A Heterogeneous Spatiotemporal Attention Fusion Prediction Network for Precipitation Nowcasting

Dan Niu [ID], Hongshu Che [ID], Chunlei Shi [ID], Zengliang Zang [ID], Hongbin Wang [ID], Xunlai Chen [ID], and Qunbo Huang [ID]

Abstract—Precipitation nowcasting underlying various public services from rainstorm warning to flight safety is quite important and remains challenging due to the fast change in convective weather. Although some deep learning models have been proposed to make prediction automatically, most of them just deal with a single radar echo data source, making them hard to adapt to heterogeneous and diverse data in practice. In this article, a heterogeneous spatiotemporal attention fusion prediction (HST-AFP) network is proposed for radar echo extrapolation (deterministic output) and further precipitation nowcasting, which deals with mining and fusing knowledge from multiple heterogeneous spatiotemporal (ST) data sources, including history radar echo observations and numerical weather prediction data. With the help of the proposed attention-based ST diffusion module, the multiencoder is designed to extract information from both dense ST tensor and sparse ST tensor. On the other hand, the fusion decoder achieves very deep trainable residual fusion prediction by integrating scalewise attention fusion module and deep residual spatial and temporal attention mechanism. It can adaptively blend multisource ST features and rescale the multiscale temporalwise and spatialwise features for better prediction. Experiments in a real-world dataset of South China show that compared with the ingenious recurrent-neural-network-based methods and newly proposed UNet-based methods, our HST-AFP network can handle complex input with heterogeneity in both space and time domains, and performs better on the precipitation nowcasting metrics as well as requires remarkable shorter forecast time.

I. INTRODUCTION

PRECIPITATION nowcasting has long been an important problem in the field of weather, which supplies very short range forecast of rainfall intensity in a local region based on radar echo maps, observation data, as well as the numerical weather prediction (NWP) models [1], [2]. Such predictions facilitate effective planning, crisis management, and the reduction of losses to life and property [3], [4], [5]. Due to relevant dynamical processes and the inherent atmosphere complexities, precipitation nowcasting is quite challenging and has emerged as a hot research topic [6], [7], [8], [9].

The existing precipitation nowcasting methods can be roughly divided into two classes [10], [11], including NWP-based methods and radar-echo-extrapolation-based methods. The NWP approaches conduct a complex and meticulous simulation for the physical equations in the atmosphere model. However, solving such equations is time consuming and usually requires several hours, even on modern supercomputers [12]. Moreover, NWP methods are also very sensitive to small perturbations in initial conditions, boundary conditions, and round-off errors [11], [13]. NWPs tend to provide poor forecasts for precipitation at 0–2 h lead time. Thus, the faster and more accurate extrapolation-based methods are widely adopted [14], [15], [16].

While conventional optical-flow-based radar extrapolation methods (e.g., real-time optical flow by variational methods for echoes of radar (ROVER), currently used in the Hong Kong Observatory [17]) have seen considerable success, they have certain limitations due to assumptions of Lagrangian persistence and smooth motion fields. There have been efforts to relax these assumptions by incorporating specific mechanisms (such as cell life cycles or convergence lines). However, there has been no fully general solution short of a costly data assimilation cycle [18], [19]. Moreover, the optical-flow-based extrapolation methods do not make full use of the vast amount of existing radar echo datasets. Recently, machine learning techniques, especially deep learning methods, have been explored as a way to fill this gap, which can capture complex nonlinear spatiotemporal (ST) patterns and combine heterogeneous data sources for prediction [20], [21], [22], [23], [24], [25]. In principle, it can weaken

Dan Niu is with the School of Automation, Southeast University, Nanjing 210096, China, and also with the Key Laboratory of Measurement and Control of CSE, Ministry of Education, Nanjing 210096, China (e-mail: danniu1 @163.com).

Hongshu Che was with the School of Automation, Southeast University, Nanjing 210096, China. He is now with Alibaba Group, Hangzhou 311121, China (e-mail: hsche1222@gmail.com).

Chunlei Shi is with the School of Automation, Southeast University, Nanjing 210096, China (e-mail: 1056540071@qq.com).

Zengliang Zang is with the Institute of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China (e-mail: zzlqxxy@163.com).

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).

Xunlai Chen is with the Shenzhen Key Laboratory of Severe Weather in South China, Shenzhen Meteorological Bureau, Shenzhen 518040, China (e-mail: cxlxun@163.com).

Qunbo Huang is with the 93110 Troops, People's Liberation Army, Beijing 100843, China (e-mail: 451453753@qq.com).

