DiffREE: Feature-Conditioned Diffusion Model for Radar Echo Extrapolation

WU Qi-liang
Nanjing University of Information Science and Technology
WANG Xing
wswql121138@163.com
Nanjing University of Information Science and Technology
ZHANG Tong
Nanjing University of Information Science and Technology
MIAO Zi-shu
Baotou Meteorological Bureau
YE Wei-liang
Nanjing University of Information Science and Technology
LI Hao
Nanjing University of Information Science and Technology

Research Article

Keywords: Deep learning, Short-term forecasts, Radar echo extrapolation, Diffusion model, Conditional encoding

Posted Date: April 19th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4270187/v1

Additional Declarations: No competing interests reported.

DiffREE: Feature-Conditioned Diffusion Model for Radar Echo Extrapolation

WU Qi-liang[1], WANG Xing[1,2*†], ZHANG Tong[3], MIAO Zi-shu[4], YE Wei-liang[5], LI Hao[5]

[1]School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, Jiangsu province, China.
[2*]National Demonstration Center for Experimental Atmospheric Science and Environmental Meteorology Education, Nanjing University of Information Science and Technology, Nanjing, Jiangsu province, China.
[3]School of atmospheric sciences, Nanjing University of Information Science and Technology, Nanjing, Jiangsu province, China.
[4]Baotou Meteorological Bureau, Baotou, Inner Mongolia Autonomous Region, China.
[5]School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, Jiangsu province, China.

*Corresponding author(s). E-mail(s): wswql12138@163.com;
Contributing authors: 20211249545@nuist.edu.cn;
†These authors contributed equally to this work.

Abstract

Deep learning techniques for radar echo extrapolation and prediction have become crucial for short-term precipitation forecasts in recent years. As the extrapolation leading time extends, radar echo intensity attenuates increasingly, and the forecast performance on strong echoes declines rapidly. These are two typical characteristics contributing to the current inaccurate results of radar extrapolation. To this end, we propose a novel diffusion radar echo extrapolation (DiffREE) algorithm driven by echo frames in this study. This algorithm deeply integrates the spatio-temporal information of radar echo frames through a conditional encoding module, and then it utilizes a Transformer encoder to automatically extract the spatio-temporal features of echoes. These features serve as inputs to the conditional diffusion model, driving the model to reconstruct the current radar echo frame. Moreover, a validation experiment demonstrates that the proposed method can generate high-precision and high-quality forecast images of

radar echoes. To further substantiate the model performance, the DiffREE algorithm is compared with the other four models by using public datasets. In the radar echo extrapolation task, the DiffREE demonstrates a remarkable improvement in the evaluation metrics of critical success index, equitable threat score, Heidke skill score and probability of detection by 21.5%, 27.6%, 25.8%, and 21.8%, respectively, displaying notable superiority.

Keywords: Deep learning, Short-term forecasts, Radar echo extrapolation, Diffusion model, Conditional encoding

1 Introduction

Short-term extreme heavy rainfall is a major weather disaster affecting human lives and socio-economics. Analyzing the features of weather radar echoes and performing the extrapolation based on these features are commonly used for nowcasting[1–3]. Currently, the predominant radar extrapolation methods in operations are based on algorithms such as cross-correlation, centroid tracking and optical-flow analysis[4, 5]. However, these methods do not adequately consider factors such as atmospheric circulation patterns, topographical conditions, humidity and water vapor content that influence the occurrence of short-term extreme heavy rainfall. Therefore, accurate forecasts are challenging for severe weather events with rapid evolution and unstable movement trends.

In recent years, some studies have attempted to improve forecast models on the precipitation occurrence process by applying deep learning techniques to radar echo extrapolation and heavy rainfall forecasts[6–8]. Deep learning-based strategies can primarily be categorized into two types: one is based on the Convolutional Neural Network (CNN), and the other is based on the Recurrent Neural Network (RNN)[9–12]. The former excels in modeling the spatial representation of radar echoes but has limited capability in modeling the temporal evolution[13]. The latter is adept at capturing the temporal correlations of radar echoes during motion processes but cannot analyze the spatial correlations between radar echo images. Recent research has proposed hybrid models that combine the CNN and RNN to address the deficiencies mentioned above. However, these models still struggle to overcome the influence of weak sparse echoes and unstable weather processes. Under complex weather conditions, the forecast accuracy and stability still do not meet the requirements of operational applications. Furthermore, as extrapolation leading time increases, radar echo intensity diminishes increasingly, and the forecasting performance for strong echoes deteriorates rapidly. Consequently, the accuracy of long-term forecasts (>1 hour) and short-term heavy rainfall predictions is limited.

This study introduces a diffusion radar extrapolation algorithm based on generative networks to address the abovementioned issues. Differing from the traditional autoregressive forecast models based on the CNN and RNN, this model proposed in this study employs a forward noising process to gradually transform the complex distribution of radar echo motion into unstructured noise, and then it utilizes historical

radar echo observations as conditional guidance to drive a reverse denoising process. In this way, the original data distribution is gradually restored, which in turn mitigates the error accumulation typically caused by deep learning autoregression. Ultimately, this approach enables the prediction of future radar echoes.

This work makes the following three main contributions:

- Introducing a radar echo extrapolation method based on a diffusion model. This model can reliably forecast complex radar echo distributions (dense or sparse) and motion trends (steady or non-steady). It surpasses pixel-to-pixel autoregressive extrapolation algorithms in terms of effectiveness.
- A novel image encoder incorporating position information is established based on a Vision Transformer(ViT)[14]. It can transform the past radar echo frames from pixel space into feature space. This encoder, serving as a guiding condition for the diffusion model, can offer more precise prior knowledge of echo motion trends.
- The proposed diffusion radar echo extrapolation (DiffREE) model is compared with four advanced methods by using public datasets, demonstrating its notable advantage in performance.

2 Related Work

From the perspective of deep learning, radar echo extrapolation can be considered as the prediction of spatio-temporal sequences. Prediction models can be divided into two primary categories. The first category comprises techniques that rely on the CNNs and their corresponding variations. In contrast, the second category involves approaches constructed based on the RNNs and their derivatives [15–18].

2.1 Convolutional Neural Network

The CNN was initially proposed by Lecun et al. (1998) at New York University. The CNNs have a remarkable capability to effectively extract and reconstruct image features, and they also exhibit excellent resilience in accurately identifying image shifts, scaling and distortions. As research in this field progresses, several spatio-temporal forecast models based on the CNNs have been developed. Agrawal et al [19]. employed the U-Net model to predict echoes using abstract features, and this model demonstrates superior performance on one-hour forecasts compared with optical flow techniques. Another notable model is the small attention-Unet, introduced by Kevin et al [20]., which is an efficient CNN based on the U-Net architecture. It incorporates attention modules and depth-wise separable convolutions to extract spatial features during precipitation development, improving computational efficiency and prediction accuracy. Furthermore, Tao et al [21]. introduced an advanced approach that combines the encoder and decoder components of the U-Net model. This approach effectively facilitates the extraction of multi-scale features and essential precipitation features, addressing gradient vanishing and consequently improving the accuracy of precipitation estimation and the ability to identify precipitation patterns.

