Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

trajPredRNN+: A new approach for precipitation nowcasting with weather radar echo images based on deep learning

Chongxing Ji [a,b,*], Yuan Xu [a,c]

[a] Faculty of Applied Sciences, Macao Polytechnic University, Macao, 999078, China
[b] School of Artificial Intelligence, Dongguan City University, Dongguang, Guangdong, 523109, China
[c] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China

ARTICLE INFO

Keywords:
Radar echo map
Deep learning
Precipitation nowcasting
PredRNN
GAN

ABSTRACT

: Short-term rainfall prediction is a crucial and practical research area, with the accuracy of rainfall prediction, particularly for heavy rainfall, significantly impacting people's lives, property, and even their safety. Existing models, such as ConvLSTM, TrajGRU, and PredRNN, exhibit limitations in capturing fine-grained appearances due to insufficient memory units or addressing positional misalignment issues, thereby compromising the accuracy of model predictions. In this study, we propose trajPredRNN+, an innovative approach that integrates the trajectory segmentation model and the PredRNN deep learning model to address both limitations in nowcasting precipitation using weather radar echo images. By incorporating attention mechanisms, the model demonstrates an enhanced focus on short-term and imminent heavy rainfall events. To ensure improved stability during training, a residual network is introduced. Lastly, a more rational and effective training loss function is proposed, encompassing weight mechanism, SSIM index, and GAN loss. To validate the proposed model, we conducted a comparative experiment and an ablation experiment using the radar echo map dataset obtained from the Shenzhen Meteorological Bureau. The results of these experiments demonstrate that our model has achieved significant improvements across multiple key performance indicators.

1. Introdution

Accurate short-term rainfall prediction is crucially important, especially within the next 0–6 h. Short-term flood forecasting that includes precipitation prediction has long been a central focus in meteorological services. Precise weather forecasts enable efficient outdoor activity planning and timely alerts for floods or traffic incidents. Radar data along with surrounding precipitation patterns and meteorological information are frequently utilized to enhance the effectiveness of short-term precipitation forecasts [1].

Conventional prediction methods, such as Numerical Weather Prediction (NWP), utilize fluid dynamics equations and thermodynamic equations to incorporate atmospheric variables encompassing wind speed, pressure, and temperature. By effectively solving a system of equations, these methods can accurately forecast forthcoming weather conditions. However, due to the reliance on physical equations for simulating a complex atmospheric model in order to predict precipitation, NWP may not exhibit efficacy within 0–6 h and is thus unsuitable for short-term precipitation forecasting [2].

Another commonly employed traditional approach is the optical flow method, which leverages temporal pixel changes and

[*] Corresponding author. Faculty of Applied Sciences, Macao Polytechnic University, Macao, 999078, China.
E-mail address: jichongxing@dgcu.edu.cn (C. Ji).

https://doi.org/10.1016/j.heliyon.2024.e36134
Received 7 May 2024; Received in revised form 6 August 2024; Accepted 9 August 2024
Available online 10 August 2024

correlation between adjacent frames to establish correspondences between previous and current frames, thereby estimating object motion information across consecutive frames. This technique generates a two-dimensional vector field that encapsulates instantaneous velocity vectors for each pixel, facilitating identification and estimation of moving targets [3]. However, the optical flow method solely focuses on computing the two echo images before and after without exploiting historical sequence information from radar echo images. In terms of long-term prediction, it entails substantial computational workload while lacking timeliness [4].

In recent years, significant advancements have been made in the field of meteorology through the application of deep learning techniques. Particularly, remarkable progress has been achieved in short-term and imminent rainfall prediction. Deep learning is highly acclaimed by researchers for its capability to forecast future rainfall distribution and intensity based on analysis of historical short-term rainfall data. Currently, two primary categories of neural network models are employed for spatiotemporal sequence forecasting: one utilizes the convolutional neural network (CNN) architecture to generate sequential images, while the other employs recurrent neural networks (RNN) for predicting such sequences. Researchers like Shi and Wang have proposed several fundamental deep learning models including ConvLSTM [5], ConvGRU [6], TrajGRU [7], TrajLSTM [8], PredRNN [9] along with various variant models such as PredRNN++ [10], E3DLSTM [11], MIM [12], RainPredRNN [13], PFST-LSTM [14], SmaAt-UNet [15]. The experimental results demonstrate that the deep learning-based radar echo extrapolation method surpasses state-of-the-art numerical and optical flow methods in short-term rainfall prediction.

Since its proposal in 2014 [16], Generative Adversarial Networks (GAN) have exhibited remarkable performance across diverse domains [17–19]. Numerous researchers have integrated GAN networks into the field of short-term rainfall prediction to address challenges related to image clarity and enhance model prediction accuracy [20,21]. The fundamental principle of GAN involves a minimax game between two modules, namely the generator and the discriminator (as depicted in Formula 1). Throughout this game, both the discriminator and generator iteratively update their respective loss values, ultimately achieving a dynamic equilibrium process. This technique is commonly employed in prediction models to improve the quality of predicted images [22,23].

$$min_G \; max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1 - D(G(z)))] \tag{1}$$

Moreover, in recent years, Resnet and CBAM technologies have garnered substantial attention across diverse fields. Resnet allows for deeper and more robust model training, thus augmenting the effectiveness of training [24–27]. Furthermore, the CBAM attention mechanism enables targeted analysis of pivotal components within the graph structure [28–31].

The objective of this study is to further enhance the model architecture and refine the training of the loss function based on prior research, aiming to augment the predictive performance of the model. Our contributions can be summarized as follows.

(1) The proposed TrajPredRNN + model simultaneously addresses two limitations of convolutional recurrent neural networks: the absence of a memory cell for preserving fine-grained spatial appearances and the issue of position misalignment.
(2) By incorporating channel attention and spatial attention mechanisms into residual networks, our model effectively directs its focus towards regions exhibiting high echo values, such as heavy rainfall. This leads to enhanced predictive performance overall, particularly in areas characterized by pronounced echo values.
(3) In contrast to the prevailing use of MSE or MAE as loss functions, In the article proposed a more reasonable training loss function, which integrating weight mechanism, SSIM index, and generative adversarial network loss.
(4) Through comparative experiments demonstrates TrajPredRNN + model has achieved a significant enhancement in predictive performance. Furthermore, the efficacy of various improvement measures proposed in this study was substantiated through ablation experiments.

2. Related work

The conventional approach for short-term rainfall prediction involves utilizing the radar echo extrapolation method, which entails feeding past radar sequence echoes $X_{1:t}\left(X_{1:t} \in R^{t\times C\times W\times H}\right)$ into a prediction model for future radar echo sequence diagrams $X_{t+1:T}\left(X_{t+1:T} \in R^{(T-t)\times C\times W\times H}\right)$. Here, C, W, and H respectively represent the number of channels, length, and width of the image. In this section, we will provide an overview of several classic prediction models and related modules. The spatiotemporal sequence forecasting problem is to predict the most likely length-K sequence in the furture given the previous J observations which include the current one [19]:

$$\widetilde{X}_{t+1}, \ldots, \widetilde{X}_{t+K} = \underset{X_{t+1},\ldots,X_{t+K}}{arg\,max} p(X_{t+1}, \ldots, X_{t+K} | \widehat{X}_{t-J+1}, \widehat{X}_{t-J+2}, \ldots, \widehat{X}_t) \tag{2}$$

2.1. *Convolutional LSTM (ConvLSTM) and convolutional GRU (ConvGRU)*

The limited ability of LSTM and GRU models to incorporate dynamic spatial information, in addition to temporal dependencies between frames, hinders their exclusive reliance for short-term and imminent rainfall prediction. This limitation arises from the inadequate capability of both models to retain spatial information while processing sequential input data, rendering them unsuitable for predicting sequences with spatial dependence. To address this constraint, Shi and Ballas et al. pioneered the integration of convolutions with spatial memory capabilities into LSTM and GRU models, resulting in the development of ConvLSTM [5] and ConvGRU [6] models. These models employ an encoding-decoding ConvRNN structure for spatiotemporal prediction tasks (Fig. 1).

