# An Improved Deep Learning Model for High-Impact Weather Nowcasting

Shun Yao , Haonan Chen , *Senior Member, IEEE*, Elizabeth J. Thompson, and Robert Cifelli

*Abstract*—Accurate nowcasting (short-term prediction, 0–6 h) of high-impact weather, such as landfalling hurricanes and extreme convective precipitation, plays a critical role in natural disaster monitoring and mitigation. A number of nowcasting approaches have been developed in the past few decades, such as optical flow and the tracking radar echoes by correlation system. Most of these mainstream operational techniques are based on radar echo map extrapolation, which determines the velocity and direction of precipitation systems using historical and current radar observations. However, the skill of the traditional extrapolation method decreases rapidly within the first hour. In order to improve nowcasting skill, recent studies have proposed using deep learning methods, such as convolutional recurrent neural network and trajectory gate recurrent unit. But none of these methods focuses on high-impact weather events, and the deep learning models trained based on general precipitation events cannot meet the demand of accurate warnings and decision-making at the scales required for high-impact weather events, such as hurricanes. Using multiradar observations, this article introduces the idea of self-attention and develops a self-attention-based gate recurrent unit (SaGRU) to enhance its generalization capability and scalability in predicting high-impact weather events. In particular, two types of high-impact weather systems, namely, landfalling hurricanes and extreme convective precipitation events, are investigated. Three models are trained based on hurricane events, heavy rainfall (i.e., nonhurricane) events, and all events combined in the southeast United States during 2015 and 2020. The impacts of different data sources on the nowcasting performance are quantified. The evaluation results of nowcasting products show that our SaGRU performs very well in predicting hurricane-induced rainfall. In the new methodology, the data from nonhurricane events are shown to provide useful information in enhancing the nowcasting performance during hurricane events as the model trained by combining all the hurricane and nonhurricane events has the best performance. In addition, this article quantifies the impact of the sequence length of input radar observations on the nowcasting performance, which shows that five consecutive observations are sufficient to obtain a stable model, and even two consecutive observations can produce reasonable results.

*Index Terms*—Deep learning, high-impact weather, precipitation nowcasting, weather radar.

Shun Yao and Haonan Chen are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: s.yao@colostate.edu; haonan.chen@colostate.edu).

Elizabeth J. Thompson and Robert Cifelli are with the NOAA Physical Sciences Laboratory, Boulder, CO 80305 USA (e-mail: elizabeth.thompson@noaa.gov; rob.cifelli@noaa.gov).

## I. INTRODUCTION

AS ONE of the most typical high-impact weather phenomena, hurricane refers to tropical cyclones with maximum sustained surface winds reaching 74 miles/h [1], which often produces severe/serious hazards, such as storm surge, floods, strong winds, and hurricane-spawned tornadoes. Unfortunately, the risk of extensive damage and loss of life caused by hurricanes is increasing due to the growth of population, changing climate, and urbanization [2]. For instance, hurricane Harvey during August 25 and September 4, 2017 impacted 13 million people with over 100 fatalities. About 135 000 homes were damaged or destroyed, and the total damage was $125 billion [3]. In less than a week, the storm poured a year of rainfall over Houston and most of southeastern Texas. Two flood-control reservoirs had burst, causing water levels to rise throughout the Houston area. Therefore, the accurate nowcasting of hurricane intensity and subsequent rainfall is critical in high-impact weather studies and operational applications of weather radar and/or satellite observations.

Conventionally, the operational precipitation nowcasting strategies based on radar measurements attempt to predict future radar echo maps through leveraging extrapolation methods, which can be roughly classified into three categories: centroid tracking methods, tracking radar echoes by correlation (TREC), and optical flow [4], [5], [6]. The centroid tracking algorithms detect isolated storms at the current moment and try to link each storm across two successive time steps, then forecast storm progress using the centroid of the identified storm. As the storm was condensed into a centroid cell, it is easier for tracking and predicting massive and strong echoes. The other advantage of the centroid-type method is that it can provide physical information about each storm, such as storm area, top, and volume [7]. However, when the echoes are fused or split, the nowcasting accuracy decreases rapidly. In contrast, TREC estimates a motion field based on correlation analysis. It subdivides a radar image into numerous square boxes of equal size. Each box can be represented as a two-dimensional array containing reflectivity intensity values. Then the correlations between corresponding boxes at two consecutive radar images are calculated. The motion vector is calculated as the space shift that results in the highest correlation coefficient. Finally, these motion vectors can be used for nowcasting. The TREC algorithm involves calculating the spatial optimal correlation coefficients for two adjacent moments and then establishing a fitting relationship for all radar echoes. This method can effectively track stratiform rainfall systems, but the performance
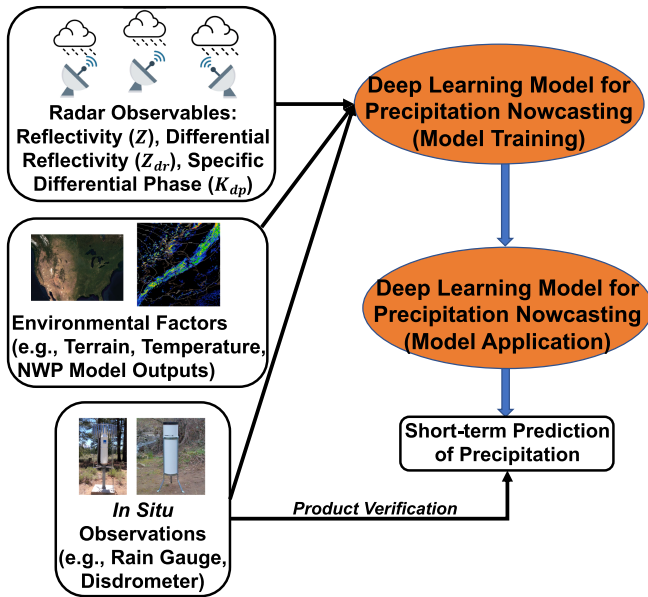
Fig. 1.   General concept of a deep learning precipitation nowcasting system. The input of the system can be radar reflectivity ($Z$), differential reflectivity ($Z_{dr}$), specific differential phase ($K_{dp}$), rain gauge data and/or environmental factors, such as terrain and NWP model outputs. The output is the prediction of precipitation.

is much lower in predicting strong convective processes with fast-evolving echoes. Similarly, the optical flow approaches estimate a motion field (optical flow), but in a different way. Based on the principle of image pixel intensity conservation, the optical flow method assumes that the reflectivity intensity remains unchanged in the two adjacent frames. Essentially, it calculates the motion information of the reflectivity between adjacent frames by using the pixel change in the image sequence and the correlation between adjacent frames to discover the relationship of the previous frame and the current frame. Then, future echo maps can be extrapolated using semi-Lagrangian advection after the optical flow has been achieved. A major disadvantage of optical flow is that it cannot predict the initiation, growth, and decay of the storms.

For high-impact weather events, such as severe convective rain and hurricanes, it is difficult to use these conventional approaches to produce reliable nowcasting products since the inherent complexity of the changing atmospheric state and non-linear cloud dynamics is high during such events and the assumption of stationarity between frames in the abovementioned methods is not valid [8], [9]. With the great success of deep learning techniques in a variety of fields, including geosciences and remote sensing research (e.g., [10], [11], [12], [13], [14], [15]), recent studies have proposed using this machine learning approach to tackle the precipitation nowcasting problem since the nonlinearity of machine learning can better model the spatiotemporal variability of precipitation [16], [17], [18], [19], [20]. As shown in Fig. 1, deep learning precipitation nowcasting can be performed based on polarimetric weather radar measurements, i.e., radar reflectivity ($Z$), differential reflectivity ($Z_{dr}$), and specific differential phase ($K_{dp}$). In situ measurements, as well as environmental factors, such as terrain feature,

temperature, and numerical weather prediction (NWP) model outputs, can also be incorporated into the deep learning-based nowcasting frameworks.

To date, most of the deep learning nowcasting models rely on recurrent neural networks (RNN) since the radar echo extrapolation can be viewed as a sequence-to-sequence problem. A typical example is the convolutional long short-term memory (ConvLSTM) model developed by Shi et al. [17], which modeled precipitation nowcasting as a spatiotemporal serial prediction issue that can be solved with the sequence-to-sequence learning framework. However, training a practical model is difficult because of a large number of parameters in ConvLSTM. A more simplified convolutional gated recurrent unit (ConvGRU) model has been proposed for echo map extrapolation [21], which utilizes convolution kernels to deal with local neighborhood sets and reduce the number of parameters. Shi et al. [18] improved the nowcasting model using trajectory gated recurrent unit (TrajGRU), which carries out trajectory convolution between different time steps to capture the structure of spatiotemporal variations for recurrent connections. However, these features are estimated with the local receptive field and only provide sparse spatial dependencies thus can not obtain long-range dependencies efficiently. Compared to the trajectory convolution, the self-attention module is capable of capturing the global spatial variations with a single layer [22]. Besides, the features at the current time step can benefit from aggregating relevant features in the past. Lin et al. [23] introduced the self-attention memory (SAM) module into the ConvLSTM. However, their SAM and the inherent part of ConvLSTM have very high computational complexity in high-resolution input, which cannot meet the demand of nowcasting high-impact weather based on high-resolution radar data. As such, we develop the self-attention-based GRU as a backbone of the deep learning model designed in this study.

Since the nowcasting model extracts features from the training dataset and then performs prediction using the learned features, data distribution, diversity, and quality are critical for deep learning. In fact, data are often considered the most important part of modern machine learning techniques. Unfortunately, few of the previous studies focused on quantifying the nowcasting performance for the model trained with diverse features during different types of precipitation events. In addition, the studies on extreme weather events, such as hurricanes are still rare, although some of the previous studies paid special attention to convective precipitation events [7], [17], [18], [24], [20]. As a result, the existing models do not have sufficient capacity for hurricane nowcasting due to the fast evolving of associated precipitation. The radar reflectivity of hurricanes is continually changing, and there are significant radial and azimuthal flow components in tropical cyclones, which impact the convective structure, suggesting that the nowcasting model must learn the features including both the movement, structure, and strength varieties of tropical cyclones at the same time.