2.2 Recurrent Neural Network

The RNN has emerged as a crucial tool for sequence prediction, with notable achievements in various fields such as machine translation, voice recognition and video captioning. It also has significant applications in forecasting weather patterns over time and space. Shi et al [22]. introduced a Convolutional Long Short-Term Memory (ConvLSTM) model, expanding the utilization of LSTM models to pictures. Subsequently, the Trajectory Gated Recurrent Unit model was introduced to integrate recursive convolutional structures with conventional precipitation nowcasting techniques relying on optical flow technology, improving forecast accuracy.

Simultaneously, Wang et al [23]. from Tsinghua University identified the specific limitations in the layer-independent memory mechanisms of the ConvLSTM. They proposed the Spatial-Temporal LSTM (ST-LSTM) and the Predictive RNN (PredRNN) model as alternative approaches, remarkably enhancing the capacity to capture short-term dynamics in radar echo images. However, the PredRNN encounters challenges related to gradient vanishing, which hinders its effectiveness in capturing long-term properties within deep networks. In 2018, this research team introduced an improved model, PredRNN++[24], which utilizes the innovative Casual LSTM unit and the Gradient Highway Unit to capture the short-term dynamics better and mitigate gradient vanishing.

In 2019, researchers recognized that many existing RNNs designed for spatio-temporal prediction struggle to effectively utilize differential signals, necessitating assistance in capturing complex spatio-temporal variations. Thus, differential concepts were introduced to address this limitation, and the Memory in Memory (MIM) model [25] was proposed, dramatically enhancing the accuracy of radar echo extrapolation. Vincent et al. (2020) introduced the PhyDNet model, which was developed based on the ConvLSTM framework. The model utilizes convolutional computations to estimate the partial differential equations representing physical principles presented by the PhyCell (an iterative physical unit). The PhyDNet model renders it suitable for the extrapolation of radar echo images[26].

Currently, some models effectively combine the CNNs with RNNs to capture spatial correlations among echoes across different regions and investigate the temporal evolution of echoes. In stable convective processes, these models outperform conventional radar echo extrapolation methods regarding forecast accuracy, but they often encounter two inherent challenges that limit the practicality of radar echo prediction. Firstly, it is essential to note that radar echo intensity decays noticeably in long-term forecasts. This attenuation is particularly pronounced when the forecast leading time of severe heavy rainfall events is one hour or more, resulting in low prediction accuracy. Furthermore, accurately identifying and predicting the intensity and spatial distribution of strong echoes in recently developed convective systems also faces major challenges, which leads to continuous and frequent missing reports of extreme precipitation events.

3 Materials and Methods

In this section, we present the overview of the proposed diffusion model and then show in detail the radar echo extrapolation method based on the diffusion model.

3.1 Diffusion Model

The diffusion model typically refers to a class of machine learning algorithms that involve gradually transforming complex distributions into unstructured noise and learning to reverse this process to recover the data distribution. These algorithms are commonly adopted for various tasks, including image generation, audio generation and image super-resolution[27–29].

Assuming that x_0 is a sample obtained from the data with the distribution P_{data}, a forward diffusion process accumulates T times Gaussian noise z on x_0 from $t = 0$ to $t = T$. The noise accumulation at each step is only dependent on the previous step and is calculated by Eq. 1,2.

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t} z \tag{1}$$

$$q_t(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t} x_{t-1}, (1-\alpha_t) I) \tag{2}$$

where $\alpha_t = 1 - \beta_t$, $\beta_t \in (0, 1)$ for $t = 1$ to T , and z follows a normal distribution $z \sim \mathcal{N}(0, I)$.

The process can be viewed as a Markov process. Therefore, x_t at any moment can be directly sampled from x_0 though Eq. 3,4.

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t} z \tag{3}$$

$$q_t(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t) I) \tag{4}$$

where $\bar{\alpha}_t = \prod_{i=1}^{T} \alpha_i$.

The reverse diffusion process involves reversing the forward diffusion process and obtaining the complex data distribution from the pure noise x_T, as shown in Eq. 5,6.

$$p_t(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \tag{5}$$

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t)} x_t \tag{6}$$

where $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. Since x_0 is not obtainable in the reverse process, an equivalent expression can be derived through Eq. 3, and it can be substituted using the formula Eq. 6, as shown in Eq. 7.

$$\mu_t(x_t, t) = \frac{1}{\sqrt{a_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{a}_t}} z_t(x_t, t) \right) \tag{7}$$

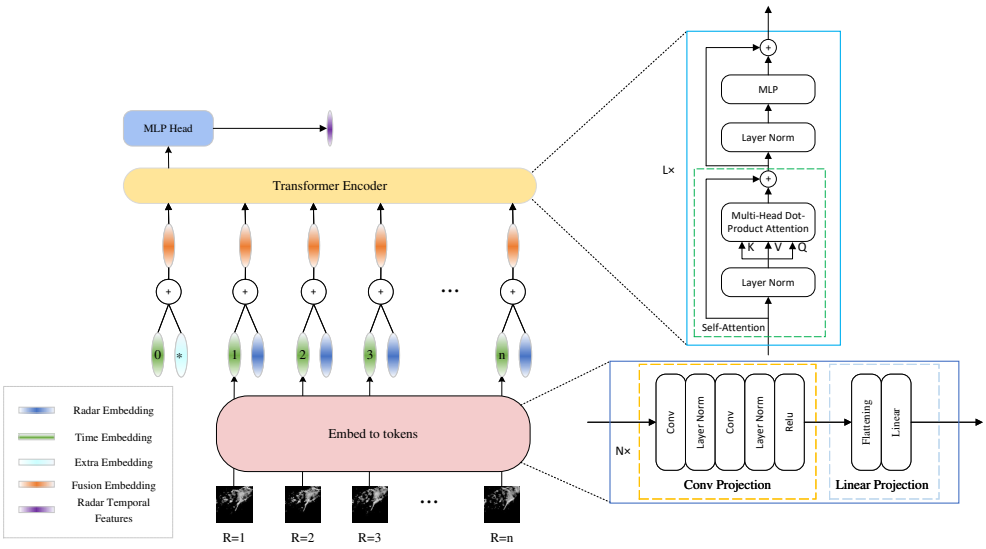

Fig. 1 Structure schematic of the condition encoding.

where $z_t(x_t, t)$ represents the target that needs to be predicted by a time-conditioned neural network parameterized by θ during the reverse process. This network is capable of reversing the process from noise to data. The loss function for the neural network is expressed by Eq. 8.

$$L(\theta) = \mathbb{E}_{t,x_0\sim p_{\text{data}},\epsilon\sim\mathcal{N}(0,I)}\left[\|z - z_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}z|t)\|_2^2\right] \tag{8}$$

3.2 Conditional Encoding

Traditional extrapolation methods typically use the single-frame images of radar echoes from the past directly as the input condition or consider multiple consecutive images as the input channels of neural network models. However, these approaches have some drawbacks. Firstly, single-frame images are difficult to accurately capture the temporal characteristics of radar echoes because of the lack of time-series information. Secondly, using images from different moments to represent channels does not fully exploit the ability of the model to investigate the temporal evolution of echoes. Additionally, using images directly as conditions also imposes tremendous computational burdens.