The initial state and unit output of the prediction network are obtained from the final state of the encoding network. Both networks

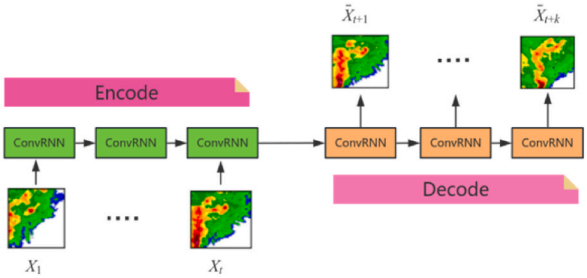

Fig. 1. Example of ConvRNN.

consist of multiple stacked ConvLSTM layers. To generate final predictions, all states within the prediction network are connected and passed through an 1 × 1 convolutional layer, taking into account that the prediction target shares identical dimensions with the input data. The equations governing ConvLSTM are described as follows:

$$\begin{aligned}
& g_t = tan\,h\left(W_{xg}{}^*X_t + W_{hg}{}^*\mathscr{H}_{t-1} + b_g\right) \\
& i_t = \sigma(W_{xi}{}^*X_t + W_{hi}{}^*\mathscr{H}_{t-1} + W_{ci} \odot C_{t-1} + b_i) \\
& f_t = \sigma\left(W_{xf}{}^*X_t + W_{hf}{}^*\mathscr{H}_{t-1} + W_{cf} \odot C_{t-1} + b_f\right) \\
& C_t = f_t \odot C_{t-1} + i_t \odot g_t \\
& o_t = \sigma(W_{xo}{}^*X_t + W_{ho}{}^*\mathscr{H}_{t-1} + W_{co} \odot C_t + b_o) \\
& \mathscr{H}_t = o_t \odot tan\,h(C_t)
\end{aligned} \tag{3}$$

Here, " *″ and ″⊙" represent separately convolution operator and the Hadamard product. The gates i_t, f_t, o_t represent separately input gates, forgetting gates, and output gates determine how to handle and retain long-term dependencies and ignore short-term memory. The input X_1 ,. . ,X_t output C_1 , . . ,C_t and the hidden states $\mathscr{H}_1$,. . ,$\mathscr{H}_\mathrm{t}$ are all three three-dimensional tensor, $X \in \mathbb{R}^{P\times M\times N}$, where $\mathbb{R}$ denotes the domain of the observed features, we observe a dynamical system over a spatial region represented by an $M\times$ N grid which consists of M rows and N columns. Inside each cell in the grid, there are P measurements which vary over time.

2.2. TrajLSTM and TrajGRU

The limitation of ConvLSTM and ConvGRU lies in their assumption that the spatial positions of pixels on input images remain unchanged during feature processing at different time steps. However, for radar echo images used in this study, there exists a 6-min interval between consecutive images. During this 6-min period, significant changes in the pixel positions of the front and rear radar echo maps have occurred (Figs. 2 and 3) due to atmospheric convection and other factors.

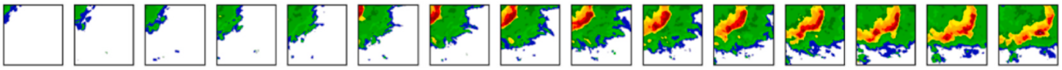

Fig. 2. A series of radar echo images, each with a time interval of 6 min.

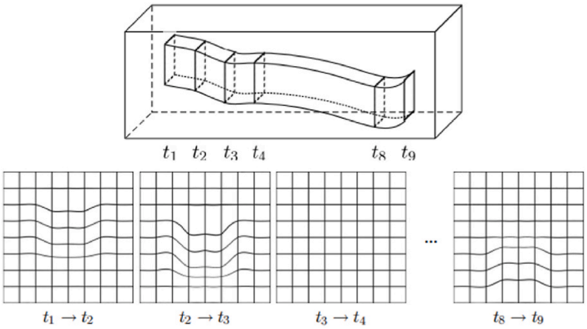

Fig. 3. The illustration depicts the warp obtained through polyharmonic interpolation, showcasing the propagation of spatio-temporal information pertaining to a moving object over an extended duration [8].

Consequently, Shi and Feng et al. proposed trajGRU [7] and TrajLSTM [8] models that incorporated trajectory structure to address the issue of positional misalignment. The traj structure is achieved by introducing a convolutional layer while constraining the output dimension, followed by iterative adjustment of convolution parameters to capture intricate optical flow movements. Essentially, it emulates the output settings of classical optical flow algorithms, where all traj structures represent the cumulative effect of adding optical flow vectors to the previous frame indicating positions in the image at specific points in time. Since positions are typically discrete and non-differentiable, interpolation methods can be utilized to obtain eigenvalues corresponding to position indices as feedback learning targets. The main formulas for TrajGRU are presented as follows:

$$
\begin{gathered}
\mathrm{U_t}, \mathrm{V_t} = \gamma(\mathrm{X_t}, \mathscr{H}_{\mathrm{t-1}}) \\
\mathrm{Z_t}\left(\mathrm{W_{xz}}{*}\mathrm{X_t} + \sum\nolimits_{\mathrm{l=1}}^{\mathrm{L}} \mathrm{W_{hz}^{l}}{*}\mathrm{warp}(\mathscr{H}_{\mathrm{t-1}}, \mathrm{U_{t,l}}, \mathrm{V_{t,l}})\right)\mathscr{R}_{\mathrm{t}} = \sigma\left(\mathrm{W_{xr}}{*}\mathrm{X_t} + \sum\nolimits_{\mathrm{l=1}}^{\mathrm{L}} \mathrm{W_{hr}^{l}}{*}\mathrm{warp}(\mathscr{H}_{\mathrm{t-1}}, \mathrm{U_{t,l}}, \mathrm{V_{t,l}})\right)\mathscr{H}'_{\mathrm{t}} \\
= \mathrm{f}\left(\mathrm{W_{xh}}{*}\mathrm{X_t} + \mathscr{R}_{\mathrm{t}} \odot \left(\sum\nolimits_{\mathrm{l=1}}^{\mathrm{L}} \mathrm{W_{hh}^{l}}{*}\mathrm{warp}(\mathscr{H}_{\mathrm{t-1}}, \mathrm{U_{t,l}}, \mathrm{V_{t,l}})\right)\right)\mathscr{H}_{\mathrm{t}} = (1 - \mathrm{Z_t}) \odot \mathscr{H}'_{\mathrm{t}} + \mathrm{Z_t} \odot \mathscr{H}_{\mathrm{t-1}}
\end{gathered}
\tag{4}
$$

where optical flow fields $\mathrm{U_t}$ and $\mathrm{V_t}$, capable of storing dynamic connections γ between network layers, are utilized to establish dynamic connections between input information $\mathrm{X_t}$ and $\mathscr{H}_{\mathrm{t-1}}$ through the warp function. Here, L denotes the total number of connections.

The TrajGRU and TrajLSTM models effectively address the issue of position misalignment. However, they lack the preservation of fine-grained spatial appearance units and restrict the transfer of spatial states to their respective time steps. Consequently, there is no sharing of spatial memory units among different time steps, which hinders the utilization of each time step's spatial distribution information during the learning process.

2.3. PredRNN

The PredRNN architecture, proposed by Wang et al. [9], introduces a zigzag structure and incorporates the Spatiotemporal Memory Unit (ST-LSTM) (Fig. 4).

In spatiotemporal prediction learning, the detailed information of the original input sequence should remain unchanged. If you want to look to the future, you need to learn from the features extracted from different levels of convolutional layers. Therefore, a zigzag structure was constructed (in the right half of Fig. 4), with the orange arrow representing the feedforward direction of the LSTM storage unit. All LSTMs share a unified memory, which is updated along the zigzag direction. This structure breaks through the limitation of memory units in ConvLSTM and TrajGRU that can only be updated horizontally, and information can only be transmitted upwards from hidden states.

