Geosci. Model Dev., 15, 5407–5419, 2022
https://doi.org/10.5194/gmd-15-5407-2022

Geoscientific Model Development
Open Access
EGU

RAP-Net: Region Attention Predictive Network for precipitation nowcasting

Zheng Zhang[1,★], **Chuyao Luo**[1,★], **Shanshan Feng**[1], **Rui Ye**[1], **Yunming Ye**[1], **and Xutao Li**[1]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China
★These authors contributed equally to this work.

Correspondence: Xutao Li (lixutao@hit.edu.cn)

Received: 21 January 2022 – Discussion started: 11 February 2022
Revised: 4 May 2022 – Accepted: 7 May 2022 – Published: 15 July 2022

Abstract. Natural disasters caused by heavy rainfall often cause huge loss of life and property. Hence, the task of precipitation nowcasting is of great importance. To solve this problem, several deep learning methods have been proposed to forecast future radar echo images, and then the predicted maps are converted to the distribution of rainfall. The prevailing spatiotemporal sequence prediction methods apply a ConvRNN structure, which combines the convolution and recurrent neural network. Although ConvRNN methods achieve remarkable success, they do not capture both local and global spatial features simultaneously, which degrades the nowcasting in regions of heavy rainfall. To address this issue, we propose a Region Attention Block (RAB) and embed it into ConvRNN to enhance forecasting in the areas with heavy rainfall. Besides, the ConvRNN models find it hard to memorize longer historical representations with limited parameters. To this end, we propose a Recall Attention Mechanism (RAM) to improve the prediction. By preserving longer temporal information, RAM contributes to the forecasting, especially in the moderate rainfall intensity. The experiments show that the proposed model, Region Attention Predictive Network (RAP-Net), significantly outperforms state-of-the-art methods.

1 Introduction

Precipitation nowcasting has vital influence in the fields of transportation, agriculture, tourism, industry, and city alarming. Due to the higher spatial and temporal resolution of the radar echo image, it is effective for forecasting the distribution of rainfall by predicting the future radar echo maps and converting each pixel to the rainfall intensity according to the Z–R relationship (Shi et al., 2017). Therefore, precipitation nowcasting is often defined as a spatiotemporal prediction problem.

Traditional approaches to precipitation nowcasting are motion-field-based methods. The specific process can be briefly divided into three parts. First, the motion field is estimated by variational radar echo-tracking methods such as optical flow (Woo and Wong, 2017). Second, the future radar reflectivities are advected by a semi-Lagrangian advection scheme under the assumption of stationary movement. Third, the performance of the forecasts is evaluated by comparing them to ground truth. However, these methods do not exploit abundant historical observations.

To overcome this limitation, some deep learning-based methods have been proposed to handle precipitation nowcasting (Shi et al., 2017; Ayzel et al., 2020; Li et al., 2021). They usually build a mapping from previous observations to future maps by learning from the abundant historical radar data. Generally, the prevailing approaches utilize the structure of ConvRNN, which combines the convolution neural network (CNN) and recurrent neural network (RNN). Furthermore, to enhance the spatiotemporal representation ability, other types of neural networks such as the spatial transformer network (STN; Shi et al., 2017), deformable convolution network (DCN; Wu et al., 2021), and attention module (Lin et al., 2020) are introduced in the ConvRNN unit and obtain better performance.

However, existing ConvRNN models confront the following three drawbacks: (1) the convolution employed in the current input only extracts the local features instead of the large-scale representation due to a fixed kernel size. It may lead

to useful information beyond the visual field of convolution not being captured, and this thus degrades the performance. (2) The convolution applied in the previous hidden states only transmits local previous representations to the current states, which means historical spatial information cannot be fully used. (3) The update process of the temporal memory limits the long-term spatiotemporal representation preservation. Thus, the information including high echo reflectivity is easily dropped. Although some remedial solutions (Wang et al., 2018b; Luo et al., 2021) based on the attention mechanism are proposed, they are hard to apply in large-scale inputs and long-term predictions due to the limitation of space occupation.

To address the first two problems, we propose a Region Attention Block (RAB) and embed it into the input and hidden state, respectively. It simultaneously exploits the global spatial representation and preserves the local feature. RAB classifies each feature map into equal-sized tensors, and each tensor gatherers a similar semantic. Then, the attention module is used to interact with the contents of all semantics. To this end, the large-scale feature map can be captured from the global view and, meanwhile, maintain local representations. Therefore, the large-scale spatial feature of the current input and previous hidden states can be preserved. Moreover, to capture the long-term spatiotemporal dependency of representation without increasing parameters, we present the Recall Attention Mechanism (RAM) to retrieval all historical inputs. More rainfall information is captured by this component. By combining these modules, the performance for heavy and moderate rainfall can be significantly improved. In brief, the main contributions of the paper are summarized as follows:

1. We first propose a new attention method, namely Region Attention Block (RAB), to capture both global and local spatial features simultaneously to improve the spatial expressivity of feature maps.

2. We embed the RAB into current inputs and previous hidden states to obtain the large-scale spatial information from the global view and persevere different semantics at the same time. For the same echo with a large-scale size and long-range movement between the adjacent time, more useful spatial information can be extracted, which leads to more accurate predictions in those regions with heavy rainfall.

3. We propose the Recall Attention Mechanism (RAM) to retrieval all historical inputs with limited parameters. The representation of moderate and heavy rainfall intensity can be preserved in the predicted unit.

2 Related work

Traditional methods (Pulkkinen et al., 2019) mainly focus on estimating the motion field between the adjacent radar maps, and then the next prediction can be extrapolated based on this movement. Here, the motion field describes the direction and distance of each pixel that needs to be moved at the next moment. To obtain the movement, tracking radar echoes by correlation (TREC; Wang et al., 2013) divides the whole radar map into serval equal-sized boxes and calculates the motion vector of each pair's box center by searching the highest correlation between boxes at the adjacent time steps. Another type of approach is the optical flow-based method (Woo and Wong, 2017). It calculates the motion field under a pixel level, based on the assumption that the brightness of pixels remains unchanged. Upon the idea, many algorithms (Ryu et al., 2020) are developed to apply the radar maps with the large movement vector. However, the invariant brightness assumption conflicts with the realistic movements of hydrometeors, and massive historical data are utilized.

To overcome it, many deep learning-based methods (Wang et al., 2017, 2019; Trebing et al., 2021) are proposed to predict the radar sequence without the above unreasonable assumption. Most of the methods commonly exploit the structure of ConvRNN. It combines convolution neural network (CNN) and recurrent neural network (RNN) to preserve the spatiotemporal feature of the historical sequence. Furthermore, Wang et al. added a spatial memory in a predicted unit (Wang et al., 2017, 2018a, b) and an attention mechanism in temporal memory (Wang et al., 2018b) to enhance the spatiotemporal representation ability of the short term and long term, respectively. Although these methods have a remarkable performance, the visualization of their predictions is usually blurry due to the loss function and the architecture of the model (Shouno, 2020). To handle the issue, a generative adversarial network (GAN; Tian et al., 2019; Xie et al., 2020; Zheng et al., 2022) has been introduced in the ConvRNN model to improve predictive clarity. Nevertheless, the non-convergence and collapse problem would cause a negative influence on the prediction. Our proposed method is different from existing deep learning methods in two aspects. In the short term, the proposed RAB can simultaneously exploit local and global spatiotemporal representations. In the long term, the RAM can effectively recall all historical observations with limited space occupancy.