In addition, hurricanes are less common compared to heavy rainfall events, indicating that there may not be sufficient data to train a mature hurricane nowcasting model solely based on hurricane observations. Since the high-impact convective
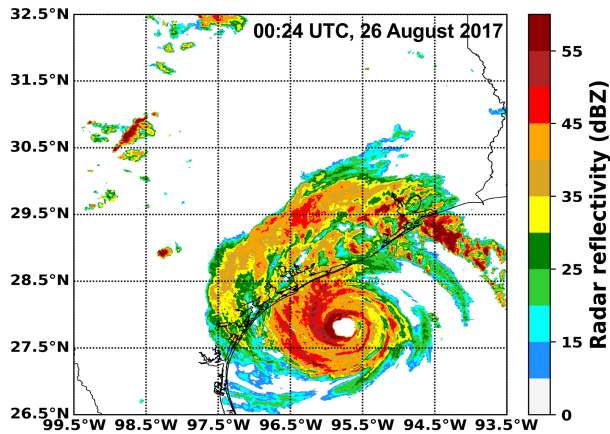
Fig. 2.    Selected study domain during hurricane Harvey event at 00:24 UTC, 26 August 2017. The color bar stands for radar composite reflectivity.

precipitation events are also of our interest, and this type of event is relatively common, this study will quantify the impact of applying the model trained based on one set of intense rainfall events to a different type of high-impact weather events. In particular, we will investigate how to adjust the deep learning model for adaptive applications based on radar observations not only for hurricanes but also for (nonhurricane) extreme convective precipitation.

The main contributions of this article include the following.

1) We develop a self-attention-based gate recurrent unit (SaGRU) model for nowcasting high-impact weather events.
2) Radar data collected from heavy convective rainfall events in South Texas from 2015 to 2020, and 22 hurricane events over the United States during the same period are selected to train the deep learning models to quantify the impact of data sources on nowcasting performance.
3) We quantify the impact of the sequence length of input radar observations on the model performance, which can serve as a guideline for precipitation nowcasting research.

The rest of this article is organized as follows. Section II describes the study domain, dataset, and nowcasting methodology used in this article. Section III details the application products during high-impact weather events and quantifies the nowcasting performance of the adapted deep learning model. In Section IV, a thorough discussion of the nowcasting performance is provided. Finally, Section V concludes this article.

## II. STUDY DOMAIN, DATASETS, AND METHODOLOGY

### A. Study Domain

South Texas is selected for our study domain, which covers an area of about $600 \times 600$ km ranging from $26.5°N–32.5°N$ latitude to $93.5°W–99.5°W$ longitude. This area includes Greater Houston region, one of the most populous metropolitan regions in the United States. Fig. 2 shows the specific study domain along with an example of the radar reflectivity map collected during hurricane Harvey at 00:24 UTC, 26 August 2017.

This region is within the humid subtropical climate zone, a typical climatology in Southern United States. During most of the year, prevailing winds are from the south and southeast, bringing heat and moisture [25]. The majority of South Texas areas receive ample rainfall in general, more than 60 inches (1500 mm) annually [26]. In addition, spring supercell thunderstorms sometimes bring tornadoes to the region, even though it is not in the Tornado Alley like much of Northern Texas. As a result, South Texas experiences a wide range of natural weather hazards, including urban fash flooding, high winds, tornadoes, and hailstorms. Furthermore, due to the flat terrain and low-permeability clay-silt prairie soils, flooding can easily be exacerbated, and there have been more flood-related deaths and property damage in this study domain than that in any other regions in the United States [27]. Accurate monitoring and prediction of the rapidly changing meteorological conditions in such a region is necessary for emergency management and decision-making. Therefore, it is an ideal location to study high-impact weather events and produce precipitation nowcasting.

### B. Dataset

As mentioned, this study uses the radar reflectivity mosaic data for deep learning-based precipitation nowcasting. Composite radar reflectivity images are produced at 6-min resolution using the National Weather Service (NWS) Weather Surveillance Radar—1988 Doppler (WSR-88D) systems in this region. Spatially, the reflectivity images are created at regular 1-km resolution grids, which means the number of pixels for the single image is $600 \times 600$. The three-dimensional data indicate precipitation patterns and their movement, and it is ideal for sequence modeling. In particular, we utilize the radar data collected during heavy precipitation events over this study domain from 2015 to 2020. The training and validation datasets are randomly selected from 2015 to 2020 (except 2017): 1348 days of data are used for training the deep learning model, 104 days are used as validation data to optimize the model parameters, and 290 days of precipitation data during 2017 are used for the independent test.

In addition, 22 hurricane events are used, which made landfall during 2015 and 2020. Here, it should be noted that the hurricane events are not limited to the region of South Texas, i.e., all the major hurricane events over the United States during 2015 and 2020 are included. Similar to the heavy precipitation events, 21 hurricane events are used for training and validation whereas hurricane Harvey is selected for independent test. In summary, we trained three models based on heavy precipitation events, hurricane events, and all events combined, respectively, to quantify the impact of data sources on the nowcasting performance.

### C. Methodology

In this section, the deep learning model utilized in this study is detailed, including data preprocessing, model structure, the essential components in model training and testing, as well as the nowcasting performance evaluation metrics.

*1) Data Preprocessing:* First, since most of the days are characterized by clear air (i.e., no rain), the learned features
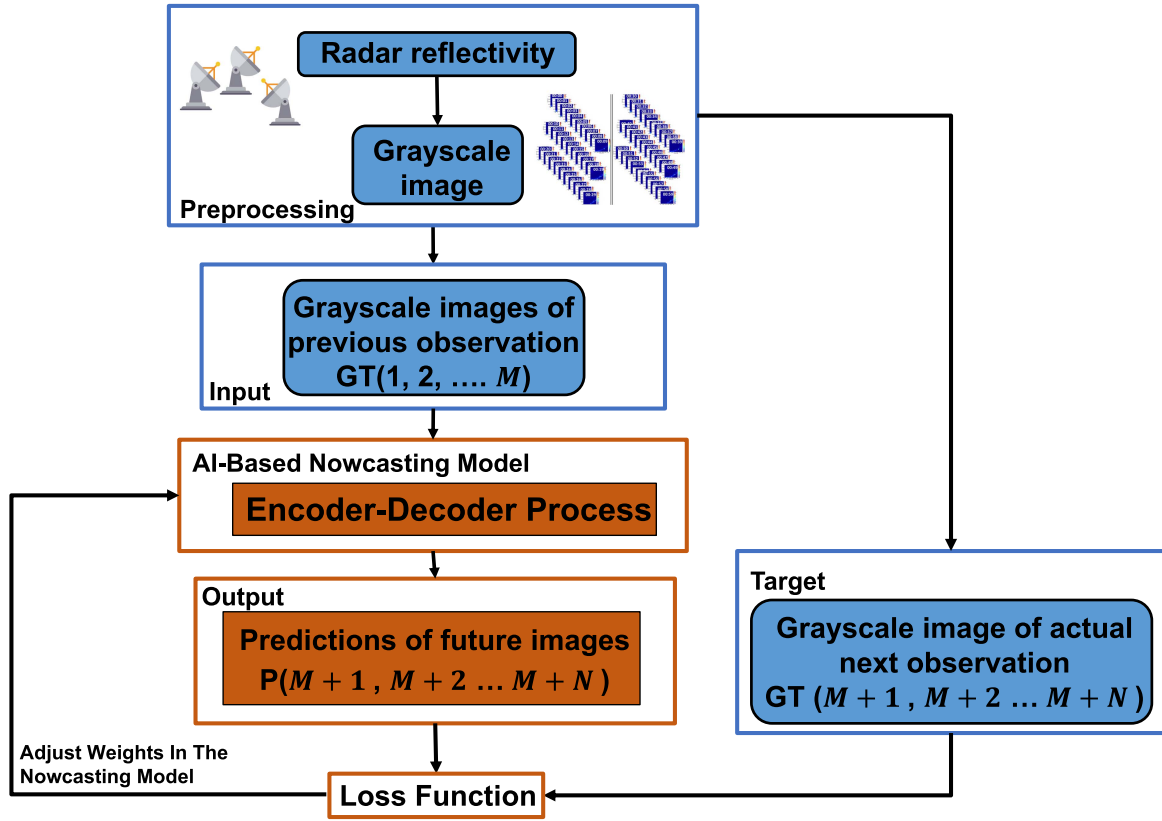
Fig. 3. Overall framework of the applied high-impact weather nowcasting system. Essentially, the system is trained to predict future radar reflectivity echo maps based on few previous observations. $M$ is the number of previous observations and $N$ represents the length of the predictions of future images.

will be dominated by these nonrain days if the model is trained based on all the data. To this end, only the days with rainfall occurring during 2015 and 2020 are selected and used, as suggested by [18] and NWS forecasters (personal communications). In addition, the filter process has taken into account the radar image sequences rather than a single image, since we need to ensure each data sequence contains adequate data samples with strong reflectivity for training a reliable sequence model. Since we also investigate the impact of the input sequence length on the nowcasting performance, the number of frames ranges from 32 to 40. Hence, the whole dataset is split into numerous sequences by a moving sequence sliding window from the start time (starting point of past observations) to the end time (nowcasting lead time). The number of grid points that have reflectivity higher than 35 dBZ is summed up for each sequence, and then divided by the sequence length to get the average number of qualified grid points for this sequence. If the average number is larger than 50, this sequence of radar reflectivity data will be selected for machine learning. After filtering all the sequences, the training dataset contains 463 602 sequences, and 31 883 sequences are used as the validation set to optimize the model parameters. In the testing stage, 87 571 sequences are used to evaluate the capacity of the trained models. Furthermore, the 87 571 testing samples are split into heavy precipitation events and hurricane events since a major goal is to quantify the nowcasting performance of the trained models during different precipitation events. All radar reflectivity

data are transformed to [0,1] gray-level pixels by min–max normalization.

