ED-DRAP: Encoder–Decoder Deep Residual Attention Prediction Network for Radar Echoes

Hongshu Che, Dan Niu, Zengliang Zang, Yichao Cao, and Xisong Chen

Abstract—**Precipitation nowcasting is quite important and fundamental. It underlies various public services ranging from rainstorm warnings to flight safety. In order to further improve the prediction accuracy for the spatiotemporal sequence forecasting problem, we propose an encoder–decoder deep residual attention prediction network, which adaptively rescales the multiscale sequence- and spatial-wise features and achieves very deep trainable residual prediction by integrating global residual learning and local deep residual sequence and spatial attention blocks (RSSABs). Experiments in a real-world radar echo map dataset of South China show that compared with the ingenious PredRNN++, TrajGRU methods, and newly proposed Unet-based methods, our ED-DRAP network performs better on the precipitation nowcasting metrics, as well as occupies small GPU memory.**

Index Terms—**Deep residual prediction, encoder–decoder, precipitation nowcasting, sequence and spatial attention.**

I. Introduction

PRECIPTATION nowcasting is to supply timely and precise rainfall intensity and range prediction for a local region over a relatively short period of time (e.g., 0–2 h) [1]. It has great application prospects in aerospace, agriculture, transportation, such as generating society-level emergency rainfall alerts [1]–[3]. Due to relevant dynamical processes and inherent atmosphere complexities, it has become a quite challenging and hot research field. Traditional precipitation forecasting mainly relies on the numerical weather prediction (NWP) approaches [3], which are based on the complex and meticulous equations of fluid and thermo dynamics by taking wind speed, pressure, temperature, and other factors into account. However, NWP method usually uses supercomputers for calculations, which require plenty of computing power and takes the burden of calculation time [4]–[5].

Another classical method for precipitation nowcasting is radar echo extrapolation-based method [6]. The conventional methods are optical flow-based methods (e.g., ROVER, currently used in the Hong Kong Observatory [6]), which estimate the convective cloud movements from the observed radar echo maps and then predict the future radar echo maps using semi-Lagrangian advection. However, two important assumptions used in optical flow-based methods restrict the performance: 1) no rapid nonlinear changes exist in the motion and 2) the total intensity remains constant [7]. In the radar echo extrapolation, the cloud motion is nonlinear and highly dynamic and then the assumptions will be violated. Moreover, the optical flow extrapolation methods do not make full use of the vast amount of existing radar echo data.

Recently, supervised deep learning techniques present potential for precipitation nowcasting, which can be formulated as a spatiotemporal sequence forecasting problem [8]–[10]. First, some advanced methods based on the recurrent neural network (RNN) and long short-term memory (LSTM) model provide some useful insights to solve this problem [11], [12]. Shi *et al.* [13] extended the LSTM and proposed the convolutional LSTM (ConvLSTM) model, which utilizes convolutional structures in both the input-to-state and state-to-state transitions, and outperforms the previous models for the grid-wise precipitation nowcasting. In addition, considering the cloud motion patterns like rotation and scaling, they also further proposed the trajectory GRU (TrajGRU) model [14] that did not use a location-invariant filter in ConvLSTM, but employed a subnetwork to actively learn the state-to-state connection structures. Moreover, a predictive RNN (PredRNN) and the improved one (PredRNN++) were presented in [15] and [16], in which the gradient highway units worked seamlessly with the causal LSTMs to adaptively capture the short and long-term dependencies and obtained better prediction results than ConvLSTM and TrajGRU models in some real datasets. However, these RNN-based methods face important problems that gradients vanishing always exist and network training requires a large amount of computing memory.