The PredRNN model makes another significant contribution through the author's innovative design of the ST-LSTM unit, which seamlessly incorporates a dedicated storage component. The newly introduced spatiotemporal memory module *M*, depicted in orange, expands upon the LSTM architecture and facilitates the transfer of spatiotemporal memory across layers. The white component represents the time memory module *C*, responsible for horizontal time flow transfer. The formulation of ST-LSTM is presented as follows:

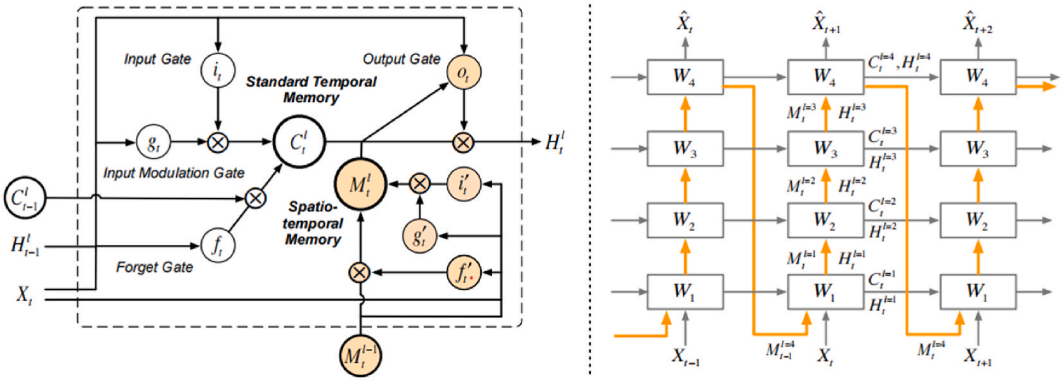

Fig. 4. ST-LSTM (left) and PredRNN (right). The orange circles in the ST-LSTM unit denotes the differences compared with the conventional ConvLSTM. The orange arrows in PredRNN denote the spatiotemporal memory flow, namely the transition path of spatiotemporal memory M_t^l in the left [9].

$$\begin{aligned}
&g_t = tan\,h\left(\left(W_{xg}{}^{*}X_t + W_{hg}{}^{*}\mathscr{H}_{t-1}^{l} + b_g\right)\right.\\
&i_t = \sigma\left(W_{xi}{}^{*}X_t + W_{hi}{}^{*}H_{t-1}^{l} + b_i\right)\\
&f_t = \sigma\left(W_{xf}{}^{*}X_t + W_{hf}{}^{*}H_{t-1}^{l} + b_f\right)\\
&C_t^l = f_t \odot C_{t-1}^{l} + i_t \odot g_t\\
&g_t' = tan\,h\left(\mathrm{W}_{\mathrm{xg}}'{}^{*}\mathrm{X}_\mathrm{t} + \mathrm{W}_{\mathrm{mg}}{}^{*}\mathscr{M}_{\mathrm{t}}^{\mathrm{l-1}} + \mathrm{b}_\mathrm{g}'\right)\\
&i_t' = \sigma\left(W_{xi}'{}^{*}X_t + W_{mi}{}^{*}M_t^{l-1} + b_i'\right)\\
&f_t' = \sigma\left(W_{xf}'{}^{*}X_t + W_{mf}{}^{*}M_t^{l-1} + b_f'\right)\\
&M_t^l = f_t' \odot M_t^{l-1} + i_t' \odot g_t'\\
&o_t = \sigma\left(W_{xo}{}^{*}X_t + W_{ho}{}^{*}H_{t-1}^{l} + W_{co}{}^{*}C_t^l + W_{mo}{}^{*}M_t^l + b_o\right)\\
&H_t^l = o_t \odot tan\,h\left(\mathrm{W}_{1\times1}{}^{*}\left[\mathrm{C}_\mathrm{t}^\mathrm{l}, \mathscr{M}_\mathrm{t}^\mathrm{t}\right]\right)
\end{aligned} \tag{5}$$

Where $\mathscr{H}_t^l$ and $\mathscr{M}_t^l$ are the hidden state and spatial memory state of the l th level at t time. C_t^l is a time memory unit.

The PredRNN model addresses the challenge of effectively incorporating asynchronous sharing of spatial appearance to enhance the preservation of intricate spatial appearance characteristics. However, it encounters a position alignment issue that becomes particularly evident in scenarios involving changes in image position, such as movement, rotation, and scaling.

2.4. Other related models

Based on the aforementioned five models, several enhanced versions have been introduced. For instance, PredRNN++, E3DLSTM, MIM, PFST-LSTM, SmaAt-UNet, and RainPredRNN [10–15] primarily focus on two aspects for improvement. Firstly, they optimize the network architecture by incorporating gradient highways and UNet models. Secondly, enhancements are achieved through the introduction of novel memory units such as Causal LSTM, MIM (Memory In Memory), pseudo flow modules, etc. For example the RainPredRNN model combines the UNet network structure with the PredRNN_v2 [32] structure to enhance short-term rainfall prediction. By leveraging the contracting expansion ability of UNet, their model significantly reduces computational complexity and improved the accuracy of short-term rainfall prediction.

3. Proposed model

3.1. trajPredRNN+

Building upon the frameworks of PredRNN and Traj structure, our study employs an encoder-decoder architecture wherein the encoder incorporates three layers of convolutional downsampling to extract spatiotemporal information, while the decoder comprises three layers of convolutional upsampling to effectively leverage this transmitted spatiotemporal information for predicting subsequent sequences. The encoder's top layer in the previous time step and the subsequent time step are connected through deconvolution operations, while the decoder is linked via convolution operations (Fig. 5). This design facilitates the transmission of memory units containing detailed spatial features in both vertical and horizontal directions, enabling the sharing of memory units across different time steps. Consequently, this model exhibits enhanced capability to learn spatial appearances.

The prediction of two future frames $\overline{\mathrm{X}}_{\mathrm{t+1}}, \overline{\mathrm{X}}_{\mathrm{t+2}}$ is achieved by employing three recurrent neural networks, given the two input frames $\mathrm{X}_{\mathrm{t-1}}, \mathrm{X}_\mathrm{t}$, To ensure accurate spatial awareness within observations, the spatial coordinates Traj are concatenated with the input frame. By utilizing convolutional structures for state transition and iteratively adjusting convolutional parameters to capture optical flow vectors, the value update of the Traj structure is derived from its previous state combined with its optical flow vector. This enables the acquisition of positional information for each point in the subsequent graph frame. Due to the discreteness and non-differentiability of position domain values, an interpolation method is employed to obtain corresponding feature values for feedback learning.

Incorporating pre-operation feature vectors into the post-operation feature vectors during convolution and deconvolution operations significantly enhances our model's training efficacy, even with its limited three-layer architecture. This phenomenon can be attributed to the inherent capability of residual network structures in effectively capturing intricate details and contextual information within shallow neural network models, thereby augmenting the network's expressive power. Moreover, this architectural design facilitates seamless information flow while mitigating excessive gradient reduction during transmission, resulting in improved training stability. The impact of incorporating or excluding residual networks is further demonstrated through subsequent ablation experiments conducted in this article. Additionally, we performed supplementary experiments on cross layer residual networks and observed no superior outcomes compared to single-layer connections.

The primary objective of the Convolutional Block Attention Module (CBAM) is to enhance the perception capability of the model by integrating channel attention and spatial attention into convolutional neural networks (CNNs), thereby improving performance without increasing network complexity. CBAM aims to address the limitations of traditional CNNs in processing information with varying scales, shapes, and orientations. Specifically, channel attention enhances feature representation across different channels, while spatial attention extracts crucial information from various spatial positions (Fig. 6).

The channel attention weights are obtained through the application of the Sigmoid activation function. Subsequently, these

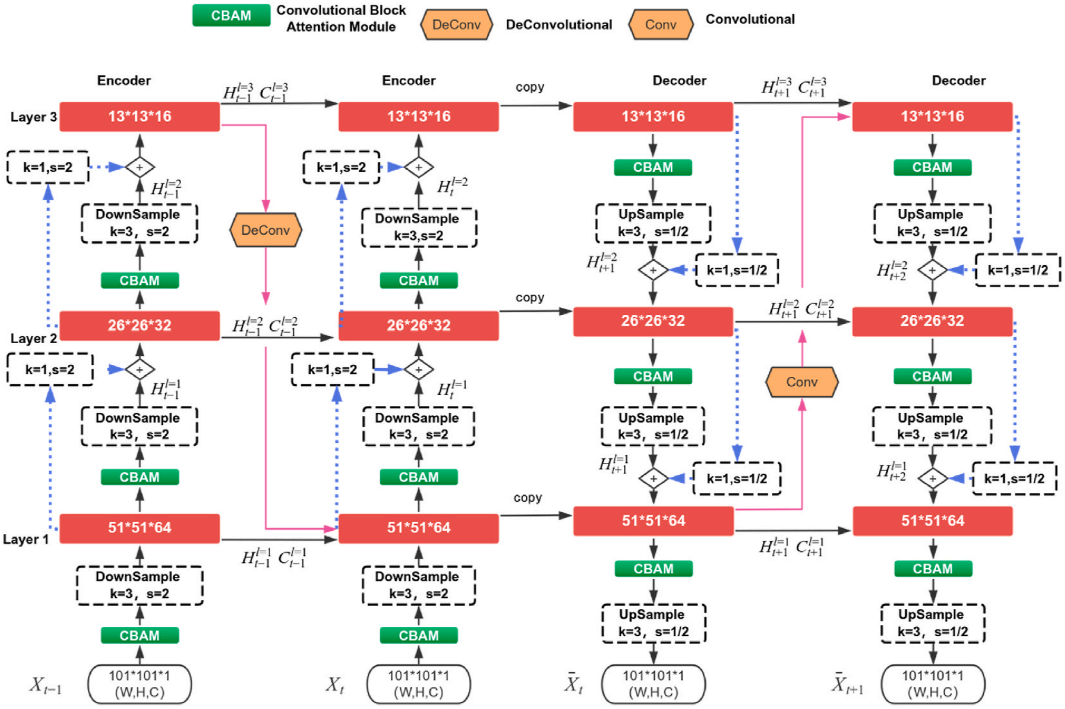

Fig. 5. The trajPredRNN + model was enhanced by incorporating Resnet and CBAM modules, as illustrated in the figure with a blue dashed arrow representing the newly introduced Resnet mechanism, while the green square symbolizes the newly integrated CBAM module.

