RainfallNet: A Dual-Source of Spatial-Channel Attention Fusion Network for Precipitation Nowcasting

RainfallNet: A Dual-Source of Spatial-Channel Attention Fusion Network for Precipitation Nowcasting

Junhao Huang[1], Dan Niu[1,2,*], Zengliang Zang[3], Xisong Chen[1,2], Xiaobin Pan[3]

[1]School of Automation, Southeast University, Nanjing 210096, China
[2]Key Laboratory of Measurement and Control of CSE, Ministry of Education
[3]Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China

E-mail: danniu1@163.com

Abstract. Accurate rainfall prediction is conductive to human life and disaster prevention. Meanwhile, deep learning methods are confirmed to be helpful to improve the accuracy of weather prediction. A novel data-driven neural network is proposed in this work referred as RainfallNet which introduces fusion module based on both radar echo observations and numerical weather prediction (NWP) data. The architecture of the network includes three elements: (1) dual encoders to extract the spatio-temporal feature of the radar echo images and NWP data respectively, (2) parallel attention mechanism combining channel attention and spatial attention to reveal the contribution of each data source and (3) combined loss function joining structural similarity loss, mean square error and mean absolute error with different weight for each rainfall level to further increase the meteorologically assessment metrics. The experiments on South China dataset demonstrate the effectiveness of our model, achieving superior performance on meteorologically assessment metrics over most existing algorithms.

1. Introduction

Weather forecast is related to people's well-being, social harmony and the sustainable development of the economy. In particular, accurate precipitation nowcasting can provide basic guarantee for regional life safety and production development. In recent years, radar echo maps which reflect the intensity and distribution of the rainfall takes a crucial role in precipitation nowcasting [1]. However, conventional extrapolation of radar echo maps including optical flow [2], centroid tracking [3], usually has only short term validity and just simply exploits the motion vectors.

In contrast, since the Recurrent Neural Network (RNN) makes a progress in the time series prediction, a number of variations of RNN are invented to solve sequence prediction. Long Short Term Memory network was designed by Horchreiter et al. (1997) to avoid the long-term dependency problem [4]. Cho et al. (2014) proposed an Encoder-Decoder model to generate variable time step prediction [5]. Shi et al. (2015) constructed the network called ConvLSTM [6] to capture spatio-temporal motion pattern which utilized convolution operator instead of hadamard product in the state-to-state and input-to-state transitions on the basis of

the work of Horchreiter and Cho. Whereas the convolutional recurrent structure in ConvLSTM is location-invariant and it could not capture the motion of rainfall effectively. Wang et al (2017) proposed to design a new spatio-temporal LSTM (ST-LSTM) unit with a zigzag connection structure which transfer the memory both horizontally over states and vertically across layers. ST-LSTM invents dual memory cells including temporal cell and spatio-temporal cell to effectively model the shape deformations and motion trajectories [7]. Nevertheless, ST-LSTM still faces the inevitable vanishing gradient problem. TrajGRU [8] can actively learn the location-variant structure for recurrent connections by generating the neighborhood set with parameterized learning subnetwork for every location. Based on ST-LSTM, PredRNN++ [9] introduced gradient highway unit to capture long-term video dependencies. E3D-LSTM [10] was introduced to integrates 3D convolutions into RNNs to recall the stored memory effectively on video prediction. Memory In Memory [11] exploited the differential signals to model stationary and non-stationary properties, handling high-order non-stationary.

The above mentioned recurrent models were designed to mainly solve video prediction task. However, precipitation nowcasting cannot just rely on video-like radar echo frames which based on recent observations. On the one hand, the existing models which rely solely on radar observation image are hard to make precision long-term prediction because of the chaotic character of the climatic system [12]. On the other hand, with the development of global/regional assimilation and prediction system (GRAPES) developed by the Chinese Academy of Meteorological Sciences (CAMS), huge amount of simulated meteorological data can be obtained by four-dimensional variational assimilation system based on non-hydrostatic grid model [13]. In addition, as a new generation of the national operational NWP system, GRAPES has commonly been used for both operation and research including tropical cyclones [14], sea fog [15], etc. However, GRAPES cannot accurately predict minute-level and high-resolution regional precipitation nowcasting [16]. Thus, the marriage between radar echo, NWP data and deep learning brings a new opportunity to precipitation nowcasting.

In this paper, to address these problems, we propose an end-to-end dual-source neural network model called RainfallNet to forecast regional precipitation. Motivated by the recent success of attention mechanism, we introduce a fusion method based on spatial attention and channel attention. We also conduct experiments with real-world radar echo data and NWP data of South China which validated our model design.