At the same time, some full-convolutional networks with faster training speed [9] instead of convolutional-recurrent architectures-based methods are presented for weather prediction [17]–[19]. A shallow U-Net model on a fusion of rainfall radar images and wind velocity was presented in [20], which transformed rain nowcasting into a classification problem and enabled significant improvement than the optical flow methods. In [21], an efficient SmaAt-UNet equipped with attention modules and depthwise-separable convolutions was proposed and it gave better prediction performance while only using less trainable parameters than the original UNet. In [22], a SE-ResUNet network combing U-Net, ResNet, and SE attention was proposed for rainfall prediction in Beijing and obtained comparable better results. However, compared

Manuscript received October 3, 2021; revised December 6, 2021; accepted January 4, 2022. Date of publication January 7, 2022; date of current version January 24, 2022. This work was supported in part by the National key Research and Development Program of China under Grant 2018YFC1506905, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20202006, and in part by Zhishan Youth Scholar Program of SEU, Key Research and Development Program of Jiangsu Province under Grant BE2019052. *(Corresponding author: Dan Niu.)*

Hongshu Che, Dan Niu, Yichao Cao, Xisong Chen are with the Key Laboratory of Measurement and Control of CSE, Ministry of Education, and School of Automation, Southeast University, Nanjing 210096, China (e-mail: danniu1@163.com).

Zengliang Zang is with the Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China.

Digital Object Identifier 10.1109/LGRS.2022.3141498

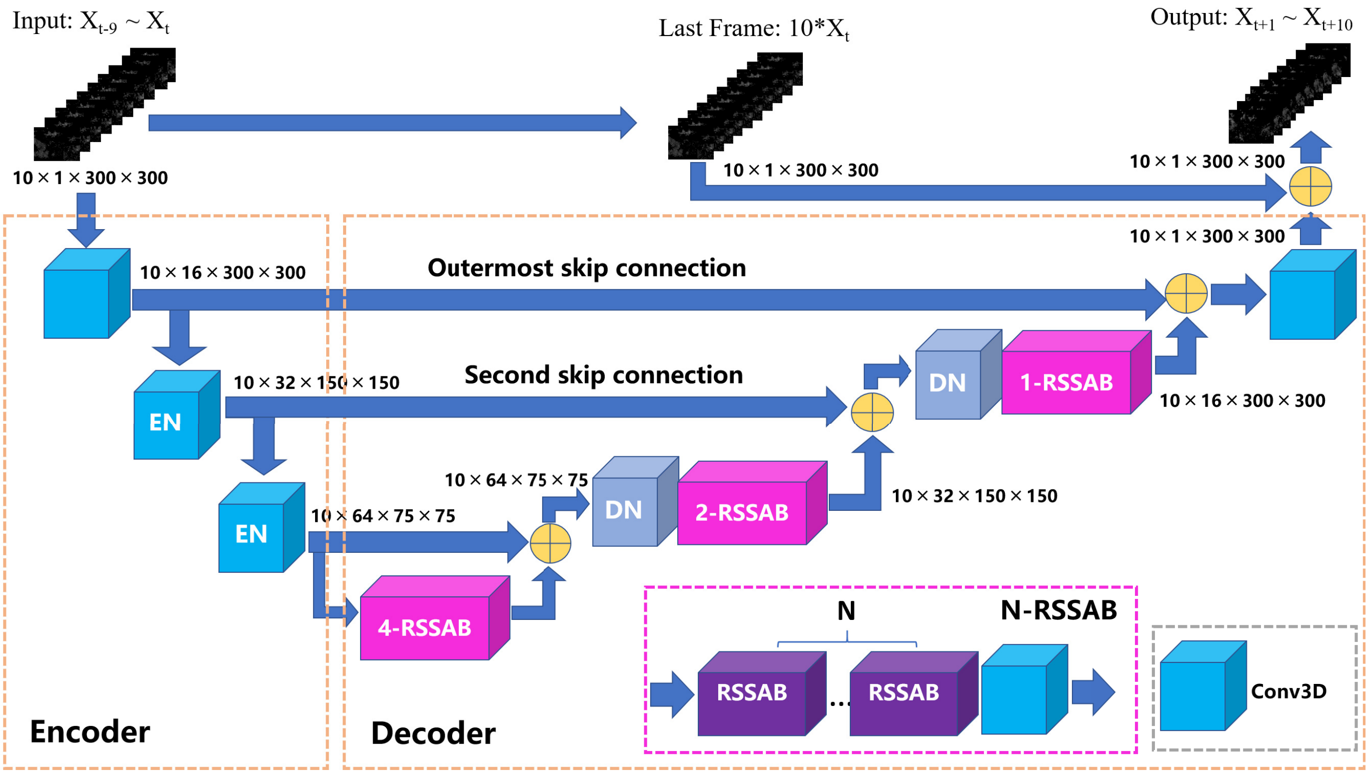

Fig. 1. Architecture of ED-DRAP. ⊕ denotes the element-wise addition.

with the prediction accuracy with RNN-based approaches, there remains room for performance improvement in the fully convolutional network methods, which were mainly used for classification purposes.