As a consequence, the ViT is adopted as the conditional encoder for radar echoes. The ViT is a deep learning model that utilizes a self-attention mechanism and can handle time-series data. However, the ViT typically divides a single image into multiple patches for inputs. Since the radar echo sequence consists of multiple images, it cannot be directly input into the ViT. Hence, we modify the existing ViT architecture in this study. Specifically, convolutional blocks are introduced in the input stage of the ViT and applied sequentially to radar echo sequences. This approach serves a dual purpose.

On the one hand, it alleviates the computational burden of the ViT caused by the large image size. On the other hand, it contributes to the transformation of the radar echo sequence from the pixel domain to the feature domain, which is beneficial for the ViT to extract crucial feature information from the radar echo sequence, enhancing its feature-capturing capability, as shown in Fig. 1.

The method proposed in this study assumes that there are n radar echo frames from past moments, denoted as $R = \{R^i\}_{i=1}^n$. We first perform a convolution operation on these frames R to extract spatial features of the radar echoes. Subsequently, the spatial features are linearly flattened, and then the flattened features are input into a fully connected layer to generate fixed-length low-dimensional radar embeddings. Simultaneously, time information is encoded into low-dimensional time embeddings. These radar embeddings and time embeddings are directly summed to fuse the spatio-temporal information of radar echo frames. In this way, the tokens required by the Transformer encoder[30] can be obtained. To further enhance the representational ability of the model, we introduce an additional zero-position embedding with learning ability. After passing through the Transformer encoder, the tokens generated by this zero-position embedding can learn to summarize the information of all the subsequent tokens. Finally, the conditional information h can be obtained through a multi-layer perceptron. The calculation is shown in Eq. 9.

$$h = \mathrm{MLP}(\mathrm{VIT}(R)) + b \tag{9}$$

where VIT represents the feature extraction operation, and b denotes the weight offset. The abovementioned processing steps can fuse the spatio-temporal information of radar echo frames into the conditional information required by the Transformer encoder. This ViT-based approach is better suited to capture the temporal characteristics of radar echoes and has lower computational burdens.

3.3 Conditional Diffusion

In this research, a set of data is formed by combining the past radar echo frames R (n frames in total) and the current radar echo frames $X = \{x^i\}_{i=1}^s$ (s frames in total). The objective is to noise the current frames and use the number of noising steps t for position encoding. Therefore, a Transformer encoder is employed for position encoding, calculated by Eq. 10.

$$\overrightarrow{e_t}^{(i)} = \begin{cases} \sin\left(\frac{1}{10000^{\frac{2k}{d}}} \cdot t\right), & \text{if } i = 2k \\ \cos\left(\frac{1}{10000^{\frac{2k}{d}}} \cdot t\right), & \text{if } i = 2k+1 \end{cases} \tag{10}$$

where $\overrightarrow{e_t} \in \mathbb{R}^d$ represents the embedding corresponding to the noising steps t, with a d-dimension. i denotes the dimension index of this embedding. By inputting $\overrightarrow{e_t}$ into the two pairs of the fully connected layer and activation layer, the output results are embedded as conditions into each stage of the denoising network. The denoising network adopts a U-net network structure, consisting of multiple residual connection blocks, as shown in Fig. 2. Each residual connection block comprises multiple sets of two-dimensional convolutional layers, adaptive global normalization layers and activation layers. The U-net network takes the current radar echo frame x_0' after the noising

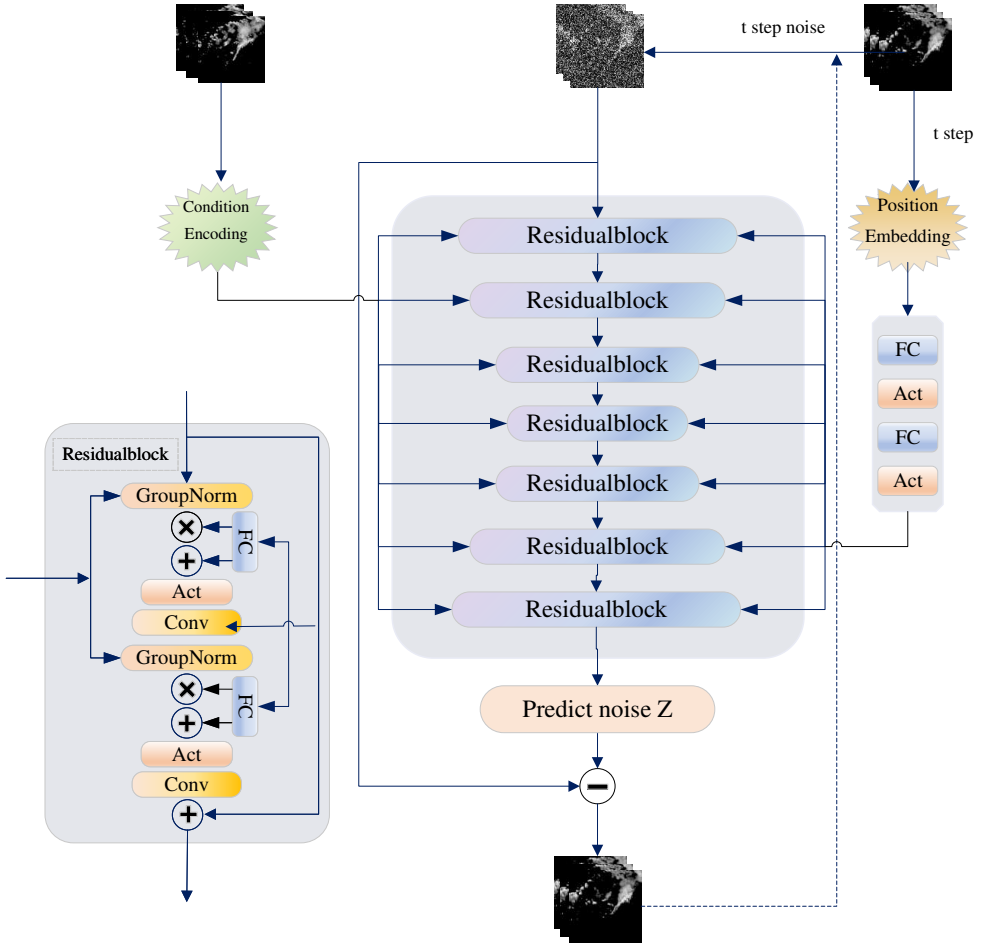

Fig. 2 Structure schematic of condition diffusion.

process as an input. Moreover, the outputs of the condition encoder, including the feature vector h and position encoding embedding $\overrightarrow{e_t}$ of past radar echoes, are used as additional inputs in this study. The ultimate goal is to predict the noise z at the current moment by using the abovementioned inputs. The loss function $L_{\text{pred}}(\theta)$ is defined as shown in Eq. 11.

$$L_{\text{pred}}\ (\theta) = \mathbb{E}_{t,[\mathbf{p},\mathbf{x}_0]\sim p_{\text{data}}\ ,\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathrm{I})}\left[\|\boldsymbol{Z} - \mathrm{Z}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{Z} \mid \mathbf{h}, t\right)\|^2\right] \tag{11}$$

where p_{data} represents the data distribution, p and x_0 denote the samples obtained from the data distribution, and ϵ indicates the Gaussian noise with a mean of zero and a unit covariance matrix. Additionally, $\bar{\alpha}_t$ represents the weighting coefficients used to balance the data before and after the noising process and measure the accuracy

of the prediction results. This loss function can evaluate the prediction accuracy by calculating the expected value over multiple samples and the square of the Euclidean distance between the predicted noise Z and the actual noise Z_θ. Finally, the denoised prediction target is obtained by subtracting the expected noise Z from the input x'_0 after the noising process.

