: ground truth frames; prediction by the optical flow; prediction by ConvGRU; prediction by ConvLSTM; prediction by GA-ConvGRU; prediction by GAN-argcPredNet v1.0; prediction by GAN-argcPredNet v2.0. The circular and rectangular regions represent heavy rainfall prediction.

To evaluate the quality of generated images objectively, the experiment also calculated the MSE and MSSIM metrics. Table

335 2 shows that the optical flow method has the worst score. In deep learning models. GAN-based models have better scores in

both ~~MAE~~ MSE and MSSIM, with GAN-argcPredNet v2.0 achieving the best score. Compared to GAN-argcPredNet v1.0,

the ~~MAE~~ MSE metric of GAN-argcPredNet v2.0 decreases by 2.3 %, while the MSSIM metric increases by 1.25 %. In the

non-GAN models, ConvLSTM scores better.

Table 2: The MSE and MSSIM scores of each model. The perfect score of MSE is 0 while the perfect score for MSSIM is 1. Bold
340 **represents the best score.**

Model	MSE$\times10^2$	MSSIM
optical flow	0.296	0.751
ConvLSTM	0.230	0.775
ConvGRU	0.243	0.763
GA-ConvGRU	0.223	0.782
GAN-argcPredNet v1.0	0.218	0.797
GAN-argcPredNet v2.0	**0.213**	**0.807**

4.2 Ablation study

In the ablation study, we investigated the effects of STIC Attention and CS Attention. STIC-GAN and CS-GAN were constructed by adding the STIC Attention module only in the generator and the CS Attention module only in the discriminator. To precisely observe the score differences, the experiment recorded the average score of each metric in a tabular format. Table 3 shows that both CS-GAN and STIC-GAN have better scores than GAN-argcPredNet v1.0. At the threshold of 0.5 mm h^{-1}, CS-GAN achieves the best score in the FAR metric, but other metrics still perform worse than GAN-argcPredNet v2.0. At thresholds of 5, 10, and 30 mm h^{-1}, the metric scores of STIC-GAN exceed those of CS-GAN, which is closer to the GAN-argcPredNet v2.0.

Table 3: The average scores of POD, FAR, CSI, HSS and Bias under different threshold. All the time steps were used to compute the scores. Bold represents the best score.

Model	Threshold = 0.5 mm h^{-1}					Threshold = 2 mm h^{-1}				
	POD	FAR	CSI	HSS	Bias	POD	FAR	CSI	HSS	Bias
GAN-argcPredNet v1.0	0.568	0.237	0.490	0.627	0.720	0.496	0.297	0.419	0.559	0.681
CS-GAN	0.569	**0.221**	0.496	0.633	0.735	0.529	0.289	0.443	0.588	0.726
STIC-GAN	0.570	0.241	0.490	0.627	0.739	0.512	0.289	0.433	0.576	0.700
GAN-argcPredNet v2.0	**0.589**	0.229	**0.507**	**0.645**	**0.755**	**0.533**	**0.286**	**0.448**	**0.593**	**0.730**

Model	Threshold = 5 mm h^{-1}					Threshold = 10 mm h^{-1}				
	POD	FAR	CSI	HSS	Bias	POD	FAR	CSI	HSS	Bias
GAN-argcPredNet v1.0	0.408	0.372	0.338	0.476	0.615	0.279	0.487	0.231	0.345	0.480
CS-GAN	0.411	0.381	0.340	0.478	0.629	0.283	0.482	0.235	0.351	0.490
STIC-GAN	0.433	0.379	0.354	0.496	0.667	0.317	0.477	0.258	0.383	0.560
GAN-argcPredNet v2.0	**0.442**	**0.361**	**0.365**	**0.510**	**0.668**	**0.335**	**0.476**	**0.268**	**0.398**	**0.600**

Model	Threshold = 30 mm h^{-1}				
	POD	FAR	CSI	HSS	Bias
GAN-argcPredNet v1.0	0.165	0.644	0.135	0.215	0.365
CS-GAN	0.171	0.637	0.141	0.224	0.383
STIC-GAN	0.194	0.632	0.156	0.248	0.457
GAN-argcPredNet v2.0	**0.196**	**0.625**	**0.158**	**0.252**	**0.461**

Fig. ~~12~~ 13 shows that CS-GAN and STIC-GAN retain better echo intensity and shape compared to GAN-argcPredNet v1.0. For heavy rainfall, STIC-GAN shows better prediction with echoes that are closer to GAN-argcPredNet v2.0. In rectangular regions, STIC-GAN accurately predicts rainfall events until 30 minutes. However, in the circular regions, STIC-GAN overestimates the echo intensity compared to GAN-argcPredNet v2.0.