In this letter, we propose an encoder–decoder deep residual attention prediction network (ED-DRAP) for precipitation nowcasting, which achieves a very deep trainable residual prediction network and adaptively learns more useful sequence-wise and spatial-wise features. Experimental results show that the proposed ED-DRAP network outperforms the traditional optical flow method, two newly proposed Unet-based methods and two ingenious RNN-based methods at precipitation nowcasting metrics. Overall, our contributions are threefold: 1) we propose an encoder–decoder deep residual attention prediction network for precipitation nowcasting; 2) global residual learning and local residual in residual structure are integrated to obtain very deep trainable residual prediction network and ease the flow of multiscale spatiotemporal feature information; and 3) deep residual sequence and spatial attention mechanism are proposed to adaptively rescale multiscale spatiotemporal features for guiding a high-to-low level residual prediction. High-level features in the decoder, which have integrated the global discriminative spatiotemporal representation, could guide the update of the low-level features for better prediction.

II. Encoder–Decoder Deep Residual Attention Prediction Network

In essence, precipitation nowcasting is a spatiotemporal sequence forecasting problem with the sequence of past radar maps as input and the sequence of future radar maps as output [13]. At a given timestamp t, the model generates the most likely K-step predictions based on the previous J observations including the current one. However, such learning problems are nontrivial due to the high dimensionality of the spatiotemporal sequences especially for multistep predictions, unless the spatiotemporal features of the data are captured well. In this letter, the encoder–decoder structure is adopted as our main network architecture, as shown in Fig. 1. The past radar echo maps are concatenated as the input of model. The encoder part processes the whole input echo sequence and extract multiple scale spatiotemporal representations by stacked downsampling layers (Conv3D), which halves the image size and doubles the number of feature maps, respectively. The encoders are subsequently followed by the same amount of decoders. In the decoder part, deep residual sequence and spatial attention block (RSSAB) are proposed to allow our decoder network to concentrate on more useful frames (time steps) and spatial regions. RSSAB can adaptively rescale and combine the multiscale spatiotemporal features to guide the sequence prediction process with the global and local deep residual learning.

A. Encoder

Since the input is radar echo map or feature map sequences, 3-D convolution (Conv3D) is employed to extract spatiotemporal features in the encoder. The encoding path starts with a Conv3D layer to extract low-level spatiotemporal features $ST_{E,0}$, which is further employed for the outermost skip connection. Then, the encoder is composed of two consecutive cells

$$ST_{E,i+1} = EN_i\left(ST_{E,i}\right) \tag{1}$$

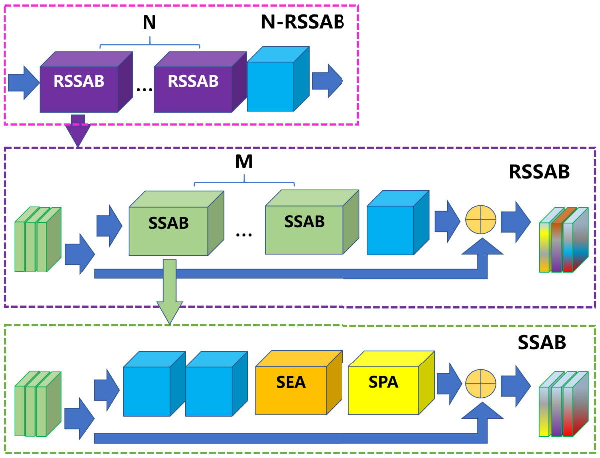

Fig. 2. N-stacked deep RSSAB.

