Precipitation Nowcasting with Star-Bridge Networks

Yuan Cao¹, Qiuying Li¹, Hongming Shan¹, Zhizhong Huang¹, Lei Chen², Leiming Ma², and Junping Zhang¹

¹Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China
²Shanghai Observatory, Shanghai, China

January 24, 2022

Abstract

Precipitation nowcasting, which aims to precisely predict the short-term rainfall intensity of a local region, is gaining increasing attention in the artificial intelligence community. Existing deep learning-based algorithms use a single network to process various rainfall intensities together, compromising the predictive accuracy. Therefore, this paper proposes a novel recurrent neural network (RNN) based star-bridge network (StarBriNet) for precipitation nowcasting. The novelty of this work lies in the following three aspects. First, the proposed network comprises multiple sub-networks to deal with different rainfall intensities and duration separately, which can significantly improve the model performance. Second, we propose a star-shaped information bridge to enhance the information flow across RNN layers. Third, we introduce a multi-sigmoid loss function to take the precipitation nowcasting criterion into account. Experimental results demonstrate superior performance for precipitation nowcasting over existing algorithms, including the state-of-the-art one, on a natural radar echo dataset.

1 Introduction

Nowcasting precipitation plays a crucial role in agriculture, flood alerting, daily life of citizens, transportation, and so on, which is gaining increasing attention in the artificial intelligence community. Predicting the short-term rainfall intensity in a region is challenging as it relies on a number of meteorological factors such as temperature, humidity, wind, pressure in the region enclosing the clouds and the ground surface. For example, the involved convective precipitation occurs in the form of showers of high intensity and short duration, which is one of the most challenging precipitation forms. Traditional numerical weather prediction (NWP) models suffer from inferior prediction performance in terms of accuracy and resolution.

Given the huge amount of video-like radar echo data provided by the operational weather radar networks, the literature proposed several deep learning-based methods in the form of video prediction for precipitation nowcasting. With the consecutive 6-minute-interval radar echo frames as the input and output data, Shi et. al first dealt with the issue of precipitation nowcasting based on the framework of recurrent neural network(RNN) [1]. More specifically, they proposed encoder-decoder based convolutional long short-term memory (ConvLSTM), which is broadly used in various video tasks. Then Shi et. al [2] utilized optical flow warping to refine the convolutional region of ConvLSTM. Remarkably, predictive RNN (PredRNN) [3] and its variant PredRNN++ [4] achieved the state-of-the-art performance by adding a zigzag memory connection across different LSTM layers.

Although video prediction methods can be directly applied to the precipitation nowcasting task, we highlight that the precipitation nowcasting has following significant differences from video prediction. First, precipitation is highly influenced by temperature, atmosphere, wind, humidity and others apart from radar echo images data, which means that precipitation nowcasting is more complicated than video prediction; see Fig. 1. Second, precipitation has various forms in terms of different intensities and duration, which is more challenging. Third, in addition to nowcasting precipitation, weather criterion such as critical success index (CSI) indicates that the predicted rainfall intensity above some predefined threshold is more important than those below it. Currently
existing methods, however, directly adopted the video prediction methods for precipitation nowcasting, which compromises the prediction performance.

To address these problems, this paper proposes a novel recurrent neural network (RNN) based star-bridge network (*StarBriNet*) for precipitation nowcasting. More specifically, our proposed network 1) includes multi-column structure to process different rainfall intensities and duration separately; 2) contains a star-shaped information bridge to enhance the information flow across RNN layers; and 3) uses a novel multi-sigmoid loss function to optimize the parameters by taking weather criterion into account. We evaluate our *StarBriNet* on a Radar Echo dataset of east China in comparison with other baseline algorithms, including the state-of-the-art one.

The contributions of this paper are summarized as follows.

1. We propose to use multi-column RNNs to deal with different intensities and duration of the rainfall separately.
2. We propose a star-shaped information bridge to enhance the information flow across RNN layers.
3. We introduce a new multi-sigmoid loss to take the precipitation nowcasting criterion into account.
2 Preliminaries

Compared to video prediction tasks, the precipitation nowcasting task of radar echo data is more difficult and complicated because the ‘rain clouds’ on radar echo images vary in the moving speed, density, and contour.

Problem definition Let $S_{1:T} = [I_1, I_2, \ldots, I_T]$ be a radar echo sequence of length $T$, with each frame $I_t \in \mathbb{R}^{H \times W}$ of size $H \times W$, $\forall t \in \{1, \ldots, T\}$. The goal of precipitation nowcasting is to predict the future radar echo sequence of length $L > 0$ subsequent to current radar echo sequence $1 : T$, which can be formulated as follows:

$$
\hat{S}_{T+1:T+L} = \arg \max_{S_{T+1:T+L}} p(S_{T+1:T+L} | S_{1:T})
$$

where $\hat{S}_{T+1:T+L}$ represents the predicted radar echo sequence of length $L$.