*2) Deep Learning Model Architecture:* The work flow of the proposed deep learning model for high-impact weather nowcasting is shown in Fig. 3. As mentioned, the radar reflectivity images are first transformed to grayscale images, as described in Section II-C1, before being fed into the nowcasting model. The precipitation nowcasting system utilizes previous $M$ steps of radar observations to predict the future $N$ steps (at 6 min intervals). For one iteration, we have a sequence of consecutive real observations as ground truths GT(1, 2, ... $M$, $M$+1, ... $M$+$N$). Then this sequence is split into two parts in the training stage. The previous GT(1, 2. . . $M$) observations are fed into the nowcasting model to get the predictions P($M$+1, ... $M$+$N$) of future observations. In the training stage, the predictions P($M$+1, . . . $M$+$N$) are compared with ground truth GT($M$+1, . . . $M$+$N$) to adjust the weights in the nowcasting model for the next iteration. It should be emphasized that various values of $M$ have been tested to quantify the required number of past radar observations for training and applying the nowcasting model. There is a tradeoff issue here since a larger number of steps $M$ requires more parameters and time to train the model, although we would expect more precipitation information content from a longer observation sequence. On the other hand, a shorter sequence (i.e., smaller $M$) may contain less features of precipitation, but it is easier to train. Results from multiple experiments for different $M$ are described in Section IV-B. $N$ is an adjustable variable up
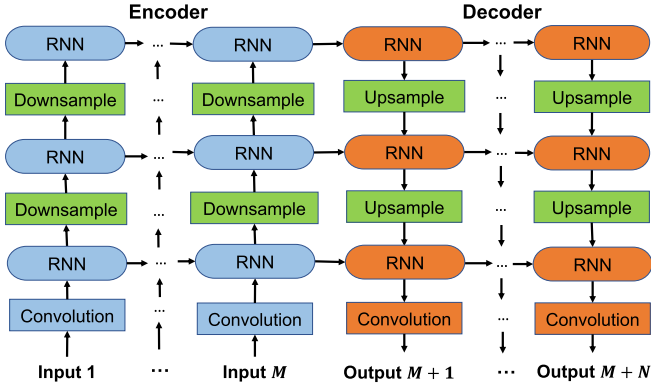
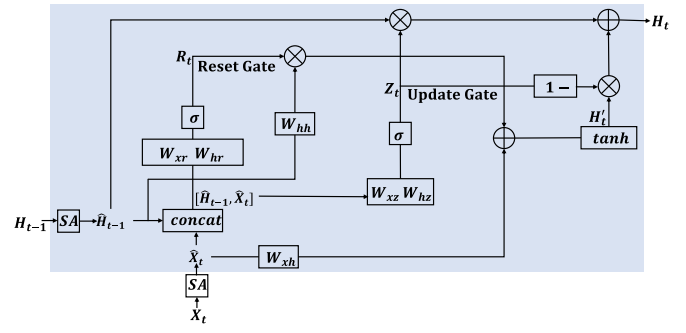Fig. 4.    Structure of encoder–decoder module, illustrated in Fig. 3.



Fig. 5.    Basic unit of RNN (SaGRU). Current Input $X_t$ and previous hidden states $H_{t-1}$ are served as input, reset gate $R_t$ and update gate $Z_t$ are used to control the current hidden state $H_t$.

to 30. That is, the deep learning model can produce nowcasting results for a lead time up to 3 h. For the results in Section III, $M = 5$ and $N = 30$ are primarily used to quantify the nowcasting performance, i.e., previous 30-min radar observations are used to predict observations of future 3 h.

Fig. 4 shows the overall structure of the expanded encoder–decoder structure, which includes four main parts: the RNN, up-sample, down-sample, and convolution. Multiple layers of RNN were stacked to build an encoder–decoder structure, resulting in an end-to-end trainable model. The encoder part extracts hidden states from previous radar echo map observations, while the decoder part uses the hidden states to forecast future echo maps. In particular, the hidden state of the RNN serves as input to the next level to extract the spatiotemporal information of different levels. Down-sample and up-sample are implemented by convolution and deconvolution, respectively. The updating of the low-level states could be guided by the high-level states, which have captured the global spatiotemporal correlations. Furthermore, low-level states could have an impact on the nowcasting. The initial hidden state of encoder and the initial input of forecaster are 0 and the final output is regressed through a convolution layer.

The choice of the RNN unit is flexible. Originally, we used TrajGRU [18] as the baseline for nowcasting. Contrary to the abovementioned ConvLSTM and ConvGRU with fixed local neighborhood sets in the convolutional kernels, TrajGRU can dynamically determine the location-dependent spatiotemporal patterns. With the adaptive neighborhood in kernels, TrajGRU generates a flow field from the current input $X_t$ and previous hidden states $H_{t-1}$, and then warps $H_{t-1}$ through bilinear sampling. The output of the TrajGRU base unit $H_t$ is given as follows:

$$\boldsymbol{u}_t, \boldsymbol{v}_t = \boldsymbol{\gamma}\left(\boldsymbol{X}_t, \boldsymbol{H}_{t-1}\right) \tag{1a}$$

$$\boldsymbol{Z}_t = \sigma\left(\boldsymbol{W}_{xz} * \boldsymbol{X}_t + \sum_{l=1}^{L} \boldsymbol{W}_{hz}^l * \text{warp}(\boldsymbol{H}_{t-1}, \boldsymbol{u}_{t,l}, \boldsymbol{v}_{t,l})\right) \tag{1b}$$

$$\boldsymbol{R}_t = \sigma\left(\boldsymbol{W}_{xr} * \boldsymbol{X}_t + \sum_{l=1}^{L} \boldsymbol{W}_{hr}^l * \text{warp}(\boldsymbol{H}_{t-1}, \boldsymbol{u}_{t,l}, \boldsymbol{v}_{t,l})\right) \tag{1c}$$

$$\boldsymbol{H}'_t = \boldsymbol{f}\left(\boldsymbol{W}_{xh} * \boldsymbol{X}_t + \boldsymbol{R}_t \circ \left(\sum_{l=1}^{L} \boldsymbol{W}_{hh}^l * \text{warp}(\boldsymbol{H}_{t-1}, \boldsymbol{u}_{t,l}, \boldsymbol{v}_{t,l})\right)\right) \tag{1d}$$

$$\boldsymbol{H}_t = (\boldsymbol{1} - \boldsymbol{Z}_t) \circ \boldsymbol{H}'_t + \boldsymbol{Z}_t \circ \boldsymbol{H}_{t-1} \tag{1e}$$

where $u_t$, $v_t$ are the flow fields that store the local connection structure. $\gamma$ is a one-hidden-layer convolutional neural network. $L$ is the total number of allowed links. $W$ denotes the weights of the convolutional kernel, $*$ is the convolution operation and $\circ$ is the Hadamard product. The warp function selects the positions pointed out by $u_{t,l}$, $v_{t,l}$ from $H_{t-1}$ and responsible to dynamically determine the recurrent connections. $\sigma$ is sigmoid function and $f$ is Leaky ReLU function.

However, even the experimental results in [18] reveal that TrajGRU captures spatiotemporal correlations better than conventional extrapolation algorithms and some other deep learning-based algorithms, it still have the deficiency that cannot capture effective long-range dependencies. The success of self-attention on computer vision tasks [28], [29], [30] demonstrates its efficiency in aggregating major features across all spatial locations. It can identify long-range spatiotemporal correlations by calculating the pairwise relationships between various feature map positions using a binary relation function. Following that, these relations can be used to determine the attended features. As such, an SaGRU is developed to capture the global spatiotemporal features of the high-impact weather in this article. The SaGRU model is constructed by cascading self-attention module and the standard ConvGRU. Contrary to TrajGRU, SaGRU uses self-attention module to aggregate features from the current input $X_t$ and previous hidden states $H_{t-1}$. Then, the output of the SaGRU $H_t$ is obtained by the update gate $Z_t$, the reset gate $R_t$, and the aggregated features $\hat{H}_{t-1}$, as shown in Fig. 5. The model is formulated as follows:

$$\hat{\boldsymbol{X}}_t = \boldsymbol{SA}\left(\boldsymbol{X}_t\right), \hat{\boldsymbol{H}}_{t-1} = \boldsymbol{SA}\left(\boldsymbol{H}_{t-1}\right) \tag{2a}$$

$$\boldsymbol{Z}_t = \sigma\left(\boldsymbol{W}_{xz} * \hat{\boldsymbol{X}}_t + \boldsymbol{W}_{hz} * \hat{\boldsymbol{H}}_{t-1} + \boldsymbol{b}_z\right) \tag{2b}$$

$$\boldsymbol{R}_t = \sigma\left(\boldsymbol{W}_{xr} * \hat{\boldsymbol{X}}_t + \boldsymbol{W}_{hr} * \hat{\boldsymbol{H}}_{t-1} + \boldsymbol{b}_r\right) \tag{2c}$$
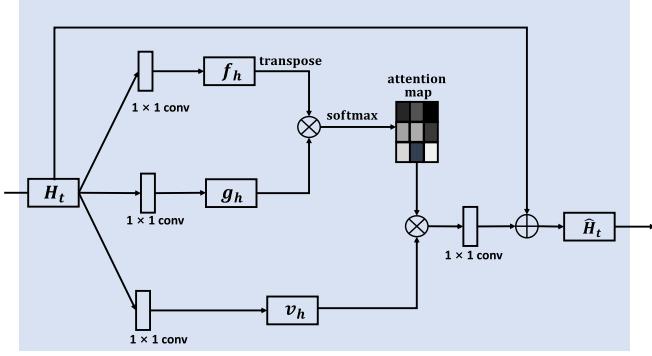
Fig. 6. Proposed self-attention module for SaGRU. $H_t$ is the image features from the hidden layer at the time step $t$. $f_h$, $g_h$, and $v_h$ are the three different feature spaces based on the $1 \times 1$ convolution on the $H_t$. $\hat{H}_t$ is the output.

$$H'_t = \tanh\left(W_{xh} * \hat{X}_t + R_t \circ \left(W_{hh} * \hat{H}_{t-1} + b_h\right)\right) \tag{2d}$$

$$H_t = (1 - Z_t) \circ H'_t + Z_t \circ \hat{H}_{t-1} \tag{2e}$$

where SA represents the self-attention module. $\hat{X}_t$ and $\hat{H}_{t-1}$ are the features aggregated from $X_t$ and $H_{t-1}$ through self-attention modules. In particular, the location at attention module aggregates the input feature by calculating a weighted sum across all locations at each time step. This allows the long-range spatiotemporal dependencies can be captured during propagation cross our encoder–decoder structure.