Digital Object Identifier 10.1109/JSTARS.2023.3310361

the Lagrangian persistence assumption and design more flexible models, which take advantage of more varied predictability sources.

In essence, precipitation nowcasting can be formulated as an ST sequence forecasting problem, where previous radar map sequence is the input and the sequence of a fixed number of future radar maps is output [10]. Some progress has been made by utilizing deep learning techniques for precipitation nowcasting [18], [26]. First, the pioneering recurrent neural network (RNN) and the long short-term memory (LSTM) encoder–decoder framework proposed in [27], [28], and [29] provide a general framework for tackling this problem. Klein et al. [30] proposed a new neural network layer called "dynamic convolutional layer" to short-term forecast the location and intensity of rain and snow. Shi et al. [10] extended the LSTM by adopting convolutional structures in both the input-to-state and state-to-state transitions and designed the convolutional long short-term memory (ConvLSTM) model, which gave more accurate predictions than the fully connected LSTM and ROVER algorithm. However, the convolutional recurrence structure in ConvLSTM-based models is location invariant; they further proposed the trajectory gated recurrent unit (TrajGRU) [1] model, which used a subnetwork to output the state-to-state connection structures before state transitions and actively learned the location-variant structure for recurrent connections. Moreover, Wang et al. [31] presented a predictive recurrent neural network (PredRNN) in the light of the idea that ST predictive learning should memorize both spatial appearances and temporal variations in a unified memory pool. To alleviate the gradient propagation difficulties in the PredRNN and provide alternative quick routes for the gradient flows, the improved one (PredRNN++) was proposed [32], where the gradient highway units working seamlessly with the causal LSTMs enabled the model to adaptively capture the short-term and long-term dependencies and outperform the previous models (ConvLSTM and TrajGRU) in some real datasets. However, a gradient vanishing problem in these RNN-based methods always exists, and network training requires a large amount of computational resources (especially GPU memory) [16], [18]. They require memory-bandwidth-bound computation, which often limits their applications [33].

In this case, some pure convolutional architectures with faster training speed and less computation memory instead of mixed convolutional recurrent architectures are explored for weather ST prediction [6], [33], [34], [35], [36], [37]. The time dimension can be handled as part of a convolutional architecture. Han et al. [21] proposed a convolutional neural network (CNN) method to nowcast convective storms. They divide the study domain into many position-fixed small boxes and turn the nowcasting problem into a classification problem. Moreover, the UNet-based models are proposed for precipitation nowcasting based on the radar echo sequences [34], [35], [36], [37]. The UNet architecture was first proposed for medical image segmentation, but it has been employed in various domains due to its flexibility and easy to extend [38], [39]. In [37], an efficient SmaAt-UNet was proposed, which equipped with depthwise-separable convolutions and attention modules. It can obtain better prediction performance while using less trainable parameters than original UNet. In [36], an SE-ResUNet network was proposed to predict

rainfall dynamics for the city of Beijing, China. It combined the strengths of UNet, ResNet, and Sqeeze-and-Excitation attention and enabled significant performance improvement.

Besides, the abovementioned deep learning methods just only deal with single radar echo data source. Unlike most computer vision tasks, weather prediction can also obtain multiple meteorological information sources, such as NWP data and ground or satellite measurements [40], [41]. It is clear that the NWP data can supply important prior prediction knowledge for meteorological parameters [42], [43], [44]. Could we combine the advantages of NWP simulation forecasts and historical radar echo observations to further enhance radar echo extrapolation ability and improve precipitation nowcasting accuracy? The expert systems that synthesize multisource data based on the predefined rules [45], [46], [47] and some RNN-based deep learning fusion networks (typical ConvLSTM-based LightNet [48] and LightNet+ [49] for lighting forecasting) have been proposed. In this article, we propose a CNN-based fusion prediction network framework, which can mine knowledge from multiple heterogeneous ST data sources. It merges radar echo history observation data with meteorological forecasts from NWP (even with restricted useful information due to low ST resolution in this article) to further improve the precipitation nowcasting. This architecture is flexible enough to add relevant multisource inputs, which is an interesting property for data fusion. In detail, a heterogeneous spatiotemporal attention fusion prediction network (HST-AFP) is proposed for precipitation nowcasting. It achieves a very deep trainable residual attention fusion prediction network and adaptively extracts and rescales more useful spatialwise and temporalwise fusion features from multiple heterogeneous ST data sources (NWP forecasts and radar echo observations, and more if supplied). Experimental results show that the proposed HST-AFP network can effectively mine complementary information distributed across two heterogeneous data sources and further enhance the precipitation nowcasting performance.