Convolutional LSTMs Significant progress has been made with ConvLSTMs towards video-based tasks like action detection [5], video object detection [6, 7], and video prediction [4, 8] since 2015. Apart from CNN-based methods, researchers also dig into RNN-based methods for video prediction. Convolutional LSTM [1] is an effective one among various RNN methods. Following the long short-term memory model, ConvLSTM utilizes its recurrent neural network architecture to memorize temporal information in a video sequence and extracts the spatial feature maps by using convolutional operation. We utilize a simple version of ConvLSTM, the key equations of ConvLSTM are summarized as follows:

$$
f_t, i_t, o_t = \sigma(W_{f} \odot [x_t, h_{t-1}] + b_{f})
$$
$$
c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{c} \odot [x_t, h_{t-1}] + b_{c})
$$
$$
h_t = o_t \odot \tanh(c_t),
$$

where $\odot$ denotes the convolution operator, $\odot$ denotes the element-wise product, $b$ is bias term of convolution kernel, $c_{t-1}$ is the cell memory of last time-step, $h_{t-1}$ is the output of last time-step, and $\sigma$ stands for the sigmoid function.

3 Methods

In this section, we present our proposed StarBriNet in detail. Subsections 3.1, 3.2, and 3.3 introduce the Multi-Column RNNs, the star-shaped information bridge to enhance the information flow across RNN layers, and a multi-sigmoid loss function to take the precipitation nowcasting criterion into account, respectively. We use an encoder-decoder structure for precipitation nowcasting shown in Fig. 2. Similar to [9], the initial cell memory and
cell output of decoding network are directly copied from the last state of the encoding network. Input sequence is fed into a Resize network consisting of 3 convolutional layers to reduce the size of feature maps. Based on the computed mean and first-order derivative (changing rate) of rain fall intensity, the output of the Resize network passes through one of the three StarBri LSTM encoders that correspond to light, moderate, and heavy rainfall intensities. We call these three RNN encoders as multi-column RNNs. Then, the cell states of StarBri LSTM encoder are transferred to a StarBri LSTM decoder as the initial states. Another Resize network consisting of three deconvolution layers up-samples the predicted frames to its original size. Instead of widely-used loss functions $L_1$ and $L_2$, we introduce a multi-sigmoid loss, which takes precipitation nowcasting criterion into account.

3.1 Multi-Column RNNs

We visualize the rainfall intensity and its changing rate of each sample sequence in Fig. 3. The diversity of data distribution is the key problem of precipitation nowcasting task. Therefore, we employ a divide-and-conquer strategy to address the diverse distribution. To this end, we propose a multi-column RNN structure in our encoder-decoder networks. Each column, which is an RNN encoder in this case, is explicitly designed to process one particular range of rainfall intensity.

As shown in Fig. 2, our proposed architecture consists of three StarBri-LSTM encoders for light, moderate, and heavy rainfall intensities, respectively. We separate the whole precipitation dataset into three classes with respect to the rainfall intensity and its first order derivative along time axis of each sequence sample. For such a data partition, rainfall intensity and changing rate in each class are less divergent. Each encoder consumes one particular intensity level in both training and testing phase to learn a specified circumstances, for example the light rain encoder only processes samples located in the small green box in Fig. 3. Encoded features are transferred to StarBriLSTM decoder as the initial cell state. Furthermore, we also make explorations on 1-encoder 3-decoder and 3-encoder 3-decoder structures. The performances of these variants are slightly lower than the one in Fig. 2. Parameter numbers will be 3 times larger but the floating-point operations are almost the same as the single-
3.2 Star-Shaped Information Bridge

We first utilize the Convolutional LSTM (ConvLSTM) network as our basic building block. Traditional multi-layer LSTMs usually take the last output $\hat{I}_{t-1}$ as the input of the first layer at next time-step. A disadvantage of this strategy is that it may bring accumulative errors to predictions in testing stage, because $\hat{I}_{t-1}$ is different from the ground truth input $I_{t-1}$. But this connection passes more information across time steps which benefits the back-propagation, and this is the main reason why researcher still use it. We further expand feature transfer by introducing a novel Star-Shape information bridge to add more information from the last time-step to make the feature flow in multi-layer ConvLSTM more robust. More specifically, we concatenate output of all ConvLSTM layers and pass it to a $1 \times 1$ convolution layer and split the output of convolution layer to all ConvLSTM layers of next time-step by a residual connection to their inputs. Fig. 4 demonstrate the star-shape information bridge binded to a 2 layer decoder. It is hard to train the star-shape structure because of gradient exploding problem. Therefore, we attach the group normalization after each convolution layer and greatly relieve this hard-training problem. We also add group normalization layer to every convolution layer in ConvLSTM which directly improve our prediction performance by a large margin.