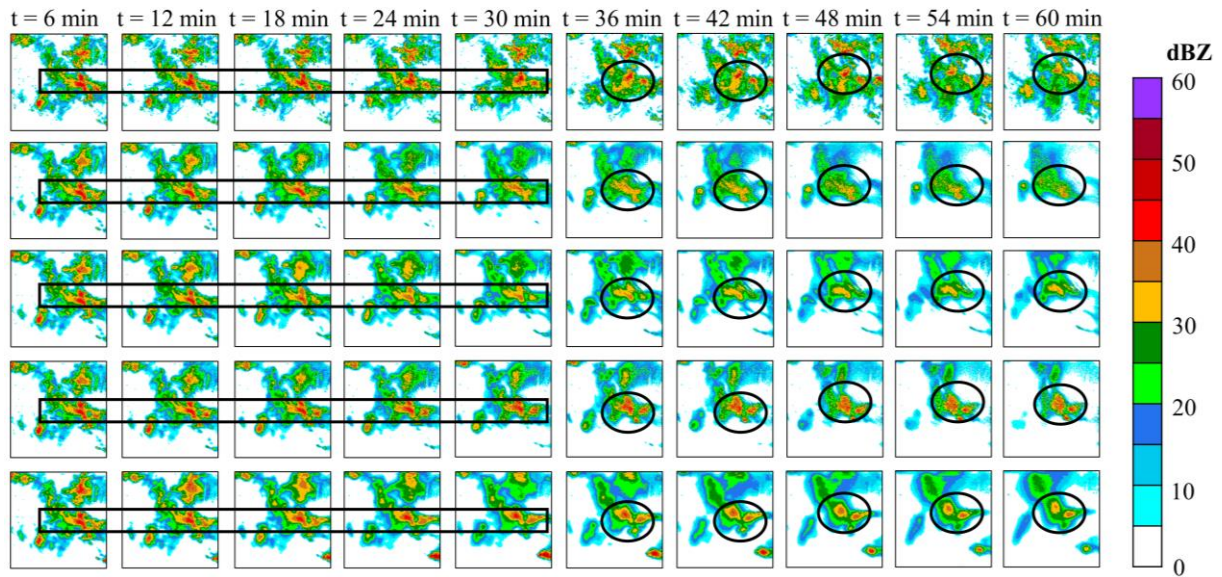
t = 6 min   t = 12 min   t = 18 min   t = 24 min   t = 30 min   t = 36 min   t = 42 min   t = 48 min   t = 54 min   t = 60 min

355

Figure 1213: The example of radar echo extrapolation. From top to bottom: ground truth frames; prediction by GAN-argcPredNet v1.0; prediction by CS-GAN; prediction by STIC-GAN; prediction by GAN-argcPredNet v2.0. The circular and rectangular regions represent heavy rainfall prediction.

5 Discussions and conclusions

5.1 Discussions

In the comparison study, the Bias Score of the optical flow is second only to GAN-argcPredNet v2.0. The result proves that, compared to other deep learning models, the number of rainfall events predicted by the optical flow method is closer to real rainfall events. However, the poor FAR Score indicates that there are many false predictions among these predicted events. Considering other metrics as well, most deep learning models perform better than the optical flow method (Fig. 8-9 and Table 2). In deep learning models, the three GAN-based models have higher prediction accuracy and better image quality compared to the two non-GAN models. This indicates that GAN structure has more advantages with its powerful image generation capability. Although ConvLSTM and ConvGRU sometimes have lower FAR scores, according to metrics such as POD, this is due to the fact that they fail to predict a large number of rainfall events. As the number of predicted rainfall events decreases, the false alarm rate also decreases. Combined with the Bias Score, ConvLSTM and ConvGRU generally predict a lower frequency of rainfall than other models. This indicates that they suffer from ~~more severe~~severer blurring effect. GA-ConvGRU and GAN-argcPredNet v1.0 improve the prediction accuracy with GAN structure, but the phenomenon of echo attenuation also cannot be ignored.