3 Proposed method

3.1 Problem definition

The precipitation nowcasting task can be defined as the spatiotemporal sequence prediction problem (Shi et al., 2017). Based on historical observations $X_{0:t}$, it aims to forecast the future radar echo images $\overline{X}_{t+1:T}$ that have maximum probability with ground truth $X_{t+1:T}$, as follows:

$$\overline{X}_{t+1:T} = \arg\max P(X_{t+1:T} \,|\, X_0, X_1, \ldots, X_t). \quad (1)$$

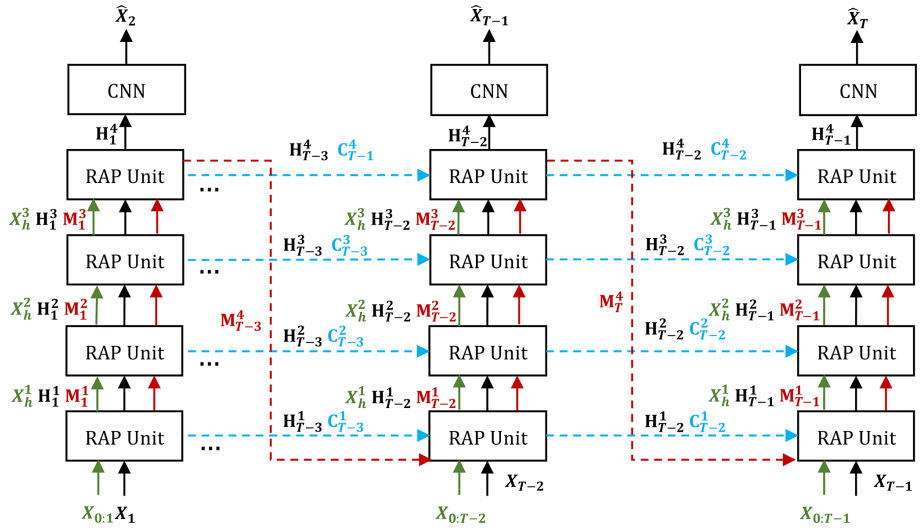


Figure 1. The overall architecture of the Region Attention Predictive Network (RAP-Net).

In this paper, t and T are set to 5 and 15, respectively, which means that 10 continuous radar maps need to be predicted according to five historical images.

3.2 Overall architecture

The overall architecture of the proposed model RAP-Net is presented in Fig. 1. It utilizes the structure of PredRNN (Wang et al., 2017) and stacks several RAP-Units to generate the predictions from timestamp 2 to T. At any timestamp t, the model predicts a radar map $\hat{X}_{t+1}$ at the next timestamp $t+1$, according to the current radar map X_t and historical radar sequence $X_{0:t}$. The red and blue arrows denote the delivery direction of the spatial memory M and temporal memory C, respectively. These two memories preserve spatial and temporal representations, respectively. Different from the PredRNN, RAP-Net exploits dissimilar data flow to transmit long-term spatiotemporal information X_h^l which preserves all historical representations. Besides, we notice that the majority of ConvRNN models (Wang et al., 2017, 2018a, b, 2019; Lin et al., 2020) employ similar architecture. Hence, the difference lies in their units instead of the employed architecture. In the experiment section, we will discuss and analyze the performance of different units utilized by the existed methods.

The internal structure of Region Attention Predictive Unit (RAP-Unit) is shown in Fig. 2. The inputs include the current input X_t^l, previous hidden state H_{t-1}^l, temporal memory C_{t-1}^l, spatial memory M_t^{l-1}, and long-term historical representation X_h^{l-1}. According to Fig. 1, RAP-Net consists of four stacked RAP-Units. At the bottom layer, X_h^{l-1} represents all historical inputs $X_{0:t}$, while at other layers, X_h^{l-1} is the output of the previous layer. The outputs of RAP-Unit are the current hidden state H_t^l, spatial memory M_t^l, temporal memory C_t^l, and new long-term representation X_h^l. The details of the calculation are presented according to the following formulas:

$$
\begin{aligned}
&X'^{l}_{t} = \mathrm{RAB}(X_t^l),\\
&H'^{l}_{t-1} = \mathrm{RAB}(H_{t-1}^l),\\
&i_t = \sigma(W_{xi} * X'^{l}_{t} + W_{hi} * H'^{l}_{t-1} + b_i),\\
&g_t = \tanh(W_{xg} * X'^{l}_{t} + W_{hg} * H'^{l}_{t-1} + b_g),\\
&f_t = \sigma(W_{xf} * X'^{l}_{t} + W_{hf} * H'^{l}_{t-1} + b_f),\\
&i'_t = \sigma(W'_{xi} * X'^{l}_{t} + W_{mi} * M_t^{l-1} + b'_i),\\
&g'_t = \tanh(W'_{xg} * X'^{l}_{t} + W_{mg} * M_t^{l-1} + b'_g),\\
&f'_t = \sigma(W'_{xf} * X'^{l}_{t} + W_{mf} * M_t^{l-1} + b'_f),\\
&C_t^l = i_t \circ g_t + f_t \circ C_{t-1}^l,\\
&M_t^l = i'_t \circ g'_t + f'_t \circ M_t^{l-1},\\
&o_t = \sigma(W_{xo} * X'^{l}_{t} + W_{ho} * H'^{l}_{t-1} + W_{co} * C_t^l + \\
&\quad W_{mo} * M_t^l + b_o),\\
&H_t^l = o_t \circ \tanh(W_{1\times1} * [X'^{l}_{t}, M_t^k]),\\
&H_t^l, X_h^l = \mathrm{RAM}(H_t^l, X_h^{l-1} * W_l),
\end{aligned}
\tag{2}
$$

where the $*$ and $\circ$ symbols denote the convolution and Hadamard product, respectively. i_t, g_t, f_t, i'_t, g'_t, and f'_t indicate various gates, which can be viewed as intermediate variables. Here, RAB and RAM are the Region Attention Block and Recall Attention Mechanism, respectively.

3.3 Region Attention Block

To address the issue, we expect that these patches can be divided adaptively, and those elements with similar semantic relationships are classified into the same patch, as

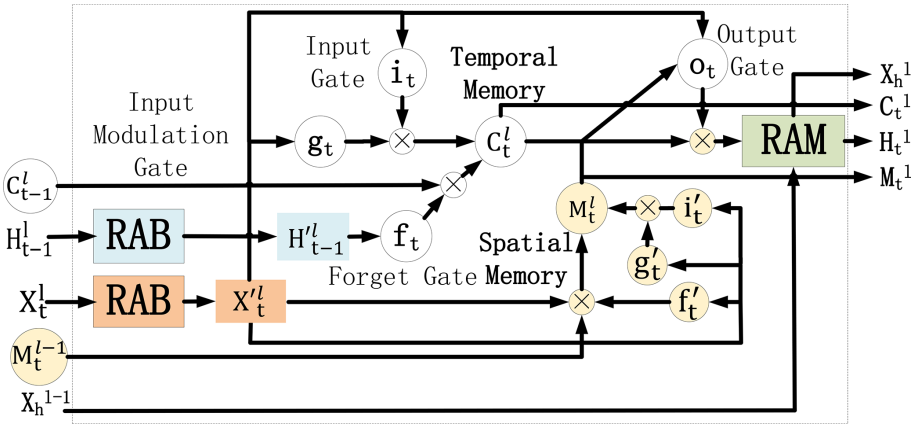
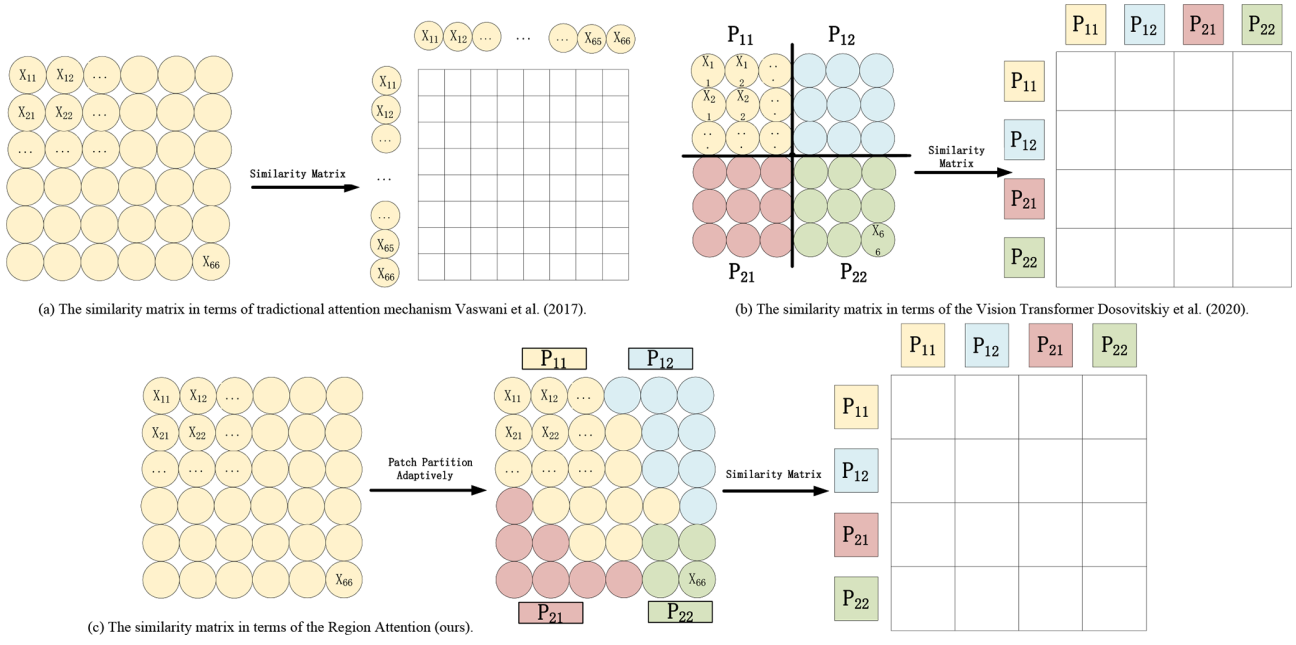