The contributions of this article are summarized as follows.

1) To achieve precipitation nowcasting by accurate radar echo extrapolation, we propose an HST-AFP network to mine knowledge from multiple heterogeneous ST data sources, where multiscale residual learnings are integrated to construct a very deep ST fusion prediction network.

2) An attention-based spatiotemporal diffusion (ASTD) module is proposed to convert sparse ST tensor into a spatialwise and temporalwise dense form and employed for merging heterogeneous data with ST resolution distinction and from different periods (the past and the future).

3) For achieving a high-to-low-level residual fusion prediction, scalewise attention fusion (SWAF) module and deep residual spatiotemporal attention (DRSTA) mechanism are proposed to adaptively rescale and blend the multisource and multiscale ST discriminative features to update the lower level features.

II. PRELIMINARY

Weather radar is one of the best instruments to monitor the precipitation system. The intensity of radar echo is related to the size, shape, state of precipitation particles, and the number of
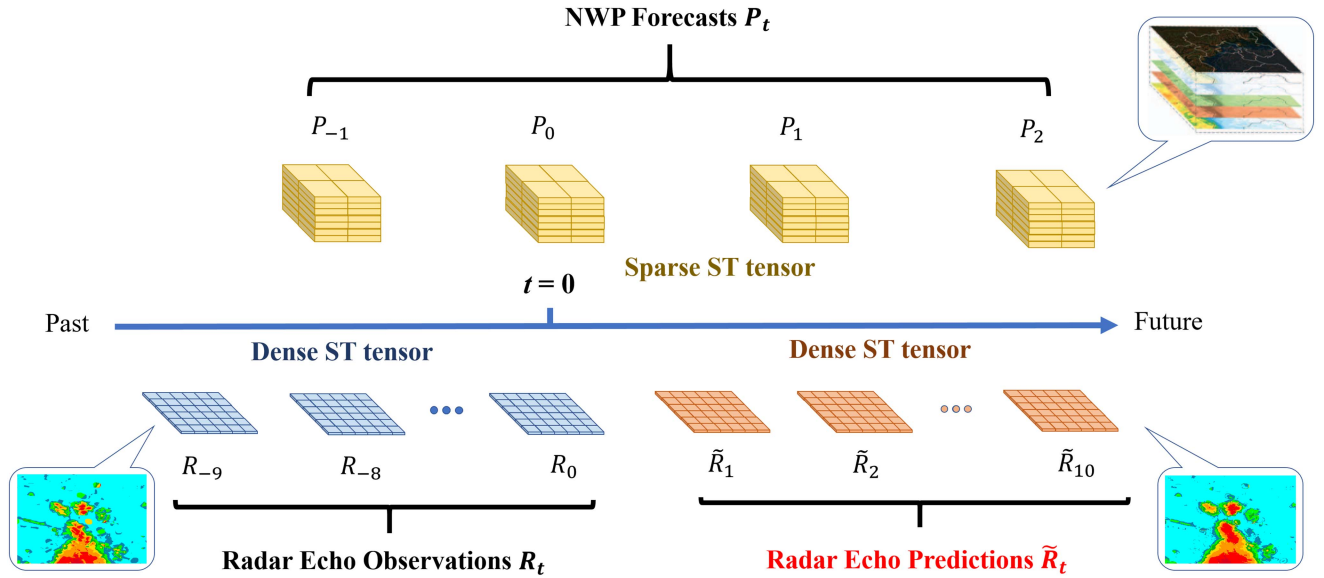
Fig. 1. Heterogeneous ST structure of data in our task. On the one hand, the radar echo history observation data and the NWP forecast data describe weather situations for different periods (the past and the future). On the other hand, they have different spatial and temporal resolutions. We seek to extract information from these heterogeneous data and produce a precipitation nowcasting.

particles per unit volume [21]. The intensity and distribution of precipitation in a weather system can be judged by the radar echo map. Actually, the rainfall rate values (mm/h) can be calculated by the radar reflectivity values using the widely used Z–R relationship. R is the radar reflectivity values and Z is the rain rate values. It means that if we can predict the future radar echo images, the goal of precipitation nowcasting can be achieved. In this article, the heterogeneous input data are first introduced, and then, the problem to be solved is defined.

A. Heterogeneous ST Data

As shown in Fig. 1, the model inputs comprise two types of heterogeneous ST data: NWP forecast data (P) and radar echo observation data (R).