GAN-argcPredNet v2.0 achieves the highest prediction accuracy and quality, suggesting its effectiveness in curbing echo attenuation compared to other models. This is because the echo intensity and shape are better maintained by suppressing the

375   blurring effect of rain distribution and reducing the negative bias. According to the scores at the thresholds of 10 and 30 mm h^{-1}, it can be observed that GAN-argcPredNet v2.0 exhibits the best improvement in predicting heavy rainfall.

GAN-argcPredNet v2.0 maintains the evolution trend well, but there are still some special cases (Fig. ~~9~~10). There are some false predictions on the bottom right corner, and the rain area near the top right goes out of the domain. This is because in the input sequence (Fig. ~~13~~14), the rainfall on the bottom right corner shows a growing trend, and the rain area near the top right

380   shows a trend of moving towards the upper left. Our model retains the information and maintains the trend through the attention mechanism, but there are still some deviations from Ground-truth. This may be overcome by increasing the dataset and strengthening training on special cases. Also, the strong echoes outside the center of the echo images are not as well represented as the main features during the extrapolation. This is because in the input sequence, the other strong echoes have a smaller spatial range, making them more susceptible to attenuation during extrapolation. ~~Although our model can curb the attenuation,~~

385   ~~its effectiveness is not as significant as in the central region. However, compared to other models, the overall result of GAN-argcPredNet v2.0 is more competitive.~~ In Figs. 10-11, GAN-argcPredNet v2.0 has a clear advantage for out-of-center region curb the attenuation compared to other deep learning models. And there are fewer incorrect predictions compared to the optical flow. This indicates that GAN-argcPredNet v2.0 focuses on the main features and also pays more attention to the minor features with rapid growth than other models. Overall, GAN-argcPredNet v2.0 is capable of curbing echo attenuation, the center region

390   is more significant, and the overall performance is more competitive.

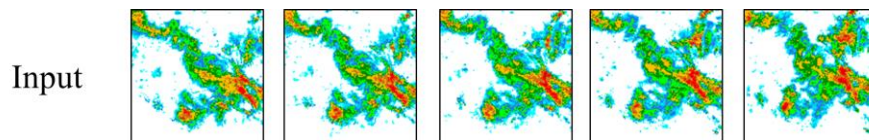



Figure ~~13~~14: The input radar images sequence. The example in Fig. 9 is extrapolated based on these five images.

In the ablation study, we find that the scores of CS-GAN and STIC-GAN are mostly better than GAN-argcPredNet v1.0 (Table 3 and Fig. 12). This indicates that both STIC Attention and CS Attention can curb echo attenuation and improve prediction

395   accuracy. As the threshold increases, the scores of STIC-GAN become closer to that of GAN-argcPredNet v2.0. This is because STIC Attention focuses more on the rapidly evolving areas of the radar images. CS Attention helps GAN-argcPredNet v2.0 achieve better overall performance.

5.2 Conclusions

The study improves precipitation nowcasting by reducing echo attenuation. By avoiding blurring or maintaining the intensity,

400   GAN-argcPredNet v2.0 curbs the attenuation and improves the rainfall prediction accuracy, especially for heavy rainfall. STIC Attention suppresses the blurring effect of rain distribution and reduces the negative bias, allowing the generator to generate more accurate images. CS Attention enables the discriminator to better guide the generator to maintain echo intensity and shape. Meanwhile, the model is designed based on the generative adversarial structure, which achieves high-quality radar echo extrapolation.

405 In practice, a predictive software has been developed based on our model. After the software accesses the radar data and establishes a prediction task, rainfall prediction results are output as dataset. Then the dataset can be fed into the urban flood warning system. The improvement of rainfall prediction has a positive impact on flood prediction and urban-operation safety. Overall, GAN-argcPredNet v2.0 is a spatiotemporal process enhancement model based on GAN, which achieves more accurate rainfall prediction.

410 Future work can be considered from two aspects. The prediction accuracy of the model proposed in the study still has room for improvement. ~~False predictions may be reduced by increasing the dataset and strengthening training on special cases~~ Given the proven efficacy of data assimilation in numerous fields, exploring the integration of data assimilation techniques with other meteorological variables, such as temperature, to study multi-modal models represents a crucial direction for precipitation nowcasting. High-resolution prediction is often limited by hardware, such as graphics card. Therefore, it is possible to reduce

415 the need for hardware by optimizing the algorithm complexity and the number of parameters.