4 Experimental Data and Preparations

The dataset used in this experiment is the observation data from the meteorological radars and automatic stations in Jiangsu Province, spanning from April to September during 2019–2021. The radar dataset is obtained from the networking mosaics of multiple S-band meteorological radars, which has undergone quality control, covering the entire area of Jiangsu province. The data values range from 0 to 70 dBZ, with a horizontal resolution of 0.01° (approximately 1 km) and a temporal resolution of 6 minutes. The training, validation and test datasets for the experiment have 20000, 2000 and 2000 sequences, respectively. Each sequence has 40 consecutive radar echo images, and the grid size of each image is 480 × 560 pixels. To simplify the computation, the images are resized to 128 × 128 pixels, and the 40 images in each sequence are split into 20 odd-numbered and 20 even-numbered images. During training, the first ten images are used as conditions, and the last ten are considered targets. Predictions are made for the subsequent two hours based on the data from the previous two hours.

The representative models used for the comparative experiment in this study include the MIM, PredRNN, ConvLSTM, PhyDNet and SmatUnet. In order to obtain rigorous results, all models have a typical architecture of 4-layer stacked structures and 64 feature mapping units. The batch size and learning rate are set to 64 and 0.0001, respectively. The Adam optimizer is used, and the loss functions for all models are root mean square errors (RMSEs). The experiments are conducted on four Tesla V-100 32GB graphic processing units.

5 Performance Evaluation Metrics

To accurately gauge the forecasting precision of radar echo extrapolations, the experiments were designed to provide objective, quantifiable assessments grounded in image structural similarity and meteorological evaluation standards.

5.1 Image Quality Metrics

The Structural Similarity Index (SSIM) is a quantitative measure of the similarity between two images by considering their brightness, contrast and structural features. It is widely used in the fields of image processing and computer vision to assess image quality or compare different image processing algorithms. A higher SSIM value indicates better image quality[31]. The calculation is as follows.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{12}$$

In Eq. 12, μ_x and μ_y represent the mean values of x and y, σ_x^2 and σ_y^2 denote the variances of x and y, and σ_{xy} indicates the covariance between x and y. c_1 and c_2 are constants added for numerical stability.

The Mean Squared Error (MSE) is a widely employed metric for assessing the average squared discrepancy between predicted and actual values in regression tasks. Owing to the distinct characteristics of radar echoes, we have devised an enhanced variant of MSE known as ThrMSE, which incorporates distinct weights for radar echo values in different regions. The computation methodology is as follows:

$$w_{ij} = \begin{cases} 0.2, & \text{if } y_{ij} \leq 10 \\ 0.4, & \text{if } 10 < y_{ij} \leq 30 \\ 0.6, & \text{if } 30 < y_{ij} \leq 40 \\ 0.8, & \text{if } 40 < y_{ij} \leq 50 \\ 1, & \text{if } 50 < y_{ij} \leq 70 \end{cases} \tag{13}$$

$$\text{ThrMSE}(x, y) = \frac{1}{mn} \sum_{i=0}^{m} \sum_{j=0}^{n} (y_{ij} - x_{ij})^2 \cdot w_{ij} \tag{14}$$

In Eq. 13,14, y_{ij} represents the actual radar echo value at grid position (i, j), x_{ij} represents the forecasted radar echo value at the same grid position, and w_{ij} denotes the loss weight for the radar echo value at that specific grid position.

5.2 Meteorological Evaluation Standards

Within the meteorological domain, there is a pronounced emphasis on accurately forecasting vital echo locations and intensities. Key metrics employed for extrapolative forecasting assessment include Critical Success Index (CSI), Heidke Skill Score (HSS), Equitable Threat Score (ETS), and Probability of Detection (POD) [32]. Specifically:

- CSI focuses on the correspondence between forecasted and actual observed events, representing the probability of successful event prediction. A higher CSI score signifies superior model performance.

$$\text{CSI} = \frac{TP}{TP + FN + FP} \tag{15}$$

- HSS underscores the proportion of correct predictions once purely random forecasts are excluded. HSS values range between -1 and 1, where 1 denotes perfect forecasting. A negative score implies performance inferior to random prediction. A higher HSS indicates commendable extrapolative capabilities.

$$\text{N} = (TP + FN) \times (FN + TN) \tag{16}$$

$$\text{M} = (TP + FP) \times (FP + TN) \tag{17}$$

$$\text{HSS} = \frac{2 \times (TP \times TN - FN \times FP)}{\text{N} + \text{M}} \tag{18}$$

- ETS is pivotal in gauging the performance of convective-scale forecasts. An ETS score reflects the enhanced proficiency of precipitation forecasts meeting specific thresholds relative to random predictions.

$$\mathrm{E} = \frac{(TP + FN) \times (TP + FP)}{TP + FN + FP + TN} \tag{19}$$

$$\mathrm{ETS} = \frac{TP - \mathrm{E}}{TP + FN + FP - \mathrm{E}} \tag{20}$$

- POD quantifies the proportion of accurately forecasted precipitation areas relative to actual ones, illustrating the model's adeptness at detecting precipitation events.

$$\mathrm{POD} = \frac{TP}{TP + FN} \tag{21}$$

where TP stands for the number of hits where radar echoes are both predicted and observed, FN represents the number of false alarms where radar echoes are predicted but not observed, TN indicates the number of missing alarms where radar echoes are not predicted and but observed, and FP signifies the number of correct rejections where radar echoes are not predicted and not observed.

6 Analysis of Experimental Results

To assess the effectiveness of the proposed *DiffREE* model in radar echo extrapolation tasks, a series of experiments are conducted, and the evaluation is performed through both quantitative and qualitative analyses. Four commonly used meteorological metrics, namely CSI, ETS, HSS and POD, are employed in the evaluation processes. The performance of the *DiffREE* model is assessed at different radar echo thresholds and compared with the baseline models by a complete test dataset. The results of the experiment are presented in Table 1,2.

Table 1 Critical Success Index (CSI) and Equitable Threat Score (ETS).

Model	CSI					ETS				
	$\tau =$ 10	$\tau =$ 20	$\tau =$ 30	$\tau =$ 50	Average	$\tau =$ 10	$\tau =$ 20	$\tau =$ 30	$\tau =$ 50	Average
MIM	0.310	0.352	0.375	0.148	0.296	0.224	0.266	0.358	0.137	0.246
PhyDNet	0.263	0.324	0.341	0.142	0.268	0.176	0.253	0.329	0.113	0.218
PredRNN	0.294	0.321	0.380	0.105	0.275	0.220	0.254	0.362	0.102	0.234
ConvLSTM	0.294	0.321	0.380	0.105	0.275	0.222	0.254	0.362	0.102	0.235
SmaAtUNet	0.287	0.351	0.414	0.138	0.298	0.210	0.273	0.391	0.127	0.250
DiffREE	0.418	0.405	0.453	0.173	0.362	0.338	0.337	0.432	0.168	0.319

From Table 1,2, it can be found that the proposed *DiffREE* model achieves the best CSI, ETS, HSS and POD scores at all thresholds for the dataset in Jiangsu Province, suggesting that the *DiffREE* model consistently outperforms the baseline models at

Table 2 Heidke Skill Score (HSS) and Probability of Detection (POD).