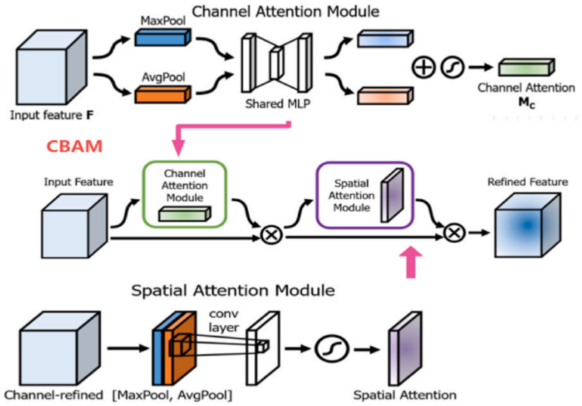

Fig. 6. The CBAM module comprises two components: the channel attention CAM and spatial attention SAM modules. The channel attention component involves global max pooling and global average pooling, fully connected layers, sigmoid activation, and attention weighting. Meanwhile, the spatial attention component undergoes max pooling and average pooling, connection and convolution, sigmoid activation, and attention weighting [28].

attention weights are multiplied with the original feature map to generate a channel feature map that emphasizes relevant channels for the given task while suppressing irrelevant ones. The formulation for channel attention is expressed as follows:

$$M_c(F) \quad = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma\left(\mathrm{W}_1\left(\mathrm{W}_0\left(F^c_{avg}\right)\mathrm{W}_1\left(\mathrm{W}_0\left(\mathrm{F}^c_{\max}\right)\right)\right)\right) \tag{6}$$

where σ denotes the sigmoid function, $W_0 \in \mathbb{R}^{C/r \times C}$, and $W_1 \in \mathbb{R}^{C \times C/r}$. Note that the *MLP* weights, W_0 and W_1, are shared for both inputs and the ReLu activation function is followed by W_0.

The spatial attention mechanism employs max pooling and average pooling operations along the channel dimension of the input radar echo feature map, enabling the capture of contextual features at multiple scales. These features are concatenated along the channel dimension to create a comprehensive feature map encompassing diverse contextual information. Spatial attention weights that effectively highlight salient regions within the image are generated after subjecting this feature map to convolutional layers. These weights are then constrained between 0 and 1 using the Sigmoid activation function. Finally, by applying these obtained spatial

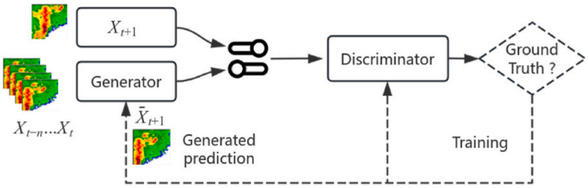

Fig. 7. Introducing the GANloss loss map, the model generates a prediction map by inputting n radar echo maps, which is then utilized alongside the real radar echo map as an input for the discriminator to discern between genuine and counterfeit instances. Subsequently, gradient feedback is employed to update their respective parameters. The discriminator's loss is referred to as GANloss.

attention weights to the original feature map, important areas can be effectively emphasized while mitigating the influence of less significant regions. The formulation of spatial attention is expressed as follows:

$$M_s(F) = \sigma\left(f^{3\times3}([AvgPool(F); MaxPool(F)])\right) = \sigma\left(f^{3\times3}\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \tag{7}$$

where σ denotes the sigmoid function, and $f^{3\times3}$ represents a convolution operation with the filter size of 3 × 3 in our model. $M_s(F) \in R^{H\times W}$ mean convolution layer to generate a spatial attention map. $F^s_{avg} \in \mathbb{R}^{1\times H\times W}$ and $F^s_{\max} \in \mathbb{R}^{1\times H\times W}$ represent separately average-pooled features and max-pooled features across the channel.

The ultimate improved feature is acquired through element-wise multiplication between the output features of both channel attention and spatial attention modules within CBAM, leading to enhanced attention. This amplified feature acts as input for subsequent network layers to efficiently attenuate noise and extraneous information while retaining essential information. The introduction of CBAM aims to enhance the model's ability to capture crucial information in the image, such as regions with high echo intensity that indicate intense rainfall in radar echo maps.

3.2. Optimization of training loss function

In line with the loss functions employed by numerous existing models, we propose incorporating weight mechanisms and Generative Adversarial Network (GAN) losses based on mean squared error (MSE), absolute error (MAE), and structure similarity index measure (SSIM) to enhance the scientific rigor and effectiveness of the model training loss function.

The prediction model functions as a generator, while the discriminator evaluates the accuracy of the predicted echo by comparing it with characteristics exhibited in genuine radar echoes (Fig. 7).

This process simultaneously generates losses for both the generator and discriminator, enabling the refinement of predicted graphics based on actual radar echo maps by leveraging gradient feedback. The integration of each round of model training plays a crucial role in rectifying and fine-tuning the generated structural graphics. The specific methodology employed in this section is elaborated upon in the experimental section.

4. Experiments

4.1. Dataset

The CIKM AnalytiCup 2017 competition radar echo map dataset, available for download at https://tianchi.aliyun.com/dataset/1085/, is utilized in this study. This dataset consists of radar maps obtained by the Shenzhen Meteorological Bureau, covering a target location and its surrounding area represented as an m × m grid. Each grid point records the radar reflectivity factor value Z, ranging from small to large values. To facilitate measurement, dBZ is employed to quantify this value: $dBZ = 10\log(Z/Z_0)$, where $Z_0 = 1\ mm^6/m^3$. The dataset includes radar maps captured at different time intervals of 6 min over a total of 15 time spans and at various altitudes with an interval of 1 km, ranging from 0.5 km to 3.5 km for a total of four altitudes. Based on the latitude and longitude coordinates of the target location, each radar map encompasses an area of 101 × 101 square kilometers. Subsequently, the data is transformed into image format with dimensions of 101 × 101 pixels per image, representing a resolution of one square kilometer per pixel. The conversion formula between dBZ and pixs is $pixel = \left\lfloor 255 \times \frac{dBZ+10}{95} \right\rfloor$. It is worth noting that the radar reflectivity formulas provided in this study are derived from stormfall intensity values (mm/h), which have been determined based on the data characteristics of Shenzhen Meteorological Bureau. However, it should be emphasized that the specific conversion formula needs to be established according to the unique data source itself. For instance, when utilizing the HKO-7 dataset provided by Hong Kong Meteorological Bureau, the Z-R relationship can be expressed as $dBZ = 10\log a + 10b\log R$, where R represents rain-rate level and is associated with a value of 58.53 and b value of 1.56. The raw logarithmic radar reflectivity factors are converted into pixel values through a linear transformation $pixel = \left\lfloor 255 \times \frac{dBZ+10}{70} + 0.5 \right\rfloor$, The values are constrained within the range of 0–255 b y means of clipping. We partitioned the radar echo images into three distinct sets: a training set consisting of 8000 image sequences, a validation set comprising 2000 sequences, and a test set containing 4000 sequences. Each sequence is composed of fifteen consecutive radar echo images.

Table 1
Rain rate statistics in the Shenzhen Meteorological Dataset.

Rain Rate (mm/h)	Proportion (%)	Rainfall Level
$0 \leq x \leq 20$	82	Below Light Rain
$20 \leq x < 35$	10.08	Light to Moderate
$35 \leq x < 45$	5.40	Moderate to Heavy
$45 \leq x$	2.52	Heavy to rainstorm

4.2. Parameter setting and evaluation metrics

The models presented in this article were trained using the GeForce RTX 4070 Ti SUPER, operating on the Ubuntu platform. PyTorch was employed as the programming framework, while Adam optimizer with a uniform learning rate of 0.001 was utilized. Additionally, an early termination training strategy was implemented (Table 1).

We performed segmented statistical analysis on the pixel value distribution of the dataset, and the resulting data is presented in the table below.