Code and data availability

The radar data used in the paper comes from Guangdong Meteorological Administration. Due to the confidentiality policy, we only provide a sequence of 12 images. If you need to access more data, please contact Kun Zheng (ZhengK@cug.edu.cn) and Qiya Tan (ses_tqy@cug.edu.cn). The GAN-argcPredNet v2.0 model is open source. You can find the source code from

420 https://doi.org/10.5281/zenodo.7505030.

Supplement

Table S1-8 are provided in Supplement.

Author contributions

Kun Zheng and Qiya Tan were responsible for developing models and writing manuscripts; Huihua Ruan, Jinbiao Zhang,

425 Cong Luo, Siyu Tang Yunlei Yi, Yugang Tian and Jianmei Cheng were responsible for data screening and preprocessing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by Science and Technology Planning Project of Guangdong Province, China [grant No.2018B020207012].

References

Austin, G. and Bellon, A.: The use of digital weather radar records for short-term precipitation forecasting, Quarterly Journal of the Royal Meteorological Society, 100, 658-664, 1974.

Ayzel, G., Scheffer, T., and Heistermann, M.: RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting, Geoscientific Model Development, 13, 2631-2644, 10.5194/gmd-13-2631-2020, 2020.

Ballas, N., Yao, L., Pal, C., and Courville, A.: Delving deeper into convolutional networks for learning video representations, arXiv preprint arXiv:1511.06432, 2015.

Bowler, N. E., Pierce, C. E., and Seed, A.: Development of a precipitation nowcasting algorithm based upon optical flow techniques, Journal of Hydrology, 288, 74-91, 10.1016/j.jhydrol.2003.11.011, 2004.

Chen, H. G., Zhang, X., Liu, Y. T., and Zeng, Q. Y.: Generative Adversarial Networks Capabilities for Super-Resolution Reconstruction of Weather Radar Echo Images, Atmosphere, 10, 10.3390/atmos10090555, 2019.

Chung, K.-S. and Yao, I.-A.: Improving radar echo Lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases, Monthly Weather Review, 148, 1099-1120, 2020.

Courbariaux, M., Bengio, Y., and David, J. P.: Binaryconnect: Training deep neural networks with binary weights during propagations, Advances in neural information processing systems, 28, 2015.

Dixon, M. and Wiener, G.: TITAN: Thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology, Journal of Atmospheric and Oceanic Technology, 10, 785-797, 1993.

Foresti, L., Sideris, I. V., Nerini, D., Beusch, L., and Germann, U.: Using a 10-year radar archive for nowcasting precipitation growth and decay: A probabilistic machine learning approach, Weather and Forecasting, 34, 1547-1569, 10.1175/WAF-D-18-0206.1, 2019.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative adversarial networks, Communications of the ACM, 63, 139-144, 2020.

Gulcehre, C., Moczulski, M., Denil, M., and Bengio, Y.: Noisy activation functions, International conference on machine learning, 3059-3068, 2016.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C.: Improved training of wasserstein gans, Advances in neural information processing systems, 30, 2017.

Han, L., Fu, S. X., Zhao, L. F., Zheng, Y. G., Wang, H. Q., and Lin, Y. J.: 3D Convective Storm Identification, Tracking, and Forecasting-An Enhanced TITAN Algorithm, Journal of Atmospheric and Oceanic Technology, 26, 719-732, 10.1175/2008jtecha1084.1, 2009.

460   Inoue, T. and Misumi, R.: Learning from Precipitation Events in the Wider Domain to Improve the Performance of a Deep Learning-based Precipitation Nowcasting Model, Weather and Forecasting, 37, 1013-1026, 10.1175/WAF-D-21-0078.1, 2022.

Kingma Diederik, P. and Adam, J. B.: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Liu, Y., Xi, D.-G., Li, Z.-L., and Hong, Y.: A new methodology for pixel-quantitative precipitation nowcasting using a pyramid Lucas Kanade optical flow approach, Journal of Hydrology, 529, 354-364, 2015.