Model	HSS					POD				
	$\tau =$ 10	$\tau =$ 20	$\tau =$ 30	$\tau =$ 50	Average	$\tau =$ 10	$\tau =$ 20	$\tau =$ 30	$\tau =$ 50	Average
MIM	0.355	0.408	0.375	0.175	0.328	0.505	0.490	0.448	0.188	0.408
PhyDNet	0.330	0.403	0.439	0.163	0.334	0.482	0.492	0.397	0.150	0.380
PredRNN	0.365	0.401	0.476	0.154	0.349	0.420	0.489	0.499	0.142	0.387
ConvLSTM	0.354	0.307	0.470	0.160	0.323	0.415	0.424	0.316	0.174	0.332
SmaAtUNet	0.356	0.400	0.395	0.172	0.331	0.439	0.530	0.370	0.165	0.376
DiffREE	0.490	0.485	0.528	0.254	0.439	0.586	0.572	0.595	0.234	0.497

different radar echo thresholds. Specifically, when the threshold is set to τ=10 and τ=20 (corresponding to an echo between 10 dBZ and 30 dBZ), the *DiffREE* model improves by 25%, 37%, 27%, and 12% on average relative to the best baseline model. When the threshold is τ=30 (corresponding to an echo between 30 dBZ and 50 dBZ, which is in the range with the most concentrated echo intensity), the *DiffREE* model shows improvements of 9%, 10%, 11%, and 19% relative to the best baseline model. Even when the threshold reaches τ=50 (corresponding to a strong echo between 50 dBZ and 70 dBZ), the *DiffREE* model still exhibits improvements of 17%, 22%, 45%, and 25% relative to the best baseline model. These results demonstrate the effectiveness and superiority of the *DiffREE* model in radar echo extrapolation tasks, substantially enhancing the prediction performance at various radar echo thresholds.

To assess the model robustness over time in radar echo extrapolation tasks, we calculate the CSI, ETS, HSS and POD values for the two-hour predictions of each model based on the echo in the previous two hours at the thresholds of τ=10, τ=20, τ=30 and τ=50, as shown in Fig. 5. The results indicate that the curves for the *DiffREE* model are consistently higher than the other curves, indicating that the *DiffREE* model achieves the highest score at all time steps. Furthermore, the curve slope for the *DiffREE* model is relatively smaller than that of the other models, suggesting that the *DiffREE* model has more stable performance. An inherent drawback of radar echo extrapolation algorithms is the accumulation of errors over time, remarkably affecting the accuracy and quality of the long-term sequence of echo extrapolation results. Hence, the proposed *DiffREE* model leverages conditional encoding to fuse past radar echo frames with positional information and extract efficient feature embeddings, avoiding the complex and redundant invalid information that arises from the direct use of images as conditions. Moreover, in contrast to the traditional image-to-image generation methods, this study employs the feature conditions of past radar echo frames as guidance to progressively reconstruct Gaussian noise, thus mitigating error propagation.

Through the methods described above, the *DiffREE* model can enhance the robustness of radar echo extrapolation and effectively reduce the impact of error accumulation over time. The SSIM and ThrMSE are utilized to evaluate the effectiveness and superiority of the *DiffREE* model in radar echo generation. As shown in Fig. 4, as the leading time increases, the *DiffREE* model consistently outperforms the baseline models in terms of the SSIM and ThrMSE scores, demonstrating that the

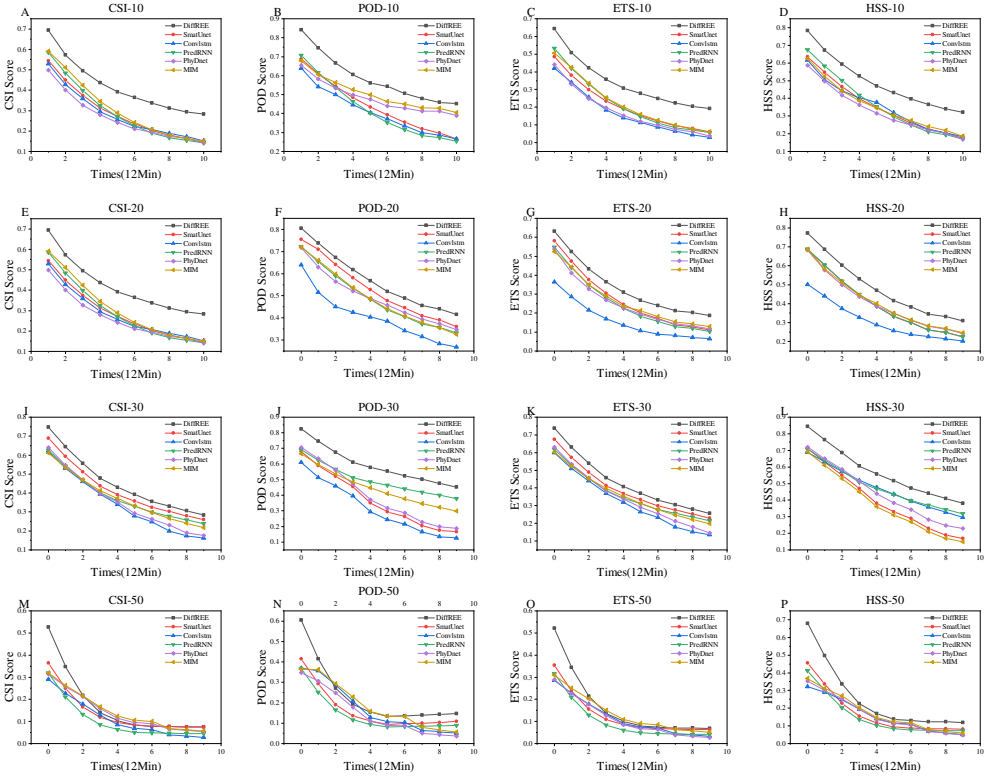

Fig. 3 Quantitative evaluation of the model performance at the two-hour forecast leading time: the critical success index (CSI), equitable threat score (ETS), Heidke skill score (HSS) and probility of detection (POD) values at thresholds $\tau = 10$, $\tau = 20$, $\tau = 30$, and $\tau = 50$

proposed approach in this study achieves higher accuracy and image quality in the image generation of radar echo extrapolation.

For a more nuanced qualitative analysis, Figures 5 and 6 showcase extrapolated images generated by all models for two test sequences from the Jiangsu station dataset. Figures 5 and 6 depict a layout where the initial column contains historical radar echo images, the second column displays the presently targeted radar echo image, and the remaining columns exhibit radar echo images forecasted by distinct models. The figure's color bar elucidates the mapping relationship between radar echo dBZ values and their corresponding colors.

A weak echo case presented in Fig. 5 indicates that as time progresses, all models capture the overall motion trend of the radar echo, but the baseline models perform unsatisfactorily in terms of echo intensity and feature distribution. In contrast, the proposed *DiffREE* model effectively captures fine echo features. Although there may be higher or lower echo intensities at specific points, the overall performance remains stable, and the image quality surpasses that of baseline models.