1) NWP Forecast Data (P): The NWP forecast product data with 3-km spatial resolution and 1-h temporal resolution are from Global/Regional Assimilation and Prediction System (GRAPES), which is a new-generation NWP system developed by the China Meteorological Administration. In order to reduce the memory cost, the region covering 300 km × 300 km of the Pearl River Delta is selected, covering longitude ranges from 112° to 115° E and latitude from 22° to 25°N. This area is located in Southeastern China. NWP forecast data are composed of a grid with 100 rows and 100 columns, while each grid cell corresponds to a scope of 3 km × 3 km in the real world. Each data grid carries simulation results of different meteorological parameters. The GRAPES performs hourly numerical simulation at 0:00 and 12:00 (UTC) per day, and each simulation covers the next 24 h. Considering compute resources and GPU memory costs of model training, some parameters that are closely related to precipitation are selected: cr, rain, and rh with five height channels from 600 to 1000 hPa. We concatenate the seven parameter channels at time t and form a comprehensive NWP forecast data P_t.

2) Radar Echo Observation Data (R): The radar echo dataset used in this article is a subset of the three-year weather radar intensities provided by the Guangdong Meteorological Bureau from 2017 to 2019. The radar CAPPI reflectivity images, which have resolution of 300 × 300 pixels and also cover a 300 km × 300 km area. It is obtained from some S-band radars, which are located at Guangzhou, Shenzhen, Shaoguan, etc. The spatial range is the same as that of NWP forecast data. However, the radar echo observation interval is 12 min, and the spatial resolution is 1 km. It is clear that the two types of data have different ST resolution (12 min versus 1 h, 1 km versus 3 km), and they are heterogeneous in spatial and temporal (history observations from the past versus simulation of the future).

B. Problem Formulation

Precipitation nowcasting is to blend the past observed radar echo sequence and the future NWP simulation data to forecast a fixed length of the future radar echo maps in a local region. In real applications, the GRAPES supplies hourly numerical simulation and the radar echo maps are taken from weather radars every 12 min, and nowcasting is usually done for the following 2 h, i.e., to predict the 10 frames ahead.

Suppose that the current moment is $t = 0$. Given the NWP forecast $P = [P_t]_{t=-q}^{m}$ (real 4-D sparse tensor) from previous q hours to future hours, the radar echo observation $R = [R_t]_{t=-l+1}^{0}$ (3-D dense tensor) for the previous l observations including the current one. Our target is to predict the most likely length-k radar echo sequence in the future $\tilde{R} = [\tilde{R}_t]_{t=1}^{k}$ (3-D dense tensor), where P, R, and $\tilde{R}$ share the same xy scope. Specifically, our goal is to find a mapping such that

$$\min_{f} \operatorname{loss}\left([\tilde{R}_t]_{t=1}^{k}, [R_t]_{t=1}^{k}\right)$$