Based on the distribution characteristics of pixel values, we propose the following weight value segmentation formula.

$$w(x) = \begin{cases} 1, & x < 20 \\ 10, & 20 \leq x < 35 \\ 20, & 35 \leq x < 40 \\ 30, & x \geq 45 \end{cases} \tag{8}$$

The effectiveness of various weight allocation strategies was evaluated, however, they did not yield superior training outcomes. The selection of weight values primarily hinges on addressing inherent data imbalances.We separately trained prediction models with and without weight mechanism, and the results of our predictions revealed that the weighted model exhibited significantly enhanced accuracy in high echo regions compared to its unweighted counterpart. The underlying cause of this phenomenon lies in the limited representation of high echo pixel values (indicative of heavy rainfall areas) within the dataset, resulting in an imbalanced distribution that hampers the model's ability to adequately capture and learn the distinctive characteristics associated with these regions. To tackle this issue, one potential approach involves the manual generation of high echo values using advanced techniques such as Generative Adversarial Networks (GAN) or the Synthetic Minority Over-sampling Technique (SMOTE). However, in the context of radar echo maps depicting short-term rainfall, artificially generated samples may not accurately represent real-world data instances and can potentially hinder model training due to dynamic changes in pixel positions, including rotation, movement, scaling, generation, and disappearance within small areas. Therefore, we introduced a weighting mechanism during training to assign higher importance to regions with limited data distribution. Incorporating this weight mechanism led to significant enhancements in predictive performance within high echo areas.

The training loss function of our model incorporates four indicators, namely MSE, MAE, SSIM, and GANloss, with a weighted mechanism. The formulations for MSE and MAE are presented as follows:

$$\mathrm{W\_MSE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{i}=1}^{101}\sum_{\mathrm{j}=1}^{101} \mathrm{w}_{\mathrm{n,i,j}}\left(\mathrm{x}_{\mathrm{n,i,j}} - \widehat{\mathrm{x}}_{\mathrm{n,i,j}}\right)^2 \tag{9}$$

$$\mathrm{W\_MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{n}=1}^{\mathrm{N}}\sum_{\mathrm{i}=1}^{101}\sum_{\mathrm{j}=1}^{101} \mathrm{w}_{\mathrm{n,i,j}}\left|x_{n,i,j} - \widehat{x}_{n,i,j}\right| \tag{10}$$

where N represents the total number of frames, The weight assigned to the (i,j)th pixel in the n_th frame is denoted as Formula 8.

The training loss function incorporates the SSIM index to enhance the model's ability to scientifically assess the similarity between the predicted radar echo map and the actual radar echo map during training. The formula is presented as follows:

$$\mathrm{SSIM}\ (x,y) = l(x,y)\cdot c(x,y)\cdot s(x,y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\cdot\left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right)\cdot\left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right) \tag{11}$$

Here, $l(x,y)$, $c(x,y)$ and $s(x,y)$ respectively represent the brightness similarity, brightness similarity, and structure similarity. μ_x and μ_y are means of x and y, σ_x and σ_y are standard deviations of x and y, σ_{xy} is the cross correlation of x and y. C_1, C_2 and C_3 are small positive constants for avoiding zero division and numerical instability. The aforementioned items exclusively pertain to a specific local region within the image, rather than encompassing its entirety. In the SSIM value range of $(-1, 1)$, a higher value closer to 1 indicates a stronger similarity between the two images. This contrasts with the MSE indicator, where a lower value signifies greater dissimilarity. To conform to this convention and ensure a normalized range of (0, 1), we modify it as (1-SSIM)/2. The GAN training loss can manifest as either the generator loss or the discriminator loss, and in our code, the discriminator loss was opted. Similar to W_MSE, weightage is assigned to SSIM and GANloss in our final training loss formula.

$$\mathrm{LOSS} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{101}\sum_{j=1}^{101} w_{n,i,j}(\mathrm{W\_MSE} + \mathrm{W\_MAE} + (1 - SSIM)\ /\ 2 + GANloss) \tag{12}$$

We utilize a threshold τ to discretize the pixel values in both the prediction and ground-truth data. Subsequently, true positives (TP) are computed when the prediction is 1 and truth is 1, false negatives (FN) occur when the prediction is 0 but truth is 1, false positives (FP) arise when the prediction is 1 but truth is 0, and true negatives (TN) are determined when both prediction and truth are 0. These metrics are evaluated using well-established meteorological scoring functions: Heidke skill score (HSS), critical success index (CSI), probability of detection (POD), and false alarm rate (FAR). The calculation formula for each metric is as follows:

$$\begin{aligned} HSS &= \frac{2^*(FN^*TN - FN^*FP)}{(TP+FN)^*(FP+FN) + (TP+FP)^*(FP+TN)} \\ CSI &= \frac{TP}{TP+FP+FN} \\ POD &= \frac{TP}{TP+FP} \\ FAR &= \frac{FN}{TP+FN} \end{aligned} \tag{13}$$

Where HSS quantifies the accuracy of predictions by considering random correctness, CSI measures accuracy in rare or unpredictable events, POD represents the probability of correctly forecasting short-term precipitation, while FAR indicates the proportion of incorrect predictions among total precipitation forecasts.

Specifically, the thresholds of 5, 35, and 45 dBZ were selected for evaluation purposes. In terms of assessing model predictive performance, a lower FAR indicates superior results, while higher values of HSS, CSI, POD are indicative of better performance. Furthermore, to evaluate the effectiveness of the model, MSE and SSIM were incorporated. A lower MSE signifies improved prediction capabilities of the model, whereas an SSIM value closer to 1 suggests better performance.

4.3. Results and analysis

Our model demonstrates superior performance across various dBZ thresholds, particularly at the highest thresholds. By examining the average Structural Similarity Index (SSIM) and Mean Squared Error (MSE) values for each model (Fig. 8), it can be calculated that our model exhibits a significant improvement of 8.9 % and 7.6 %, respectively, compared to the best-performing model among the other five models, as well as a remarkable enhancement of 17.7 % and 30.9 %, respectively, compared to the least performing model.

By referring to Table 2, we calculate the evaluation index values of radar echo prediction maps under various thresholds. Our approach demonstrates superior performance in terms of HSS, CSI, POD, and FAR metrics compared to even the highest-performing

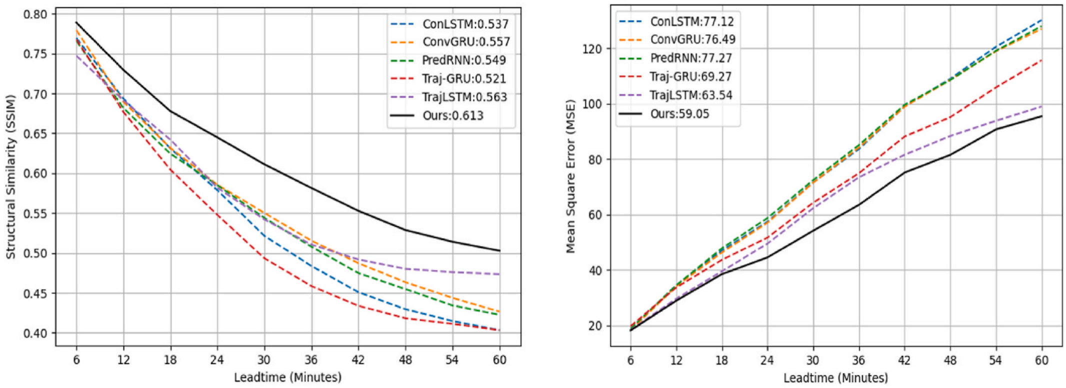

Fig. 8. The performance variations in relation to different nowcast lead times are assessed based on SSIM and MES scores. The model name is presented in the top column, followed by the average score of 10 frames of radar echo maps. The curve depicts the sequential values of SSIM or MES scores for each predicted frame (1-10) of radar echo map.

Table 2
The mean values of HSS, CSI, POD, and FAR for each model at various thresholds, with the best result highlighted in bold within a specific setting.