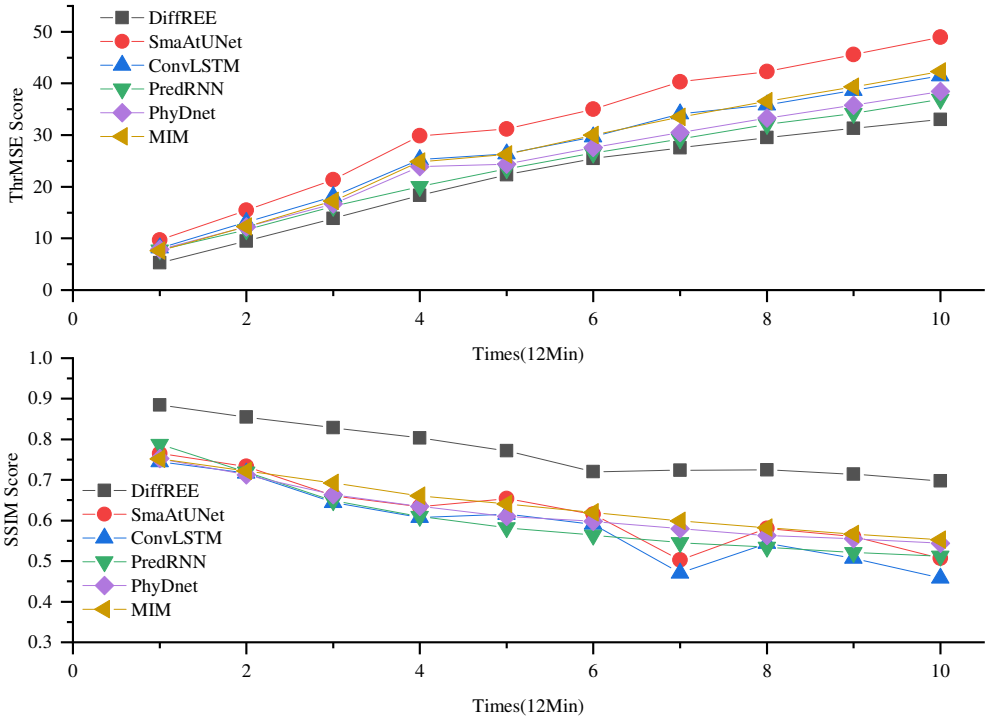

Fig. 4 Radar echo image performance evaluation scores for SSIM and ThrMSE.

Figure 6 shows a strong echo case, where the strong echo region highlighted by the red box is closely related to potential severe convective weather. Thus, the accurate extrapolation in this strong echo region is crucial. From the visualized radar echo predictions, it can be found that the *ConvLSTM* model predicts the general echo outlines but fails to predict the echo intensity effectively. The *PredRNN* model provides better predictions for the yellow region , but there is an apparent echo dissipation in the prediction area over time. Compared with the *PredRNN*, the *MIM* shows some improvement in echo intensity prediction and mitigating echo dissipation. However, it still fails to predict the strong echo in the focus region. Finally, the *SmaAtUNet* is the best-performing model among all baseline models, exhibiting a noticeable improvement in predicting echo trends, shapes and intensity. However, there is still a notable echo dissipation problem, preventing compelling predictions within the focus region.

Compared with the other models, the *DiffREE* model proposed in this research can predict radar echo boundaries almost consistently with the observed images, demonstrating remarkable accuracy in prediction. With the efficient feature extraction capability of the conditional encoding and the image reconstruction ability of the diffusion model, we are able to model radar echoes more accurately and effectively mitigate the echo dissipation problem in the focus region.

Several comparative models are designed as a part of the experiment to validate further the impact of the conditional encoding on the performance of the diffusion

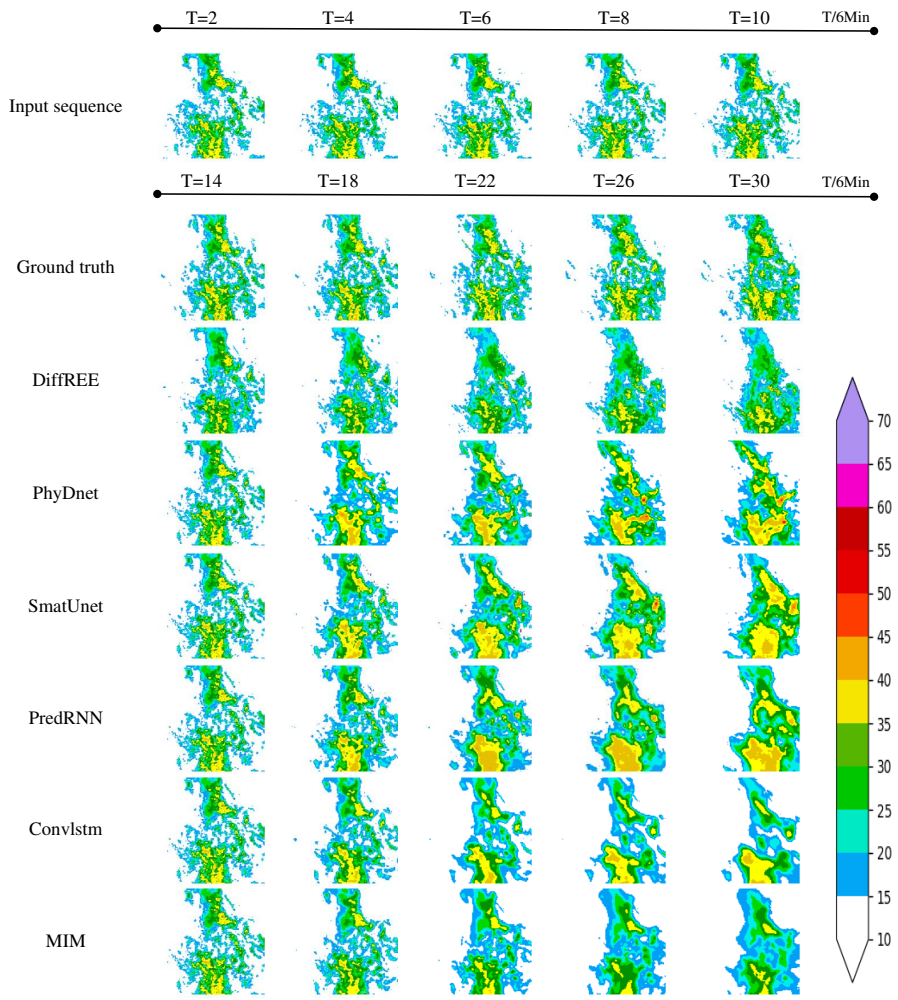

Fig. 5 Visualization for the results from the DiffREE model and the baseline models in the extrapolation of a weak radar echo case. The color bars display the mapping between the echo values and the colors.

model and find a relatively optimal approach. The first is the *DiffREE_Base* model, which directly takes past frames as conditions and does not use the conditional encoding scheme. The second model is the *DiffDREE_CNN* model, which applies convolution for feature extraction on the past frames as conditions. Furthermore, the *DiffDREE_CNN_TIME* model is designed, which adds temporal information as a condition based on the convolutional extraction of past frames. Finally, the *DiffREE* model, which adopts the condition encoding scheme, is also included in the experiment.

