HYBRID WEIGHTING LOSS FOR PRECIPITATION NOWCASTING FROM RADAR IMAGES

Yuan Cao[1] *Lei Chen*[2] *Danchen Zhang*[3] *Leiming Ma*[2] *Hongming Shan*[4*]

[1] School of Computer Science, Fudan University
[2]Shanghai Central Meteorological Observatory [3]University of Pittsburgh
[4]Institute of Science and Technology for Brain-inspired Intelligence, Fudan University

ABSTRACT

Precipitation nowcasting is gaining increasing attention in the signal processing community. Existing deep learning-based studies focus on designing an effective model architecture, neglecting the influence of the severe imbalanced distribution of rainfall data that can compromise the predictive accuracy on heavy rainfall intensities. To address the uneven distribution of precipitation nowcasting data, we propose a novel data reweighting strategy, termed Hybrid Weighting, which hybrids reweighting and non-weighting strategies together, boosting the precipitation nowcasting performance. Experimental results on two natural radar echo benchmark datasets demonstrate the superior performance of our proposed approach for precipitation nowcasting over existing loss functions on high rainfall intensities, without degenerating on low rainfall intensities compared with state-of-art reweighting methods.

***Index Terms*—** Precipitation nowcasting, reweighting, data imbalance, weather radar echo extrapolation

1. INTRODUCTION

Nowcasting precipitation is an important task for weather nowcasting and has received increasing attention in the computer vision and signal processing communities. With the vast amount of radar echo data provided by the contemporary operational weather radar networks, many deep learning-based methods have been developed for precipitation nowcasting in recent years [1]. For example, ConvLSTM [2] and PredRNN++ [3] are the most well-known precipitation nowcasting models, which have demonstrated superior prediction performance than optical flow based models [4] and numerical weather prediction based methods [5, 6].

Though these models can often deliver promising nowcasting performance on low to medium rainfalls, they show unsatisfying forecasts on heavy and extreme rain rates. The reason is that rainfall data usually have long-tailed distribution, overwhelmed by rain-less and drizzle data, as shown in Fig. 1. Such severe imbalanced distribution of rainfall data challenges the predictive accuracy of current models. We also observe that several widely-used loss functions such as cross-entropy [2, 7, 8, 9], mean absolute error (MAE), and mean squared error (MSE) [10, 3] do not take the imbalanced distribution of rainfall data into account as they usually assume the data are uniformly distributed. However, few studies dive into the data imbalance in precipitation nowcasting.

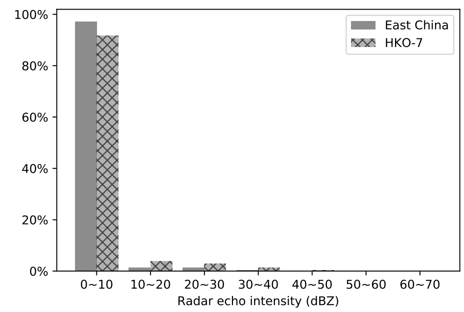

Fig. 1. Distribution of different radar echo intensity levels. East China is collected from the north temperate zone, and HKO-7 is collected from the tropics.

To mitigate this problem, we conduct experiments on several state-of-art data reweighting strategies, i.e., Focal loss [11], Weighted Mean-Square Error (WMSE) [12, 13], Distributional Ranking (DR) loss [14]. These reweighting strategies share the same idea of reducing majority sample weights to boost minority sample weight. This leads to boosting the performance on heavy rainfalls but compromised performance on low rainfall intensities.

In this work, we propose a novel loss function, Hybrid Weighting, to handle the imbalance problem in precipitation data. We combine the reweighting and non-weighting loss functions together, aiming to balance highlighting minority and sacrificing majority. In addition, we propose a new reweighting loss function, where we force the total weight factors of each class to be equivalent and use sample/pixel

*Corresponding Author: hmshan@fudan.edu.cn
This work was supported in part by National Key Research and Development Program of China (No. 2018YFB1305104), the Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01) and ZJLab, the Shanghai Center for Brain Science and Brain-inspired Technology, National Natural Science Foundation of China (No. 42105001, 41975069), and Shanghai Municipal Science and Technology Project (No. 18DZ1200404).

level weight to handle the imbalance.

Experimental results of three models on two real-world datasets show that the proposed loss function dramatically improves the performance compared to the unweighted ones and reweighting strategies without downgrading the performance on low rainfall intensities.