where $ST_{E,i}$ $(i = 0, 1)$ is the input feature map sequence. $ST_{E,i+1}$ $(ST_{E,1}, ST_{E,2})$ represent the extracted different scale features, which are utilized for the second and innermost skip connections. $EN_i(\cdot)$ denotes downscaling process and is a succession of a bilinear downsampling layer (Conv3D) followed by a batch-norm layer and a rectifier linear unit (ReLU). The Conv3D layers halve the input image size in the sequence and double the number of feature maps, respectively. Batch-norm helps the network to train faster and to be more stable. The ReLU layer enables the network to model nonlinear relations. By the encoder part, the resulting small-, middle-, and large-scale spatiotemporal feature maps and receptive fields are obtained, which will be combined in the decoder part via global skip-connections. In this work, global skip connection and local deep residual learning are integrated to ease and combine the flow of multiscale spatiotemporal feature information for generating better prediction. The local deep residual learning is introduced by the following stacked deep RSSAB.

B. Stacked Deep RSSAB

In this work, stacked deep RSSAB is proposed to construct a very deep trainable decoder network and adaptively learn more useful sequence-wise and spatial-wise features. Residuals in residual structure [23] are introduced, where the local long and short skip connections ease the flow of spatiotemporal information. As illustrated in Figs. 1 and 2, N-Stacked deep RSSAB contains N RSSAB and a Conv3D layer. The RSSAB further contains *M-sequence* and spatial attention block (SSAB), a Conv3D, and introduces local long skip connection (LLS), since simply stacking many SSABs would fail to achieve better performance and LLS can ease the flow of information across SSABs. Furthermore, each SSAB contains two Conv3D, 3-D sequence attention (SEA), and 3-D spatial attention (SPA) blocks and introduces local short skip connection (LSS), which further allows the main network to learn residual information. Such residual in residual structure allows to train a very deep network and combines deep spatiotemporal features for high prediction performance.

Instead of treating all features fairly, sequence and SPA modules are proposed for recognizing and rescaling discriminative features adaptively across feature sequences and spatial regions by modeling the interdependencies. As shown in Fig. 3, the 3-D SEA contains a 3-D average pooling, two Conv3D, and sigmoid. And the 3-D SPA contains Conv3D and sigmoid. To capture dependencies from the aggregated information by average pooling, the simple sigmoid is utilized to learn a nonlinear relationship.

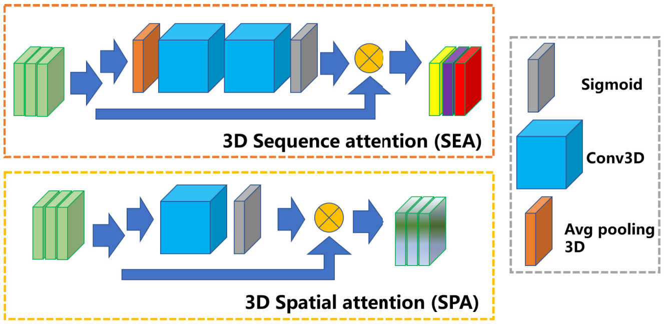

Fig. 3. SEA and 3-D SPA block.

As a result, the SEA makes the different radar echo frames (time steps) from sequence have different impacts on future forecasting. And the SPA expresses the important differences of each three-dimensional pixel of the receptive field in the spatial dimension for future forecasting. More useful spatiotemporal features can be concentrated and discriminative learning ability can be enhanced by such sequence and SPA mechanism.

C. Decoder

In the decoder, we first employ four-stacked RSSAB and a Conv3D layer to extract high-level discriminative features from the deep feature representation $ST_{E,2}$. Then the extracted features are incorporated with the original deep representations $ST_{E,2}$ by the innermost skip connnection and high-level residual prediction $ST_{D,2}$ is obtained. High-level information could guide the subsequent low-level prediction process. Similar to the encoding part, two consecutive upsample cells are constructed

$$ST_{D,i} = RS_i\left(DN_i\left(ST_{D,i+1}\right)\right) + ST_{E,i} \tag{2}$$

where $ST_{D,i+1}$ $(i = 0, 1)$ is the higher level residual prediction. $RS_i(DN_i(ST_{D,i+1}))$ represents the extracted features, which will be combined by the second and outermost skip connections. $DN_i(\cdot)$ denotes upscaling process and is composed of three operations: a DeConv3D, Batch Normalization (BN), and ReLU. $RS_i(\cdot)$ represents stacked RSSAB to adaptively rescale more useful spatiotemporal features. The global skip-connections in the decoder part are enabled to use multiple scale features of the input sequence for restoring the lost information and generating the predicted output radar echo sequence.