Fig. 6 shows the details of the self-attention module. The image features $H_t \in \mathbb{R}^{C \times N}$ from previous layers are transformed into three feature spaces $f$, $g$, $v$ to calculate the dependencies across different image regions, where $f_h = W_f H_t \in \mathbb{R}^{\hat{C} \times N}$, $g_h = W_g H_t \in \mathbb{R}^{\hat{C} \times N}$, $v_h = W_v H_t \in \mathbb{R}^{\hat{C} \times N}$. Here, $C$ and $\hat{C}$ are the number of channels and we choose $\hat{C} = C/8$ for memory efficiency. $N$ is the number of feature locations from the previous hidden layer. $W_f$, $W_g$, and $W_v$ are the learnable weight matrices, which are implemented as $1 \times 1$ convolutions. We transpose $f_h$ and perform matrix multiplication to calculate the similarity scores between the $i$th point and the $j$th point as follows:

$$s_{i,j} = f(h_i)^T g(h_j). \tag{3}$$

After the softmax operation, the similarity scores are normalized as follows:

$$\beta_{i,j} = \frac{\exp(s_{i,j})}{\sum_{i=1}^{N} \exp(s_{i,j})}, i,j \in \{1, 2, \ldots, N\}. \tag{4}$$

The attention of the input features is calculated with a weighted sum at all locations and the output of the attention layer is att $= \{\text{att}_j \in \mathbb{R}^{\hat{C}}\}, j \in \{1, 2, \ldots, N\}$, where

$$\text{att}_j = \sum_{i=1}^{N} \beta_{i,j} v(h_i). \tag{5}$$

Then, the output of the attention layer will be added back to the input feature map. Therefore, the final output is given by

$$\hat{H}_t = W_h \text{att} + H_t. \tag{6}$$

3) *Hyperparameters and Loss Function:* In this study, we use a three-layer encoder–decoder architecture with the number of filters for the RNNs set to 64, 128, and 128, respectively. For the first RNN layer, the $X_t$ is $120 \times 120$ vector since the kernel size is $7 \times 7$, padding is 1 and strides are 5 for the first convolution layer. The first down-sampling is implemented by the convolutional layer with $5 \times 5$ kernel size, padding 1, strides 3, and the kernel size is $3 \times 3$, padding is 1 and strides are 2 for the second down-sampling. Thus, the $X_t$ for second RNN and third RNN layer is $40 \times 40$ and $20 \times 20$. Similarly, the first and second up-sampling is implemented by deconvolution with $5 \times 5$ kernel size, padding 1, strides 3, and $4 \times 4$ kernel size, padding 1, strides 2. For the self-attention module in SaGRU, the kernal size of convolutions is $1 \times 1$ as mentioned in Section II-C2. All the models are optimized by the Adam optimizer with learning rate of $10^{-4}$ and momentum of 0.5 [31]. The learning rate of each parameter group decays by gamma 0.5 once the number of epochs reaches the milestones of 10 000, 30 000, 90 000. The training batch size is set as 3 and the maximum iteration is set to 300 000. All experiments are implemented using the PyTorch platform [32].

It should be pointed out that since the frequencies of different rainfall intensities are highly imbalanced, especially during high-impact weather events, such as hurricanes, this research utilizes two weighted loss functions termed balanced mean squared error (B-MSE) and balanced mean absolute error (B-MAE), defined as follows:

$$w(x) = \begin{cases} 1 & 0 \leq x < 10 \\ 1 & 10 \leq x < 20 \\ 5 & 20 \leq x < 30 \\ 10 & 30 \leq x < 35 \\ 30 & 35 \leq x < 40 \\ 30 & 40 \leq x < 50 \\ 35 & x > 50 \end{cases} \tag{7a}$$

$$\text{B-MSE} = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{i=600} \sum_{j=1}^{j=600} w_{n,i,j} (x_{n,i,j} - \hat{x}_{n,i,j})^2 \tag{7b}$$

$$\text{B-MAE} = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{i=600} \sum_{j=1}^{j=600} w_{n,i,j} |x_{n,i,j} - \hat{x}_{n,i,j}| \tag{7c}$$

where $x$ and $\hat{x}$ represent the real and predicted reflectivity, respectively; $N$ is the sample number; $w_{n,i,j}$ is the weight corresponding to the $(i, j)$th reflectivity value in the $n$th training data.

4) *Model Evaluation:* To evaluate the performance of precipitation nowcasting products, this article adopts four widely used metrics, namely, Heidke skill score (HSS), critical success index (CSI), probability of detection (POD), and false alarm rate (FAR). The values of POD, FAR, HSS, and CSI are all between 0 and 1. Higher POD, HSS, CSI, or lower FAR indicate better nowcasting performance. Since HSS and CSI are more integrated metrics, they are direct indicators of the model capacity. In addition, for better interpretation of the nowcasting performance at different rainfall intensities, the evaluation metrics are computed using a number of reflectivity thresholds, including
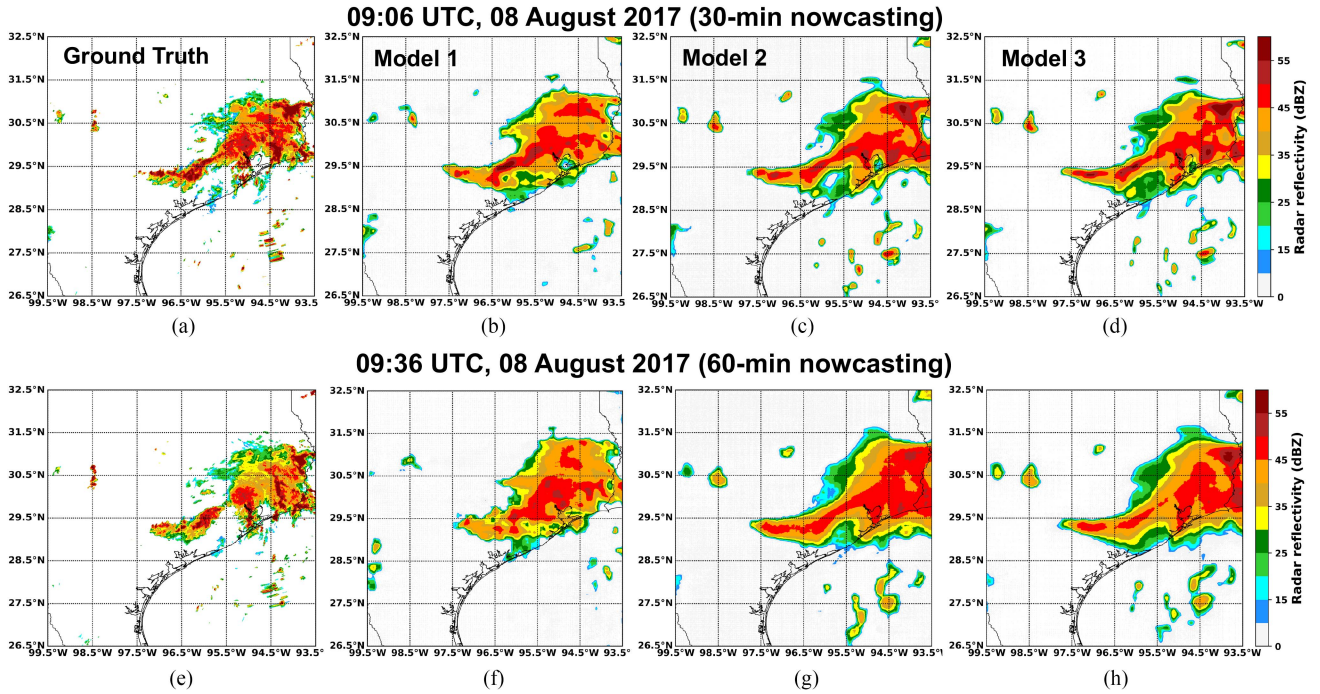
Fig. 7. Example of 30-min (valid at 09:06 UTC) and 60-min (valid at 09:36 UTC) nowcasting results issued at 08:36 UTC on 08 August 2017: (a), (e) Radar observation (ground truth). (b), (f) Hurricane model. (c), (g) Nonhurricane model. (d), (h) Combined model.

20, 30, 40 dBZ. For each threshold, we compare the nowcasting product with the corresponding ground truth by transforming both reflectivity fields into binary matrices. In particular, if the reflectivity at a grid pixel is higher than the threshold, "1" is assigned to this grid pixel, otherwise a "0" will be assigned. Then the evaluation matrices are derived based on the two binary matrices

$$\text{HSS} = \frac{2(\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP}+\text{FN})(\text{FN}+\text{TN}) + (\text{TP}+\text{FP})(\text{FP}+\text{TN})} \quad (8a)$$

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad (8b)$$

$$\text{POD} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8c)$$

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \quad (8d)$$

where true positive (TP) is the number of grid points which are assigned "1" for both nowcasting product and corresponding ground truth; false negative (FN) is the number of grid points which are assigned "0" for the nowcasting product, but "1" for the ground truth; true negative (TN) is the number of grid points which are assigned "0" for both nowcasting product and corresponding ground truth; and false positive (FP) represents the number of grid points, which are assigned "1" for the nowcasting product, but "0" for the ground truth.

## III. EXPERIMENTAL RESULTS

As mentioned, heavy rainfall events occurred in South Texas and 22 hurricane events that made landfall in the U.S. during 2015–2020 are used in this study. In particular, the heavy rainfall events and hurricane Harvey in 2017 are selected for testing, while other data are utilized for model training. To quantify the influence of different training data sources on high-impact weather nowcasting performance, three models are trained based on hurricane events, (nonhurricane) heavy rainfall events, and all events combined, respectively.