2. RELATED WORK

2.1. Deep learning for precipitation nowcasting

Precipitation nowcasting models are usually categorized into three classes: numerical weather prediction methods, extrapolation-based methods, and deep learning-based methods [15]. In this paper, we focus on the last one due to its superior performance.

Regarding precipitation nowcasting as a spatiotemporal sequence forecasting problem, in 2015, Shi *et al.* first proposed Convolutional Long Short-Term Memory (ConvLSTM) [2] to directly predict the future rainfall intensities based on the past radar echo maps. Subsequently, various ConvLSTM-like variants have been proposed for precipitation forecasting. Wang *et al.* added a zigzag memory connection across different LSTM layers [3]. Shi *et al.* utilized optical flow warping to refine the convolutional region of ConvLSTM [12]. In addition, Ayzel *et al.* adopt U-Net [16], and Tian *et al.* incorporated adversarial loss to generate more realistic predictions [17].

Meanwhile, many studies proposed to incorporate various extra information to help nowcasting. For instance, Franch *et al.* incorporated orographic features to improve the performance of weather nowcasting [13]. Lebedev *et al.* utilized satellite imagery as well as the output of Global Forecast System [7]. Li *et al.* used automatic weather stations data and satellite data [18]. However, most studies assume that data is subject to a uniform distribution, contradicting the highly imbalanced data distribution shown in Fig. 1.

2.2. Reweighting methods for imbalanced distribution

Resampling and reweighting are two common strategies for addressing the data imbalance problem. The former oversamples the minority classes or undersamples the majority classes on training data. However, radar echo sequences can only be sampled at frame or sequence level, but imbalance also happens at the pixel level, which is unsuitable for solving by resampling. Thus, we focus on reweighting methods.

To the best of our knowledge, Ronneberger *et al.* utilized Weighted Cross-Entropy (WCE) for class imbalance [19]. Franch *et al.* adopted WMSE to re-balance between classes [13]. Both of them share the same idea that class weight is negatively correlated to the proportion of each class. Besides the inter-class imbalance, focal loss further considers intra-class imbalance [11]. In addition to the classes' weight, focal loss heuristically down-scales the weights of well-trained samples. Compared to focal loss, DR loss [14] reformulates the classification task as a ranking problem and proposes a refined inter- and intra-class weight assignment strategy that further improves the learning performance. In this study, we will use these imbalance solutions as baselines.

3. METHOD

3.1. Problem Statement

Given a sequence of k radar echo images, $\mathcal{S}_{1:k}$, collected in the past one hour, the goal of precipitation nowcasting is to predict a sequence of next m radar echo images, $\mathcal{S}_{k+1:k+m}$, which can be formulated as

$$\widehat{\mathcal{S}}_{k+1:k+m} = \underset{\mathcal{S}_{k+1:k+m}}{\arg\max}\, p(\mathcal{S}_{k+1:k+m}|\mathcal{S}_{1:k}), \tag{1}$$

where $\widehat{\mathcal{S}}_{k+1:k+m}$ represents the predicted radar echo sequence of length m. In this study, $k = m = 10$. For simplicity, we remove the subscript.

To facilitate the reweighting strategy, following [8], we convert the precipitation nowcasting task from the original regression task to a classification problem. Specifically, we map the original radar echo range, $0-70$ dBZ, into 7 classes, each of which corresponds to a 10 dBZ interval, as shown in Fig. 1. Among them, the first two classes covering $0-20$ are rainless; class $20-30$ is drizzle; class $30-40$ is moderate rain; the last three classes, $40-70$, are heavy and extreme rains. Notably, we focus on predicting heavy and extreme rainfall in the last three classes, as they have much less data than the first four classes.

As above mentioned, nowcasting model predicts n classes, where $n = 7$ as shown in Fig. 1, and each class is indexed by t, where $t \in \{1, \dots, n\}$. For each pixel i, we predict its class with a vector $\boldsymbol{P}_i$ of size n, each of which corresponds to a class. Prediction score on class t of pixel i is represented as $p_{i,t} \in \boldsymbol{P}_i$, ranging from 0 to 1. To solve imbalance problem, we expect to learn a weight for each pixel i, which is denoted as w_i, ranging from 0 to 1. Further, we constrain the pixels' weight set $\boldsymbol{W}_t$ satisfying $\boldsymbol{W}_t = \{w_i | \sum_{i\in\mathcal{S}^t} w_i = 1, \forall\, i, w_i \geq 0\}$.