Figure 2. The internal structure of the Region Attention Predictive Unit (RAP-Unit).



(a) The similarity matrix in terms of tradictional attention mechanism Vaswani et al. (2017).

(b) The similarity matrix in terms of the Vision Transformer Dosovitskiy et al. (2020).

(c) The similarity matrix in terms of the Region Attention (ours).

Figure 3. The calculation process of a similarity matrix, based on three different attention methods.

shown in Fig. 3c. To realize this idea, we propose the Region Attention Block (RAB), whose structure is illustrated in Fig. 4. First, a convolution and softmax layer are employed in the input feature map $F^i \in R^{B\times C\times H\times W}$ to generate $F^c \in R^{B\times N\times H\times W}$ to distinguish the N classifications. Second, the original input F^i is split into N groups of feature maps $F^n \in R^{N\times B\times C\times H\times W}$ by the split module, following this formula:

$$\text{Split}(F^i, F^c) = \text{Concatenate}(\{F^i_{j,k,h,w} \cdot F^c_{j,n,h,w} | 1 < n < N, n \in Z\}, \text{axis} = 0). \quad (3)$$

These groups denote various semantic information extracted from different positions. Third, $F^{qk} \in R^{B\times N\times c\times h\times w}$ is convolved by F^n to further exploit the feature of F^n and reduce the parameters, where the c, h, and w are smaller than C, H, and W. Besides, $F^v \in R^{N\times B\times C\times H\times W}$ are outputted by a convolution layer applied in F^n to preserve the original information. Fourth, three different convolutions are used to generate query Q_s, key K_s, and value V_s, based on F^{qk} and F^v. After flattening, $Q_s \in R^{B\times N\times c*h*w}$, $K_s \in R^{B\times N\times c*h*w}$ and $V_s \in R^{B\times N\times C*H*W}$ are fed into the spatial attention function to obtain $F^a \in R^{B\times N\times C\times H\times W}$ that have been interacting with the local representation from different regions. The output after the attention function (Fu et al., 2019) is as follows:

$$\text{Attention}(Q_s, K_s, V_s) = \text{softmax}\left(\frac{f(Q_s, K_s^T)}{\sqrt{d_k}}\right) V_s, \quad (4)$$

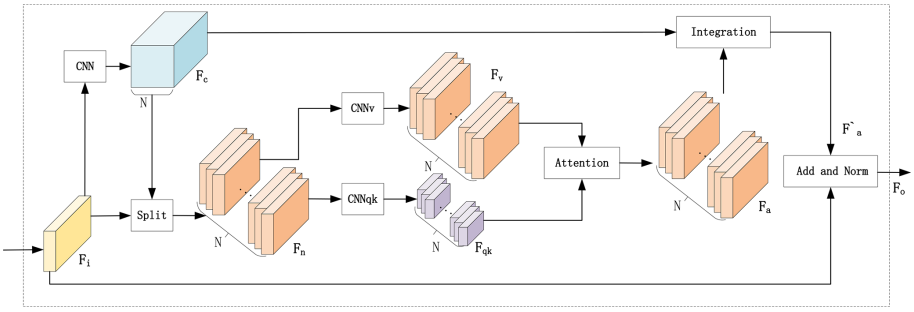


Figure 4. The structure of Region Attention Block (RAB).

where f denotes the dot product, and d_k is the dimension of key K_s. $f(Q_s, K_s^T) \in R^{B\times N\times N}$ is the similarity matrix of various semantics in different regions. Fifth, an integration module is utilized to integrate F^a based on the F^c and obtains the result $F^{a'}$ from this equation, as follows:

$$\begin{aligned} F^{a'} &= \text{Integration}(F^a, F^c) \\ &= \sum_{n=1}^{N} F^a_{j,k,h,w} \cdot F^c_{j,n,k,h,w}, \end{aligned} \tag{5}$$

where, $F^{a'}$ has the same size as input feature map F^i. Finally, the structure of ResNet (He et al., 2016) is introduced to deeply exploit the spatial feature and achieve the final result $F^o \in R^{B\times N\times C\times H\times W}$. In summary, the calculation process is described by the following formulas:

$$\begin{aligned} F^c &= \text{softmax}(F^i * W_c), \\ F^n &= \text{Split}(F^i, F^c), \\ F^{qk} &= F^n * W'_{qk}, \\ F^v &= F^n * W'_v, \\ F^a &= \text{Attention}(F^{qk} * W_q, F^{qk} * W_k, F^v * W_v), \\ F^{a'} &= \text{Integration}(F^a, F^c), \\ F^o &= F^i + F^{a'}. \end{aligned} \tag{6}$$

The traditional attention mechanism calculates, in Fig. 3a, the similarity between different pixels, and the attention manner of the vision transformer, in Fig. 3b, compares different regions in a fixed location. Different from both mechanisms, the attention similarity from region attention (ours), in Fig. 3c, compares the difference between regions with a flexible size and position. Due to the irregular shape of radar echo and the different distributions, RAB can capture the correlation between the different radar echoes better. Therefore, the introduction of this block can improve the spatiotemporal ability of the model, especially since the information of radar echoes with high reflectivity is more easily extracted because they have a more stable appearance and shape. RAM has more of a contribution to improving the performance in these regions with heavy rainfall.

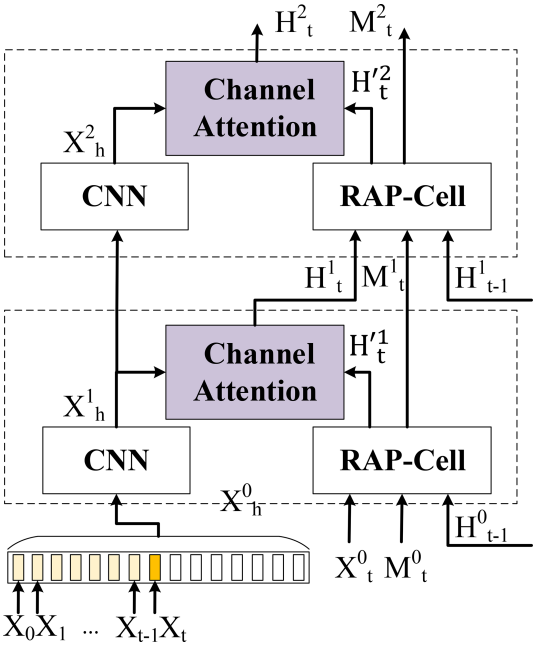


Figure 5. The manner of embedding the Recall Attention Mechanism (RAM) into the proposed predicted unit. Here, the RAP-Cell is the RAP unit without RAM.

