FsrGAN: A Satellite and Radar-Based Fusion Prediction Network for Precipitation Nowcasting

Dan Niu, Yinghao Li, Hongbin Wang, Zengliang Zang, Mingbo Jiang, Xunlai Chen, and Qunbo Huang

***Abstract*—Precipitation nowcasting refers to the prediction of small-scale precipitation events at minute and kilometer scales within the upcoming 0 to 2 h, which significantly impacts both human activities and daily life. However, prevailing deep learning models have primarily focused on a single radar echo data source, limiting their ability to effectively capture intricate and rapidly evolving precipitation patterns. Thus, meteorological satellite is considered to supplement radar echo data. To achieve a comprehensive integration of multisource data with enhanced details, a two-stage fusion satellite and radar GAN-based prediction network (named FsrGAN) is proposed. In the first stage, we design a satellite-radar fusion prediction network known as FsrNet. This network employs an encoder-fusion-decoder architecture, where a novel spatial-channel attention (SCA) is proposed to enhance the filtering and fusion of multisource and multiscale features. In the second stage, we introduce a GAN-based network (FusionGAN) that also mines the complementary information of satellite images to sharpen the first-stage predicted radar maps with more details. Experiments are conducted on meteorological dataset in the Yangtze River Delta (YRD) region. The test results exhibit the notably superior performance of our model in terms of image quality and precipitation forecasting metrics in comparison to traditional optical flow-based methods and some well-known deep learning methods (ConvLSTM, ConvGRU, TrajGRU and PredRNN++). More importantly, our fusion model using satellite and radar data demonstrates the ability to predict convective initiation.**

***Index Terms*—Generative adversarial network, multisource, precipitation nowcasting, spatiotemporal fusion.**

Manuscript received 30 September 2023; revised 12 December 2023 and 19 February 2024; accepted 9 March 2024. Date of publication 18 March 2024; date of current version 27 March 2024. This work was supported in part by the Beijige foundation of NJIAS under Grant BJG202305, in part by Open Project of the High Impact Weather Key Laboratory of CMA, Key Laboratory of Smart Earth under Grant KF2023YB03-05, in part by National Natural Science Foundation of China under Grant 42005120 and Grant 62374031, in part by the Beijige foundation of NJIAS under Grant BJG202209, in part by the National key R&D Program of China under Grant 2022YFC3003904, Grant 2019YFE0110100, and Grant 2018YFC1506905. *(Corresponding authors: Hongbin Wang; Zengliang Zang.)*

Dan Niu and Yinghao Li are with the School of Automation, Southeast University, Nanjing 210096, China.

Hongbin Wang is with the Key Laboratory of Transportation Meteorology of China Meteorological Administration, Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210041, China (e-mail: wanghb@cma.gov.cn).

Zengliang Zang is with the High Impact Weather Key Laboratory of CMA, and College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China (e-mail: zzlqxxy@163.com).

Mingbo Jiang is with the Key Laboratory of Smart Earth, Beijing 100094, China.

Xunlai Chen is with the Shenzhen Key Laboratory of Severe Weather in South China, and Shenzhen Meteorological Bureau, Shenzhen 518040, China.

Qunbo Huang is with the 93110 Troops, PLA, Beijing 100843, China.

Digital Object Identifier 10.1109/JSTARS.2024.3376987

I. Introduction

PRECIPITATION nowcasting is a high-resolution forecast of rainfall intensity in the next few hours (0–2 h) for a specific region [1] and is one of the important issues in weather forecasting [2]. It is closely related to human life and affects decision-making in various industries, such as agricultural production [3], [4], transportation [5], [6], and aviation control [7]. Given the complexity of atmospheric dynamical processes and the dynamic characteristics of precipitation events, precipitation nowcasting has been a challenging research hotspot in meteorology [8], [9], [10].

Traditional rainfall forecasting methods, also known as numerical weather prediction (NWP) [11], are based on hydrodynamic and thermodynamic equations describing complex atmospheric motions and predicting future atmospheric states. However, NWP is usually very sensitive to perturbations in initial and boundary conditions, which leads to the inability to provide accurate 0–2 h precipitation forecasts [12]. In addition, the computational cost of NWP is high and time-consuming to solve even on modern supercomputers [13]. With the development of observational techniques, faster methods based on radar echo extrapolation have become the mainstream method for precipitation nowcasting [14], [15], [16]. The optical flow based radar echo extrapolation method has high computational efficiency [17] and is now widely adopted in the short-term precipitation nowcasting system. It estimates the convective motion by obtaining the future radar echo maps and then predicts the rainfall area. However, the two important assumptions of Lagrangian persistence and smooth kinematic field [18] in the optical flow methods have some limitations, which make it difficult to cope with the complex and highly dynamic nonlinear precipitation variations [19].

The rise of artificial intelligence has triggered the creation of a series of deep learning-based radar echo extrapolation methods [20], [21], [22], which do not rely on complex meteorological knowledge, but are trained to predict rainfall by existing radar echo datasets. Precipitation nowcasting can be treated as a video prediction problem [23], where future radar echo sequences are predicted based on past radar echo sequences. First, some approaches based on recurrent neural networks (RNNs) provided useful insights [24], [25]. Shi et al. [19] extended the input-to-state and state-to-state transitions in LSTM to convolutional structures and proposed the convolutional LSTM (ConvLSTM) model, which outperforms the optical

flow method and fully connected LSTM in grid precipitation nowcasting. Considering that the convolutional in ConvLSTM model is position invariant, they further proposed the trajectory gated recursive unit (TrajGRU) model [7] to adapt to the rotational and scaling motion patterns of clouds by actively learning different positional structures of the recursive connections. In [26], Wang et al. proposed ST-LSTM based on LSTM units and applied it to predictive recurrent neural networks (PredRNN). To alleviate the difficulty of gradient propagation in PredRNN, they further proposed an improved PredRNN++ [27] with a gradient unit module to adaptively capture short-term and long-term dependencies. In addition, memory-in-memory (MIM) networks [28] and MotionGRU units [29] are also proposed to capture complex spatiotemporal motions more precisely. However, the above approaches mainly focus on the improvement of the model structure without considering the one-sidedness of single data source in some complex tasks [30], [31].

Precipitation is also related to many other factors, such as wind speed, temperature, and humidity. Thus, in recent years, many meteorological researchers have investigated multisource data fusion methods [32], [33]. Wehbe et al. [34] proposed a geographically weighted regression algorithm. Bouget et al. [35] introduced an extrapolation prediction algorithm that fuses wind and radar data. In addition, Zhou et al. [36] designed a lightning prediction network, named LightningNet, which simultaneously uses three different sources of data, including radar data, satellite data, and lightning data to make predictions of lightning occurrence. Furthermore, LightNet+ fuses lightning data (LIG), weather research, and forecasting data (WRF), and automated weather station data to predict the lightning field at future moments [37]. For precipitation, radar can observe the internal distribution characteristics of convective storms to reflect the specific precipitation intensity at the current moment, but it is difficult to predict convective initiation because the radar is not sensitive to the movement of clouds before the precipitation process is formed. In addition to radar echo maps, meteorological satellite data can also reflect precipitation intensity [38]. Meteorological satellites are capable of observing convective initiation and the development characteristics of convective cloud tops. Therefore, satellites data can relieve the limitations of extrapolation using single radar echo data. In addition, the widely used MSE or MAE loss functions smooth the prediction results and lead to blurry extrapolation radar maps [5], [21], [25], [39]. The reasons can be attributed to two aspects. 1) MSE losses are very sensitive to outliers, which may lead to overaveraging of the prediction results [21]. 2) The predictability of the radar echo is related to the scale of the echo, and the MSE is not able to focus well on the small-scale fine structure of certain regions, leading to the loss of these details in the extrapolation process [5].