In addition, extensive experiments are performed using different sequence lengths of input radar observations in the nowcasting models, ranging from 2 to 10 time frames. This is to quantify the impact of input sequence length on the nowcasting performance, so as to provide guidelines on how many radar observations would be required to train a deep learning model for high-impact weather nowcasting. In this section, example nowcasting products based on five historical radar observation frames are illustrated to demonstrate the nowcasting performance.

Fig. 7 shows the practical examples of 30-min (valid at 09:06 UTC) and 60-min (valid at 09:36 UTC) precipitation nowcasting results during a severe convective rainfall event in the study domain issued at 08:36 UTC, August 08, 2017. In Fig. 7, both the nowcasting results from three different models and the corresponding observations are illustrated. Overall, all the three models can predict the overall pattern and distributions of rainfall. However, scrutinizing the detailed structure of the nowcasting results, it is found that the model trained purely based on hurricane data significantly underestimates the precipitation intensity during this severe convective event. For the area where the reflectivity values are larger than 45 dBZ, the patterns are inconstant with the real observation. In addition, some small rainfall regions are missed by this model [see Fig. 7(b) and (f)].
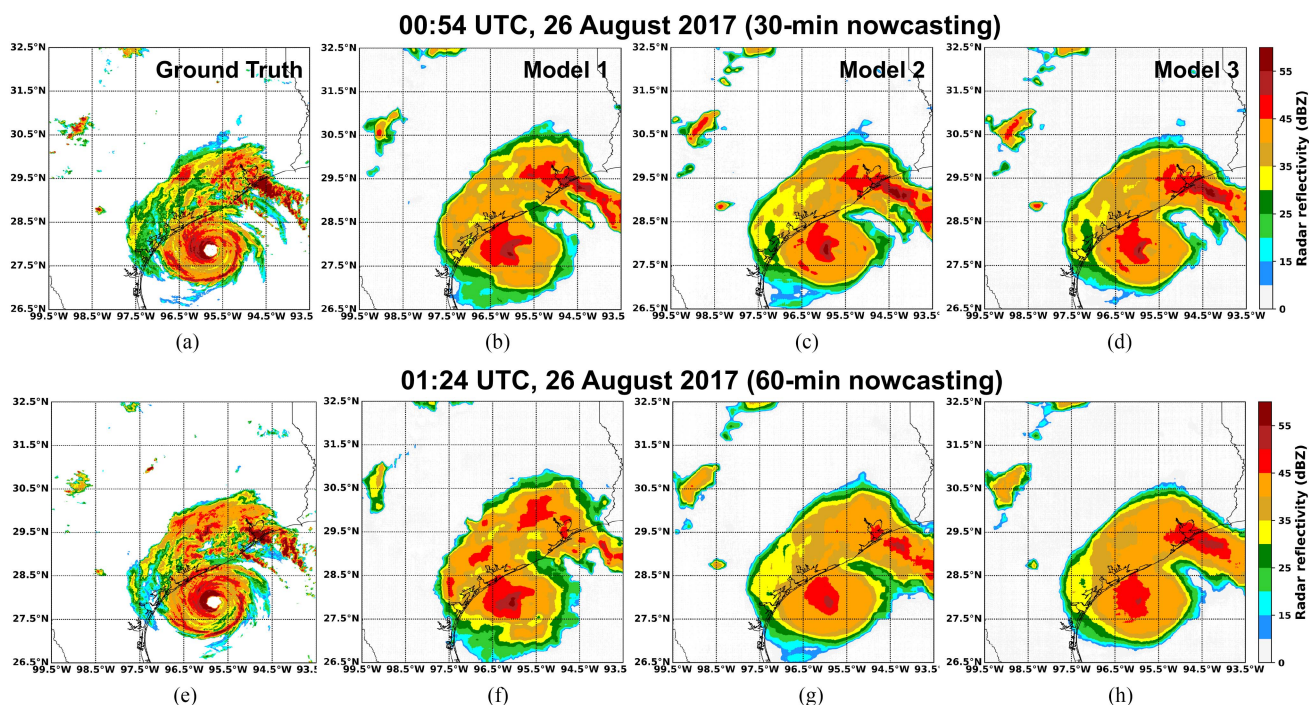
Fig. 8. Example of 30-min (valid at 00:54 UTC) and 60-min (valid at 01:24 UTC) nowcasting results issued at 00:24 UTC on 26 August 2017: (a), (e) Radar observation (ground truth). (b), (f) Hurricane model. (c), (g) Nonhurricane model. (d), (h) Combined model.

This is likely due to the insufficiency of hurricane data (only 21 events) in capturing heavy rain features in this particular domain. In addition to the limited amount of data for model training, location representation could be an issue that limits the nowcasting performance as most of the hurricane events were spanning a much larger domain beyond the State of Texas. In contrast, the model trained based on nonhurricane events predicts the precipitation distribution more precise than the hurricane model, especially at lead times of 60 min. However, compared to the model trained based on combined events, it still has a deficiency of underestimation when reflectivity values are larger than 50 dBZ [see Fig. 7(g) and (h)]. In general, the combined model can not only capture the precipitation patterns, but also predict the precipitation intensity well.

Fig. 8 illustrates the nowcasting products at lead times of 30 min (valid at 00:54 UTC) and 60 min (valid at 01:24 UTC) during hurricane Harvey issued at 00:24 UTC, August 26, 2017. It is found that all the three models can capture the structure of the tropical cyclone and predict the overall distribution of precipitation intensity at lead times of 30 min. Being trained on the hurricane events, the hurricane model can provide more plausible details in terms of the cyclone structure, especially near the eye wall relative to the other nowcasting models. However, it tends to underestimate the precipitation intensity, produce wrong pattern in the outer spiral rainband and still miss some small rainfall regions around the tropical cyclone. Surprisingly, although the models trained using nonhurricane data and combined data tend to provide a smoother structure of the cyclone, both can predict higher rainfall intensity near the outer rainband, which is more consistent with real observations in Fig. 8(a) and (e).

The quantitative evaluation results of the 60-min nowcasting products using the three models based on all the test data during heavy rain events are summarized in Fig. 9, where the best nowcasting skill scores are indicated in bold. In order to highlight the nowcasting performance for different rainfall intensities, three thresholds, namely, 20, 30, and 40 dBZ, are applied in calculating the skill scores. Fig. 9 indicates that during (nonhurricane) heavy rain events, the model trained using hurricane data has a rather poor performance, especially for nowcasting convective cores, which have reflectivity higher than 40 dBZ. This is consistent with the examples shown in Figs. 7 and 8. The models developed using nonhurricane data or combined data have similar performance in terms of all skill scores, and both are better than the model trained using only hurricane data. In particular, the model trained using combined data has slightly better skill scores compared to the model trained using nonhurricane data. This is encouraging since including the features learned from hurricane data did not bring any negative impact on the performance of the combined model. In addition, when the reflectivity threshold increases from 20 to 40 dBZ, the performance of all models drops slightly. As expected, predicting heavy rain storm cores is more challenging than predicting weaker rain regions.

Similarly, Fig. 10 presents the nowcasting skill scores during the test hurricane event. It can be seen that the model trained using hurricane data delivers competitive results in terms of FAR, CSI, and HSS. Compared to model based on hurricane data, the model trained using nonhurricane data and combined data has slightly worse FAR, especially when the reflectivity threshold is low. Considering that FAR is an indicator of underestimating the rainy areas, the model based on hurricane data delivers a better precipitation pattern, although the intensity is underpredicted
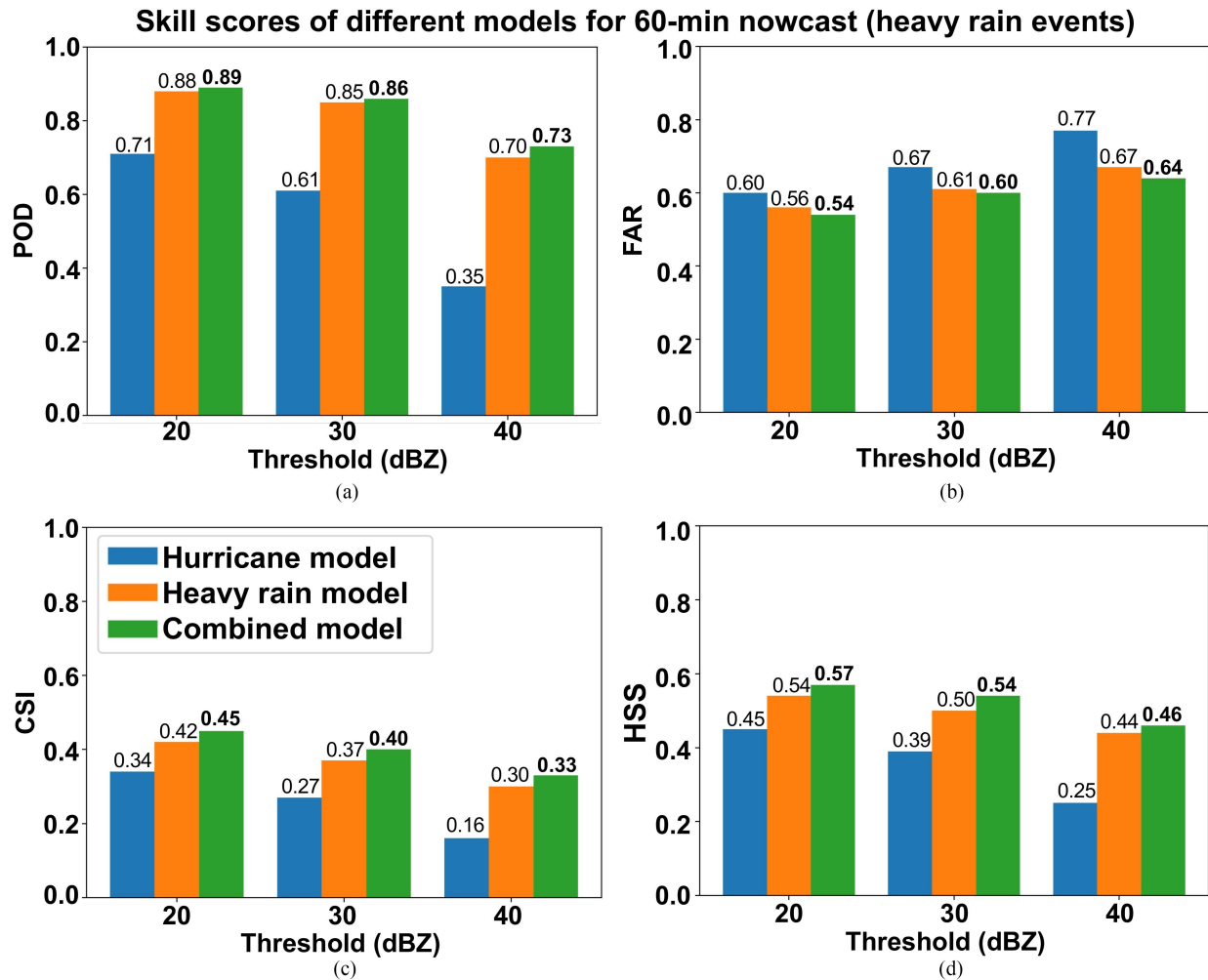
Fig. 9. Skill scores of the three models during heavy precipitation events at 60-min lead time. The best results are marked with bold face. (a) POD. (b) FAR. (c) CSI. (d) HSS.