3.4 Recall Attention Mechanism

To capture the temporal long dependencies of representation, Wang et al. (2018b) embedded the spatial attention module in the updating of temporal memory. However, it has the following two limitations: (1) it saves abundant history temporal memories, which leads to the number of parameters easily exceeding the space occupancy as the lead time goes. (2) The temporal memory has lost some information during the generation of various gates. Therefore, the preserved previous representation fails to capture all the information, and long-term spatiotemporal expressivity is limited.

To address these issues, we propose the Recall Attention Mechanism (RAM) to enhance the long-term spatiotemporal representation ability with a fixed space occupation, as Fig. 5 shows. First, we build an empty long-memory fea-

Table 1. The parameter settings of the RAP-Unit. The term "In res" and "Out res" denote the resolutions of input and output, respectively, while "Conv" is the convolution operation.

Attention type	Name	Kernel	Stride	Pad	Ch I/O	In res	Out res	Type
Region Attention Block	CNN_c	5×5	1×1	2×2	64/64	32×32	32×32	Conv
	CNN_{qk}	4×4	4×4	0×0	64/8	8×8	8×8	Conv
	CNN_v	5×5	1×1	2×2	64/64	32×32	32×32	Conv
	Lin_q	–	–	–	–	512	512	Linear
	Lin_k	–	–	–	–	512	512	Linear
Recall Attention Mechanism	CNN	5×5	1×1	2×2	14/64	32×32	32×32	Conv
RNN unit	CNN_x	5×5	1×1	2×2	64/448	32×32	32×32	Conv
	CNN_h	5×5	1×1	2×2	64/256	32×32	32×32	Conv
	CNN_m	5×5	1×1	2×2	64/192	32×32	32×32	Conv
	CNN_o	5×5	1×1	2×2	64/128	32×32	32×32	Conv
	CNN_{last}	1×1	1×1	0×0	128/64	32×32	32×32	Conv

ture map $X_h^0 \in R^{B \times T \times C \times H \times W}$ in the bottom layer and feed the current input X_t into it continually. Note that the X_h^0 contains all original previous inputs $X_{0:t} \in R^{B \times T \times C \times H \times W}$. Second, a convolution neural network is employed to extract the feature of X_h^0 and output the long-memory hidden state $X_h^1 \in R^{B \times T \times C \times H \times W}$. Last, X_h^1 and the output H'^1_t of the RAP-Cell (which can be regarded as the RAP-Net model without RAM) feed into the channel attention module to generate new hidden states, where the X_h^1 can be regarded as the key $K_c \in R^{B \times T * C \times H * W}$ and value $V_c \in R^{B \times T * C \times H * W}$, and the H'_t represents the query $Q_c \in R^{B \times C \times H * W}$. The formula of channel attention is shown as follows:

$$\text{Attention}(Q_c, K_c, V_c) = \text{softmax}(\frac{f(Q_c, K_c^T)}{\sqrt{d_k}})V_c, \tag{7}$$

where the f denotes the dot product, and d_k is the dimension of key K_c. $f(Q_c, K_c^T) \in R^{B \times T \times T * C}$ is the similarity matrix between channels of Q_s and channels of K_c. From Eq. (7), we can see that the V_c can be extracted according to the $f(Q_c, K_c^T)$, where Q_c decides how to explore the V_c by dot-producing with K_c. Therefore, the original output H'^1_t of the RAP-Cell can be regarded as query Q_c to explore long-term spatiotemporal representations X_h^1 that are the key K_c and value V_c. In this way, the new output H_t^1 has recalled all original historical representations, and long-term dependencies can be effectively preserved. Besides, the size of the long-memory feature map X_h^l is fixed at any time step because the size of X_h is predefined, and the corresponding content at different timestamps are fed into X_h. Similarly, in the lth layer, the input of the long-memory hidden state is the X_h^1. In the bottom layer, X_h^l is the result after convolution by historical input sequences $X_{0:t}$. In the other layers, the X_h^l is the result after convolution by the X_h^{l-1}. By RAM, the long-term historical representation can be delivered to the next layer.

4 Experiments

4.1 Dataset

The dataset is collected from the CIKM AnalytiCup2017 competition (available at https://tianchi.aliyun.com/competition/entrance/231596/information, Alibaba Cloud, 2022), which covers the whole area of Shenzhen, China. For convenience, we name this public dataset RadarCIKM. RadarCIKM has a training set and test set with 10 000 and 4000 sequences, respectively. There are 2000 sequences randomly sampled from the training set to build the validation set. Each sequence contains 15 continual observations within 90 min, where the spatial and temporal resolution of each map is 101×101 and 6 min, respectively. The range of each pixel is from 0 to 255, and each pixel denotes 1 km $\times$ 1 km. Moreover, the type of pixel is an integer, and each value can be converted to radar reflectivity (dBZ) by the following equation:

$$\text{dBZ} = p \times \frac{95}{255} - 10. \tag{8}$$

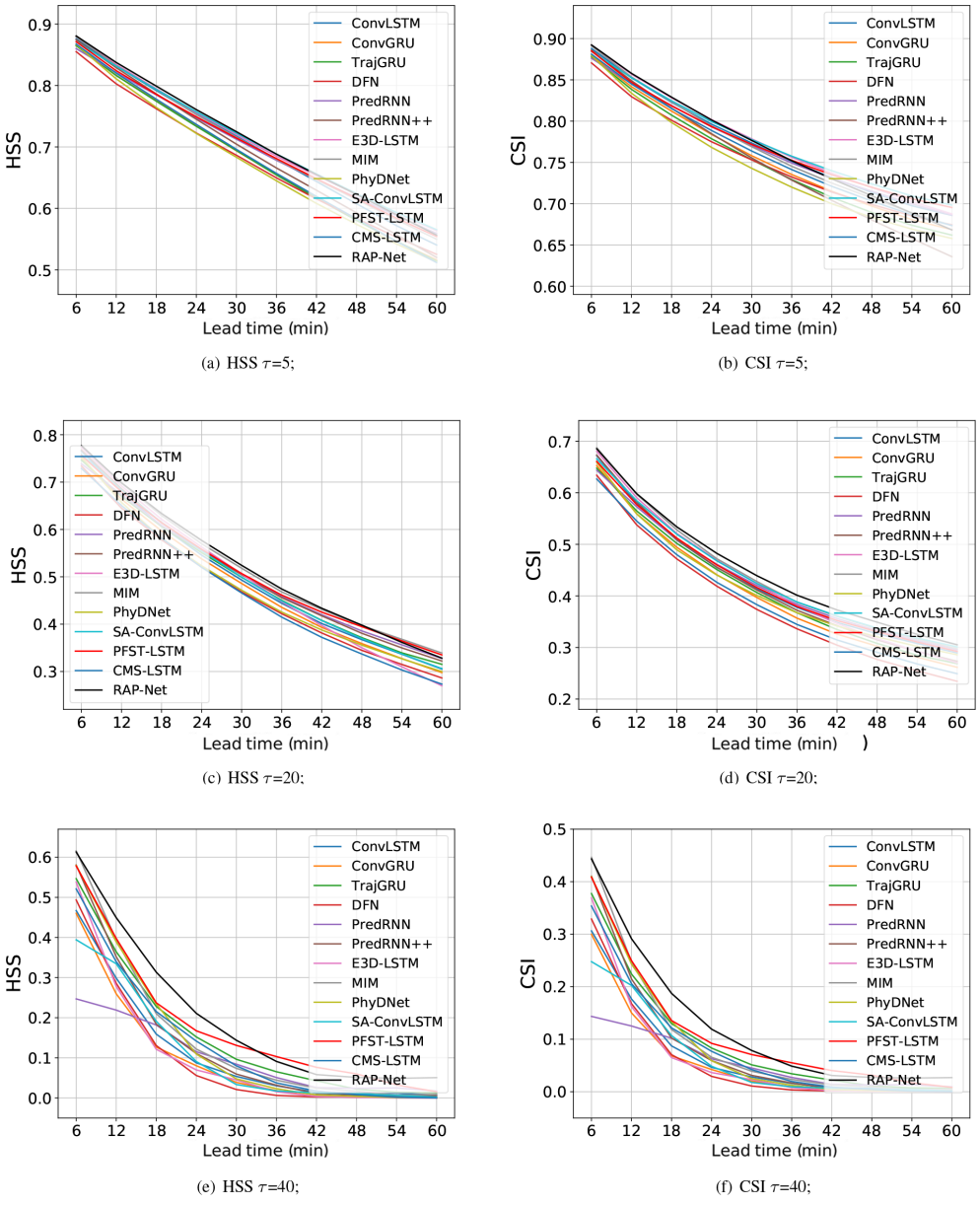
Then, the rainfall intensity can be obtained by the radar reflectivity (dBZ) and Z–R relationship as follows:

$$\text{dBZ} = 10 \log a + 10b \log R, \tag{9}$$

where the R is the rain rate level, $a = 58.53$, and $b = 1.56$.

4.2 Evaluation metrics

In this paper, in addition to common measurements such as structural similarity (SSIM) and mean absolute error (MAE) in video prediction, we also utilize the Heidke skill score (HSS) and critical success index (CSI) that are commonly used in precipitation nowcasting tasks. The HSS evaluates the fraction of correct forecasts after eliminating random predictions. The CSI measures the number of correct forecasts



(a) HSS τ=5; (b) CSI τ=5;

(c) HSS τ=20; (d) CSI τ=20;

(e) HSS τ=40; (f) CSI τ=40;

Figure 6. The HSS and CSI scores of different case lead time values (best viewed in color).

divided by the total number of occasions when the rainfall events were forecasted or observed. Specifically, the prediction and ground truth are converted to a binary matric based on a threshold τ. When the value of the dBZ is larger than τ, then it is set to 1 or otherwise to 0. Next, the number of the true positive (TP; prediction = 1; truth = 1), false negative (FN; prediction = 0; truth = 1), false positive (FP; prediction = 1; truth = 0), and true negative (TN; prediction = 0; truth = 0) are counted. Finally, the HSS and CSI can be calculated by the following formulas:

$$\mathrm{HSS} = \frac{2(\mathrm{TP}\times\mathrm{TN}-\mathrm{FN}\times\mathrm{FP})}{(\mathrm{TP}+\mathrm{FN})(\mathrm{FN}+\mathrm{TN})+(\mathrm{TP}+\mathrm{FP})(\mathrm{FP}+\mathrm{TN})}, \quad (10)$$

$$\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}. \quad (11)$$

Here, the range of HSS, CSI, and SSIM is [0,1]. The range of MAE is $[0,+\infty]$.

4.3 Parameters setting

The proposed RAP-Net takes five previous radar echo maps as inputs and outputs 10 predictions. It utilizes four layers of

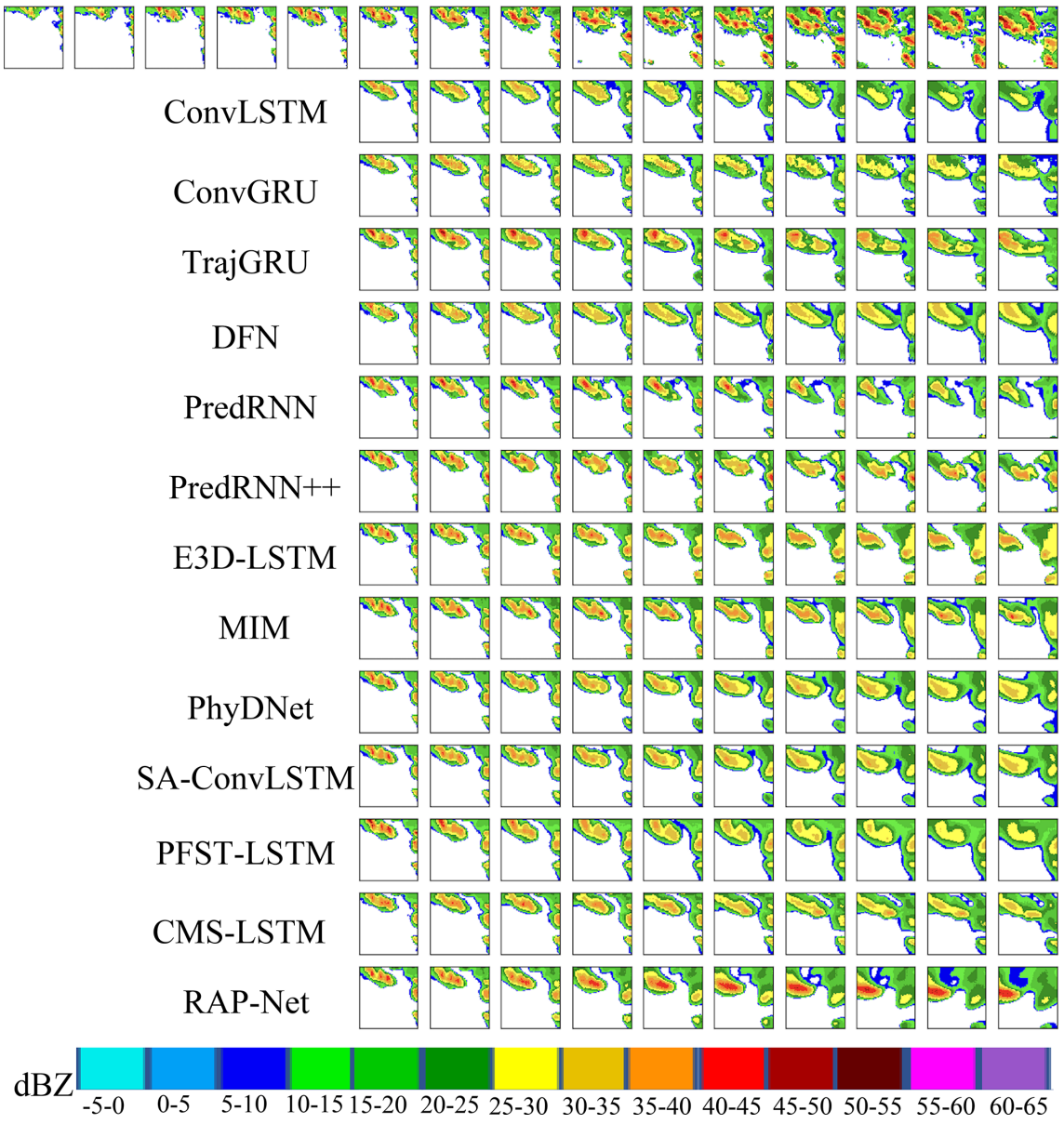
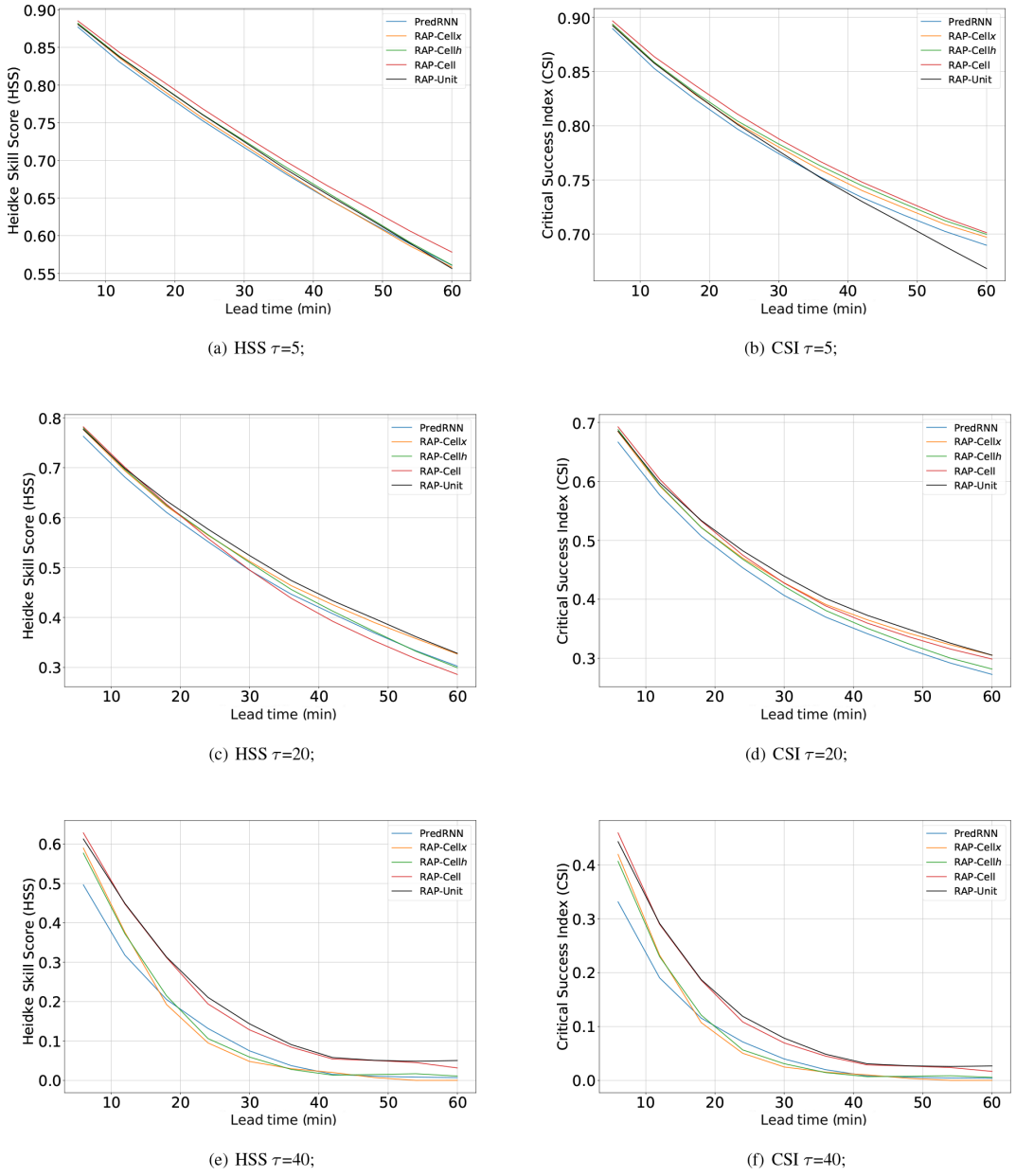