In this work, we propose a two-stage GAN-based fusion prediction model to take advantage of multisource complementary information of satellite and radar data for precipitation forecasting. Our method fuses radar and satellite data as inputs, and is divided into the fusion extrapolation stage and detail refinement stage. The first stage named FsrNet adopts a encoder-fusion-decoder structure and designs a novel spatial-channel attention (SCA) mechanism to filter and merge multisource information for radar echo extrapolation. In the second stage, a GAN-based detail enhancement model (FusionGAN) is proposed to refine the first-stage prediction and sharpen the predicted maps with more details. The contributions of this work are as follows:

1) We propose a multisource spatiotemporal prediction model (named FsrGAN), which can mine and fuse the effective precipitation complementary information from radar and satellite data to achieve the prediction of convective initiation, as well as enhance the prediction accuracy and small-scale details for the radar echo extrapolation.
2) An encoder-fusion-decoder framework is designed in the fusion extrapolation stage, which mines complementary information of multisource data for prediction. In the fusion part, a novel SCA module is proposed to adaptively fuse the multiscale spatiotemporal discriminative features of multisource data.
3) For dealing with the blurry radar extrapolation, a GAN-based detail refinement model (FusionGAN) is designed, which also fuses the complementary information of history real satellite images to sharpen the first-stage predicted radar maps with more details.

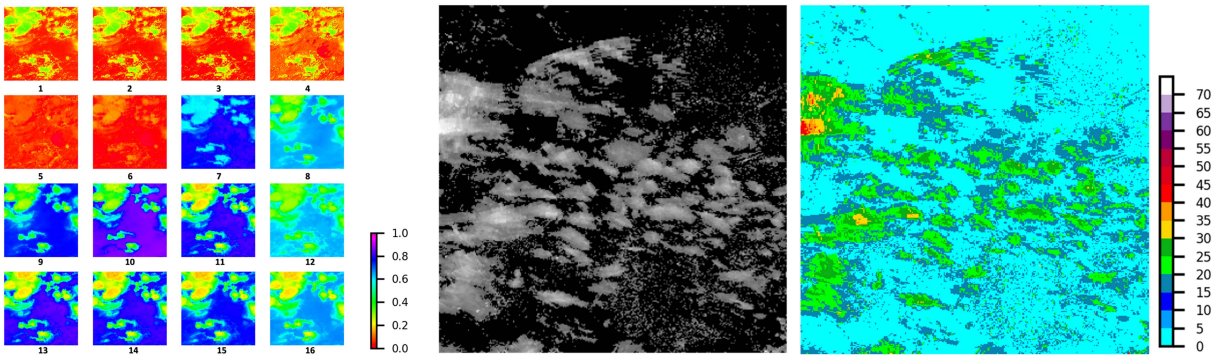
II. Preliminary

Weather radar and satellite are two important means of remote observations. While satellites offer a panoramic view, facilitating the monitoring of cloud formation and movement, they lack precision in quantifying precipitation intensity. Conversely, weather radar's perspective spans from the surface to the atmosphere, enabling the capture of radar echo intensity and thus precipitation intensity and distribution in weather systems, however, it is not sensitive to the movement of clouds. This means that satellite data can complement some of the limitations of extrapolation using radar echo data alone.

A. Multisource Spatiotemporal Data

As shown in Fig. 1, the model inputs include two types of spatiotemporal data, namely satellite data (S) and radar echo observations (R). The radar echo dataset and meteorological satellite dataset used in this article are a subset of the three-year radar echo intensities and Himawari-8 satellite cloud maps provided by the Nanjing Joint Institute for Atmospheric Sciences (NJIAS) with synchronized temporal resolution, and the data cover part of the Jiangsu province with an area of 300 km $\times$ 300 km.

Satellite Data: There are 16 observation channels in satellite data, including 6 visible light channels and 10 infrared channels. The visible light channels receive solar radiation reflected by clouds and the surface during the day, and the value of the visible light channel at night is almost 0, while the infrared channel obtains radiation information around the clock. Thus, the visible light channels (1–6) are not used in the fusion prediction. Considering the large number of satellite data, we will group the satellite channels to simplify the model inputs. We count the range of values for different channels of NJIAS satellite data from 2019 to 2021, as shown in Fig. 3. In addition,



(a) (b)

Fig. 1. Examples of visualization of multisource data. (a) Color visualizations of 16 satellite observation channels at 13:50 on July 25, 2019. (b) Visualizations of radar echo data at 01:10 on June 1, 2019 (grayscale on the left and color on the right).

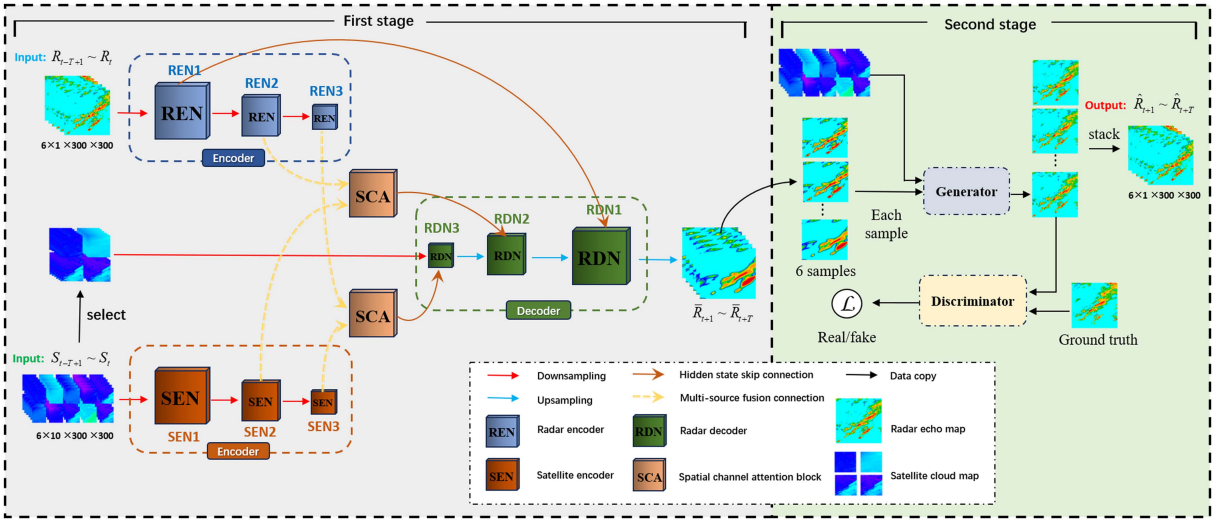


Fig. 2. Method overview. In the first stage, multiscale spatiotemporal complementary features from input radar and satellite sequences are extracted by the dual-path encoders. The multisource and multiscale feature information is further fused by the proposed SCA block, and connected to the decoder by the skip connections to generate the predicted radar echo sequence. In the second stage, the first-stage radar echo extrapolation sequence and past satellite sequence are fed into the generator, which learns against the discriminator to sharpen the blurry predictions with more details.

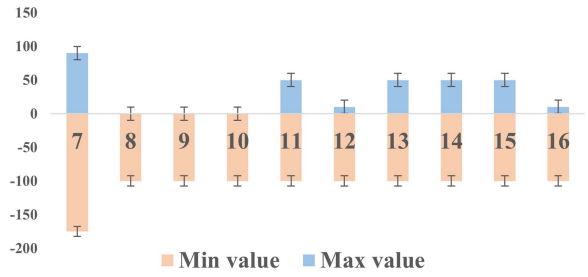


Fig. 3. Range of values for channels of satellite data.