2. Problem Description

Given previously radar echo observation data and NWP data in the past and future, our goal is to predict the most possible sequence of radar echo in the future. From mathematical perspective, the task can be considered as spatio-temporal sequence forecasting task which forecasts a certain grid area in the next time. Both observation radar echo data and GRAPES data have the same real spatial size, but due to different resolution, they may have different grid size of rows and columns. Thus, the radar observation which consists of m rows and n columns can be denoted to $X \in \mathbb{R}^{p \times m \times n}$, where p represents the number of measurements. We employ $G \in \mathbb{R}^{h \times m' \times n'}$ to denote GRAPES data which consists of m' rows and n' columns, where h represents the channels and each channel represents a parameter or a certain height level of a parameter. Hence, the spatio-temporal sequence forecasting problem can be formulated as:

$$\hat{X}_{t+1}, ..., \hat{X}_{t+l} = f\left(X_t, ..., X_{t-p}, G_{t-k}, ..., G_t, ..., G_{t+l}\right), \tag{1}$$

where l denotes the length of prediction sequence, and k, p presents the given historical GRAPES data and observations data.

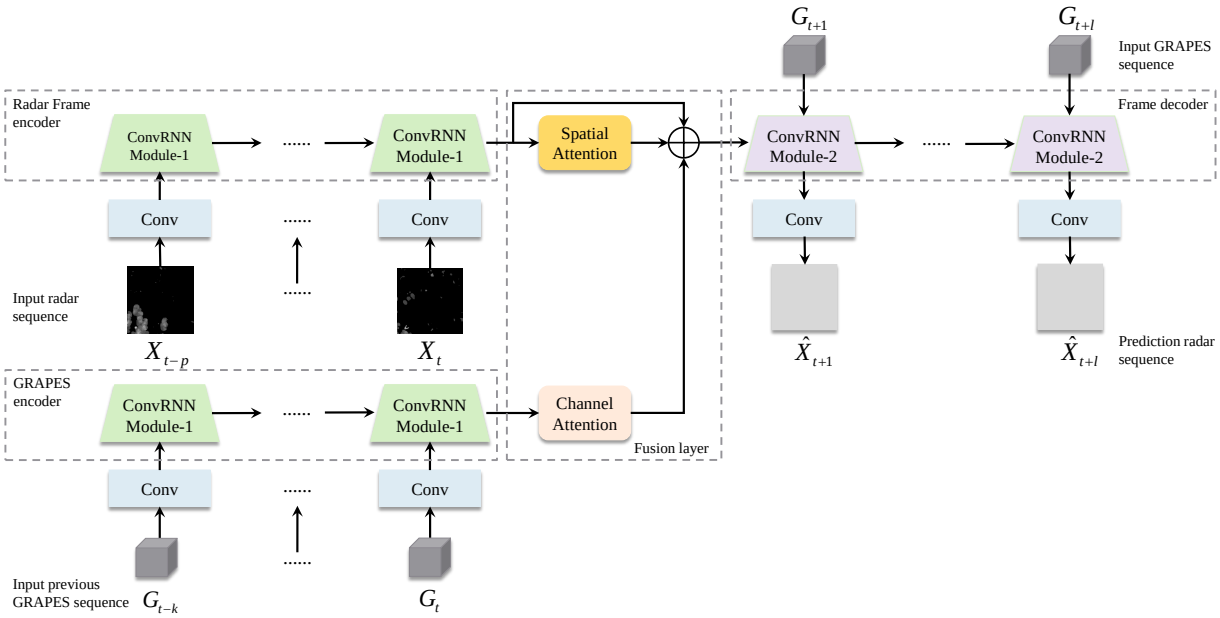

Figure 1. The architecture of RainfallNet.

3. Method

3.1. Network Architecture

Figure. 1 illustrates the architecture of our RainfallNet, where TrajGRU [8] is adopted as our ConvRNN backbone. RainfallNet consists of two encoding parts to extract spatio-temporal feature and a decoding part for predicting.

First of all, the radar observation encoder takes ten frames of radar echo image which is recorded every 12 minutes in order to capture the spatio-temporal feature. The GRAPES encoder is responsible for extracting the information of NWP data. Secondly, since historical radar echo images are able to provide short-term forecasting and GRAPES data has a ability to make long period predictions. Consequently, we introduce an attention-based fusion module to enhance long-term extrapolation ability and can also calibrate GRAPES rain data. Finally, we employ a decoder to receive output of the fusion module and the GRAPES data as input to yield the future predictions.