3.3 Multi-Sigmoid Loss

As an important criterion in precipitation nowcasting field, critical success index (CSI) skill score is defined as

$$CSI = \frac{\text{hits}}{\text{misses} + \text{hits} + \text{falsealarms}},$$

where hits denotes the number of True Positives, in other words we predict a rainning area correctly. misses denote the number of False Negatives, which means we fail to forecast a raining area. falsealarms denote the number of False Positive as shown in Table 1.

$$1 - CSI = \frac{\text{misses} + \text{falsealarms}}{\text{misses} + \text{hits} + \text{falsealarms}},$$

Intuitively, it might increase the CSI performance of forecasting by minimizing $1$-CSI directly. However, CSI is not differentiable.
Table 1: The 20dBZ CSI skill Score

| Ground Truth | Prediction |
|--------------|------------|
| ≥ 20dBZ      | False      |
| < 20dBZ      | False      |

Alternatively, we approximate 1-CSI by $L_{SSL}^i$ at classification point $c_i$:

$$L_{SSL}^i = \| \sigma(I - c_i) * s - \sigma(\hat{I} - c_i) * s \|^2_2,$$

where $s$ is the scale factor, a hyper-parameter to control the slope of the sigmoid function, and the subscript SSL is short for single sigmoid loss.\[\| \cdot \|_2\] denotes the Frobenius norm.

$L_{SSL}$ is sensitive to false alarms and misses, less sensitive to others. The multi-sigmoid loss is composed of a set of sigmoid losses \{$L_{SSL}^i\}_{i=1,2,\cdots,n}$, in which $L_{SSL}^i$ is to evaluate if $I$ gives out the correct classification for the classification point $c_i$:

$$L_{MSL} = \sum_{i=1}^{n} L_{SSL}^i,$$

where the subscript MSL is short for multi-sigmoid loss. In this work, we follow the weather forecasters’ recommendation, \{20dBZ, 30dBZ, 40dBZ\} in radar echo scale, as our critical points \{$c_1, c_2, \cdots, c_n$\} for $L_{MSL}$. Scale factor $s$ is 15 based upon experiments.

4 Experiments

4.1 Dataset and Metrics

Dataset. Our radar echo dataset contains 170,000 weather radar intensity frames collected from October 2015 to July 2018 with the minimal interval of 6 minutes by the dual polarization Weather Surveillance Radar-1988 Doppler Radar (WSR-88D) located in Shanghai. Each frame is a 501 $\times$ 501 grid of image, covering almost 501 $\times$ 501 square kilometers. We normalize the echo intensity values $R (0 \leq R \leq 70)$ into a gray scale $P$ by $P = R/70$.

Our dataset is available on Harvard Dataverse [10].

We applied a sliding window of stride 1 to generate 76,779 consecutive sequences of length 20. The training set (October 2015 to July 2017) and testing set (October 2017 to July 2018) contain 44,060 and 32,719 sequences, respectively. As a pre-processing step, we removed the rain-less frames from the training set to effectively train the network, while we kept the testing data unchanged. To fairly compare with other methods, we resized all the frames from 501 $\times$ 501 to 100 $\times$ 100.

Metrics. We report frame-wise mean square error (MSE) and CSI score for each experiment setting. Frame-wise MSE are calculated as follows,

$$MSE = \frac{1}{L} \sum_{t=1}^{L} \| I_t - \hat{I}_t \|^2_2,$$

where $t$ is the index of radar echo frame and $L$ is the length of predicted frames. Another metric is the weather criterion, CSI, which has been defined in Eq. (3). The critical point in CSI was set to be 20dBZ.
4.2 Implementation Details

The lengths of both the input and output sequence are 10. All our experiments are implemented with PyTorch and conducted on 4 NVIDIA GTX 1080Ti GPUs. We trained our model with ADMM optimizer and the learning rate is 0.002. The size of mini-batch is 32 and we stop the training after 20,000 iterations. We use two stride-2 2-D convolution layers to resize the input frames from $100 \times 100 \times 1$ to $25 \times 25 \times 64$. The kernel size is (4,4). The number of channels of Group-Normalization are 16. Same parameters are adopted for the transposed convolution layers. We use a 2-layer encoding-forecasting structure with the number of filters set to 64, 64. The kernel size of StarBri LSTM is (3,3).