Pearson correlation analysis is used to analyze the correlation between channels (7–16), and the results are shown in Table I. To simplify the subsequent inputs, the satellite channels will be grouped here. From the combination of the range of values and Pearson correlation analysis, channels (7–16) will be divided into four groups, which are (1) channel 7, (2) channels 8, 9, and

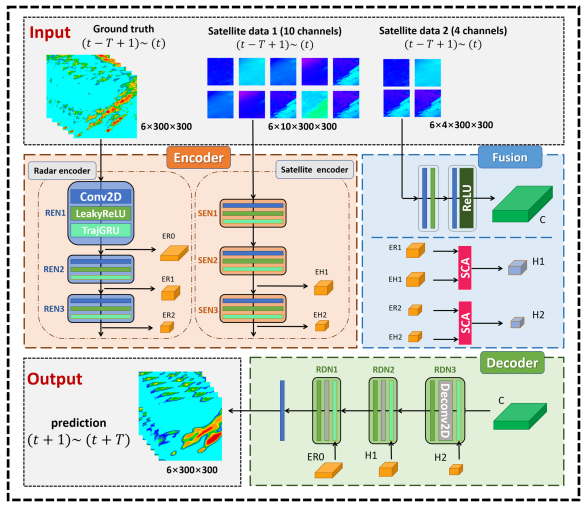


Fig. 4. Architecture of the satellite-radar fusion prediction network (FsrNet).

TABLE I
CORRELATION COEFFICIENT MATRIX OF SATELLITE CHANNEL 7–16

channel	7	8	9	10	11	12	13	14	15	16
7	1.000	0.723	0.758	0.767	0.865	0.859	0.870	0.856	0.835	0.813
8	0.723	1.000	0.981	0.907	0.778	0.752	0.784	0.787	0.794	0.807
9	0.758	0.981	1.000	0.967	0.837	0.821	0.842	0.845	0.854	0.872
10	0.767	0.907	0.967	1.000	0.897	0.892	0.897	0.903	0.916	0.940
11	0.865	0.778	0.837	0.897	1.000	0.994	0.999	0.998	0.993	0.980
12	0.859	0.752	0.821	0.892	0.994	1.000	0.993	0.992	0.987	0.977
13	0.870	0.784	0.842	0.897	0.999	0.993	1.000	0.999	0.994	0.980
14	0.856	0.787	0.845	0.903	0.998	0.992	0.999	1.000	0.998	0.987
15	0.835	0.794	0.854	0.916	0.993	0.987	0.994	0.998	1.000	0.994
16	0.813	0.807	0.872	0.940	0.980	0.977	0.980	0.987	0.994	1.000

10, (3) channels 11, 13, 14, and 15, and (4) channels 12 and 16. The different satellite channels in each group have similar range of values and high correlation, as well as similar visualizations.

Radar Data: The radar echo dataset used in this work is a subset of the three-year weather radar intensities provided by NJIAS from 2019–2021. The radar echo maps have resolution of 300 × 300 pixels and also cover a 300 km × 300 km area. They are captured at a frequency of 10 min. It means a spatial resolution of 1 km and a temporal resolution of 10 min. Given that different satellite channels have different spatial resolutions (0.5 km, 1 km or 2 km), a data preprocessing process including screening and interpolation is implemented to achieve harmonized temporal and spatial resolutions between satellite and radar data.

B. Problem Formulation

Short-term precipitation forecasts typically refer to prediction of precipitation over the upcoming 1–2 h. The radar echo map reflects regional precipitation intensity by acquiring the magnitude of the radar echo. Consequently, the extrapolation of radar echo serves as a valuable tool for short-term precipitation prediction. The radar echo extrapolation constitutes a form of time-series prediction. The historical radar echo maps are employed as inputs to the model, which subsequently predicts future radar echo images. In this work, fusion prediction by multisource data is proposed, combining both satellite channels and radar echo maps for precipitation forecasting.

Using the tensor $R_t, S_t \in \mathbb{R}^{C\times W\times H}$ to denote the radar echo data and the satellite data observed at time t, respectively, where C, W, and H denote the number of channels, the width, and the height of the time series data. Thus, the radar echo sequence can be represented as a tensor sequence $\{R_{t-T+1}, \ldots, R_{t-1}, R_t, \ldots, R_{t+T}\}$. Let $X = \{R_{t-T+1}, \ldots, R_t\}$ is the past radar observations as the model inputs and $Y = \{R_{t+1}, \ldots, R_{t+T}\}$ is the future real maps. Similarly, let $S = \{S_{t-T+1}, \ldots, S_t\}$ is the input past satellite observations. In this work, we train a neural network parameterized as θ for multisource fusion prediction, and the radar echo extrapolation problem can be described as

$$\hat{Y} = \underset{Y}{\operatorname{argmax}} \quad P(Y \mid X, S; \theta) \tag{1}$$

where P is the conditional probability, $\hat{Y} = \{\hat{R}_{t+1}, \ldots, \hat{R}_{t+T}\}$ denotes the final predicted radar sequence of length T by the model.

III. METHOD

A general overview of the methodology is shown in Fig. 2. We first propose the FsrNet to explore and mine the complementary roles of satellite data and radar echoes in precipitation forecasting. Considering the blurry problem of the predicted radar echo maps, we use the prediction results of FsrNet ($\{\bar{R}_{t+1}, \ldots, \bar{R}_{t+T}\}$) as a pre-extrapolation and further refine the image details through the GAN-based network (FusionGAN) in the second stage ($\{\hat{R}_{t+1}, \ldots, \hat{R}_{t+T}\}$). Since the two data sources have the same temporal and spatial resolution, the two input sequences are processed separately using similar-structured encoders. The multiscale representations are extracted by downscaling blocks and then fed into the multisource fusion (MSF) module. The MSF module employs a novel SCA to filter and fuse the spatiotemporal complementary features of the radar and satellite data, which update the lower level features for radar echo prediction. In the second stage, a detail refinement network based on the GAN model is designed to extract and fuse satellite data features using multiscale blocks to refine the details of the first-stage extrapolation results.

A. First Stage: FsrNet

Fig. 4 illustrates the detailed structure of FsrNet, which consists of three parts, i.e., the multisource encoder module (MSE), the MSF, and the fusion prediction decoder module (FPD), which are denoted with orange, blue, and green background in Fig. 3, respectively.

Multisource Encoder: The encoder is divided into two parts, which encode and extract the multiscale features of the radar and satellite data separately. The radar encoder and the satellite encoder have similar structures, consisting of three stacked REN/SEN layers. The REN/SEN layer consists of a succession of a bilinear downsampling layer (Conv2D) followed by a leaky rectifier linear unit (LeakyReLU) and trajectory gated recurrent unit (TrajGRU [7]). The Conv2D layers reduce the input image size and double the number of feature maps. The LeakyReLU

layer enables the network to model nonlinear relations. TrajGRU [7] is employed for effectively capturing spatiotemporal correlations and features. Note that the parameters in REN and SEN are different since there is only one channel in the radar echo data, whereas there are ten channels in the satellite data. The different scale spatiotemporal features ER_0, ER_1, ER_2 and EH_1, EH_2 will be extracted by the three REN/SEN layers. Then, ER_0 will be used as the initial state of the RDN1 layer of the decoder. ER_1, ER_2 and EH_1, EH_2 will be input to the fusion module, and then the fused middle- and large-scale spatiotemporal features H_1, H_2 are connected to RDN2 and RDN3 layers of the decoder by global skip connections, respectively.