Finally, a $1 \times 1 \times 1$ 3-D convolution layer is utilized for $ST_{D,0}$ and outputs the residual prediction for the last frame X_t

$$X_{t+1}, \ldots, X_{t+10} = X_t + \text{Conv}_{1\times1\times1}\left(ST_{D,0}\right). \tag{3}$$

It will be demonstrated that this operation could further improve the whole model's ability for precipitation nowcasting.

TABLE I

SKILL SCORES ($R >= 20$ dBZ)

Models	Skill metrics			
	CSI↑	***HSS↑***	***POD↑***	***FAR↓***
OpticalFlow [6]	0.490	0.563	0.628	0.322
TarjGRU [14]	0.564	**0.644**	0.695	0.260
PredRNN++ [16]	0.554	0.628	0.724	0.306
SmaAt-Unet [21]	0.541	0.621	0.677	0.279
SE-ResUNet [22]	0.536	0.619	0.646	0.251
ED-DRAP-NRP	0.558	0.640	0.680	**0.250**
ED-DRAP	**0.570**	**0.644**	**0.761**	0.310

TABLE II

SKILL SCORES ($R >= 35$ dBZ)

Models	Skill metrics			
	CSI↑	***HSS↑***	***POD↑***	***FAR↓***
OpticalFlow [6]	0.317	0.441	0.455	0.519
TarjGRU [14]	0.357	0.491	0.495	**0.459**
PredRNN++ [16]	0.346	0.474	0.571	0.550
SmaAt-Unet [21]	0.321	0.446	0.456	0.499
SE-ResUNet [22]	0.348	0.480	0.535	0.519
ED-DRAP-NRP	0.354	0.487	0.514	0.486
ED-DRAP	**0.363**	**0.493**	**0.651**	0.560

TABLE III

SKILL SCORES ($R >= 45$ dBZ)

Models	Skill metrics			
	CSI↑	***HSS↑***	***POD↑***	***FAR↓***
OpticalFlow [6]	0.129	0.212	0.207	0.772
TarjGRU [14]	0.154	0.251	0.209	**0.651**
PredRNN++ [16]	0.154	0.253	**0.322**	0.783
SmaAt-Unet [21]	0.119	0.196	0.169	0.720
SE-ResUNet [22]	0.147	0.241	0.249	0.750
ED-DRAP-NRP	0.151	0.251	0.214	0.676
ED-DRAP	**0.160**	**0.258**	0.297	0.761

III. EXPERIMENTS

In this letter, the radar echo dataset is provided by Guangdong Meteorological Bureau from 2017 to 2019. The observation interval is 12 min. The spatial resolution is 1 km and the observation area is Southern China. In order to reduce the memory cost, the region covering 300 km × 300 km of the Pearl River Delta is selected, covering longitude ranges from 112° to 115°E and latitude from 22° to 25°N. For preprocessing, the radar intensities are first mapped to pixel values and 300 × 300 gray-scale radar images are obtained [13]. Then the consecutive radar images are sliced with a 20-frame-wide sliding window. Thus, each sequence consists of 20 frames, ten for the input, and ten for forecasting (2 h). The total 356 precipitation events are split into a training set of 254 samples and a test set of 102 samples. The frequencies of different rainfall levels are highly imbalanced. Thus the weighted loss function B-MSE + B-MAE is designed [14] and Adam optimizer to optimize it, with an initial learning rate of 10^{-4}.