Algorithms	HSS			CSI			POD			FAR		
	$\tau = 5$	$\tau = 35$	$\tau = 45$	$\tau = 5$	$\tau = 35$	$\tau = 45$	$\tau = 5$	$\tau = 35$	$\tau = 45$	$\tau = 5$	$\tau = 35$	$\tau = 45$
ConLSTM	0.7564	0.2993	0.0783	0.7957	0.2140	0.0477	0.8766	0.2420	0.0321	0.0766	0.6086	0.8803
ConvGRU	0.7782	0.2676	0.0397	0.8179	0.1945	0.0236	0.8557	0.2938	0.0367	0.0701	0.5439	0.8848
PredRNN	0.7657	0.3213	0.0476	0.8049	0.2322	0.0275	0.8557	0.2938	0.0367	**0.0700**	0.5439	0.8785
Traj-GRU	0.7705	0.1705	0.0171	0.8073	0.1170	0.0093	0.8550	0.1309	0.0097	0.0660	0.5674	0.8643
TrajLSTM	0.7586	0.2400	0.0587	0.7985	0.1714	0.0337	0.8493	0.1987	0.0395	0.0714	0.5887	0.8591
Ours	**0.7970**	**0.4293**	**0.1084**	**0.8329**	**0.3123**	**0.0660**	**0.8922**	**0.4090**	**0.1016**	0.0746	**0.4728**	**0.7970**

model among these five models. Specifically at thresholds of 35 dBZ and 45 dBZ, our model demonstrates significant improvements in HSS by 33.61 % and 38.44 %, CSI by 34.50 % and 38.36 %, POD by 39.21 % and 57.22 %, while also achieving a reduction in FAR by 13.07 % and 7.22 %.

The calculation results demonstrate the consistent superiority of our model over other models, particularly in high threshold scenarios where its performance is exceptional. This superiority is further supported by the trends of various indicators depicted in Figs. 9 and 10. Although the meteorological evaluation indicators HSS, CSI, and POD encompass distinct physical interpretations, their mathematical monotonicity remains consistent based on the calculation formulas. Consequently, we present a trend chart illustrating changes in HSS. The figure demonstrates a consistent superiority of our model over other models in terms of HSS score throughout the entire prediction process, with an increasingly pronounced margin as the prediction progresses. Although our FAR score may initially lag behind that of other models during the first 1–2 predictions, our model swiftly surpasses them and maintains a substantial lead as the prediction advances.

The contour accuracy of TrajLSTM and TrajGRU in predicting the distribution of pixel values is observed to be higher than that of ConvLSTM, ConvGRU, and PredRNN (Fig. 11). This can be attributed to their incorporation of a position alignment algorithm unit which dynamically tracks the positional changes of each pixel. Although PredRNN outperforms other models in high echo areas during the prediction process, its alignment capability falls short of expectations. This indirectly confirms that PredRNN excels at storing intricate appearance memory units and effectively propagating information from memory alterations throughout the entire prediction period. However, due to the absence of position alignment units, the prediction accuracy for high echo areas progressively diminishes. Our model demonstrates superior performance compared to other models regarding position alignment and preservation of high echo areas. To enable a point-to-point comparison, we selected the black boxed section from the last set of images as an evaluation template and observed that our model demonstrated superior accuracy. This accomplishment can be attributed to our successful integration of advantageous features from other models, followed by a series of effective enhancements and optimizations.

From the comparison of MES, SSIM, HSS, CSI, POD, FAR evaluation index scores and visualization graphs above, our model performs better than the other five classic models, which means that our model has higher prediction accuracy than other models.

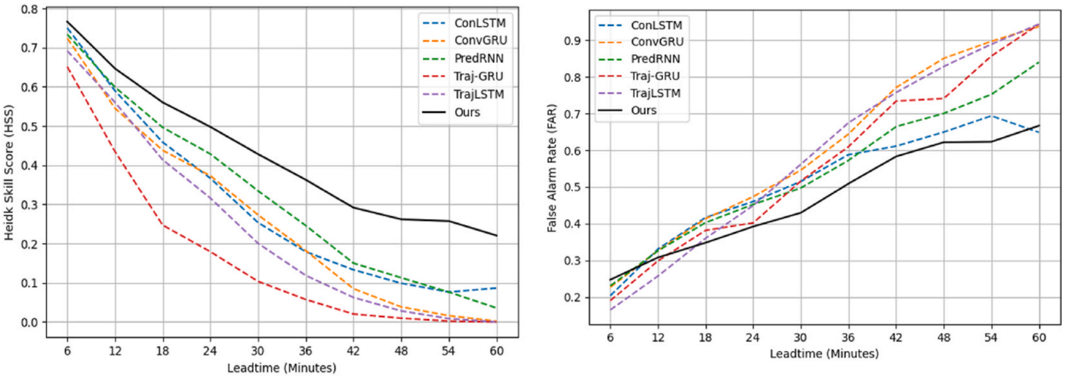

Fig. 9. The performance variations in terms of HSS and FAR scores are examined for different nowcast lead times when τ = 35. The plotted curve illustrates the sequential score values corresponding to each predicted radar echo map frame from 1 to 10.

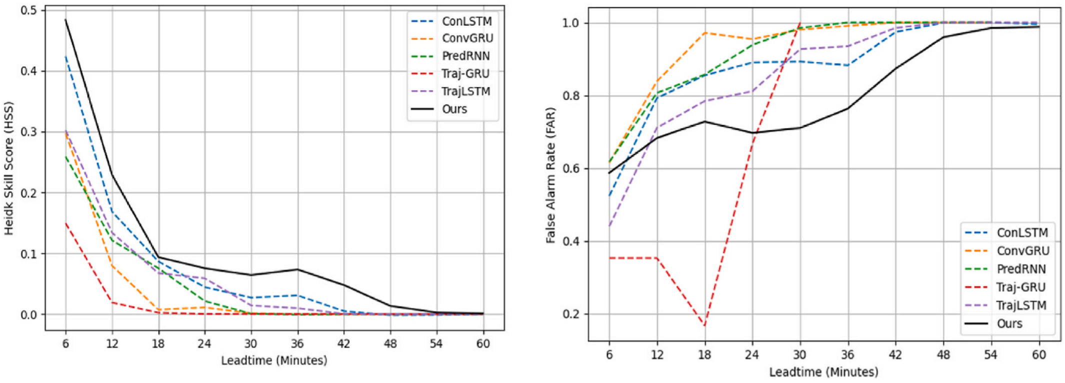

Fig. 10. The performance variations in terms of HSS and FAR scores are examined for different nowcast lead times when τ = 45. The plotted curve illustrates the sequential score values corresponding to each predicted radar echo map frame from 1 to 10.

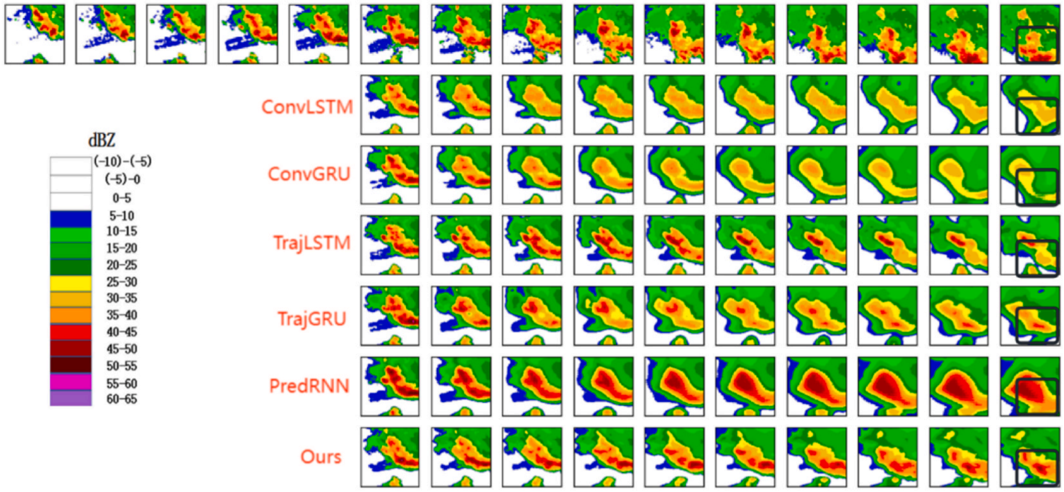

Fig. 11. The initial row comprises the first five radar echo images as inputs, while the subsequent ten images represent the actual sequence of radar echo images. Rows 2–6 showcase ten predicted radar echo maps generated by the model based on the aforementioned input set.

4.4. Ablation study

We conducted ablation experiments for validate the efficacy of our proposed three enhancement strategies. The gradual increase in the MES index observed in Fig. 12 following incremental ablations provides evidence of a progressive decline in model performance, thereby indirectly supporting the effectiveness of each module. Notably, removal of the GAN module exhibits a significant impact on SSIM, while its influence becomes more pronounced upon eliminating both CBAM and Resnet modules, this suggests that GAN contributes significantly to improving image clarity in predictions.