3.2. Attention Based Fusion Layer

In the encoder stage, we use the stacked TajGRU model as ConvRNN module layers to extract the spatio-temporal information. Nevertheless, in order to enhance not only the short-term forecasting but also long-term forecasting performance and combine the advantages of both data source, we introduce the attention based fusion module in our model.

On the one hand, The spatial attention module produces a spatial attention future map by exploiting the inter-spatial relationship of the hidden states. It mainly focuses on where is an informative part. The formulas of spatial attention as follows:

$$\mathbf{Z_s}(\mathbf{H}) = \sigma\left(\mathrm{Conv}([\mathrm{MaxPool}(\mathbf{H}); \mathrm{AvgPool}(\mathbf{H})])\right). \tag{2}$$

By using max-pooling and avg-pooling to generate two future maps and then concatenate it to pass a convolutional layer [17]. Then 2D spatial attention map is produced.

On the other hand, GRAPES parameter usually varies in different hourly forecasting and it is obvious and meaningful to set weights automatically for different channels. Consequently, the channel attention module produces a channel attention future map by utilizing the inter-channel relationship of the hidden states. Different from spatial attention, channel attention focuses on what is meaningful. The formulas of channel attention as follows:

$$\mathbf{Z_c}(\mathbf{H}) = \sigma(\mathrm{MLP}(\mathrm{MaxPool}(\mathbf{H})) + \mathrm{MLP}(\mathrm{AvgPool}(\mathbf{H}))). \tag{3}$$

We squeeze the spatial dimension of the hidden states by max-pooling and avg-pooling and feed them into a shared multi-layer perceptron (MLP) network.

As we explicitly introduce the two attention modules with a parallel arrangement, we add shortcut value (refers to hidden stats in observation encoder) to main path which is the parallel outputs of two attention modules. In short, the input hidden states of decoder is computed as:

$$\mathbf{H}_{\mathrm{decoder}} = \mathbf{H}_{\mathrm{radar\_encoder}} + \mathbf{Att_s}\left(\mathbf{H}_{\mathrm{GRAPES\_encoder}}\right) + \mathbf{Att_c}\left(\mathbf{H}_{\mathrm{radar\_encoder}}\right). \tag{4}$$

3.3. Combined Loss for Precipitation Nowcasting

We proposed a combined loss function for precipitation nowcasting to capture raining dynamics accurately. Based on statistics, heavy rain whose echo value greater than 45dbZ account for small proportion. Nevertheless, compared to light rain, predicting heavy rain accurately is more valuable due to its suddenness and destructiveness. Consequently, in order to solve such imbalanced dataset problem, we took different weight for each rainfall level. On the basis of classical and effective loss function for regression challenge like mean square error (MSE) and mean absolute error (MAE), we also introduced Structural Similarity [18] (SSIM) to our combined loss function. The formulas of combined loss as follows:

$$\mathcal{L} = \sum_{i,j}\left(w_{ij}\left(\left(\tilde{y}_{ij} - y_{ij}\right)^2 + \left|\tilde{y}_{ij} - y_{ij}\right|\right)\right) + \mathcal{L}_{ssim}, \tag{5}$$

where w is a piecewise weight function to each echo value z:

$$w\left(z\right) = \begin{cases} 1, & z < 20 \\ 2, & 20 \leq z < 35 \\ 6, & 35 \leq z < 45 \\ 60, & z \geq 45. \end{cases} \tag{6}$$

SSIM is used to measure the similarity between two images and focuses on luminance, contrast and structure. The loss function for SSIM is defined by

$$\mathcal{L}_{ssim} = \frac{1}{M}\sum_{i=1}^{M} 1 - \mathrm{SSIM}\left(\tilde{y}_i, y_i\right). \tag{7}$$

And SSIM function is defined by

$$\mathrm{SSIM}(\mathrm{x}, \mathrm{y}) = \frac{\left(2\mu_{\mathrm{x}}\mu_{\mathrm{y}} + C_1\right)\left(2\sigma_{\mathrm{xy}} + C_2\right)}{\left(\mu_{\mathrm{x}}^2 + \mu_{\mathrm{y}}^2 + C_1\right)\left(\sigma_{\mathrm{x}}^2 + \sigma_{\mathrm{y}}^2 + C_2\right)}, \tag{8}$$

where μ denotes the mean intensity, σ denotes the standard deviation, C is the specific constant.

4. Experiment

4.1. Radar Echo Dataset

The radar echo dataset provided by Guangdong Meteorological Bureau covers Pearl River Delta region from 2017 to 2019. Since its latitude and longitude are 22-25 and 112-115 degrees, each pixel of image is $3km \times 3km$ which represents 1km resolution. We normalized the echo value to $[0.0, 1.0]$. To match the range of gray scale images pixel value, we first convert to $[0, 255]$ by multiplying the factor and then also normalized it to between 0 and 1.