4.3 Experimental Results

The overall evaluation results are summarized in Table 2. We perform a detailed ablation study on multi-column encoder, star-shaped information bridge, and multi-sigmoid loss. We also evaluate 5 other nowcasting algorithms, including 1 optical flow based methods (ROVER), 4 deep learning methods (ConvLSTM, PredRNN, TrajGRU, and 3D CNN). We implement a naive 8 layer 3D CNN, the kernel size is (3,3,3). Both CSI (the higher the better) and MSE (the lower the better) increases when we adopt multi sigmoid loss, because the multi sigmoid loss focus on misses and false-alarms, those are more correlated to CSI score. Overall, the StarBriNet model outperforms the other methods. We visualize the generated radar maps in Fig. 7. We can see that all methods generate the blurriness predictions. Because we treat precipitation nowcasting as a deterministic task. Unfortunately, this task are full of uncertainty. Our method treat this sample as a moderate one and generate more rainy area than others.

4.4 Model Analysis

Scale factors of $\mathcal{L}^{\mathcal{SSL}}$. We tested several scale factors in Fig. 6 and Fig. 5. Larger scale factor leads the gradient sensitive to tiny differences at the starting phase of training which influence the training robustness. To test our hypothesis, we run an experiment of gradually increasing the scale factor from 1 to 40, with CSI score 64.27%, MSE 6.17.
Figure 6: CSI performance of different scale factors of $L_{MSL}$. X-axis stands for the number of training iteration.

**Resizing.** Shi et. al. [1] first utilize a patch-resizing trick to transfer an input frame from $100 \times 100 \times 1$ to $50 \times 50 \times 4$ when patch-size is 2. This trick cut image into small patches and stack pixels inside each patch along channel-wise directly. The GPUs memory consumption and training time are greatly reduced during training. Wang et. al. [?,4] also adopt this trick. However, combined with convolution layer in ConvLSTM layers, patch-resizing trick is equivalent to a big kernel size convolution. So we recommend to use alternate down-sampling and up-sampling strategies instead, e.g., the resize network in Fig. 2.

**Data splitting strategy.** Fig. 3 demonstrates the distribution of the whole dataset, where each point in this figure represents for a length-10 sample sequence of radar echo frames. Y-axis is the changing rate of rain intensities in each sample sequence. Positive value means the rain increased and negative means the rain is letting up. X-axis is the average rainfall intensity of each sample. Most samples lies in the green colored box, we category this as the light rain cut, blue for the moderate cut and the purple area for the heavy rain cut. This is a simple yet effective division.

Table 2: Comparison results with other models on East-China radar echo dataset. MC is Multi-Column structure in section 3.1 and MSL is multi sigmoid loss in section 3.3.

| Model           | MSE | CSI [%] |
|-----------------|-----|---------|
| ROVER           | 8.84| 58.8    |
| 3D CNN          | 6.33| 58.4    |
| ConvLSTM [1]    | 6.31| 60.0    |
| TrajGRU [2]     | -   | 63.0    |
| PredRNN [3]     | 7.03| 63.1    |
| ConvLSTM+MC     | 6.51| 62.1    |
| StarBri LSTM+MC | **6.12** | 63.8    |
| StarBri LSTM+MC+MSL | 6.18 | **64.4** |
5 Conclusion

We presented multi-column Star shape information Bridge Network (StrBriNet) for precipitation nowcasting task. A video prediction model that effectively passes feature and gradient through RNN layers. We proposed a new multi-sigmoid loss function which takes CSI score into consideration, and yield better prediction. We performed quantitative analysis on a Radar Echo dataset of east China. The results demonstrated that, compared with other method, our model outperform other state-of-the-art methods.

References

[1] Xingjian Shi, Zhourong Chen, Hao Wang, Dit Yan Yeung, Waikin Wong, and Wangchun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In Proceedings of the International Conference on Neural Information Processing Systems, pages 802–810, 2015.

[2] Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In Advances in neural information processing systems, pages 5617–5627, 2017.

[3] Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and S Yu Philip. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In Advances in Neural Information Processing Systems, pages 879–888, 2017.

[4] Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, and S Yu Philip. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In International Conference on Machine Learning, pages 5110–5119, 2018.

[5] Lin Song, Shiwei Zhang, Gang Yu, and Hongbin Sun. Tacnet: Transition-aware context network for spatiotemporal action detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[6] Hongmei Song, Wenguan Wang, Sanyuan Zhao, Jianbing Shen, and Kin-Man Lam. Pyramid dilated deeper convlstm for video salient object detection. In Proceedings of the European Conference on Computer Vision (ECCV), pages 715–731, 2018.

[7] Guanbin Li, Yuan Xie, Tianhao Wei, Keze Wang, and Liang Lin. Flow guided recurrent neural encoder for video salient object detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
[8] Yunbo Wang, Lu Jiang, Ming-Hsuan Yang, Li-Jia Li, Mingsheng Long, and Li Fei-Fei. Eidetic 3d lstm: A model for video prediction and beyond. 2018.

[9] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In International conference on machine learning, pages 843–852, 2015.

[10] Chen Lei. A Deep Learning Based Methodology for Precipitation Nowcasting with Radar, 2019.