3.2. Hybrid Weighting (HW)

As discussed in the above section, the state-of-the-art reweighting methods boost the minority samples' weight and sacrifice the weight of majority samples to some degree. This study hopes to propose a hybrid method that can balance between sacrificing the majority and highlighting the minority. Intuitively, we use a hyperparameter α to combine the non-weighting loss $\mathcal{L}_{\text{N}}$ and reweighting loss $\mathcal{L}_{\text{R}}$:

$$\mathcal{L}_{\text{hybrid}} = \mathcal{L}_{\text{N}} + \alpha\mathcal{L}_{\text{R}}. \tag{2}$$

The state-of-the-art deep learning based precipitation nowcasting models [2, 3] usually employ the following non-weighting loss function: $\sum_{t=1}^{n}\sum_{i\in\mathcal{S}^t}\ell(p_{i,t})$, where function ℓ measures the distance between predicted $p_{i,t}$ and ground-truth score 1; in this study we use $\ell(p_{i,t}) = 1 - p_{i,t}$.

Reweighting methods usually assign each pixel i with a weight w_i, and we could obtain a weighted version of loss function for t-th class:

$$\mathcal{L}_t = \sum_{i\in\mathcal{S}^t} w_i \ell(p_{i,t}). \tag{3}$$

In this study, we force the total weights for each class to be equivalent to 1. That is to say, the weights of classes with more pixels are generally smaller than that of another class with fewer pixels. In addition, our proposed reweighting strategy aims to identify the most reasonable pixel weight assignment, where higher loss samples should receive higher weights than lower loss samples. In following iterations, the prediction could be adjusted based on the weights. The weight $\boldsymbol{W}_t$ that gives maximum $\mathcal{L}_t$ should be the solution weight assignment. We let $\mathcal{L}_t^*$ be the optimal solution:

$$\mathcal{L}_t^* = \max_{\boldsymbol{W}_t} \mathcal{L}_t. \tag{4}$$

Note that trivial solutions will be derived when optimizing Eq. (4). For example, a worst-case scenario could be that the highest loss pixel has a weight equal to 1 while the other pixels' weights are 0. This makes the model sensitive to the outlier pixels and leads to compromised performance. Hence, we constrained the weight distribution $\boldsymbol{W}_t$ close to the uniform distribution $\boldsymbol{U}_t$. Then, Eq. (4) could be represented as:

$$\mathcal{L}_t^* = \min_{\boldsymbol{W}_t} - \sum_{i\in\mathcal{S}^t} w_i\ell(p_{i,t}) \text{ s.t. } w_i \in \boldsymbol{W}_t, D_{\text{KL}}[\boldsymbol{W}_t \| \boldsymbol{U}_t] \le \epsilon$$

Here, $D_{\text{KL}}[\cdot\|\cdot]$ is the Kullback-Leibler divergence operator constraining the 'distance' between $\boldsymbol{W}_t$ and $\boldsymbol{U}_t$ by a margin ϵ. According to Lagrangian duality theory [20], we can obtain the optimal constrained distribution $\boldsymbol{W}_t$ by solving the problem

$$\min_{\boldsymbol{W}_t} \sum_{i\in\mathcal{S}^t} -w_i\ell(p_{i,t}) + \lambda D_{\text{KL}}[\boldsymbol{W}_t\|\boldsymbol{U}_t] + \mu \sum_{i\in\mathcal{S}^t} w_i - 1, \tag{5}$$

where λ and μ are Lagrangian multipliers, $\lambda \ge 0$ and $\mu \ge 0$. By Karush-Kuhn-Tucker condition [20], we have an optimal solution of $\boldsymbol{W}_t$ as follows:

$$w_i = \frac{\frac{1}{|\mathcal{S}^t|}\, e^{\ell(p_{i,t})/\lambda}}{\sum_i \frac{1}{|\mathcal{S}^t|} e^{\ell(p_{i,t})/\lambda}} = \textbf{softmax}\left(\frac{\ell(p_{i,t})}{\lambda}\right), \tag{6}$$

where $|\mathcal{S}^t|$ is the pixel count in class t. Then, we could get

$$\mathcal{L}_t^* = \sum_{i\in\mathcal{S}^t} \textbf{softmax}\left(\frac{\ell(p_{i,t})}{\lambda}\right)\ell(p_{i,t}). \tag{7}$$