Multisource Fusion Module: The fusion module is represented in Fig. 4 with a blue background and is divided into two parts. In the upper right part, satellite data are downsampled and directly sent to the decoder. The comparison tests show that not all ten satellite channels but four selected channels are enough to supply satellite complementary spatiotemporal features in the fusion part. Obviously, lightweight satellite inputs are conducive to reduce the training and prediction loads. In this work, only one channel is selected in each satellite group and then lightweight four-channel satellite input is formed. In detail, channel 7, 8, 13, 16 (representing mid-infrared channel, infrared water vapor channel, infrared window channel and infrared channel, respectively) are selected based on meteorologists' suggestions and test comparisons in this work. Then, the high-level satellite features are extracted through a two-layer downscaling block, and they are used as the input unit of the TrajGRU layer in the RDN3 module to enhance the prediction ability for convective initiation. The middle and large-scale spatiotemporal features ER_1, ER_2, EH_1, and EH_2 are used to generate the multiscale spatiotemporal fusion features H_1 and H_2 by the proposed SCA and then are connected as the initial hidden state of the TrajGRU layer in the RDN1 and RDN2 modules.

Fusion Prediction Decoder: The decoder is shown in the green background of Fig. 4. It is formed by stacking three layers of RDNs, where multiscale spatiotemporal features from the radar and satellite sequence in the encoder part will be adaptively rescaled and fused to achieve a high-to-low level residual prediction. The bottom layer RDN3 takes the satellite feature map C as the unit input and the spatiotemporal fusion information H_2 as the initial hidden state of the TrajGRU layer. Similarly, the RDN2 employs the middle-scale fusion feature H_1. The RDN1 takes the ER_0 from the radar encoder as the initial hidden state. The skip connections will ease and combine the flow of multiscale spatiotemporal fusion features to recover the lost information and generate better predictions.

B. SCA Module

In this work, a novel SCA block is proposed to filter and fuse the spatiotemporal features of radar and satellite sequences, as shown in Fig. 5. EH and ER represent the multiscale spatiotemporal features from the satellite encoder and radar encoder, respectively. The output H of the SCA serves as the initial hidden state of the TrajGRU layer in the decoder part.

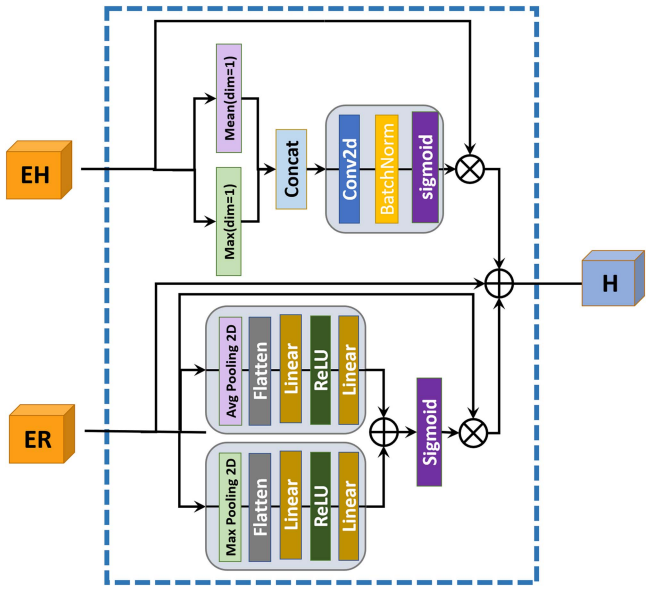


Fig. 5. Novel SCA block.

The radar and satellite data have different characteristics. For radar echo data, it is necessary to pay attention to its spatial characteristics, while satellite data are more important to extract its channel features. Instead of treating all features equally, SCA is proposed for spatialwise and channelwise weightings, and strengthens the discriminative learning ability and the representational power. For the radar feature input, the average and maximum values on the spatial scale are first calculated, and summed by a two-layer convolutional network. Average pooling is used to retain more common features while max pooling can filter out more recognizable features. Both max and average-pooled features are simultaneously used to greatly improve the representation power of networks. Next, the spatial feature map is obtained by the sigmoid activation function. As for the satellite feature data, the average and maximum values on the channel scale are calculated to obtain the channel feature map. Finally, in order to increase the weight of the radar echo features, the spatial feature map, the channel feature map, and the original radar echo feature are summed up to get the final output of SCA. Local skip connections can stabilize the network training and ease the flow of spatiotemporal information.

C. Second Stage: FusionGAN

Although FsrNet can fuse satellite and radar echo data to achieve higher precision for precipitation forecasting. However, the predicted radar echo maps still tend to be blurry due to the widely used loss functions (e.g., MSE/MAE). Therefore, we designed a GAN-based detail refinement model (named FusionGAN) in the second stage to further fuse with satellite data to achieve more small-scale detailed prediction. FusionGAN consists of a generator and a discriminator. The discriminator is only used during model training, and only the generator is used for formal prediction.

Generator: Fig. 6 illustrates the structure of the FusionGAN generator, which also can be divided into three parts: encoder,

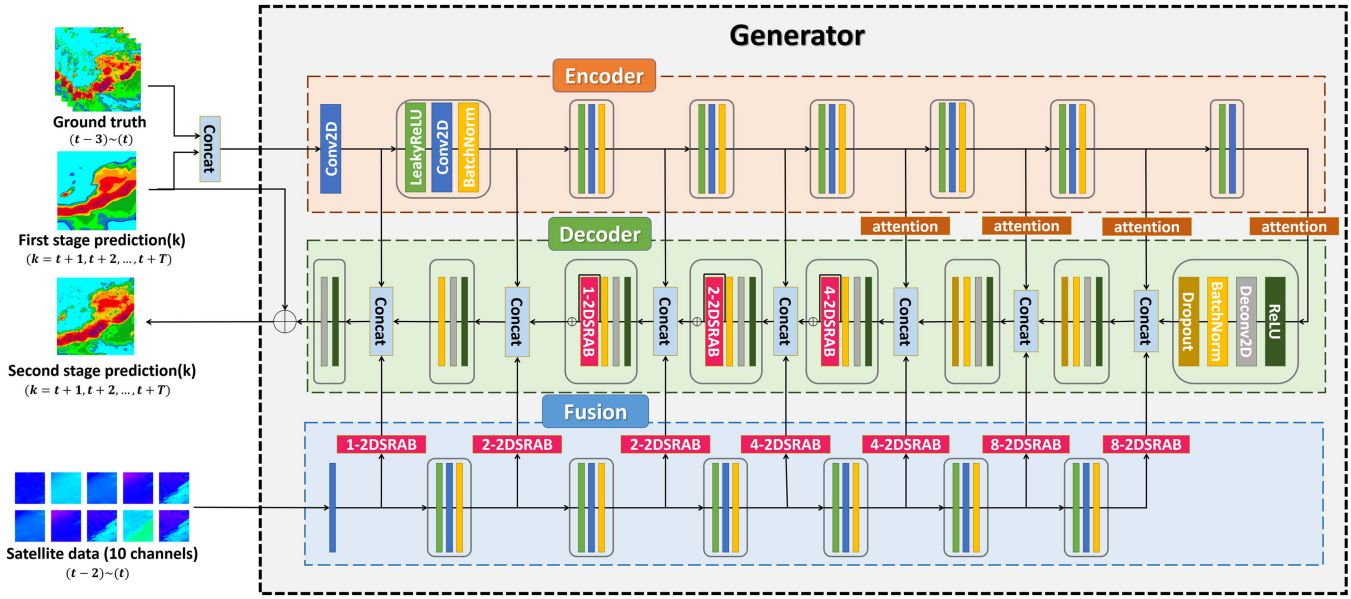


Fig. 6. Architecture of the generator in FusionGAN.

decoder, and fusion network. In this work, the first-stage prediction radar echo maps $\{\bar{R}_{t+1}, \ldots, \bar{R}_{t+T}\}$, history four real radar echo maps $\{R_{t-3} \sim R_t\}$, as well as history three satellite maps blue $\{S_{t-2} \sim S_t\}$ are input to the generator to obtain the final predicted radar echo sequence $\{\hat{R}_{t+1}, \ldots, \hat{R}_{t+T}\}$. In this case, the generator can not only enjoy relatively accurate prediction results generated by FsrNet in the first stage and focus on improving the small-scale details of the prediction maps, but also can take advantage of the spatiotemporal features of the historical real radar echo and satellite maps to further improve the prediction accuracy.