Figure 7. The first row is the reflectivity of the ground truth, and the remaining rows show the predicted reflectivity of various methods on an example from the RadarCIKM dataset (best viewed in color).

RAP-Units, as shown in Fig. 1, and the parameters setting of each RAP-Unit are shown in Table 1. The Adam optimizer is applied to train our model with a learning rate of 0.0004. Besides, the early stopping and scheduled sampling strategies are also used to optimize our model. The loss function combines the L1 and L2 to train RAP-Net. All experiments are implemented in Pytorch and conducted on NVIDIA 3090 graphics processing units (GPUs).

4.4 Result and analysis

Table 2 shows the results of all models. The best results are in boldface and the second best scores are underlined. We find that the RAP-Net achieves the smallest error and the highest structural similarity according to the MAE and SSIM. It is observed that our model outperforms other models in terms of the comprehensive performance. Besides, the proposed model has significant superiority especially for the nowcasting in heavy rainfall regions. Because the HSS and CSI keep the top position in the middle and high thresholds (20 and 40 dBZ). For the state-of-art method, PFST-LSTM (Luo et al., 2020), all measurements of it are exceeded by RAP-Net, which shows the performance of our model. Comparing with PredRNN, PredRNN++, and RAP-Net, we can see that they have a similar SSIM due to applying the same



(a) HSS τ=5; (b) CSI τ=5;

(c) HSS τ=20; (d) CSI τ=20;

(e) HSS τ=40; (f) CSI τ=40;

Figure 8. The performance changes against different nowcast lead times in terms of HSS and CSI scores in the ablation study (best viewed in color).

architecture. However, the other evaluation scores of RAP-Net are significantly higher than PredRNN and PredRNN++, which implies the benefit of RAP-Unit. Last, we notice that the SA-ConvLSTM (Lin et al., 2020) has the best HSS and CSI in the lowest threshold (5 dBZ). Nevertheless, its performance is poor in the highest threshold (40 dBZ), which implies that the RAB and RAM can improve the prediction in the area with a high radar echo compared to the traditional attention mechanism because the main difference between the RAP-Net and SA-ConvLSTM is that they introduce different attention submodules.

To show the performances of various models at different nowcasting lead times, Fig. 6 presents the HSS and CSI curves with regard to different lead times under all thresholds. We observe that both HSS and CSI scores of all models decrease as the lead time increases, which shows the difficulty of long-term predictions. Among these models, RAP-Net achieves notable superiority in the middle and late stages of the nowcasting period at the highest threshold. Especially in the last prediction, all baseline methods trend to obtain the same poor result. The RAP-Net remarkably outperforms other models. It implies that the introduction of

Table 2. Comparison results on RadarCIKM in terms of HSS, CSI, SSIM, and MAE. The best results are given in bold, and the second-best scores are underlined.

Methods	HSS ↑				CSI ↑				MAE ↓	SSIM ↑
	5 dBZ	20 dBZ	40 dBZ	avg.	5 dBZ	20 dBZ	40 dBZ	avg.		
ConvLSTM (Xingjian et al., 2015)	0.7031	0.4857	0.1470	0.4453	0.7663	0.4092	0.0801	0.4186	5.97	0.6334
ConvGRU (Shi et al., 2017)	0.6816	0.4827	0.1225	0.4289	0.7522	0.3952	0.0657	0.4043	6.00	0.6338
TrajGRU (Shi et al., 2017)	0.6809	0.4945	0.1907	0.4553	0.7466	0.4028	0.1061	0.4185	5.90	0.6424
DFN (Jia et al., 2016)	0.6772	0.4719	0.1306	0.4266	0.7489	0.3771	0.0704	0.3988	6.03	0.6268
PredRNN (Wang et al., 2017)	0.7082	0.4915	0.1639	0.4606	0.7692	0.4051	0.0901	0.4215	5.42	0.6887
PredRNN++ (Wang et al., 2018a)	0.7061	0.5047	0.1710	0.4548	0.7642	0.4176	0.0940	0.4253	5.44	0.6851
E3D-LSTM (Wang et al., 2018b)	0.7111	0.4810	0.1361	0.4427	0.7720	0.4060	0.0734	0.4171	5.51	0.6958
MIM (Wang et al., 2019)	0.7052	0.5166	0.1858	0.4692	0.7628	0.4279	0.1034	0.4313	5.47	0.6796
PhyDNet (Guen and Thome, 2020)	0.6741	0.4709	0.1832	0.4427	0.7402	0.4003	0.1017	0.4141	6.25	0.6443
SA-ConvLSTM (Lin et al., 2020)	**0.7118**	0.4861	0.1582	0.4520	**0.7725**	0.4161	0.0870	0.4252	5.71	0.6709
PFST-LSTM (Luo et al., 2020)	0.7045	0.5071	0.2218	0.4778	0.7680	0.4175	0.1257	0.4371	5.82	0.6367
CMS-LSTM (Chai et al., 2022)	0.6835	0.4605	0.1720	0.4387	0.7567	0.3788	0.0948	0.4101	5.95	0.6496
RAP-Net	0.7117	**0.5116**	**0.2293**	**0.4842**	0.7666	**0.4305**	**0.1307**	**0.4426**	**5.37**	**0.7019**