We adopt the conventional MSE loss function as the non-weighting loss. Thus, the objective function is

$$\mathcal{L}_{\text{hybrid}} = \mathcal{L}_{\text{MSE}} + \alpha \sum_{t=1}^{n} \mathcal{L}_t^*. \tag{8}$$

4. EXPERIMENT

4.1. Experimental Setup

Implementation Details The lengths of both the input and output sequence are 10. We trained our model with an ADAM optimizer, and the learning rate is 0.002. Mini-batch size is 32, and the maximum of training iterations is $20,000$. Layer Normalization is utilized in ConvLSTM cells layers in all experiments. The hidden size is 32 for all ConvLSTM cells of both models. The size of convolution filters is 3×3. All our experiments were implemented with PyTorch and conducted on 1 NVIDIA Tesla A100 GPU.

Datasets We evaluate the performance of our proposed losses on two public radar echo datasets: East China Precipitation dataset (ECP) [15] and HKO-7 dataset [12]. The first dataset contains around 170,000 weather radar intensity frames collected from October 2015 to July 2018, with an interval of 6 minutes. Each frame, covering 501km$\times$501km area centered in Shanghai. At pre-processing step, we removed the rain-less sequences (less than 5% raining area). The partition ratio of the training set, the validation and the testing set is $7:1:2$. The second dataset contains around 238,000 radar echo frames from 2009 to 2015 collected by Hong Kong Observatory. The radar reflectivity images cover a 512km$\times$512km area centered in Hong Kong. The data are recorded every 6 minutes. We use the same training, validation and testing partition setting following [12]. Images are resized to 120×120 for both dataset.

Base Models We carry out experiment with three open-sourced state-of-art models: ConvLSTM, PredRNN++, RainNet [2, 3, 16]. The first model is the most popular precipitation nowcasting model, which combines Long Short Term Memory (LSTM) and convolutional operation. They replace the dot operator in LSTM with convolution operation, which enables the LSTM to extract local features of images. The second model proposes a spatiotemporal LSTM to enhance the spatial memory transition and introduces an additional zig-zag link to transport spatial memory. RainNet is a UNet-based non-RNN model which only yields one frame each time.

Loss Functions Cross-entropy (CE) and MAE+MSE are the most basic and popular loss functions in precipitation nowcasting field [2, 3]. In WMSE, the class weight tuning range is $[1, 30]$, following suggestion in paper [13]. We adopt same hyper parameters settings as original papers for Focal loss [11] and DR loss [14].

Evaluation Metrics We evaluate the model performance on weather-oriented criterion, Critical Success Index (CSI),

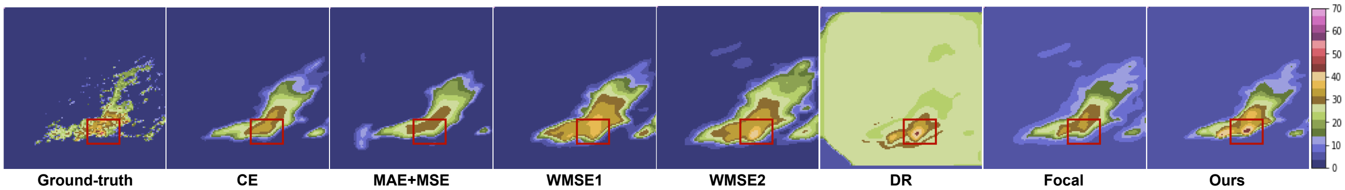

Fig. 2. Predictions of different losses in 60 minutes on ECP dataset

which can be expressed as $\mathrm{CSI}_z = \frac{\mathrm{TP}}{\mathrm{TP+FN+FP}}$, where z (dBZ) is the threshold and TP, FN, FP are True Positive, False Negative, and False Positive respectively. In this case, CSI_{20} can demonstrate the model's performance in distinguishing rainfall intensities between rain-less and rain, and CSI_{30} is for drizzle versus moderate. CSI_{40} is disposed to the performance on the heavy or extreme scenario. Chen *et al.* [15] only reported nowcasting performance of $\mathrm{CSI}_{20}, \mathrm{CSI}_{30}$. To demonstrate the heavy rain detection performance, we particularly present results on CSI_{40}. Besides, we average all three CSIs for comparison.