(see Fig. 8). The model based on nonhurricane events has slightly better POD, CSI, and HSS scores, and the skill gaps between these two models are even larger when the reflectivity threshold is higher. The model trained using combined data renders the best performance among the three models. These results indicate that precipitation features learned from heavy convective precipitation events can be used to enhance hurricane nowcasting. On the other hand, including features learned from hurricane events has no negative impact on nowcasting (nonhurricane) heavy rain events.

For completeness, Fig. 11 shows the CSI scores as a function of lead time up to three hours for the nowcasting model trained using combined data when applied to the test heavy rain and hurricane events. As expected, the performance will decrease for both heavy rain and hurricane events as the nowcasting lead time increases. For heavy rain events, the differences of CSI scores are quite small when different reflectivity thresholds are used, demonstrating that the performance of the nowcasting model is stable for different rainfall intensities. For hurricane events, the CSI score is relatively low when a reflectivity threshold of 40 dBZ is used, indicating the challenge of predicting heavy

rain bands during hurricane events. Nevertheless, when a lower reflectivity threshold is used, the CSI scores are much higher. In addition, the CSI scores during hurricane events are higher than those during heavy rain events. Even for the lead time of 180 min, the CSI score is about 0.4, which is among the best results available in the literature (e.g., [6], [7], [20]).

## IV. DISCUSSION

### A. Impact of Diverse Training Data on the Nowcasting Performance

Although the products and quantitative evaluation results demonstrated the effectiveness and superiority of deep learning in high-impact weather nowcasting, especially its capability of capturing the spatiotemporal evolution of severe precipitation systems, it should be noted that generalization capability of the nowcasting model still requires further investigation. It is well known that the performance of deep learning models is highly dependent on the quality and distribution of the training dataset. In the training stage, if some data samples are significantly different from the overall distribution of the training data, the
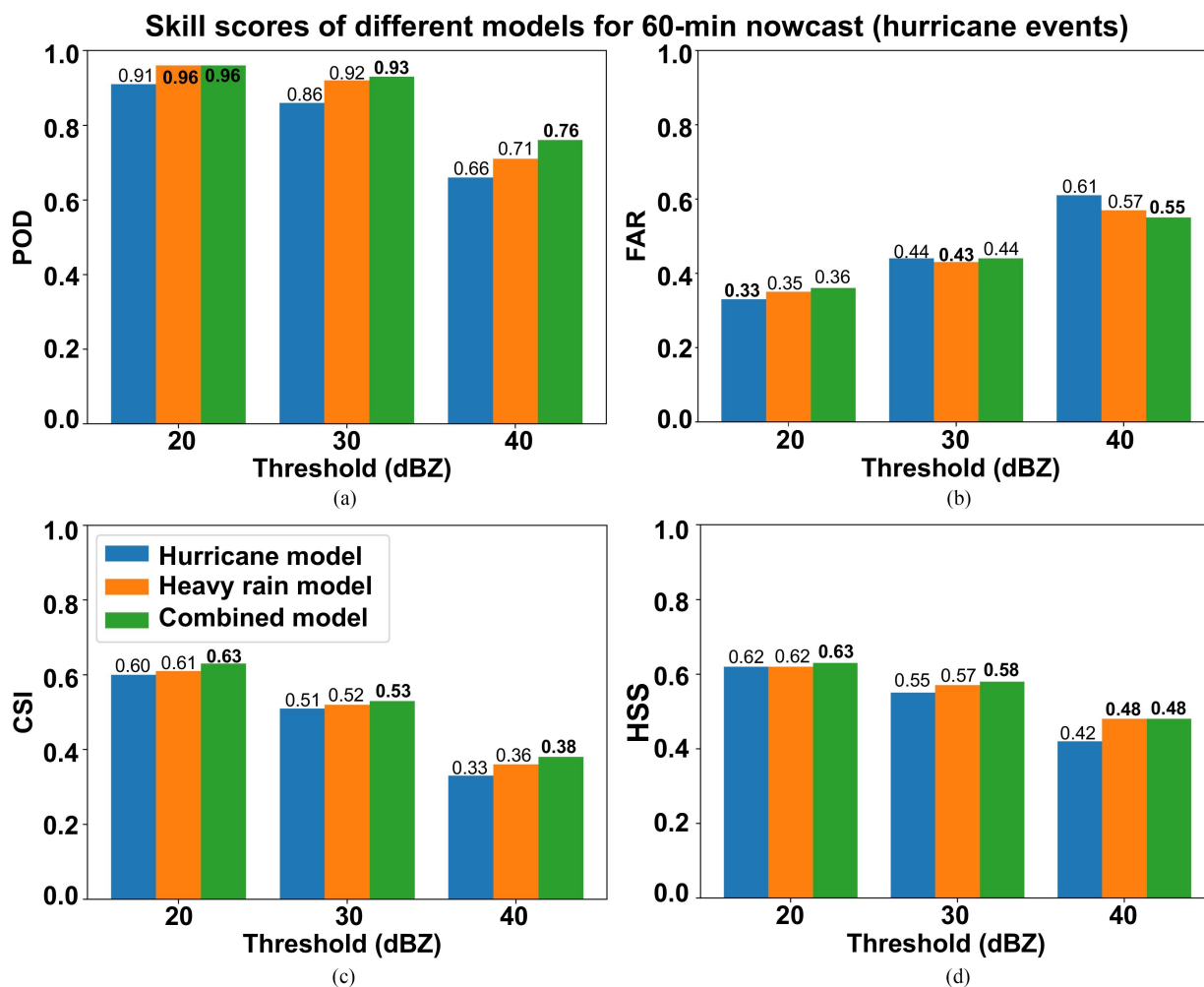
Fig. 10. Skill scores of the three models during hurricane events at 60-min lead time. The best results are marked with bold face. (a) POD. (b) FAR. (c) CSI. (d) HSS.

trained model will learn the features from these "outliers" (e.g., extreme events) resulting from natural variability and exhibit a worse performance than the one trained based on the dataset without these extreme events.

Through this article, we aim to provide a reference about how to select training data for short-term prediction of heavy rain, with an emphasis on hurricane events. It is encouraging that the model trained by combining hurricane and (non-hurricane) heavy rain events has better performance than the model trained solely based on hurricane data or heavy rain events. This is noteworthy since the hurricane events are rare compared to heavy rainfall events. There may not be sufficient data for training a mature hurricane nowcasting model only based on hurricane observations. This is demonstrated by the surprising results that the model trained using 21 hurricane events is not significantly better (in fact some of the scores are even slightly worse) than the one trained only using heavy convective precipitation events during hurricane applications. In other words, the rainfall features learned from heavy rain events can largely represent the characteristics of

rainfall associated with hurricanes, which is critical for hurricane nowcasting.

In addition, our experimental results show that the model based on combined data provides slightly better results than the model trained without including hurricane data during nonhurricane events. This is different from our assumption that involving hurricane events in the training stage may compromise the model capacity for applications during nonhurricane events. In fact, the features learned from hurricane events could even enhance the overall nowcasting model performance. Combining data from diverse precipitation events in training the deep learning model is strongly suggested, especially when there is a lack of sufficient data for model training.