In this letter, we compare our proposed ED-DRAP network with typical optical flow-based method (ROVER [6]), two ingenious RNN-based methods (TrajGRU [14], PredRNN++ [16]), and two well-known CNN-based methods (SmaAt-Unet [21] and SE-ResUNet [22]) on a server with RTX 3090 24GB GPU and Intel 3.40 GHz CPUs (24 cores). Four commonly used precipitation nowcasting metrics, including Critical Success Index (CSI), Heidke Skill Score (HSS), Probability of Detection (POD), False Alarm Rate (FAR), are used to evaluate the prediction accuracy [14]. Moreover, in order to give an all-round performance evaluation, we calculate the skill scores for three radar reflectivity R thresholds (20, 35, and 45 dBZ) that correspond to different rainfall levels. Tables I–III show the precipitation nowcasting metric results for the 2 h prediction. Fig. 4 also presents nowcasting metric scores (CSI, HSS, FAR, POD at different thresholds) for 12–120 min lead time. Moreover, GPU memory usage for the model training with batch size 1 and the time spent during the forecast are also compared in Table IV.

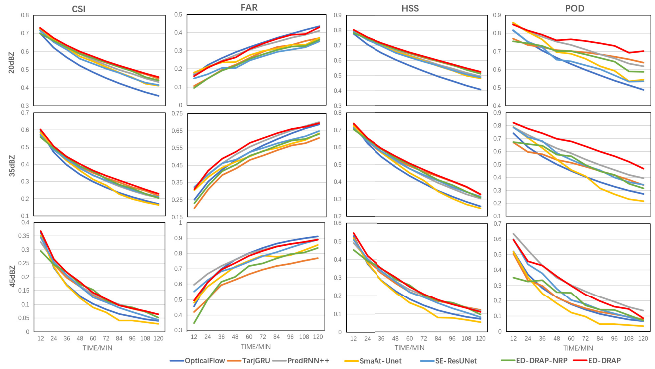

Fig. 4. Nowcasting metric scores (CSI, HSS, FAR, POD at different thresholds) for 12–120 min lead time.

TABLE IV

GPU MEMORY USAGE IN TRAINING STEP (BATCHSIZE = 1) AND THE FORECAST TIME SPENT

Models	GPU memory usage (MB)	Forecast time spent (s)
TarjGRU	4174	0.5413
PredRNN++	12966	0.4188
ED-DRAP	**3186**	**0.0147**

It is clear that the deep learning models outperform the optical flow-based model. Among the deep learning models, our proposed ED-DRAP network performs the best at nearly four metrics over the two newly proposed Unet-based methods and also two RNN-based methods. Moreover, residual prediction for the last frame X_t can further improve prediction accuracy by compared with nonresidual prediction version (ED-DRAP-NRP). At the same time, the proposed method occupies less GPU memory than two RNN-based methods, which means less training time and less computing costs.

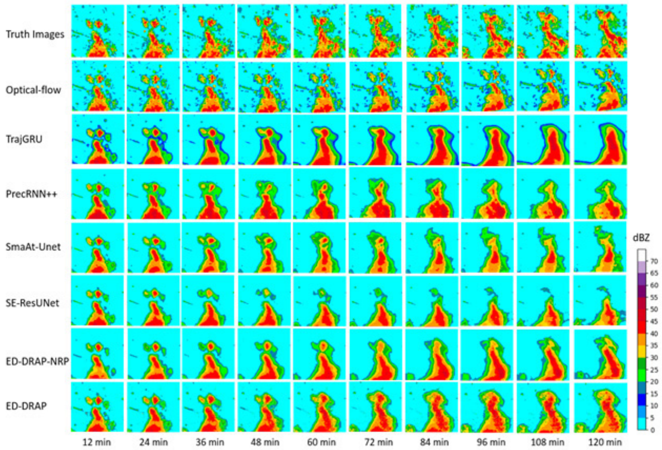

Fig. 5. Visualization comparison among the evaluated methods for 12–120 min lead times.

In addition, visualizing the comparison among the evaluated methods is shown in Fig. 5. All models predict accurately at the first moment. As the lead time increases, the models show significant differences. The OpticalFlow forecasting scale is gradually reduced, and the change of echo intensity is nearly ignored. The small-scale details in deep learning methods are gradually lost, the boundaries become smooth. Although OpticalFlow method [6] can give sharper predictions than deep learning methods, they trigger more false alarms and are less precise than deep learning methods in general. Compared with other deep learning methods, our proposed ED-DRAP method shows the best performance. The extrapolations are more realistic. As time goes by, the intensity and the position are closer to the truth images.