The results presented in Table 3 demonstrate a gradual decrease in the average HSS, CSI, and POD values of ten predicted radar echo maps as the ablation experiments progress under thresholds of 5, 35, and 45. Notably, while FAR did not exhibit a gradual increase at a threshold of 5, its value displayed a progressive increase when the threshold was raised to 35 and 45.

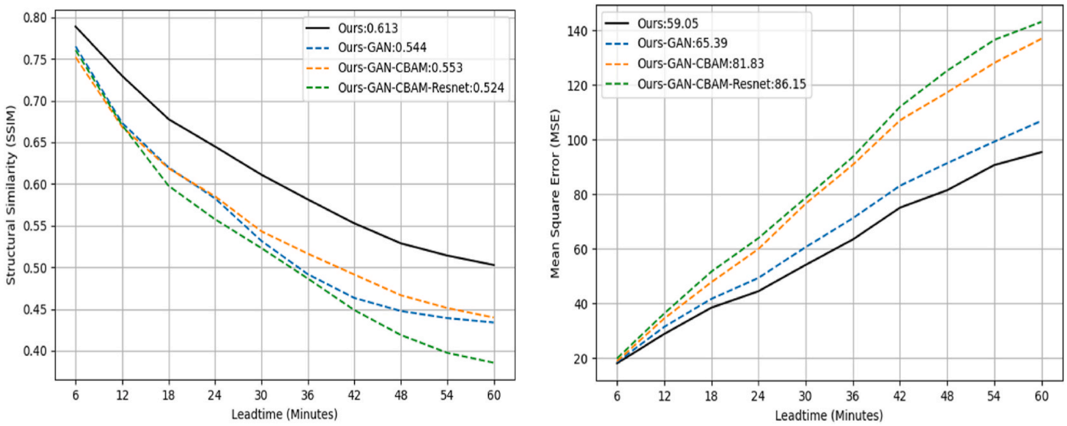

Fig. 12. The trend chart of Model SSIM and MES was generated by sequentially removing the corresponding functional modules.

Table 3
The mean scores of HSS, CSI, POD, and FAR throughout the ablation process.

Algorithms	HSS			CSI			POD			FAR		
	r ≥ 5	r ≥ 35	r ≥ 45	r ≥ 5	r ≥ 35	r ≥ 45	r ≥ 5	r ≥ 35	r ≥ 45	r ≥ 5	r ≥ 35	r ≥ 45
Ours	**0.7970**	**0.4293**	**0.1084**	**0.8329**	**0.3123**	**0.0660**	**0.8922**	**0.4090**	**0.1016**	0.0746	**0.4728**	**0.7970**
-GAN	0.7657	0.3213	0.0476	0.8049	0.2322	0.0275	0.8557	0.2938	0.0367	0.0700	0.5439	0.8041
-CBAM	0.7568	0.2895	0.0664	0.7968	0.2077	0.0382	0.8446	0.2564	0.0454	**0.0682**	0.5644	0.8372
-Resnet	0.7425	0.2848	0.0596	0.7889	0.2010	0.0345	0.8471	0.2385	0.0408	0.0822	0.5123	0.8641

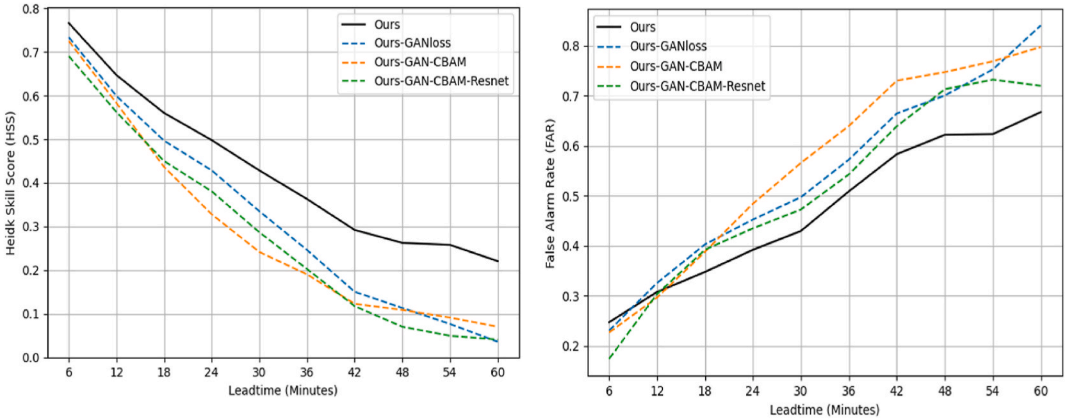

Fig. 13. The HSS and FAR trend chart was generated by iteratively removing the corresponding functional modules, with a fixed threshold value of $\tau = 35$.

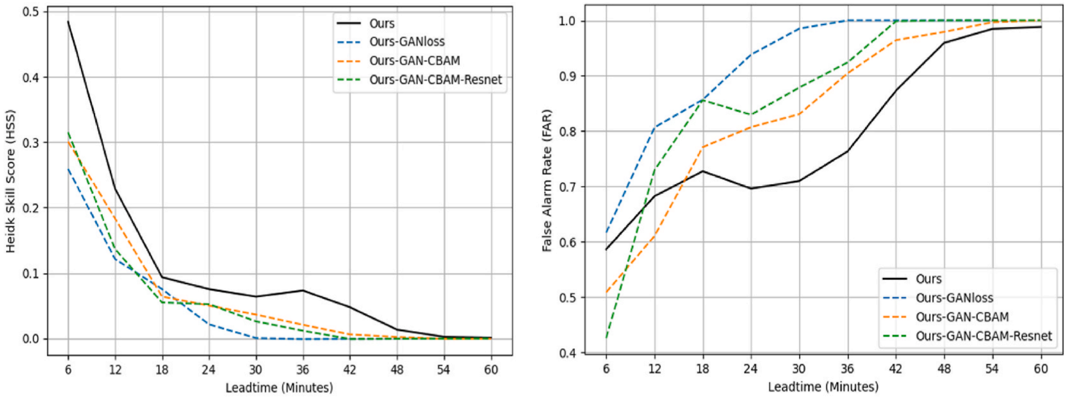

Fig. 14. The HSS and FAR trend chart was generated by iteratively removing the corresponding functional modules, with a fixed threshold value of $\tau = 45$.

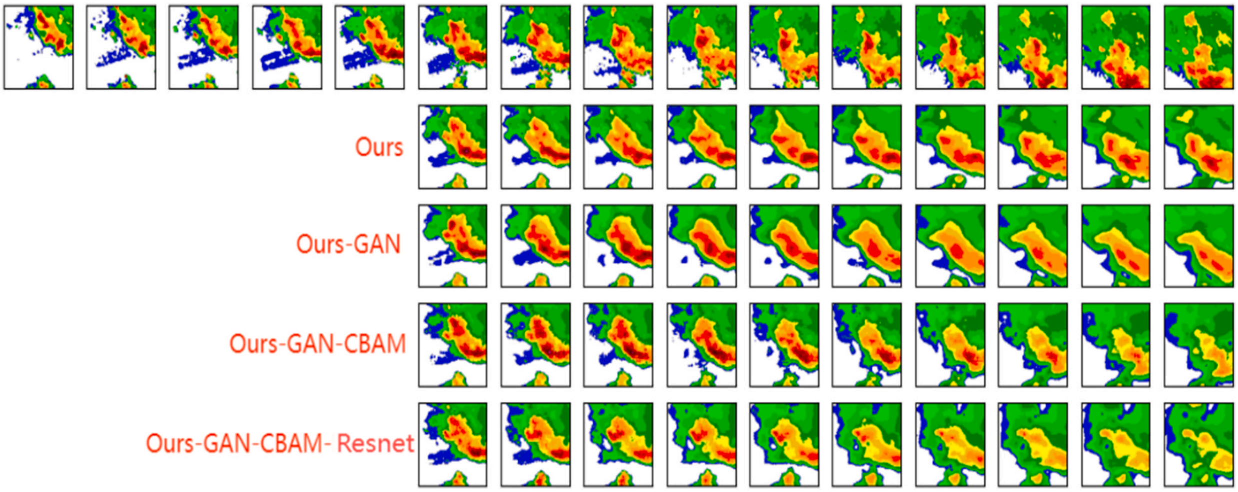

Fig. 15. The reconstructed radar echo maps were generated through an iterative process, wherein each functional module was systematically eliminated based on our proposed model.