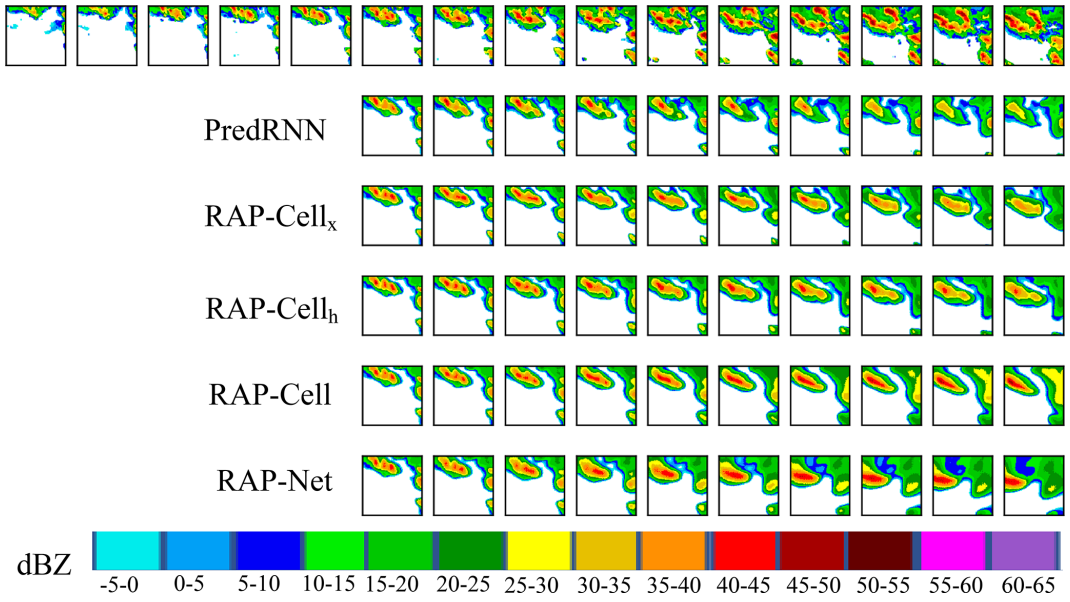


Figure 9. The first row is the reflectivity of the ground truth and the remaining rows are the predicted reflectivity of different methods on an example from the RadarCIKM dataset (best viewed in color).

RAB and RAM in the proposed model contributes to generating long-term predictions within heavy rainfall regions. Although the performance of RAP-Net would be degraded when the threshold becomes small, it still has competitiveness compared to other models.

Figure 7 shows an example of predictions from these models. The various colors denote the different ranges of reflectivity according to the color bar in the bottom of Fig. 7. From the ground truth in the first row, the rainfall event is obviously the trend of increasing the rainfall intensity. However, only our model can forecast this trend and keep the intensity of the regions. The RAP-Net can generate a high reflectivity area, which can also explain why our model can achieve the highest evaluation index HSS and CSI in the middle and high thresholds.

4.5 Ablation study

To investigate the influence of various modules, we conduct an ablation study to discuss the effectiveness of Region Attention Block to the current input and the last hidden state. The result of evaluations is shown in Table 3. RAP-Cell$_x$ and RAP-Cell$_h$ denote the PredRNN model embedding the RAB into the input and hidden state, respectively. The RAP-Cell model is the combination of RAP-Cell$_x$ and RAP-Cell$_h$ and can also be regarded as RAP-Net without RAM. The results of RAP-Cell$_x$ and RAP-Cell$_h$ are higher than PredRNN,

Table 3. Ablation results on RadarCIKM in terms of HSS, CSI, MAE, and SSIM. The best results are given in bold, and the second-best scores are underlined.

Methods	HSS ↑				CSI ↑				MAE ↓	SSIM ↑
	5 dBZ	20 dBZ	40 dBZ	avg.	5 dBZ	20 dBZ	40 dBZ	avg.		
PredRNN	0.7082	0.4915	0.1639	0.4545	0.7692	0.4051	0.0901	0.4215	5.42	0.6887
RAP-Cell$_x$	0.7102	0.5042	0.1754	0.4633	0.7747	0.4235	0.0967	0.4316	5.36	0.6965
RAP-Cell$_h$	0.7149	0.4967	0.1753	0.4623	0.7772	0.4138	0.0967	0.4292	**5.32**	0.7009
RAP-Cell	**0.7234**	0.4757	0.2283	0.4758	**0.7817**	0.4143	0.1300	0.4420	5.64	**0.7036**
RAP-Net	0.7117	**0.5116**	**0.2293**	**0.4842**	0.7666	**0.4305**	**0.1307**	**0.4429**	5.37	0.7019

which shows the advantage of introducing the Region Attention Block, especially since RAP-Cell$_h$ significantly reduces the error according to MAE. Besides, the HSS, CSI, and SSIM of the RAP-Cell have significant improvements, particularly when the threshold τ is 40 dBZ, which implies that RAB being simultaneously employed in the input and hidden state contributes to the prediction in the heavy rainfall regions. Moreover, by comparing the RAP-Cell and RAP-Net, we find that the RAM can enhance the accuracy of the nowcasting, especially in the areas with moderate intensity rainfall.

Similarly, we also plot Fig. 8 to show the experimental results of all models against different nowcast lead times. We can see that RAP-Net delivers more promising results when the threshold increases, which demonstrates the effectiveness of combining RAB and RAM in terms of long-term prediction in a high reflectivity area. The performance of RAM can be shown by comparing RAP-Cell and RAP-Net. We notice that the introduction of RAM can improve the prediction in the regions of moderate rainfall intensity. Besides, RAP-Cell$_x$ and RAP-Cell$_h$ embed RAB in the current input and the hidden state, respectively. Their performance is better than the original model PredRNN, especially in the 20 dBZ threshold. It shows the superiority of RAB.

We also show predictions of different methods for a given sample in Fig. 9. We find that RAP-Cell can generate the red area which is reflected by better evaluation indexes of HSS and CSI in the highest threshold. However, all forecasts, except for RAP-Net, have a gap in the radar echo block, which is obviously different from the ground truth. The improvement of prediction in moderate rainfall intensity regions can be owed to the embedding of RAM.

5 Conclusions

In this paper, we propose the RAP-Net to handle the precipitation nowcasting task. On the one hand, it embeds the Region Attention Block to enhance the local and global spatial representation ability simultaneously by extracting and delivering the features in ConvRNN. The improvement can significantly enhance the accuracy, especially in those regions with heavy rainfall. On the other hand, we introduce the Recall Attention Mechanism to improve the temporal expressivity in the long term. It can preserve and retrieve longer historical information and effectively enhance the performance of prediction, particularly for the moderate rainfall intensity regions. We conduct extensive experiments to evaluate the performances of most ConvRNN models. Empirically, RAP-Net can preserve regions of heavy intensity in long-term predictions. It shows the effectiveness of RAB and RAM in improving forecasting. The ablation study independently measures the influence of these two modules. The RAB is able to enhance the accuracy in the high threshold, and RAM can improve the prediction in the middle threshold.