Figure 7 illustrates the comprehensive assessment of the model performance corresponding to thresholds $\tau = 10$, $\tau = 20$, $\tau = 30$, and $\tau = 50$. From the CSI, ETS,

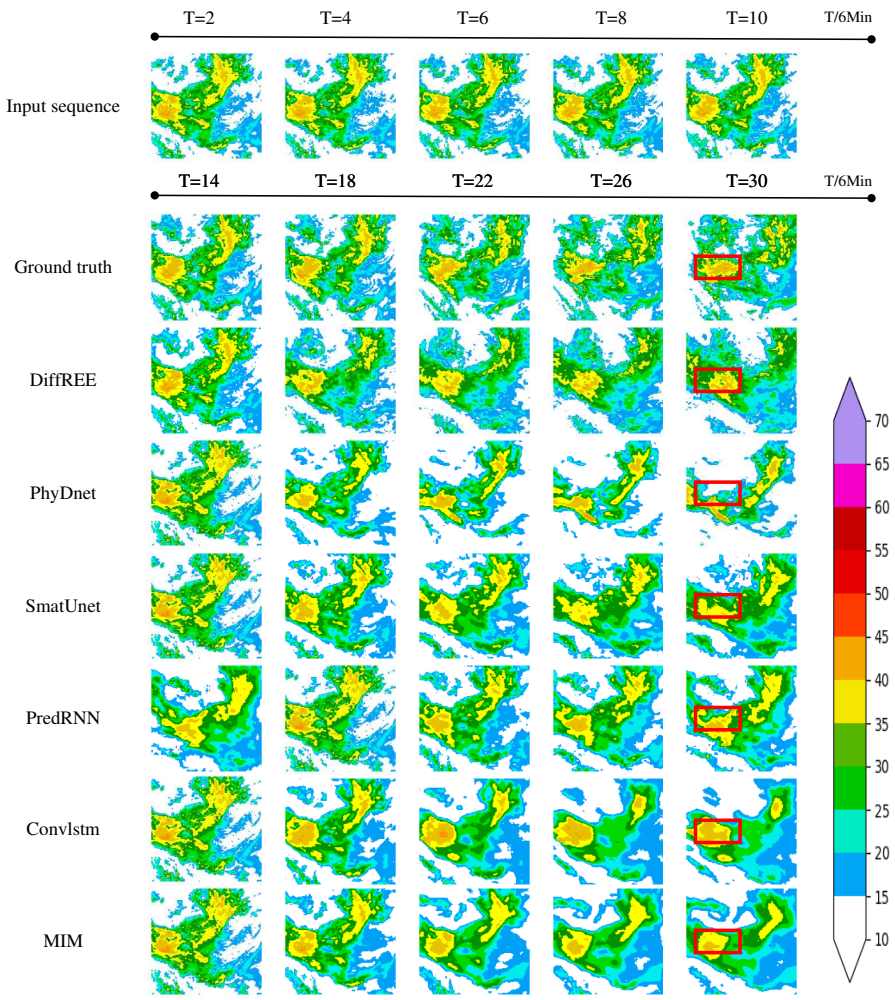

Fig. 6 Visualization for the results from the DiffREE model and the baseline models in the extrapolation of a strong radar echo case. The color bars display the mapping between the echo values and the colors.

HSS and POD values for each model at a two-hour leading time under the input condition of the past two-hour images, it can be clearly seen that the *DiffREE* model consistently outperforms the others.

These results further confirm the positive impact of the condition encoding on model performance. Compared with the other models, the *DiffREE* model achieves higher scores across different leading times and thresholds, indicating its robustness in radar echo extrapolation. This strongly supports the approach proposed in this study, emphasizing the importance of leveraging conditional encoding to integrate information from past frames and enhance the prediction capability of the model.

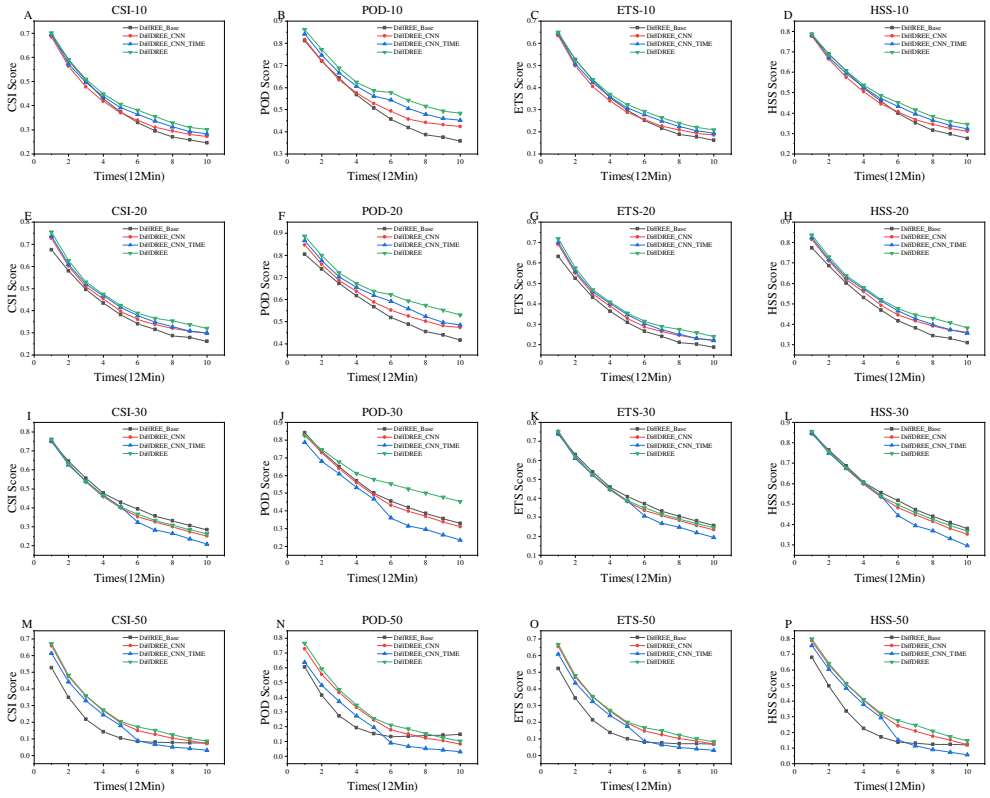

Fig. 7 Comprehensive assessment of the model performance at the two-hour forecast leading time: the CSI, ETS, HSS and POD values at specific thresholds $\tau = 10$, $\tau = 20$, $\tau = 30$, and $\tau = 50$.

In summary, the results of the experiments in this study demonstrate the outstanding performance of the *DiffREE* model on radar echo extrapolation. This achievement is of great academic and practical significance in meteorological forecasts and radar image processing, offering valuable guidance for further improving and optimizing radar echo prediction models.

7 Conclusion

This study develops a novel radar echo extrapolation model, *DiffREE*, which utilizes past radar echo frames as conditions to provide reliable and precise support for radar echo nowcasting. The proposed conditional encoding efficiently combines image-based spatial information with position-based spatio-temporal information, driving the diffusion model to reconstruct the current radar echo frame. The results of the experiments demonstrate that the model proposed in this research can accurately simulate radar echoes.