4.2. GRAPES Dataset

The GRAPES performs hourly numerical each day at 0:00 and 12:00. Each numerical computation includes the following 24 hours. We select the parameters which are closely related to rainfall: relative humidity, vertical velocity, convective rain, no convective rain. The cumulative precipitation is made up of RAINC and RAINNC. In addition, we convert it into hourly precipitation.

In the training step, in order to enrich the diversity of our dataset and improve the generalization ability, we adopted the common trick about data augmentation for each batch size. Half of samples were randomly selected and the whole selected sequence were transformed by randomly operation among horizontal flip, vertical flip and 45° rotation.

Since considering the weather condition of the Pearl River Delta region, we choose the radar data and GRAPES data from March to September and only in rainy days. Finally, our train dataset and validation dataset across two years (2017&2018) and using 2019 for testing.

4.3. Evaluation Metrics

We adopted four types of commonly used evaluation metrics in weather forecasting. Critical Success Index (CSI) also called the Threat Score (TS) which combines Hit Rate and False Alarm Ratio into one score for low frequency events. Heidke Skill Score (HSS), the reference measure is the proportion correct that would be expected by random forecasts. Probability of Detection (POD) also called Hit Rate which is the fraction of observed events that is forecast correctly. False Alarm Ratio (FAR) is the equal to the number of false alarms divided by the total number of shock forecasts [19].

4.4. Implementation Details

The implementation of the model is based on PyTorch [20] machine learning framework on Nvidia Tesla T4 GPUs. We use the Adam optimizer [21] and set the learning rate equals to 10^{-4} with step-wise learning rate decay at every 5 epoch. The batch-size is set to 4 and we use three stacked layers of ConvRNN unit (TrajGRU) both in encoder and in decoder.

4.5. Experimental Results

We compare our method with three other state-of-the-art (SOTA) methods based on machine learning, which are ConvLSTM [6], TajGRU [8], PredRNN++ [9] and OpticalFlow model [22]. In order to evaluate the performance of RainfallNet, we also conduct ablation experiment on the variation of RainfallNet which removes the fusion layer (denote as RainfallNet w/o fusion). In addition, we compute prediction metrics for three levels of rainfall. The comparison results are shown in Table 1-3. Overall, the RainfallNet model outperforms the other methods.

Table 1. Comparison results of six models (≥ 20 dBZ)

MODEL	First Hour				First Two Hours			
	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR
OpticalFlow [22]	0.579	0.661	0.714	0.252	0.490	0.563	0.628	0.322
ConvLSTM [6]	0.650	0.729	**0.788**	0.221	0.581	0.658	**0.733**	0.281
TrajGRU [8]	**0.651**	**0.730**	0.780	0.220	**0.582**	0.660	0.728	0.282
PredRNN++ [9]	0.647	0.726	0.782	0.213	0.576	0.653	0.731	0.277
RainfallNet w/o fusion	0.648	0.728	0.775	**0.204**	0.578	0.658	0.626	**0.257**
RainfallNet	**0.651**	**0.730**	0.783	0.209	**0.582**	**0.661**	0.729	0.265

Table 2. Comparison results of six models (≥ 35 dBZ)

MODEL	First Hour				First Two Hours			
	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR
OpticalFlow [22]	0.420	0.563	0.577	0.408	0.317	0.441	0.455	0.519
ConvLSTM [6]	0.474	0.619	0.666	0.408	0.378	0.515	0.554	0.501
TrajGRU [8]	0.482	0.625	0.720	0.413	0.393	0.530	0.623	0.499
PredRNN++ [9]	0.468	0.612	0.705	0.424	0.378	0.513	0.610	0.516
RainfallNet w/o fusion	0.483	0.627	**0.725**	0.413	0.398	0.536	**0.626**	0.490
RainfallNet	**0.486**	**0.630**	0.709	**0.400**	**0.399**	**0.538**	0.610	**0.336**

Table 3. Comparison results of six models (≥ 45 dBZ)