Despite the positive performance results, a few relevant issues should be considered in the general application of the developed deep learning model. First, all the three models tend to smooth and fuse the structure of tropical cyclones and underestimate the tail of heavy rainfall regions, especially at longer nowcasting lead time. This is mainly because the implementation of convolutional and pooling processes, which involves computation based
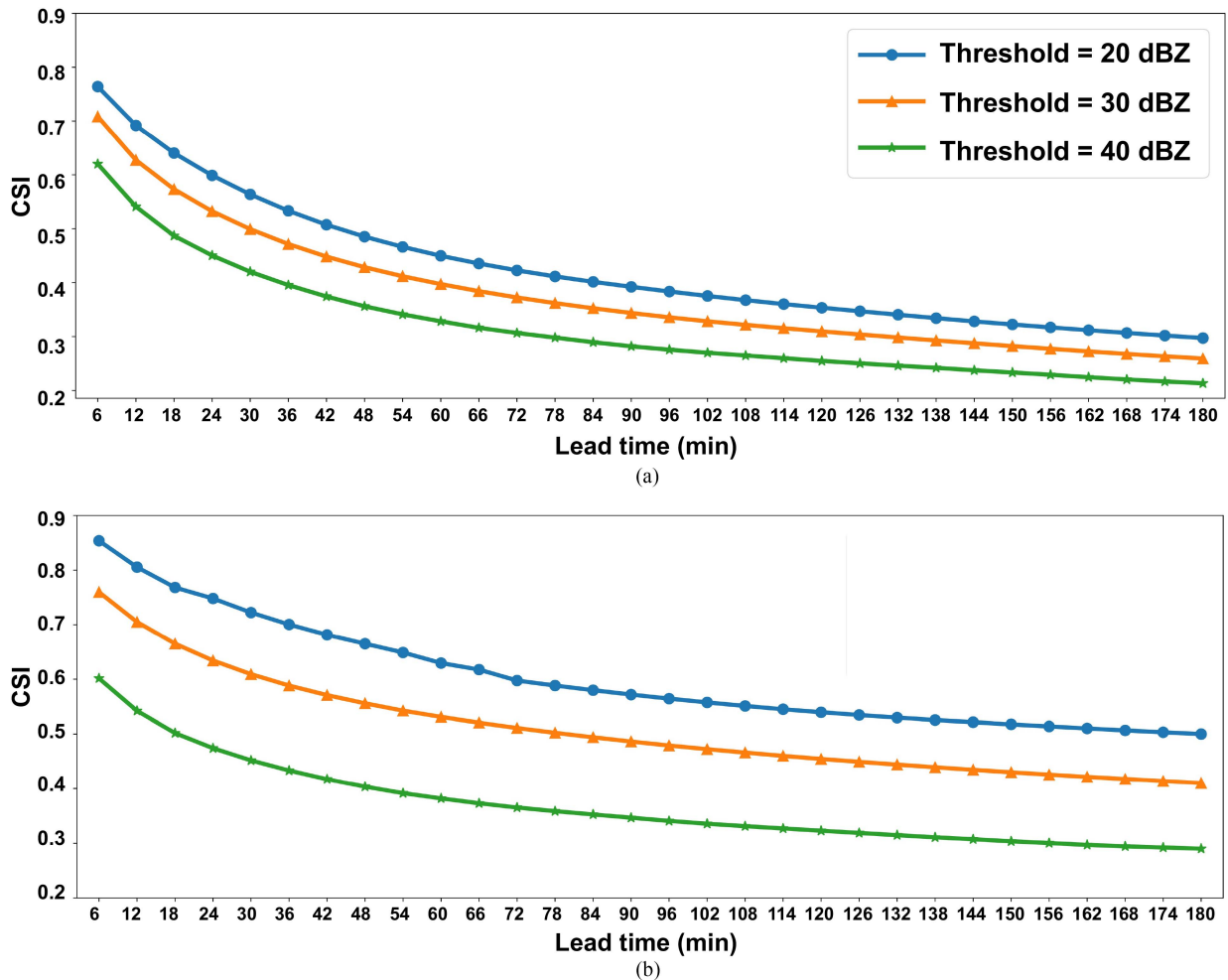
Fig. 11. CSI values of combined model with lead time up to 180 min during two high-impact weather events. (a) Heavy rain events. (b) Hurricane events.

on adjacent observations. In addition, the model always underestimates heavy rainfall regions as the lead time increases even after we incorporate the sequences containing strong precipitation echoes and adjust the weights for different precipitation intensity. A possible solution is to further increase the weights for reflectivity values larger than 35 dBZ. Further, more hurricane events should also be utilized for training, including those in other regions, in order to train a mature deep learning model that is more broadly applicable [19]. Another issue is about the model extension, i.e., the inclusion of additional factors in the nowcasting model. For example, including dual-polarization radar observables, such as $Z_{dr}$ and $K_{dp}$ can potentially improve the nowcasting performance as more precipitation microphysical information can be gleaned from polarimetric radar data [20].

### B. Impact of Radar Observation Sequence Length on the Nowcasting Performance

As mentioned in Section II-C2, it can take a long time to train a reliable nowcasting model using long sequences (i.e., large number of $M$ in Figs. 3 and 4) of past radar observations due to the high computational cost, especially when the training dataset is large. It is critical to quantify the number of past radar observations required to train the forecast model for high-impact weather applications. To provide a guideline on how many past observation frames we should use, we have trained nine models with different number of $M$, ranging from 2 to 10. Note that all the nine models are trained using combined data including hurricane and nonhurricane heavy rain events. For illustration purposes, Fig. 12 shows the skill scores of 60-min nowcasting products from the nine models when applied during hurricane events. Again, different reflectivity thresholds are used when computing the scores in order to further evaluate the model performance for predicting rainfall with different intensities. With a reflectivity threshold of 20 dBZ, it can be seen that smaller $M$ has relatively high POD but also higher FAR, indicating that the models trained based on fewer past radar observations generate too many precipitation pixels (i.e., rain area is over predicted). The two integrated metrics, CSI and HSS, both show an increasing trend as the number of past radar observations increases, suggesting that larger $M$ will produce better performance. When a reflectivity threshold of 30 dBZ is used, the POD scores exhibit a relatively flat line with some variability for $M \geq 5$, indicating that the POD is relatively insensitive to $M$ at the 30 dBZ threshold.
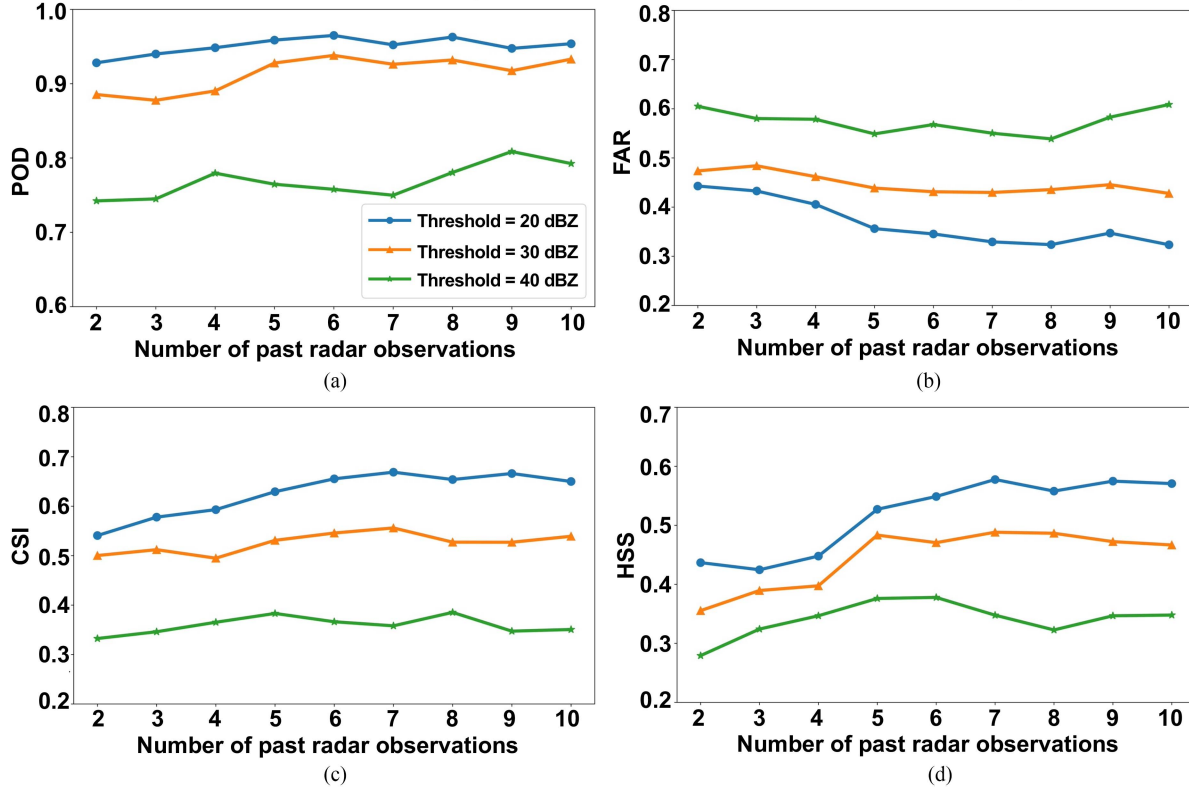
Fig. 12. Skill scores for different lengths of past radar observation sequence used for nowcasting during hurricane events. The nowcasting lead time in this example is 60-min. (a) POD. (b) FAR. (c) CSI. (d) HSS.

Similarly, FAR, HSS, and CSI are relatively stable for $M \geq 5$ while they present a growing trend when $M \leq 5$. This trend is much clearer if a reflectivity threshold of 40 dBZ is used. With a 40 dBZ threshold, it is also apparent that when $M > 8$ the model performance can get slightly worse, especially in terms of the FAR and CSI scores.

Based on these experimental results, we conclude that the nowcasting model can learn more features when more past radar observations are considered. The improvement is more significant for nowcasting weak-moderate rain with relatively low reflectivity. However, this does not mean that the longer sequence we use, the better performance we would get, especially for heavy rain nowcasting as the models trained based on more past observations are not stable. This is mainly because the life cycle of severe convective storm cells is short, and heavy rain regions can be initiated or disappear in a short amount of time due to the complex atmospheric state and the complex nonlinear cloud dynamics that occur in these events. More past radar observations contain too much of these changes, which cannot be effectively learned by the nowcasting model, not to mention that it will need more parameters and time to train. It is also encouraging that even two consecutive radar observations can yield an acceptable result, indicating that our deep learning model can learn the features of the motion field efficiently. This finding demonstrates that the deep learning model helps solve the issue of fast-changing conditions in extreme weather. In short, the sensitivity analysis suggests that five past radar observations

are the optimal choice for training a reliable deep learning model for high-impact weather nowcasting. Although the model trained based on five past observations may not provide the best performance all the time, the model is easy to train and very stable in generating reliable nowcasting results. It should be pointed out that the experiment can be extended by incorporating other precipitation systems from different climate regimes to provide more generalized guidelines, which will be included in future work.