Currently, most of the existing methods focus on radar echo maps prediction based on a single altitude layer. The variety and movement of the echo not only need to consider the previous sequence in the same layers but also need to use different altitude layers because the hydrometeors not only happen in the horizontal direction but also act in the vertical direction. For future work, we will consider integrating other layers' historical information to improve the forecasting. In detail, we intend to utilize channel attention to exploit the spatiotemporal representations and then integrate those into the RAP unit. After training, the model can adaptively extract valid spatial information from different levels. We will perform further experiments on multi-channel RAP-Net based on multi-layers of radar echo images. Besides, by a visualization of the similarity matrix in channel attention, the level which is more important for final predictions can be found out.

Code availability. The source code and pretrained model of RAP-Net are available at https://doi.org/10.5281/zenodo.5979275 (Zhang and Luo, 2022).

Data availability. The data are available at https://doi.org/10.5281/zenodo.5979275 (Zhang and Luo, 2022).

Author contributions. Conceptualization, methodology, and writing of the investigation were performed by ZZ and CL. RY contributed to the project administration and visualization. SF and XL contributed to writing, review, editing, data curation, and validation. YY contributed to supervision, visualization, investigation, resources, and funding acquisition.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. This work has been supported in part by the Shenzhen Science and Technology Program, China (grant nos. JCYJ20200109113014456 and JCYJ20210324120208022).

Financial support. This work has been supported in part by the Shenzhen Science and Technology Program (grant nos. JCYJ20180507183823045, JCYJ20200109113014456, and JCYJ20210324120208022).

Review statement. This paper was edited by Charles Onyutha and reviewed by three anonymous referees.

References

Alibaba Cloud: CIKM AnalytiCup2017 competition, Alibaba Cloud [data set], https://tianchi.aliyun.com/competition/entrance/231596/information, last access: 2022.

Ayzel, G., Scheffer, T., and Heistermann, M.: RainNet v1.0: a convolutional neural network for radar-based precipitation nowcasting, Geosci. Model Dev., 13, 2631–2644, https://doi.org/10.5194/gmd-13-2631-2020, 2020.

Chai, Z., Yuan, C., Lin, Z., and Bai, Y.: CMS-LSTM: Context-Embedding and Multi-Scale Spatiotemporal-Expression LSTM for Video Prediction, arXiv [preprint], https://doi.org/10.48550/arXiv.2102.03586, April 2022.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H.: Dual attention network for scene segmentation, in: CVPR, pp. 3146–3154, https://doi.org/10.1109/CVPR.2019.00326, 2019.

Guen, V. L. and Thome, N.: Disentangling physical dynamics from unknown factors for unsupervised video prediction, in: CVPR, pp. 11474–11484, https://doi.org/10.1109/CVPR42600.2020.01149, 2020.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: CVPR, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Jia, X., De Brabandere, B., Tuytelaars, T., and Gool, L. V.: Dynamic filter networks, NIPS, 29, 667–675, 2016.

Li, D., Liu, Y., and Chen, C.: MSDM v1.0: A machine learning model for precipitation nowcasting over eastern China using multisource data, Geosci. Model Dev., 14, 4019–4034, https://doi.org/10.5194/gmd-14-4019-2021, 2021.

Lin, Z., Li, M., Zheng, Z., Cheng, Y., and Yuan, C.: Self-attention convlstm for spatiotemporal prediction, in: AAAI, vol. 34, pp. 11531–11538, https://ojs.aaai.org/index.php/AAAI/article/view/6819 (last access: July 2022), 2020.

Luo, C., Li, X., and Ye, Y.: PFST-LSTM: A SpatioTemporal LSTM Model With Pseudoflow Prediction for Precipitation Nowcasting, IEEE J. Sel. Top. Appl., 14, 843–857, 2020.

Luo, C., Li, X., Wen, Y., Ye, Y., and Zhang, X.: A Novel LSTM Model with Interaction Dual Attention for Radar Echo Extrapolation, Remote Sensing, 13, 164, https://doi.org/10.3390/rs13020164, 2021.

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., and Foresti, L.: Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0), Geosci. Model Dev., 12, 4185–4219, https://doi.org/10.5194/gmd-12-4185-2019, 2019.

Ryu, S., Lyu, G., Do, Y., and Lee, G.: Improved rainfall nowcasting using Burgers' equation, J. Hydrol., 581, 124140, https://doi.org/10.1016/j.jhydrol.2019.124140, 2020.

Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C.: Deep learning for precipitation nowcasting: A benchmark and a new model, NIPS, https://proceedings.neurips.cc/paper/2017/hash/a6db4ed04f1621a119799fd3d7545d3d-Abstract.html (last access: July 2022), 2017.

Shouno, O.: Photo-realistic video prediction on natural videos of largely changing frames, arXiv [preprint], https://doi.org/10.48550/arXiv.2003.08635, 19 March 2020.

Tian, L., Li, X., Ye, Y., Xie, P., and Li, Y.: A generative adversarial gated recurrent unit model for precipitation nowcasting, IEEE Geosci. Remote S., 17, 601–605, 2019.

Trebing, K., Staczyk, T., and Mehrkanoon, S.: Smaat-unet: Precipitation nowcasting using a small attention-unet architecture, Pattern Recogn. Lett., 145, 178–186, 2021.

Wang, G., Wong, W., Liu, L., and Wang, H.: Application of multiscale tracking radar echoes scheme in quantitative precipitation nowcasting, Adv. Atmos. Sci., 30, 448–460, 2013.

Wang, Y., Long, M., Wang, J., Gao, Z., and Yu, P. S.: Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms, in: NIPS, pp. 879–888, https://proceedings.neurips.cc/paper/2017/hash/e5f6ad6ce374177eef023bf5d0c018b6-Abstract.html (last access: July 2022), 2017.

Wang, Y., Gao, Z., Long, M., Wang, J., and Philip, S. Y.: Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning, in: ICML, pp. 5123–5132, PMLR, http://proceedings.mlr.press/v80/wang18b.html (last access: July 2022), 2018aa.

Wang, Y., Jiang, L., Yang, M.-H., Li, L.-J., Long, M., and Fei-Fei, L.: Eidetic 3d lstm: A model for video prediction and beyond, in: International conference on learning representations, https://openreview.net/forum?id=B1lKS2AqtX (last access: July 2022), 2018b.

Wang, Y., Zhang, J., Zhu, H., Long, M., Wang, J., and Yu, P. S.: Memory in memory: A predictive neural net-

work for learning higher-order non-stationarity from spatiotemporal dynamics, in: CVPR, pp. 9154–9162, https://doi.org/10.1109/CVPR.2019.00937, 2019.

Woo, W.-C. and Wong, W.-K.: Operational application of optical flow techniques to radar-based rainfall nowcasting, Atmosphere, 8, 48, https://doi.org/10.3390/atmos8030048, 2017.

Wu, H., Yao, Z., Wang, J., and Long, M.: MotionRNN: A flexible model for video prediction with spacetime-varying motions, in: CVPR, pp. 15435–15444, https://openaccess.thecvf.com/content/CVPR2021/html/Wu_MotionRNN_A_Flexible_Model_for_Video_Prediction_With_Spacetime-Varying_Motions_CVPR_2021_paper.html (last access: July 2022), 2021.

Xie, P., Li, X., Ji, X., Chen, X., Chen, Y., Liu, J., and Ye, Y.: An Energy-Based Generative Adversarial Forecaster for Radar Echo Map Extrapolation, IEEE Geosci. Remote S., 1–5, https://doi.org/10.1109/LGRS.2020.3023950, 2020.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: NIPS, pp. 802–810, https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html (last access: July 2022), 2015.

Zhang, Z. and Luo, C.: RAP-Net: Region Attention Predictive Network for Precipitation Nowcasting, Zenodo [code and data set], https://doi.org/10.5281/zenodo.5979275, 2022.

Zheng, K., Liu, Y., Zhang, J., Luo, C., Tang, S., Ruan, H., Tan, Q., Yi, Y., and Ran, X.: GAN–argcPredNet v1.0: a generative adversarial model for radar echo extrapolation based on convolutional recurrent units, Geosci. Model Dev., 15, 1467–1475, https://doi.org/10.5194/gmd-15-1467-2022, 2022.