In the encoder part of the generator, an eight-layer convolutional network, together with BatchNorm and LeakyReLU modules are used. Similarly, the decoder also consists of an eight-layer deconvolutional network, together with a BatchNorm module, ReLU module, and Dropout. Multiscale skipping connections are used between the encoder and the decoder to capture and preserve the multiscale spatiotemporal features obtained in the encoder. The attention module in Fig. 6 employs the self-attention block in [40], which help the model to capture spatial features between different regions of the radar echo maps and enhance its generalization under different precipitation events. In order to better utilize the high-level spatiotemporal fusion representations in the decoder to guide the updating of the low-level features, the N-RSSAB block (N denotes the number of stacks) in [8] is introduced in the decoder and the fusion module, where the 3-D convolution operation in N-RSSAB is adapted with a 2-D convolution to form our N-2DSRAB, which can adaptively rescale the multiscale spatial fusion features in the decoder and fusion network to improve the performance of the residual prediction at different scales.

Discriminator: The structure of the discriminator is shown in Fig. 7. Both the refined maps by the generator and the real radar echo images are fed into the discriminator. The training objective is to determine that the true radar echo maps are true and the refined maps are false. By adopting a five-layer convolutional structure, the discriminator could obtain good discriminative performance. The last layer is composed of Conv2D and Sigmoid.

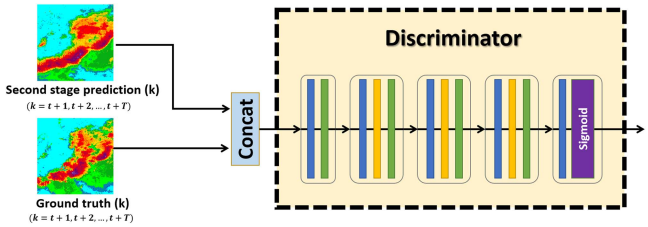


Fig. 7. Architecture of the discriminator in FusionGAN.

D. Loss Functions

First Stage Loss: To improve the prediction accuracy of heavy rainfall, we use weighted MSE and MAE as loss function in the first-stage FsrNet to enhance the prediction accuracy for heavy rainfall, which assign different weights to pixels with different precipitation intensities. The weights are defined as in (2), where dBZ is the radar echo value

$$w(z) = \begin{cases} 1, & z < 25\text{dBZ} \\ 2, & 25\text{dBZ} \le \text{z} < 35\text{dBZ} \\ 6, & 35\text{dBZ} \le \text{z} < 40\text{dBZ} \\ 10, & 40\text{dBZ} \le \text{z} < 45\text{dBZ} \\ 20, & 45\text{dBZ} \le \text{z} < 50\text{dBZ} \\ 60, & z \ge 50\text{dBZ} \end{cases} \tag{2}$$

$$\text{Loss}_{\text{stage1}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i,j}\left(w_{t,i,j}\left(\left(R_{t,i,j} - \bar{R}_{t,i,j}\right)^2 + \left|R_{t,i,j} - \bar{R}_{t,i,j}\right|\right)\right) \tag{3}$$

TABLE II
CSI AND HSS SCORES OF THE MODELS WITH THE THRESHOLDS OF 25, 35, 40, 45, AND 50 DBZ

Models	CSI↑					HSS↑				
	25dBZ	35dBZ	40dBZ	45dBZ	50dBZ	25dBZ	35dBZ	40dBZ	45dBZ	50dBZ
ConvGRU [7]	0.639	0.482	0.341	0.232	0.210	0.758	0.634	0.493	0.364	0.333
PredRNN++ [27]	0.651	0.490	0.360	0.263	0.230	0.768	0.642	0.517	0.405	0.363
TrajGRU [7]	0.653	0.508	0.356	0.248	0.202	0.770	0.659	0.511	0.387	0.321
ConvLSTM [19]	0.644	0.491	0.355	0.256	0.223	0.762	0.643	0.509	0.394	0.353
ROVER [18]	0.558	0.403	0.274	0.199	0.174	0.690	0.556	0.415	0.319	0.284
FsrGAN (No satellite)	0.691	0.533	0.425	0.324	0.281	0.789	0.671	0.580	0.482	0.438
FsrGAN (Ours)	**0.707**	**0.563**	**0.449**	**0.337**	**0.292**	**0.801**	**0.701**	**0.606**	**0.495**	**0.447**

TABLE III
COMPARISON OF IMAGE QUALITY INDICATOR RESULTS

Models	RMSE↓	SSIM↑	Sharpness↑
ConvGRU [7]	0.071	0.764	66.307
PredRNN++ [27]	0.068	0.766	67.687
TrajGRU [7]	0.067	0.771	91.346
ConvLSTM [19]	0.069	0.769	74.139
ROVER [18]	0.080	0.701	32.594
FsrGAN (No satellite)	0.065	0.776	103.325
FsrGAN (Ours)	**0.063**	**0.778**	**104.438**

TABLE IV
ABLATION STUDY FOR SCA (50 DBZ THRESHOLD)

Model	Module	CSI↑	HSS↑
TrajGRU [7]	-	0.202	0.321
FsrNet	$w/o\ SCA$	0.277	0.430
FsrNet	$SCA\ only\ to\ R\&SEN3$	0.280	0.433
FsrNet	$SCA\ only\ to\ R\&SEN2$	0.285	0.437
FsrNet	$w\ SCA$	**0.288**	**0.442**

w and w/o represent with and without, respectively.

where T is the length of the predicted radar echo sequence, $w_{t,i,j}$ is the weight of (i, j) pixels at moment t. $R_{t,i,j}$ and $\bar{R}_{t,i,j}$ are the (i, j) pixel values of the ground truth and the first-stage predicted image at moment t, respectively.

Second Stage Loss: Although the original GAN model can generate realistic results from noise, it is difficult to produce expected constraints on the output of the generator. The CGAN model can be used to better match the required image in the input of the generator. The discriminator determines whether the image is from the generator or the real image, and consequently directs the generator to produce an image that satisfies the real data distribution. They compete with each other for better performance. The loss function of CGAN is as follows:

$$Loss_{CGAN} = \mathbb{E}_{x,z}[\log D(z,x)] + \mathbb{E}_z[\log(1 - D(z, G(z)))] \tag{4}$$

where x represents the real image and z represents the conditional input of the generator. However, it is difficult to constrain the generated image on a pixel scale just using the loss of CGAN. To achieve better echo intensity prediction accuracy, the MSE and MAE functions are also incorporated into the loss function in the second stage, described as

$$\begin{aligned} Loss_{stage2} &= w_1 \left(\mathbb{E}_{x,z}[\log D(z,x)] - \mathbb{E}_z[\log(1 - D(z, G(z)))]\right) \\ &\quad + w_2(x - G(z))^2 + w_3 \left|x - G(z)\right| \end{aligned} \tag{5}$$

where w_1, w_2, w_3 denote the weights of each part. Here $w_1 = 1$, $w_2 = w_3 = 200$ are set to match the orders of magnitude.