## V. SUMMARY

Accurate nowcasting of high-impact weather, such as hurricanes and other heavy rain events, can support severe weather warnings and emergency management decision-making. Although some previous studies focused on convective precipitation nowcasting, there are very few studies that focus on high-impact weather events, such as hurricanes. This article has developed a deep learning-based nowcasting system and introduced the self-attention module to the GRU for extreme weather events. Five years of radar observations (2015–2020) over South Texas and 22 hurricane events that made landfall in the United States during 2015–2020 are used for training, validation, and testing of the deep learning models. In particular, three models trained based on hurricane events, (nonhurricane) heavy rain events, and all events combined are utilized to quantify the impact of diverse training data on the nowcasting system. In

addition, a number of experiments are conducted to investigate the required number of past radar observations for training the nowcasting models. The main conclusions of this article include the following.

1) Visually, all the three fine-tuned models can capture the precipitation patterns and distribution fairly well. However, the model trained purely based on hurricane events tends to underestimate the high-reflectivity regions during both heavy rain events and hurricane events. Nevertheless, it provides plausible details about the cyclone structure during hurricane event. The (nonhurricane) heavy rainfall and combined models can not only predict the precipitation patterns well but also the precipitation intensity, even though they tend to smooth and fuse the echoes during hurricane events.

2) The evaluation results of the three models reveal that precipitation features learned from heavy convective precipitation events can be utilized to improve hurricane nowcasting and the features learned from hurricane events have no negative impact on nowcasting (nonhurricane) heavy rain events. Therefore, it is strongly recommended that data from different precipitation systems be combined in training the deep learning model, especially when there is a dearth of data for machine learning model training.

3) The high CSI scores indicate the stable performance of the nowcasting model trained based on the combined data for lead times up to 3 h. In addition, the CSI scores are higher than those reported in the literature (e.g., [6], [7], and [20]).

4) To quantify the impact of sequence length of past radar observations for precipitation nowcasting, nine models are trained with different numbers of past radar observation (2–10). Considering the balance of computational cost and nowcasting performance, five consecutive observations are an optimal choice to yield a reliable model for nowcasting during high-impact weather events.

In future, a more detailed investigation of the nowcasting performance in the scenarios of storm initiation, growth, and decay will be performed.

## REFERENCES

[1] National Hurricane Center Public Affairs, "Hurricane basics," *Nat. Ocean. Atmospheric Admin.*, 1999, Art. no. 18 .

[2] P. Peduzzi et al., "Global trends in tropical cyclone risk," *Nat. Climate Change*, vol. 2, pp. 289–294, 2012.

[3] E. S. Blake and D. A. Zelinsky, "Hurricane harvey (AL092017)," Miami-Dade County, FL, USA, National Hurricane Center, Tropical Cyclone Report, 2018. [Online]. Available: https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

[4] M. Dixon and G. Wiener, "TITAN: Thunderstorm identification, tracking, analysis, and nowcasting-a radar-based methodology," *J. Atmos. Ocean. Technol.*, vol. 10, no. 6, pp. 785–797, Dec. 1993.

[5] R. Rinehart and E. Garvey, "Three-dimensional storm motion detection by conventional weather radar," *Nature*, vol. 273, no. 5660, pp. 287–289, 1978.

[6] N. E. Bowler, C. E. Pierce, and A. Seed, "Development of a precipitation nowcasting algorithm based upon optical flow techniques," *J. Hydrol.*, vol. 288, no. 1/2, pp. 74–91, 2004.

[7] L. Han, H. Liang, H. Chen, W. Zhang, and Y. Ge, "Convective precipitation nowcasting using U-Net model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4103508.

[8] L. Han, S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, "3 D convective storm identification, tracking, and forecasting–an enhanced titan algorithm," *J. Atmos. Ocean. Technol.*, vol. 26, no. 4, pp. 719–732, 2009.

[9] C. Radhakrishnan and V. Chandrasekar, "CASA prediction system over dallas–fort worth urban network: Blending of nowcasting and high-resolution numerical weather prediction model," *J. Atmos. Ocean. Technol.*, vol. 37, no. 2, pp. 211–228, 2020.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[12] I. Goodfellow et al., "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, pp. 2672–2680, 2014.

[13] H. Chen, V. Chandrasekar, H. Tan, and R. Cifelli, "Rainfall estimation from ground radar and TRMM precipitation radar using hybrid deep neural networks," *Geophys. Res. Lett.*, vol. 46, no. 17/18, pp. 10669–10678, Sep. 2019.

[14] Y. Zhang et al., "Deep learning for polarimetric radar quantitative precipitation estimation during landfalling typhoons in South China," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3157.

[15] H. Chen, V. Chandrasekar, R. Cifelli, and P. Xie, "A machine learning system for precipitation estimation using satellite and ground radar network observations," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 982–994, Feb. 2020.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[17] 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. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[18] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5618–5628.

[19] L. Han, Y. Zhao, H. Chen, and V. Chandrasekar, "Advancing radar nowcasting through deep transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4100609.

[20] X. Pan, Y. Lu, K. Zhao, H. Huang, M. Wang, and H. Chen, "Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model," *Geophysical Res. Lett.*, vol. 48, no. 21, 2021, Art. no. e2021GL095302.

[21] 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.

[22] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[23] Z. Lin, M. Li, Z. Zheng, Y. Cheng, and C. Yuan, "Self-attention convlstm for spatiotemporal prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 11531–11538.

[24] S. Kim, S. Hong, M. Joh, and S.-K. Song, "Deeprain: Convlstm network for precipitation prediction using multichannel radar data," in *Proc. 7th Int. Workshop Climate Informat.*, Sep. 20–22, 2017, pp. 1–4.

[25] J. W. Nielsen-Gammon, F. Zhang, A. M. Odins, and B. Myoung, "Extreme rainfall in texas: Patterns and predictability," *Phys. Geogr.*, vol. 26, no. 5, pp. 340–364, 2005.

[26] J. A. Smith, M. L. Baeck, J. E. Morrison, and P. Sturdevant-Rees, "Catastrophic rainfall and flooding in texas," *J. Hydrometeorol.*, vol. 1, no. 1, pp. 5–25, 2000.

[27] H. Sharif, T. Jackson, M. Hossain, and D. Zane, "Analysis of flood fatalities in texas," *Natural Hazards Rev.*, vol. 16, 2015, Art. no. 0 4014016.

[28] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[29] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10073–10082.

[30] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 68–80, 2019.

[31] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15.

[32] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 8026–8037, 2019.

**Shun Yao** received the B.S. degree from Lanzhou University, Lanzhou, China, in 2017, and the M.S. degree from Santa Clara University, Santa Clara, CA, USA, in 2020, both in computer science. He is currently working toward the Ph.D. degree in electrical engineering with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA.

His research interests include computer vision, deep learning, and radar and satellite remote sensing of precipitation.
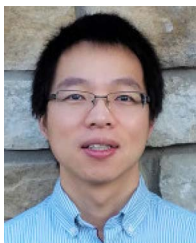
**Haonan Chen** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Colorado State University (CSU), Fort Collins, CO, USA, in 2017.

Since August 2020, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, CSU. He is also an Affiliate Faculty with the Data Science Research Institute, CSU. Before joining the CSU faculty, he was with the Cooperative Institute for Research in the Atmosphere and the NOAA Physical Sciences Laboratory, Boulder, CO, USA, from 2012 to 2020, first as a Research Student, then a National Research Council Research Associate and a Radar, Satellite, and Precipitation Research Scientist. His research interests include remote sensing and multi-disciplinary data science, including radar and satellite observations of natural disasters, polarimetric radar systems and networking, clouds and precipitation processes, big data analytics, and deep learning.

Dr. Chen was an Associate Editor for the *Journal of Atmospheric and Oceanic Technology*, *URSI Radio Science Bulletin*, and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING.

**Elizabeth J. Thompson** received the B.S. degree in meteorology from Valparaiso University, Valparaiso, IN, USA, in 2010, and the M.S. and Ph.D. degrees in atmospheric science from Colorado State University, Fort Collins, CO, USA, in 2012 and 2016, respectively.

She is a Research Meteorologist studying physical processes in the atmosphere, upper ocean, and the air-sea interface. She collects and analyzes measurements of the ocean, air-sea fluxes, and atmosphere to understand the coevolution of atmospheric and oceanic boundary layers. This has included research on precipitation microphysics, and how rain, wind, and sunlight control upper ocean stability. She is currently assessing how such ocean variability relate to the growth or inhibition of clouds and precipitation. Her research focuses on dual- and single-polarization radars, satellites, disdrometers, as well as ocean and air-sea flux instrumentation deployed on ships and autonomous platforms. She has developed algorithms for predicting near-surface ocean stability, estimating precipitation rate from radar, and classifying precipitation type in clouds with radar. Her research activities contribute to greater fundamental understanding of how the ocean and atmosphere interact via processes,such as turbulence, cloud microphysics, precipitation extremes, the global water cycle, atmospheric thermodynamics, ocean stratification, and meteorological phenomena on synoptic- and meso-scales. Her research products support the improvement and evaluation of environmental prediction models, diagnostic nowcasting tools, and operational datasets used to monitor the ocean and atmosphere.

**Robert Cifelli** received the bachelor's degree in geology from the University of Colorado Boulder, Boulder, CO, USA, in 1983, the M.S. degree in hydrogeology from West Virginia University, Morgantown, WV, USA, in 1986, and the Ph.D. degree in atmospheric science from Colorado State University, Fort Collins, CO, in 1996.

He is a Radar Meteorologist with more than 25 years of experience in precipitation research and more than 70 publications in peer-reviewed journals. Since 2009, he has been leading NOAA scientists dedicated to improving precipitation and hydrologic prediction in complex terrain and other geographic regions. He currently leads the Hydrometeorology Modeling and Applications Team in the NOAA Physical Sciences Laboratory, Boulder. A major focus of HMA is to improve understanding of physical processes associated with too much and too little water, including forcings for NOAA's National Water Model and to guide future model development. He completed a detail with the Bureau of Reclamation in 2016 through the President's Management Council Interagency Rotation Program to improve weather, climate, and water forecasts of extreme events to better meet water management needs. He also completed a separate detail with NOAA's Office of Water Prediction in 2017 to advance NOAA's hydrologic prediction capabilities.