IV. Conclusion

In this letter, we propose an encoder–decoder deep residual attention prediction network for precipitation nowcasting, which achieves very deep trainable residual prediction and eases the flow of multiscale spatiotemporal features by combing the global residual learning and local residual in residual structure. Moreover, deep residual sequence and SPA mechanism is proposed to adaptively rescale multiscale spatiotemporal features for guiding a high-to-low level residual prediction. We have shown that the proposed ED-DRAP has less GPU memory costs but has better prediction performance than the newly proposed two Unet-based and two ingenious RNN-based precipitation nowcasting methods. For future work, we will try to build an operational nowcasting system using the proposed algorithm with the Guangdong Meteorological Bureau.

References

[1] T. Gneiting and A. E. Raftery, "Weather forecasting with ensemble methods," *Science*, vol. 310, no. 5746, pp. 248–249, Oct. 2005.

[2] N. Jones, "Machine learning tapped to improve climate forecasts," *Nature*, vol. 548, pp. 379–380, Apr. 2017.

[3] G. Marchuk, *Numerical Methods in Weather Prediction*. Amsterdam, The Netherlands: Elsevier 2012.

[4] M. A. Tolstykh and A. V. Frolov, "Some current problems in numerical weather prediction," *Izvestiya Atmos. Ocean. Phys.*, vol. 41, no. 3, pp. 285–295, 2005.

[5] J. Sun *et al.*, "Use of NWP for nowcasting convective precipitation: Recent progress and challenges," *Bull. Amer. Meteorol. Soc.*, vol. 95, no. 3, pp. 409–426, Mar. 2014.

[6] W.-C. Woo and W.-K. Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 12, p. 48, Feb. 2017.

[7] L. Tian, X. Li, Y. Ye, P. Xie, and Y. Li, "A generative adversarial gated recurrent unit model for precipitation nowcasting," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 601–605, Apr. 2020.

[8] C. L. Bromberg, C. Gazen, J. J. Hickey, J. Burge, L. Barrington, and S. Agrawal, "Machine learning for precipitation nowcasting from radar images," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, BC, Canada, 2019, pp. 1–4.

[9] R. Prudden *et al.*, "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," 2020, *arXiv:2005.04988*.

[10] C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Jul. 2020, pp. 92–97.

[11] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting," *Proc. Comput. Sci.*, vol. 135, pp. 89–98, Jan. 2018.

[12] D. Niu, L. Diao, L. J. Xu, Z. L. Zang, X. S. Chen, and S. S. Liang, "Precipitation forecast based on multi-channel ConvLSTM and 3D-CNN," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, 2020, pp. 367–371.

[13] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2015, pp. 802–810.

[14] X. Shi *et al.*, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5617–5627.

[15] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 879–888.

[16] Y. Wang, Z. Gao, M. Long, J. Wang, and S. Y. Philip, "PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning," in *Proc. Mach. Learn. (ICML)*, 2018, pp. 5123–5132.

[17] M. G. Schultz *et al.*, "Can deep learning beat numerical weather prediction?" *Philos. Trans. Roy. Soc. A*, vol. 379, no. 2194, 2021, Art. no. 20200097.

[18] M. Qiu *et al.*, "A short-term rainfall prediction model using multi-task convolutional neural networks," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 395–404.

[19] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine learning for precipitation nowcasting from radar images," 2019, *arXiv:1912.12132*.

[20] V. Bouget, B. Dominique, J. Brajard, A. Charantonis, and A. Filoche, "Fusion of rain radar images and wind forecasts in a deep learning model applied to rain nowcasting," *Remote Sens.*, vol. 13, no. 2, p. 246, 2021.

[21] K. Trebing, T. Staǹczyk, and S. Mehrkanoon, "SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture," *Pattern Recognit. Lett.*, vol. 145, pp. 178–186, May 2021.

[22] K. Song *et al.*, "Deep learning prediction of incoming rainfalls: An operational service for the city of Beijing China," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2019, pp. 180–185.

[23] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.