IV. EXPERIMENTS

A. Experiments Setup

Our goal is to predict the radar echo sequence for the next hour, and the inputs to the model include the radar echo observations and satellite data for the previous hour. Since the spatial resolution is 1 km and the temporal resolution is 10 min, the radar echo input dimension is $6 \times 1 \times 300 \times 300$. The radar echo intensity is linearly transformed to pixel values and cropped to between (0, 255) by the formula $\text{pixel} = 255 \times \frac{\text{dBZ}}{70}$. Besides, the points with values smaller than 15 dBZ and larger than 70 dBZ are set to 0. Moreover, the satellite data contains 16 channels with different wavelengths, it is necessary to standardize the data magnitude on each channel and perform a linear transformation to map the raw data into the range of (0, 1). Then a 12-frame wide sliding window is used to slice the radar echo sequence and the satellite image instances respectively to obtain a 12-frame long multisource spatiotemporal sequence, where the first 6 frames are used for input and the last 6 frames are used for prediction. Our dataset is formed with a total of 246 days, and 28 953 frames for training samples and 7992 frames for test samples. Since satellite data and radar data have the same temporal and spatial resolutions, the dataset divisions are kept consistent.

B. Metrics

To achieve "no blurry" and high-accuracy radar echo prediction, we employed both image quality evaluation and prediction accuracy indexes to evaluate the performance of the proposed network.

First, two commonly used precipitation nowcasting metrics, including critical success index (CSI), and Heidecker Skill Score (HSS), will also be used to evaluate the forecast accuracy. To give an all-round evaluation of the algorithms' nowcasting

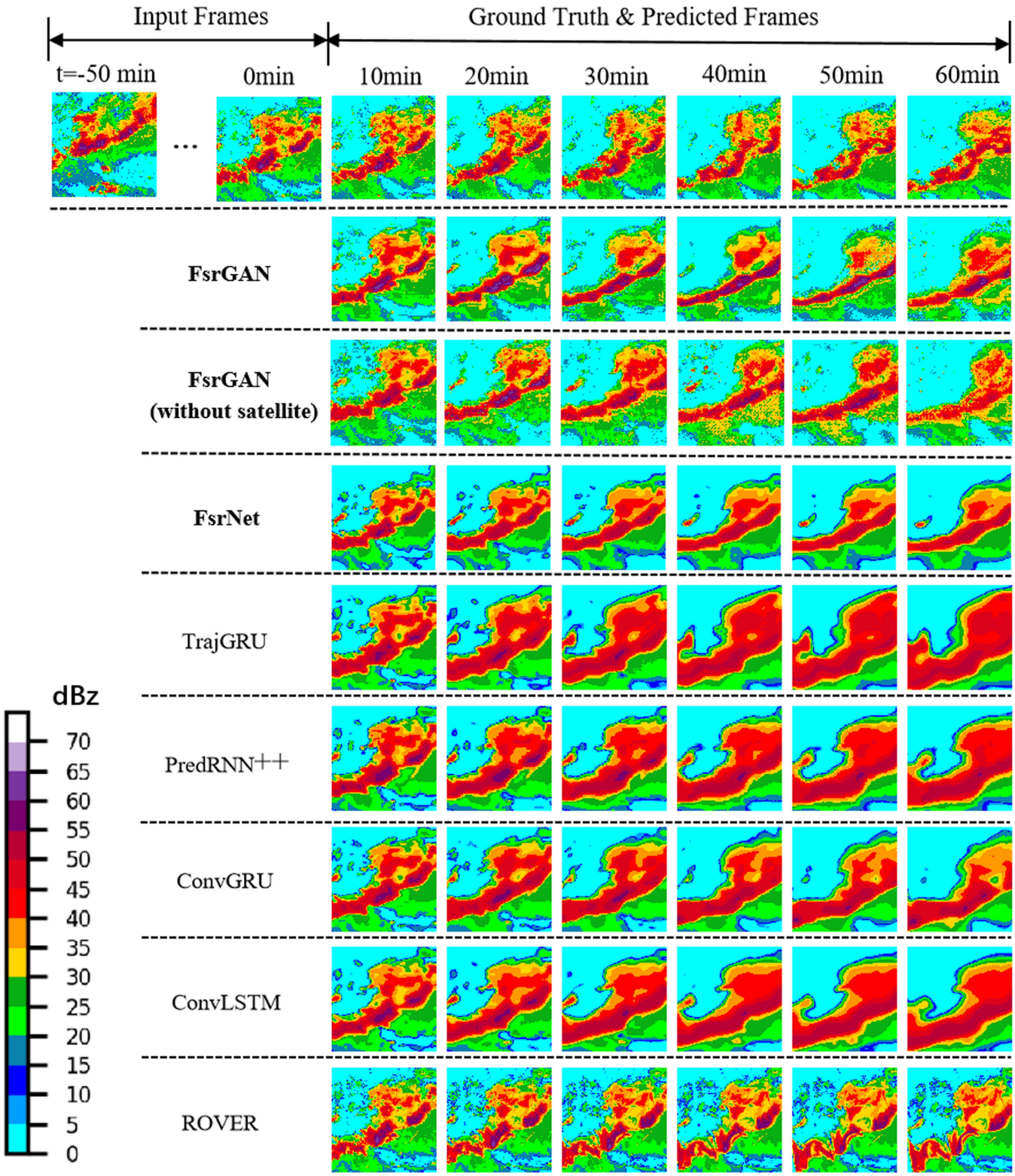


Fig. 8. Visualization comparison of different models. Our method outperforms other deep learning methods and traditional ROVER method in terms of clarity and prediction accuracy (especially in heavy rainfall). The figure shows the precipitation process starting at 19:50:00 on July 6, 2019 in some areas of Jiangsu province.