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

Luo, C., Li, X., and Ye, Y.: PFST-LSTM: A spatiotemporal LSTM model with pseudoflow prediction for precipitation nowcasting, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 843-857, 10.1109/JSTARS.2020.3040648, 2020.

470   Luo, C. Y., Li, X. T., Wen, Y. L., Ye, Y. M., and Zhang, X. F.: A Novel LSTM Model with Interaction Dual Attention for Radar Echo Extrapolation, Remote Sensing, 13, 10.3390/rs13020164, 2021.

Nerini, D., Foresti, L., Leuenberger, D., Robert, S., and Germann, U.: A reduced-space ensemble Kalman filter approach for flow-dependent integration of radar extrapolation nowcasts and NWP precipitation ensembles, Monthly Weather Review, 147, 987-1006, 2019.

475   Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S.: Activation functions: Comparison of trends in practice and research for deep learning, arXiv preprint arXiv:1811.03378, 2018.

Pan, X., Lu, Y. H., Zhao, K., Huang, H., Wang, M. J., and Chen, H. A.: Improving Nowcasting of Convective Development by Incorporating Polarimetric Radar Variables into a Deep-Learning Model, Geophysical Research Letters, 48, 10.1029/2021gl095302, 2021.

480   Pulkkinen, S., Chandrasekar, V., and Niemi, T.: Lagrangian Integro-Difference Equation Model for Precipitation Nowcasting, Journal of Atmospheric and Oceanic Technology, 38, 2125-2145, 2021.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., and Foresti, L.: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1. 0), Geoscientific Model Development, 12, 4185-4219, 2019.

485   Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S.: Skilful precipitation nowcasting using deep generative models of radar, Nature, 597, 672-+, 10.1038/s41586-021-03854-z, 2021.

Reyniers, M.: Quantitative precipitation forecasts based on radar observations: Principles, algorithms and operational systems,

490   Institut Royal Météorologique de Belgique Brussel, Belgium, 2008.

Sakaino, H.: Spatio-Temporal Image Pattern Prediction Method Based on a Physical Model with Time-Varying Optical Flow, IEEE Transactions on Geoscience and Remote Sensing, 51, 3023-3036, 10.1109/tgrs.2012.2212201, 2013.

Shi, X. J., Chen, Z. R., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, 29th Annual Conference on Neural Information Processing Systems (NIPS),

495 Montreal, CANADA, Dec 07-12, 2015.

Shi, X. J., Gao, Z. H., Lausen, L., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C.: Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model, 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, Dec 04-09, 2017.

Sutskever, I., Vinyals, O., and Le, Q. V.: Sequence to sequence learning with neural networks, Advances in neural information

500 processing systems, 27, 2014.

Tian, L., Li, X. T., Ye, Y. M., Xie, P. F., and Li, Y.: A Generative Adversarial Gated Recurrent Unit Model for Precipitation Nowcasting, IEEE Geoscience and Remote Sensing Letters, 17, 601-605, 10.1109/lgrs.2019.2926776, 2020.

Wang, X., Girshick, R., Gupta, A., and He, K.: Non-local neural networks, Proceedings of the IEEE conference on computer vision and pattern recognition, 7794-7803, 2018a.

505 Wang, Y., Jiang, L., Yang, M.-H., Li, L.-J., Long, M., and Fei-Fei, L.: Eidetic 3D LSTM: A model for video prediction and beyond, International conference on learning representations, 2018b.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing, 13, 600-612, 10.1109/TIP.2003.819861, 2004.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: Cbam: Convolutional block attention module, Proceedings of the European

510 conference on computer vision (ECCV), 3-19, 2018.

Xie, P. F., Li, X. T., Ji, X. Y., Chen, X. L., Chen, Y. Z., Liu, J., and Ye, Y. M.: An Energy-Based Generative Adversarial Forecaster for Radar Echo Map Extrapolation, IEEE Geoscience and Remote Sensing Letters, 19, 10.1109/lgrs.2020.3023950, 2022.

Zheng, K., Liu, Y., Zhang, J. B., Luo, C., Tang, S. Y., Ruan, H. H., Tan, Q. Y., Yi, Y. L., and Ran, X. T.: GAN-argcPredNet

515 v1.0: a generative adversarial model for radar echo extrapolation based on convolutional recurrent units, Geoscientific Model Development, 15, 1467-1475, 10.5194/gmd-15-1467-2022, 2022.