MODEL	First Hour				First Two Hours			
	CSI	HSS	POD	FAR	CSI	HSS	POD	FAR
OpticalFlow [22]	0.200	0.320	0.315	0.666	0.129	0.212	0.207	0.772
ConvLSTM [6]	0.244	0.382	0.379	**0.606**	0.171	0.277	0.263	**0.690**
TrajGRU [8]	0.260	0.403	0.435	0.654	0.188	0.302	0.312	0.728
PredRNN++ [9]	0.255	0.396	0.486	0.656	0.184	0.297	0.362	0.740
RainfallNet w/o fusion	0.264	0.407	**0.508**	0.652	0.192	0.308	**0.367**	0.722
RainfallNet	**0.269**	**0.414**	0.464	0.619	**0.196**	**0.314**	0.336	0.693

5. Conclusion

In this paper, we proposed a dual-source of saptial-channel attention fusion network (RainfallNet) for precipitation nowcasting. RainfallNet makes use of dual encoders to extract spatio-temporal features by the fusion layer. The fusion layer contains the spatial attention and channel attention to make better use of the GRAPES data and historical radar information. The learning procedure was constrained by combined loss based on structural similarity loss to improve the metrics further. The experimental results show that RainfallNet achieves the better performance in our real-world radar echo dataset of Pearl River Delta region in China for the next two hours prediction.

6. Acknowledgements
This work was supported by National key R&D Program of China (No. 2018YFC1506900), Zhishan Youth Scholar Program of Southeast University, National NSF of China (No. 61504027), the Key R&D Program of Jiangsu Province (No. BE201905, BE2017076), the Key R&D industrialization Program of Suzhou (No. SGC201733, SGC201854).

References
[1] Wilson JW, Crook NA, Mueller CK, Sun J and Dixon M 1998 Nowcasting thunderstorms: A status report. *Bulletin of the American Meteorological Society*, 79(10):2079-100
[2] Woo WC and Wong WK 2017 Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, 8(3):48
[3] Dixon M and Wiener G 1993 TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *Journal of atmospheric and oceanic technology*, 10(6):785-97
[4] Hochreiter S and Schmidhuber J 1997 Long short-term memory. *Neural computation*, 9(8):1735-80
[5] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv*:1406.1078
[6] Shi X, Chen Z, Wang H, Yeung DY, Wong WK and Woo WC 2015 Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28:802-10
[7] Wang Y, Long M, Wang J, Gao Z and Philip SY 2017 Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Advances in Neural Information Processing Systems*, pp. 879-888
[8] Shi X, Gao Z, Lausen L, Wang H, Yeung DY, Wong WK and Woo WC 2017 Deep learning for precipitation nowcasting: A benchmark and a new model. InAdvances in neural information processing systems, pp. 5617-5627
[9] Wang Y, Gao Z, Long M, Wang J and Yu PS 2018 Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. *arXiv preprint arXiv*:1804.06300
[10] Wang Y, Jiang L, Yang MH, Li LJ, Long M and Fei-Fei L 2018 Eidetic 3d lstm: A model for video prediction and beyond. *International Conference on Learning Representations*
[11] Wang Y, Zhang J, Zhu H, Long M, Wang J and Yu PS 2019 Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9154-9162
[12] Mihailović DT, Mimić G and Arsenić I 2014 Climate predictions: The chaos and complexity in climate models. *Advances in Meteorology*
[13] Liu Y, Zhang L and Lian Z. 2018. Conjugate gradient algorithm in the four-dimensional variational data assimilation system in GRAPES. *Journal of Meteorological Research*, 32(6):974-84
[14] Shen X, Wang J, Li Z, Chen D and Gong J 2020 Research and Operational Development of Numerical Weather Prediction in China. *Journal of Meteorological Research*, 34(4):675-98
[15] Huang H, Huang B, Yi L, Liu C, Tu J, Wen G and Mao W 2019 Evaluation of the Global and Regional Assimilation and Prediction System for Predicting Sea Fog over the South China Sea. *Advances in Atmospheric Sciences*, 36(6):623-42
[16] Cao Y, Li Q, Shan H, Huang Z, Chen L, Ma L and Zhang J 2019 Precipitation Nowcasting with Star-Bridge Networks. *arXiv preprint arXiv*:1907.08069
[17] Woo S, Park J, Lee JY and So Kweon I 2018 Cbam: Convolutional block attention module. *Proceedings of the European conference on computer vision*, pp. 3-19
[18] Wang Z, Bovik AC, Sheikh HR and Simoncelli EP 2004 Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600-12
[19] Wilks DS. 2011. Statistical methods in the atmospheric sciences. *Academic press*
[20] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L and Desmaison A 2019 Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, pp. 8026-8037
[21] Kingma DP and Ba J 2014 Adam: A method for stochastic optimization. *arXiv preprint arXiv*:1412.6980
[22] Woo WC and Wong WK 2014 Application of optical flow techniques to rainfall nowcasting. *the 27th Conference on Severe Local Storms*