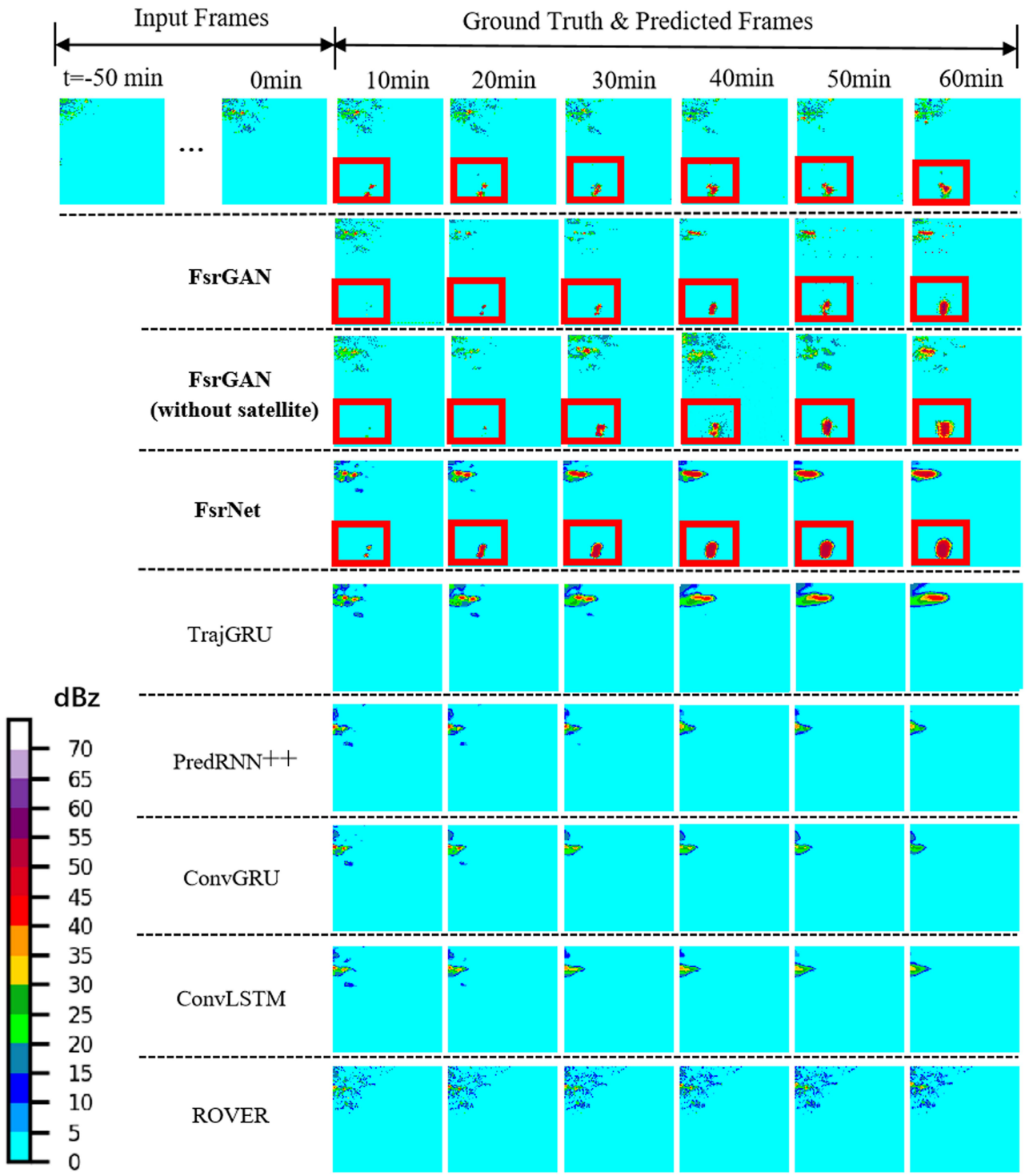


Fig. 9. Visualization comparison of convective initiation prediction by different models. The figure shows the precipitation process starting at 14:30:00 on August 1, 2019 in some areas of Jiangsu province.

performance, we evaluate the skill scores for multiple thresholds (25, 35, 40, 45, and 50 dBZ) that correspond to different rainfall levels. We first convert the pixel values of forecast or true-ground images to 0 or 1 according to the thresholds, and then calculate TP (forecast $=1$, truth $=1$), FN (forecast $=0$, truth $=1$), FP (forecast $=1$, truth $=0$), and TN (forecast $=0$, truth $=0$), respectively. Finally, the two forecast metrics are calculated as follows:

$$CSI = \frac{TP}{TP+FP+FN}$$
$$HSS = \frac{2\times(TP\times TN - FP\times FN)}{(TP+FN)(FN+TN)+(TP+FP)(FP+FN)}. \quad (6)$$

Second, to compare echo images generated by our model with the real images considering image structure and clarity, some image quality assessment metrics are used to evaluate the small-scale detail performance of the predicted echo maps. In this work, we use the root mean square error (RMSE) for evaluating the overall error, and the structural similarity index (SSIM) [41] and the sharpness difference (SD) for measuring the overall similarity between the two images.

C. Quantification Results

In the experiments, we compare our FsrGAN model with one optical-flow based model: ROVER [18], four well-known deep-learning models: ConvLSTM [19], ConvGRU, TrajGRU [7], and PredRNN++ [27] on the forecasting evaluation indexes CSI, HSS, and on the image quality evaluation indexes RMSE, SSIM, and sharpness. Moreover, we conduct ablation experiments in the FusionGAN stage to investigate the effect of satellite data on radar echo map refinement and extrapolation accuracy. "No satellite" represents that satellite data are deleted in the fusionGAN stage.

Tables II and III present the one-hour precipitation nowcasting skill scores for the different models as well as the image quality evaluation scores, where "↑," and "↓" imply that higher, and lower values are better, respectively. The best results under each metric are labeled in bold, and the second best results are underlined. As shown in Tables II, and III, the deep learning methods are significantly better than the ROVER-based optical flow method because of their nonlinear fitting advantage to learn more complex spatiotemporal patterns from the data. Among the deep learning models, the proposed FsrGAN network extracts and fuses the multiscale spatiotemporal complementary features from radar and satellite data, and performs the best at all five metrics. It is worth noting that our method have more significant improvements on heavy rainfall forecasting (45 and 50 dBZ), which usually is a more difficult task. Specifically, the CSI of the proposed FsrGAN is 28.0% higher than PredRNN++ and 32% higher than ConvLSTM at the 45 dBZ threshold. For the 50 dBZ threshold, our method also achieve 27% higher CSI than PredRNN++, and 30.9% higher than ConvLSTM. Moreover, the HSS at the 50 dBZ threshold is also much improved by 23.1% relative to that under the PredRNN++ method of 26.6% relative to that under the ConvLSTM method.

Furthermore, considering the image quality evaluation indexes, our method can also achieve the best results in RSME, SSIM, especially Sharpness metrics. It improves the Sharpness score of TrajGRU (second best) from 91.346 to 104.438 (increase of 14.3%). It is clear that our model is beneficial to generate sharper radar echo prediction with more small-scale details. In addition, the proposed FsrGAN that incorporates satellite data in the FusionGAN stage can further improve the prediction accuracy and refinement for radar echo extrapolation by comparing the FsrGAN with FsrGAN (No satellite) in Tables II and III.

A visual comparison of the radar echo maps by different methods is shown in Fig. 8. It shows the 1-h precipitation movement. It can be seen that the ROVER method has no echo intensity change and lower prediction accuracy although it retains the small-scale echo details. The predictions of other deep learning methods are relatively fuzzy and lose a lot of details. They also tend to exaggerate the forecasting scale and the echo intensity tends to be overestimated. By comparison, our method achieves good prediction performance and the echo intensity and position are more consistent with the real echo maps.

In order to verify the prediction ability for convective initiation. Fig. 9 shows a case of convective initiation. From observing the region circled by the red box, it is clear that the echo intensity of circled region on the input radar echo maps are all 0, but convective initiation occurs from the future six real radar echo maps. In this case, the echo intensity of circled region are still 0 on the predicted maps by other methods. However, the echo intensity is effectively predicted by our method on the circled region due to the complementary information from the satellite data.

D. Ablation Studies

To further explore the role of SCA in the first stage, Table IV gives the test results of FsrNet with and without SCA. It is clear that when the SCA block is removed, the prediction accuracy (CSI and HSS) will decrease significantly. In addition, only adding SCA modules to R&SEN3 (REN3 and SEN3) or R&SEN2 (REN2 and SEN2) will also weaken the accuracy performance of the model due to the loss of spatiotemporal information from multisource fusion. It is demonstrated that the fusion of multisource spatiotemporal features by the proposed SCA is crucial for enhancing the prediction accuracy.

V. Conclusion

In this article, an effective multisource data fusion short-term precipitation prediction model (FsrGAN) is proposed by fusing satellite data to complement the limitation of extrapolation using radar echo data alone. In the first-stage fusion network, the novel SCA is proposed to realize the multiscale adaptive fusion of satellite and radar echo features. In addition, a postprocessing stage based on the improved GAN model (FusionGAN) is designed to further fuse the satellite data to sharpen the first-stage predicted radar maps with more details. The experimental results show that the proposed FsrGAN is superior to other methods in terms of clarity and forecast accuracy, and also effectively realizes the prediction of convective initiation.

References

[1] S. Ravuri et al., "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, 2021.
[2] N. Jones, "How machine learning could help to improve climate forecasts," *Nature*, vol. 548, no. 7668, pp. 379–380, 2017.
[3] F. Schmid, Y. Wang, and A. Harou, "Nowcasting guidelines—a summary," *Bulletin*, vol. 2019, 2019, Art. no. 68.
[4] L. Xu, D. Niu, T. Zhang, P. Chen, X. Chen, and Y. Li, "Two-stage uagan for precipitation nowcasting," *Remote Sens.*, vol. 14, no. 23, 2022, Art. no. 5948.
[5] Z. Ma, H. Zhang, and J. Liu, "MM-RNN: A multimodal RNN for precipitation nowcasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023.