However, the reconstruction of radar echoes based on the diffusion model demands a large amount of computational resources for image diffusion, leading to more training costs and extended prediction time. Hence, future research may consider adopting a denoising diffusion implicit model, which treats the forward process as a discretized neural ordinary differential equation rather than the Markov process. Additionally, it is possible to reduce the sampling time by utilizing sparse time sequences during the generation process. These methods can help reduce the training costs and prediction time of the *DiffREE* model. Finally, further exploration of alternative solutions to enhance the accuracy of radar echo extrapolation while reducing the computational costs remains a potential avenue for research.

Acknowledgments. This work is financially supported by the National Natural Science Foundation of China (Grand No. 41805033) and the Natural Science Foundation of Jiangsu Province of China (Grand No. BK20221344). We thank Nanjing Hurricane Translation for reviewing the English language quality of this paper.

Declarations

- Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

- Data Availability Statement

The radar data at automatic weather stations used in this research are housed within the Experiment Teaching Center for Atmospheric and Environmental Sciences at Nanjing University of Information Science and Technology. The data is available at http://md.nuist.edu.cn/.

References

[1] Chen, X., Wang, M., Wang, S., Chen, Y., Wang, R., Zhao, C., Hu, X.: or extreme precipitation prediction based on the temporal and spatial generative adversarial network **13**(8), 1291 https://doi.org/10.3390/atmos13081291 . Number: 8. Accessed 2022-11-06

[2] Yu, T., Kuang, Q., Zheng, J., Hu, J.: Deep precipitation downscaling. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2021)

[3] Yao, S., Chen, H., Thompson, E.J., Cifelli, R.: An improved deep learning model for high-impact weather nowcasting. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **15**, 7400–7413 (2022)

[4] Li, L., Chen, S., Mai, X.-F.: Sub-pixel precipitation nowcasting over guangdong province using optical flow algorithm. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4638–4641. https://doi.org/10.1109/IGARSS.2017.8128034 . ISSN: 2153-7003

[5] Qiu, X., Zhang, F.: Prediction and predictability of a catastrophic local extreme precipitation event through cloud-resolving ensemble analysis and forecasting with doppler radar observations. Science China Earth Sciences **59**, 518–532 (2016)

[6] Wu, D., Wu, L., Zhang, T., Zhang, W., Huang, J., Wang, X.: Short-term rainfall prediction based on radar echo using an improved self-attention predrnn deep learning model. Atmosphere **13**(12), 1963 (2022)

[7] Tran, Q.-K., Song, S.-k.: Multi-channel weather radar echo extrapolation with convolutional recurrent neural networks. Remote Sensing **11**(19), 2303 (2019)

[8] Guo, S., Sun, N., Pei, Y., Li, Q.: 3d-unet-lstm: A deep learning-based radar echo extrapolation model for convective nowcasting. Remote Sensing **15**(6), 1529 (2023)

[9] Shi, E., Li, Q., Gu, D., Zhao, Z.: A method of weather radar echo extrapolation based on convolutional neural networks. In: MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24, pp. 16–28 (2018). Springer

[10] Sun, N., Zhou, Z., Li, Q., Jing, J.: Three-dimensional gridded radar echo extrapolation for convective storm nowcasting based on 3d-convlstm model. Remote Sensing **14**(17), 4256 (2022)

[11] Wang, Y., Yang, Z., Liu, Q., Liu, X.: An input sampling scheme to radar echo extrapolation for rnn-based models. In: 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 1–5 (2022). IEEE

[12] Liu, J., Xu, L., Chen, N.: A spatiotemporal deep learning model st-lstm-sa for hourly rainfall forecasting using radar echo images. Journal of Hydrology **609**, 127748 (2022)

[13] Bai, C., Sun, F., Zhang, J., Song, Y., Chen, S.: Rainformer: Features extraction balanced network for radar-based precipitation nowcasting. IEEE Geoscience and Remote Sensing Letters **19**, 1–5 (2022)

[14] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv. https://doi.org/10.48550/arXiv.2010.11929 . http://arxiv.org/abs/2010.11929 Accessed 2023-06-19

[15] Junchao, W., Zhibin, W., Anwei, L., Yanjiao, X., Jue, W.: Experimental study

on short-term and impending prediction of precipitation echo based on blending method of numerical prediction and radar extrapolation prediction. Journal of Arid Meteorology **40**(3), 485 (2022)

[16] Zhu, K., Chen, H., Han, L.: Mct u-net: A deep learning nowcasting method using dual-polarization radar observations. In: IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 4665–4668 (2022). IEEE

[17] Sit, M., Demiray, B.Z., Demir, I.: A systematic review of deep learning applications in interpolation and extrapolation of precipitation data (2022)

[18] Xu, L., Niu, D., Zhang, T., Chen, P., Chen, X., Li, Y.: Two-stage ua-gan for precipitation nowcasting. Remote Sensing **14**(23), 5948 (2022)

[19] Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., Hickey, J.: Machine Learning for Precipitation Nowcasting from Radar Images. arXiv. https://doi.org/10.48550/arXiv.1912.12132 . http://arxiv.org/abs/1912.12132 Accessed 2023-06-19

[20] Trebing, K., Stanczyk, T., Mehrkanoon, S.: SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture **145**, 178–186 https://doi.org/10.1016/j.patrec.2021.01.036 . Accessed 2022-11-05

[21] Tao, R., Zhang, Y., Wang, L., Cai, P., Tan, H.: Detection of precipitation cloud over the tibet based on the improved u-net. Computers, Materials & Continua **65**(3), 2455–2474 (2020)

[22] SHI, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., WOO, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems, vol. 28. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html Accessed 2022-11-06

[23] Wang, Y., Long, M., Wang, J., Gao, Z., Yu, P.S.: PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In: Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/e5f6ad6ce374177eef023bf5d0c018b6-Abstract.html Accessed 2022-11-06

[24] Wang, Y., Gao, Z., Long, M., Wang, J., Philip, S.Y.: Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In: International Conference on Machine Learning, pp. 5123–5132 (2018). PMLR

[25] Wang, Y., Zhang, J., Zhu, H., Long, M., Wang, J., Yu, P.S.: Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics, pp. 9154–9162. https://openaccess.thecvf.com/content_CVPR_2019/html/Wang_Memory_in_Memory_A_Predictive_

Neural_Network_for_Learning_Higher-Order_CVPR_2019_paper.html Accessed 2022-11-06

[26] Le Guen, V., Thome, N.: A deep physical model for solar irradiance forecasting with fisheye images, pp. 630–631. https://openaccess.thecvf.com/content_CVPRW_2020/html/w38/Le_Guen_A_Deep_Physical_Model_for_Solar_Irradiance_Forecasting_With_Fisheye_CVPRW_2020_paper.html Accessed 2022-11-06

[27] Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D.P., Poole, B., Norouzi, M., Fleet, D.J., et al.: Imagen video: High definition video generation with diffusion models. arXiv preprint arXiv:2210.02303 (2022)

[28] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models, pp. 10684–10695. https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html Accessed 2023-06-20

[29] Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Cui, B., Yang, M.-H.: Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796 (2022)

[30] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[31] Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through fsim, ssim, mse and psnr—a comparative study. Journal of Computer and Communications **7**(3), 8–18 (2019)

[32] Jing, J., Li, Q., Peng, X., Ma, Q., Tang, S.: Hprnn: A hierarchical sequence prediction model for long-term weather radar echo extrapolation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4142–4146 (2020). IEEE