The changing trends of HSS and FAR indicators during the model ablation experiment at thresholds of 35 and 45 are illustrated in Figs. 13 and 14, respectively. It can be observed from these graphs that the overall HSS gradually decreases while the FAR gradually increases. Despite a cross phenomenon between the HSS and FAR values corresponding to the radar echo maps at each moment during the ablation process, this does not significantly impact the overall trend of gradual degradation.

Regarding the radar echo visualization depicted in Fig. 15, excluding GAN leads to a reduction in image expenditure, while the absence of CBAM results in degradation of the final few high echo values within the image. Furthermore, eliminating the Resnet module intensifies distortion and fading phenomena of high echo values exhibited by the model. Consequently, there is a clear inclination towards increasing deviation from the true radar echo map in terms of overall performance. The ablation experiment verifies that GAN, CBAM, and Resnet have varying degrees of impact on the accuracy of predicted images.

5. Conclusion and future work

In this article, we propose the trajPredRNN + prediction model that addresses the limitations of trajGRU in maintaining fine-grained appearance memory units and overcomes the positional alignment issue of PredRNN. To enhance the predictive performance of our model, we incorporate an attention mechanism, Resnet structure, and improve the loss function. Comparative experiments have demonstrated the effectiveness of our proposed optimization measures compared to other classical models, while ablation experiments have validated their efficacy. TrajPredRNN + provides valuable insights for meteorological prediction research, particularly in leveraging deep learning for short-term rainfall forecasting.

In our future research, we aim to enhance the comprehensiveness of predictions by integrating diverse types of training data, such as satellite imagery, and incorporating additional parameters including temperature, humidity, wind direction, and cloud shape into the model training process. Moreover, optimizing the network topology structure is crucial, for instance, fine-tuning hyperparameters and architecture of the GAN network. Additionally, introducing the Transformer mechanism with multi-head self-attention enables a non-recurrent neural network structure that reduces computational complexity. Finally, it is imperative to improve the transferability of pre-trained models and enhance their practical applicability.

Data availability statement

All datasets available for download at https://tianchi.aliyun.com/dataset/1085/.

Funding statement

This work is a milestone achievement of the first author's pursuit of a doctoral degree in Computer Application Technology from the Macau Polytechnic University. This work was supported by Dongguan City Universtiy Youth Teacher Development Fund Project, NO:2023QJY001Z. China National College Student Innovation and Entrepreneurship Training plan Projects. NO:202313844005, 202413844003, S202413844017.

CRediT authorship contribution statement

Chongxing Ji: Writing – review & editing, Writing – original draft, Validation, Software, Data curation. **Yuan Xu:** Writing – review & editing, Funding acquisition, Supervision.

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.

References

[1] Z. Gao, X. Shi, B. Han, et al., Prediff: precipitation nowcasting with latent diffusion models[J], Adv. Neural Inf. Process. Syst. (2024) 36.
[2] M.J. Mayer, D. Yang, Calibration of deterministic NWP forecasts and its impact on verification[J], Int. J. Forecast. 39 (2) (2023) 981–991.
[3] X. Shi, Z. Huang, W. Bian, et al., Videoflow: exploiting temporal cues for multi-frame optical flow estimation[C], Proceedings of the IEEE/CVF International Conference on Computer Vision (2023) 12469–12480.
[4] M. Egelhaaf, Optic flow based spatial vision in insects[J], J. Comp. Physiol. 209 (4) (2023) 541–561.
[5] X. Shi, Z. Chen, H. Wang, et al., Convolutional LSTM network: a machine learning approach for precipitation nowcasting[J], Adv. Neural Inf. Process. Syst. (2015) 28.
[6] N. Ballas, L. Yao, C. Pal, et al., Delving deeper into convolutional networks for learning video representations, arXiv preprint arXiv:1511. (2015) 469–480, 06432.
[7] X. Shi, Z. Gao, L. Lausen, et al., Deep learning for precipitation nowcasting: a benchmark and a new model[J], Adv. Neural Inf. Process. Syst. (2017) 30.
[8] Y. Feng, L. Ma, W. Liu, et al., Spatio-temporal video re-localization by warp lstm[C], Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019) 1288–1297.
[9] Y. Wang, M. Long, J. Wang, et al., Predrnn: recurrent neural networks for predictive learning using spatiotemporal lstms[J], Adv. Neural Inf. Process. Syst. (2017) 30.
[10] Y. Wang, Z. Gao, M. Long, et al., Predrnn++: towards a Resolution of the Deep-In-Time Dilemma in Spatiotemporal Predictive learning[C]//International Conference on Machine Learning, PMLR, 2018, pp. 5123–5132.

[11] Y. Wang, L. Jiang, M.H. Yang, et al., Eidetic 3D LSTM: a model for video prediction and beyond[C]. International Conference on Learning Representations, 2018.

[12] Y. Wang, J. Zhang, H. Zhu, et al., Memory in memory: a predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics[C], Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2019) 9154–9162.

[13] D.N. Tuyen, T.M. Tuan, X.H. Le, et al., RainPredRNN: a new approach for precipitation nowcasting with weather radar echo images based on deep learning[J], Axioms 11 (3) (2022) 107.

[14] C. Luo, X. Li, Y. Ye, PFST-LSTM: a spatiotemporal LSTM model with pseudoflow prediction for precipitation nowcasting[J], IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14 (2020) 843–857.

[15] K. Trebing, T. Stańczyk, S. Mehrkanoon, SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture[J], Pattern Recogn. Lett. 145 (2021) 178–186.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial nets[J], Adv. Neural Inf. Process. Syst. (2014) 27.

[17] M. Ghiasi, Z. Wang, M. Mehrandezh, et al., Evolution of smart grids towards the Internet of energy: concept and essential components for deep decarbonisation [J], IET Smart Grid 6 (1) (2023) 86–102.

[18] W. Jiang, X. Wang, H. Huang, et al., Optimal economic scheduling of microgrids considering renewable energy sources based on energy hub model using demand response and improved water wave optimization algorithm[J], J. Energy Storage 55 (2022) 105311.

[19] S. Li, X. Fang, J. Liao, et al., Evaluating the efficiency of CCHP systems in Xinjiang Uygur Autonomous Region: an optimal strategy based on improved mother optimization algorithm[J], Case Stud. Therm. Eng. 54 (2024) 104005.

[20] C. Luo, X. Li, Y. Ye, et al., Experimental study on generative adversarial network for precipitation nowcasting[J], IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–20.

[21] R. Venkatesh, C. Balasubramanian, M. Kaliappan, Rainfall prediction using generative adversarial networks with convolution neural network[J], Soft Comput. 25 (2021) 4725–4738.

[22] H. Xie, L. Wu, W. Xie, et al., Improving ECMWF short-term intensive rainfall forecasts using generative adversarial nets and deep belief networks[J], Atmos. Res. 249 (2021) 105281.

[23] S. Choi, Y. Kim, Rad-cGAN v1. 0: radar-based precipitation nowcasting model with conditional generative adversarial networks for multiple dam domains[J], Geosci. Model Dev. (GMD) 15 (15) (2022) 5967–5985.

[24] S. Targ, D. Almeida, K. Lyman, Resnet in resnet: generalizing residual architectures, arXiv preprint arXiv:1603. (2016) 2169–2182, 08029.

[25] R. Wightman, H. Touvron, H. Jégou, Resnet strikes back: an improved training procedure in timm, arXiv preprint arXiv:2110. (2021) 759–773, 00476.

[26] M. Farooq, A. Hafeez, Covid-resnet: a deep learning framework for screening of covid19 from radiographs[J], arXiv preprint arXiv:2003 14395 (2020).

[27] D. Sarwinda, R.H. Paradisa, A. Bustamam, et al., Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer [J], Procedia Comput. Sci. 179 (2021) 423–431.

[28] S. Woo, J. Park, J.Y. Lee, et al., Cbam: convolutional block attention module[C], Proceedings of the European conference on computer vision (ECCV) (2018) 3–19.

[29] G. Magacho, E. Espagne, A. Godin, Impacts of the CBAM on EU trade partners: consequences for developing countries[J], Clim. Pol. 24 (2) (2024) 243–259.

[30] W. Wang, X. Tan, P. Zhang, et al., A CBAM based multiscale transformer fusion approach for remote sensing image change detection[J], IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 15 (2022) 6817–6825.

[31] Y. Luo, Z. Wang, An Improved Resnet Algorithm Based on cbam[C]//2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), IEEE, 2021, pp. 121–125.

[32] Y. Wang, H. Wu, J. Zhang, et al., Predrnn: a recurrent neural network for spatiotemporal predictive learning[J], IEEE Trans. Pattern Anal. Mach. Intell. 45 (2) (2022) 2208–2225.