[6] B.-Y. Yan, C. Yang, F. Chen, K. Takeda, and C. Wang, "FDNet: A deep learning approach with two parallel cross encoding pathways for precipitation nowcasting," *J. Comput. Sci. Technol.*, vol. 38, pp. 1002–1020, 2023.

[7] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5617–5627.

[8] H. Che, D. Niu, Z. Zang, Y. Cao, and X. Chen, "ED-DRAP: Encoder–decoder deep residual attention prediction network for radar echoes," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[9] R. Reinoso-Rondinel, M. Rempel, M. Schultze, and S. Tromel, "Nationwide radar-based precipitation nowcasting—A localization filtering approach and its application for Germany," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1670–1691, 2022.

[10] M. R. Ehsani, A. Zarei, H. V. Gupta, K. Barnard, E. Lyons, and A. Behrangi, "NowCasting-Nets: Representation learning to mitigate latency gap of satellite precipitation products using convolutional and recurrent neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–21, 2022.

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

[12] Z. Ma, H. Zhang, and J. Liu, "Focal frame loss: A simple but effective loss for precipitation nowcasting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6781–6788, 2022.

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

[14] T. Chen, Q. Li, X. Peng, Q. Lv, and J. Jing, "NLED: Nonlocal echo dynamics network for radar echo extrapolation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[15] J. Jing, Q. Li, L. Ma, L. Chen, and L. Ding, "REMNet: Recurrent evolution memory-aware network for accurate long-term weather radar echo extrapolation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[16] N. Sun, Z. Zhou, Q. Li, and J. Jing, "Three-dimensional gridded radar echo extrapolation for convective storm nowcasting based on 3DconvLSTM model," *Remote Sens.*, vol. 14, no. 17, 2022, Art. no. 4256.

[17] L. Sevilla-Lara, Y. Liao, F. Guney, V. Jampani, A. Geiger, and M. J. Black, "On the integration of optical flow and action recognition," in *Proc. Pattern Recognit., 40th German Conf.*, 2019, pp. 281–297.

[18] W.-c. Woo and W.-k. Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 3, 2017, Art. no. 48.

[19] 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. Annu. Conf. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

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

[21] C. Wang, P. Wang, P. Wang, B. Xue, and D. Wang, "Using conditional generative adversarial 3-D convolutional neural network for precise radar extrapolation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5735–5749, 2021.

[22] D. Niu, H. Che, C. Shi, Z. Zang, H. W. X. Chen, and Q. Huang, "A heterogeneous spatiotemporal attention fusion prediction network for precipitation nowcasting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8286–8296, 2023.

[23] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine learning for precipitation nowcasting from radar images," in *Proc. NeurIPS Workshop Tackling Climate Change Mach. Learn.*, 2019.

[24] R. Prudden et al., "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," 2020, *arXiv:2005.04988*. [Online]. Available: https://doi.org/10.48550/arXiv.2005.04988

[25] D. Niu, J. Huang, Z. Zang, L. Xu, H. Che, and Y. Tang, "Two-stage spatiotemporal context refinement network for precipitation nowcasting," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4285.

[26] 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.

[27] Y. Wang, Z. Gao, M. Long, J. Wang, and S. Y. Philip, "Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 5123–5132.

[28] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu, "Memory in memory: A predictive neural network for learning higher-order nonstationarity from spatiotemporal dynamics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9146–9154.

[29] H. Wu, Z. Yao, J. Wang, and M. Long, "MotionRNN: A flexible model for video prediction with spacetime-varying motions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15430–15439.

[30] K.-S. Chung and I.-A. Yao, "Improving radar echo lagrangian extrapolation nowcasting by blending numerical model wind information: Statistical performance of 16 typhoon cases," *Monthly Weather Rev.*, vol. 148, no. 3, pp. 1099–1120, 2020.

[31] S.-S. Yoon, "Adaptive blending method of radar-based and numerical weather prediction QPFs for urban flood forecasting," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 642.

[32] C. Bai, D. Zhao, M. Zhang, and J. Zhang, "Multimodal information fusion for weather systems and clouds identification from satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7333–7345, 2022.

[33] C. Bai, C. Zeng, Q. Ma, and J. Zhang, "Graph convolutional network discrete hashing for cross-modal retrieval," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 23, 2022, doi: 10.1109/TNNLS.2022.3174970.

[34] Y. Wehbe, M. Temimi, and R. F. Adler, "Enhancing precipitation estimates through the fusion of weather radar, satellite retrievals, and surface parameters," *Remote Sens.*, vol. 12, no. 8, 2020, Art. no. 1342.

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

[36] K. Zhou, Y. Zheng, W. Dong, and T. Wang, "A deep learning network for cloud-to-ground lightning nowcasting with multisource data," *J. Atmospheric Ocean. Technol.*, vol. 37, no. 5, pp. 927–942, 2020.

[37] Y. Geng et al., "A deep learning framework for lightning forecasting with multi-source spatiotemporal data," *Quart. J. Roy. Meteorological Soc.*, vol. 147, no. 741, pp. 4048–4062, 2021.

[38] C. Huang, C. Bai, S. Chan, and J. Zhang, "MMSTN: A multi-modal spatial-temporal network for tropical cyclone short-term prediction," *Geophysical Res. Lett.*, vol. 49, no. 4, 2022, Art. no. e2021GL096898.

[39] H.-B. Liu and I. Lee, "MPL-GAN: Toward realistic meteorological predictive learning using conditional GAN," *IEEE Access*, vol. 8, pp. 93179–93186, 2020.

[40] A. Vaswani et al., "Attention is all you need," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 1–11.

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Dan Niu was born in 1986 in Jiangsu, China. He received the Ph.D. degree in system LSI from Graduate School of Information, Production and Systems, Waseda University, Tokyo, Japan, in 2013.

He has been an Associate Professor with the School of Automation, Southeast University. His research interests include AI-based weather forecasting method and AI for EDA.

Yinghao Li was born in Sichuan, China. He received the B.E. degree in automation from Sichuan University, Chengdu, China, in 2022. He is currently working toward the M.S. degree in control science and engineering with the School of Automation, Southeast University, Nanjing, China.

His research interests include spatiotemporal prediction and remote sensing image processing.

Hongbin Wang was born in 1985 in Shanxi, China. He received the Ph.D. degree in atmospheric physics and atmospheric environment from the Lanzhou University, Lanzhou, China, in 2013.

He is the Deputy Chief of the Scientific Research Team, Nanjing Joint Institute for Atmospheric Sciences. His research interests include AI-based weather forecasting methods and satellite remote sensing.

Zengliang Zang was born in 1977 in Jiangsu, China. He received the Ph.D. degree in atmospheric science from PLA University of Science and Technology, China, in 2005.

He has been a Professor with the College of Meteorology and Oceanography, National University of Defense Technology, Hunan, China. His research interests include numerical simulation and data assimilation for atmospheric chemistry model.

Mingbo Jiang was born in 1978 in Shandong, China. He received the B.S. degree in atmospheric sounding and observing from the PLA University of Science and Technology, China, in 2001.

He has been an Associate Researcher with the Beijing Institute of Applied Meteorology, Beijing, China. His research interests include meteorological observation and meteorological support.

Xunlai Chen was born in 1979 in Henan, China. He received the Ph.D. degree in meterology from the Department of Atmospheric Science, Sun Yat-sen University, China, in 2007.

He has been a Senior Engineer of Meteorology, Shenzhen Meteorological Bureau, Shenzhen, China. His research interests include AI for nowcasting and numerical weather prediction.

Qunbo Huang was born in 1988 in Anhui, China. He received the Ph.D. degree in computer science and technology from National University of Defense Technology, Changsha, China, in 2018.

He has been an Engineer with the Unit of 93110 of PLA, Beijing, China. His research interests include numerical simulation and satellite data assimilation.