Table 1. Results of CSIs on two benchmark datasets

Loss	East China Precipitation				HKO-7				
	Avg	20dBZ	30dBZ	40dBZ	Avg	20dBZ	30dBZ	40dBZ	
CE	0.333	0.592	0.408	0.000	0.291	0.532	0.340	0.000	ConvLSTM
MAE+MSE	0.343	0.611	0.418	0.000	0.311	0.548	0.385	0.000	
WMSE	0.335	0.601	0.405	0.000	0.348	0.488	0.348	**0.208**	
DR	0.371	0.562	0.394	0.156	0.341	0.518	0.357	0.148	
Focal	0.387	0.605	0.416	0.138	0.356	0.512	0.390	0.166	
Ours	**0.412**	**0.623**	**0.426**	**0.187**	**0.381**	**0.555**	**0.397**	0.190	
CE	0.341	0.605	0.418	0.000	0.316	0.556	0.392	0.000	PredRNN++
MAE+MSE	0.346	0.611	0.425	0.000	0.343	**0.565**	0.397	0.068	
WMSE	0.342	0.602	0.425	0.000	0.344	0.498	0.369	0.163	
DR	0.314	0.494	0.339	0.108	0.346	0.515	0.369	0.155	
Focal	**0.417**	0.616	**0.447**	**0.188**	0.371	0.560	0.406	0.148	
Ours	0.411	**0.623**	0.433	0.176	**0.402**	0.564	**0.428**	**0.213**	
CE	0.322	**0.592**	0.373	0.000	0.334	0.535	0.379	0.088	RainNet
MSE+MAE	0.334	0.588	0.392	0.020	0.349	**0.555**	0.397	0.094	
WMSE	0.318	0.503	0.253	0.156	0.292	0.419	0.267	0.190	
DR	0.318	0.370	0.404	**0.179**	0.359	0.529	0.380	0.169	
Focal	0.380	0.588	0.395	0.156	0.374	0.543	0.392	0.186	
Ours	**0.391**	0.587	**0.410**	0.175	**0.380**	0.547	**0.399**	**0.193**	

4.2. Main Results

In this study, we tried 7 loss functions with two precipitation nowcasting models on two datasets, as shown in Table 1. The CSIs are mean scores of 10 predicted images.

First, with CSI_{40} being or approaching 0, both non-weighting methods, CE and MAE+MSE, fail to forecast heavy rainfall intensities under the influence of the overwhelming low dBZ data. Second, inter-class rebalanced methods, WMSEs and DR, show inferior performance on CSI_{20} and CSI_{30} and better performance on CSI_{40}. This indicates that inter-class reweighting strategy may sacrifice performance on majority class to obtain prediction accuracy improvement on minority class. Please note that CSI_{40} of WMSE on HKO-7 are better than that on ECP is due to that HKO-7 has more heavy and extreme rainfall data than ECP. Besides, as shown in Fig. 2, DR loss also performs worse on low rainfall intensities, and we can infer DR fails to fit majority classes on ECP. Third, Focal loss performs better than all other losses, except ours. This demonstrates the effectiveness of its intra-class reweighting strategy. Finally, our proposed HW loss provides a clear improvement on CSI_{40} without degenerating results on CSI_{20} and CSI_{30}. Only DR and our proposed HW managed to predict pixels above 40 dBZ (colored red) in Fig. 2.

Table 2. Hybrid factor α on ConvLSTM on ECP

α	CSI			
	Avg	20dBZ	30dBZ	40dBZ
1	0.330	0.519	0.337	0.132
0.1	0.360	0.582	0.381	0.118
0.01	**0.412**	0.623	**0.426**	**0.187**
0.001	0.394	0.627	0.420	0.136
0.0001	0.391	0.625	0.420	0.128
0.00001	0.369	**0.628**	0.416	0.064

4.3. Effects of hyperparameters

To find a good balance for our hybrid method, different α are tested, and results are demonstrated in Table 2. Best performance is obtained with $\alpha = 0.01$, and we adopt it in all the above experiments. In the training phase, we find $\mathcal{L}_{\mathrm{R}}$ is often roughly 100 times bigger than $\mathcal{L}_{\mathrm{MSE}}$, and $\alpha = 0.01$ makes both values in the same range. Besides, model is relatively insensitive to hyper parameter λ in Eq. (7) while $\lambda \in [0.5, 20]$. We only report results with the best performed setting $\lambda = 6$.

5. CONCLUSION

In this paper, we find that the rainfall intensity imbalance problem diminishes the performances of precipitation nowcasting models. To mitigate this influence, we hybrid non-weighting loss and a new proposed reweighting loss together. Experimental results demonstrate that our proposed loss function improves the performance of minority rainfall classes without degenerating the majority classes. Next, we will target extreme rainfall events nowcasting above 50 dBZ.

6. REFERENCES

[1] Zhihan Gao, Xingjian Shi, Hao Wang, Dit-Yan Yeung, Wangchun Woo, and Waikin Wong, "Deep learning and the weather forecasting problem: Precipitation nowcasting," *Deep learning for the Earth Sciences: With Applications and R, Second Edition*, pp. 218–239, 2021.

[2] 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*, 2015, pp. 802–810.

[3] 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*, 2018, pp. 5110–5119.

[4] Wangchun Woo and Waikin Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 3, pp. 48, 2017.

[5] Juanzhen Sun, Ming Xue, James W Wilson, Isztar Zawadzki, Sue P Ballard, Jeanette Onvlee-Hooimeyer, Paul Joe, Dale M Barker, Ping-Wah Li, Brian Golding, et al., "Use of nwp for nowcasting convective precipitation: Recent progress and challenges," *Bulletin of the American Meteorological Society*, vol. 95, no. 3, pp. 409–426, 2014.

[6] Morris L Weisman, Christopher Davis, Wei Wang, Kevin W Manning, and Joseph B Klemp, "Experiences with 0–36-h explicit convective forecasts with the wrf-arw model," *Weather and Forecasting*, vol. 23, no. 3, pp. 407–437, 2008.

[7] Vadim Lebedev, Vladimir Ivashkin, Irina Rudenko, Alexander Ganshin, Alexander Molchanov, Sergey Ovcharenko, Ruslan Grokhovetskiy, Ivan Bushmarinov, and Dmitry Solomentsev, "Precipitation nowcasting with satellite imagery," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2680–2688.

[8] Shreya Agrawal, Luke Barrington, Carla Bromberg, John Burge, Cenk Gazen, and Jason Hickey, "Machine learning for precipitation nowcasting from radar images," *arXiv preprint arXiv:1912.12132*, 2019.

[9] Roope Tervo, Joona Karjalainen, and Alexander Jung, "Short-term prediction of electricity outages caused by convective storms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 8618–8626, 2019.

[10] Benjamin Klein, Lior Wolf, and Yehuda Afek, "A dynamic convolutional layer for short range weather prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[12] Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Waikin Wong, and Wangchun Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Advances in Neural Information Processing Systems*, 2017, pp. 5617–5627.

[13] Gabriele Franch, Daniele Nerini, Marta Pendesini, Luca Coviello, Giuseppe Jurman, and Cesare Furlanello, "Precipitation nowcasting with orographic enhanced stacked generalization: Improving deep learning predictions on extreme events," *Atmosphere*, vol. 11, no. 3, pp. 267, 2020.

[14] Qi Qian, Lei Chen, Hao Li, and Rong Jin, "Dr loss: Improving object detection by distributional ranking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12164–12172.

[15] Lei Chen, Yuan Cao, Leiming Ma, and Junping Zhang, "A deep learning-based methodology for precipitation nowcasting with radar," *Earth and Space Science*, vol. 7, no. 2, pp. e2019EA000812, 2020.

[16] Georgy Ayzel, Tobias Scheffer, and Maik Heistermann, "RainNet v1. 0: a convolutional neural network for radar-based precipitation nowcasting," *Geoscientific Model Development*, vol. 13, no. 6, pp. 2631–2644, 2020.

[17] Lin Tian, Xutao Li, Yunming Ye, Pengfei Xie, and Yan Li, "A generative adversarial gated recurrent unit model for precipitation nowcasting," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 4, pp. 601–605, 2019.

[18] Dawei Li, Yudi Liu, and Chaohui Chen, "Msdm v1. 0: A machine learning model for precipitation nowcasting over eastern china using multisource data," *Geoscientific Model Development*, vol. 14, no. 6, pp. 4019–4034, 2021.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[20] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.