$$\text{s.t. } [\tilde{R}_t]_{t=1}^{k} = f\left([P_t]_{t=-q}^{m}, [R_t]_{t=-l+1}^{0}\right). \quad (1)$$
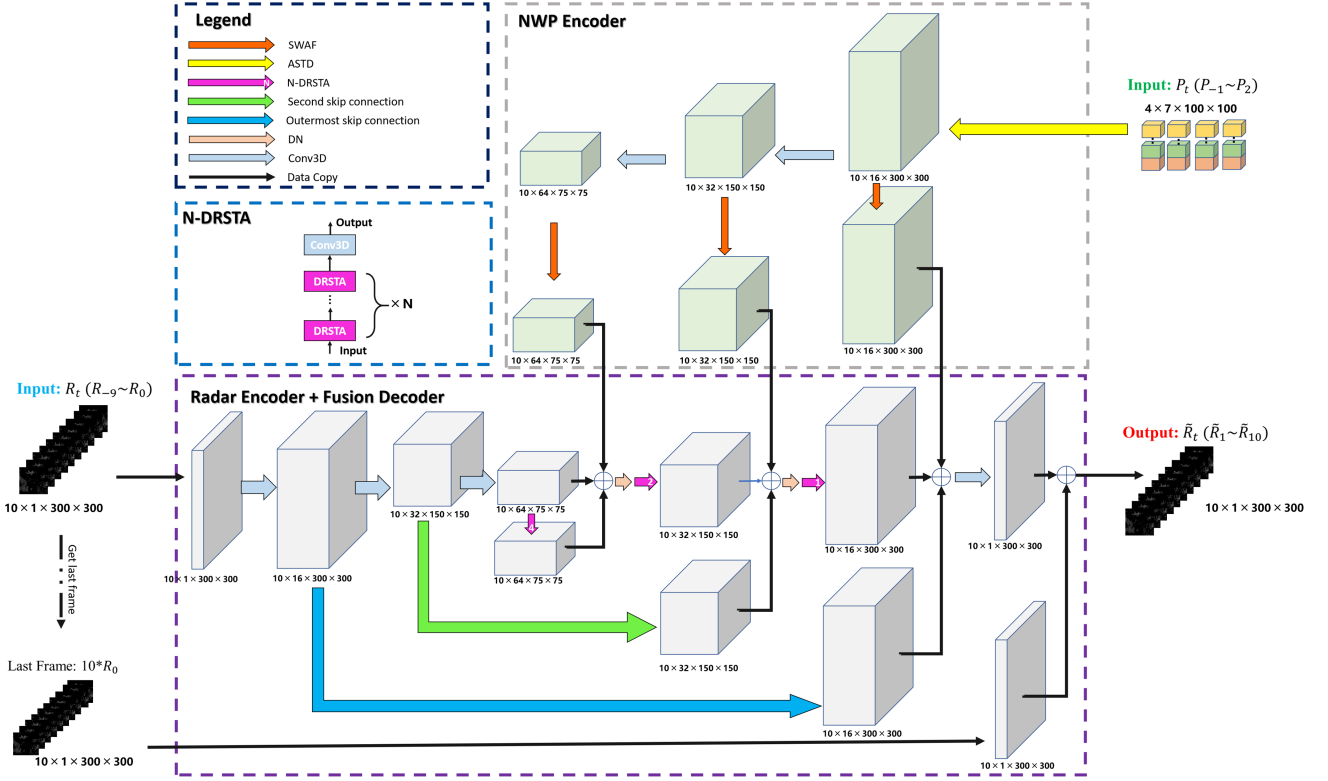
Fig. 2. Architecture of multisource deep residual attention fusion prediction network. $\oplus$ denotes the elementwise addition.

Here, we should emphasize the heterogeneous ST structure of our data, as illustrated in Fig. 1. On the one hand, the NWP forecast data (P) and the radar echo observation data (R) describe weather situations for different periods (the future and the past). On the other hand, the sparse structure of $(P,$ 3 km and 1 h) distinguishes it from the dense radar input and predicted radar output tensors $(R$ and $\tilde{R}$, 1 km and 12 min). We seek to mining knowledge from these heterogeneous data.

III. HST-AFP NETWORK

Considering ST heterogeneity and the high dimensionality of the multisource ST sequences especially for multistep predictions, such prediction problem is nontrivial, unless the ST features of the multisource heterogeneous data are captured and fused well. This section presents the architecture of the proposed HST-AFP network, as illustrated in Fig. 2. The main network architecture is the multiencoder and fusion decoder structure. The past radar echo maps and the NWP forecast data are concatenated as the inputs of model. Since the two data sources have different temporal and spatial resolution, and they come from different time domains (history observations versus future simulations), two encoders (NWP forecast Encoder and Radar echo Encoder) are designed, respectively, to process the two input sequences and extract multiple scale ST representations by stacked downscaling block, which usually halves the image size and doubles the number of feature maps. In the NWP forecast Encoder, the ASTD module is first proposed to deal with ST resolution distinction. In the decoder part, SWAF and

DRSTA modules are proposed to allow our decoder network to fuse discriminatively multilevel trend information from NWP forecast data, and concentrate on more useful frames (time steps) and spatial regions from radar echo sequence. DRSTA working with SWAF adaptively rescale and blend the multisource and multiscale ST features, which guide the update of the lower level features and sequence prediction process with the multiscale deep residual learning.

A. Multiencoder

A multisource deep residual attention fusion prediction network introduces the NWP forecast encoder and the radar echo encoder to convert the two source raw data into feature maps in a unified space for further combination.

1) NWP Forecast Encoder: It is responsible for encoding the NWP forecast data. We stack cr, rain, and rh with five height channels along the z-direction for each time slot. These meteorological parameters constitute the input $[P_t]_{t=-q}^{m} = P_{-1}, P_0, P_1, P_2$ of the NWP encoder. It is noted that not only the data from future 2-h prediction but also the data at the previous 1 h and the current time stamp are selected to supply more weather and variation trend information. P_t is a sparse tensor due to low spatial resolution (3 km) and low temporal resolution (1 h). However, the prediction output is the radar echo map sequence with high spatial resolution (1 km) and high temporal resolution (12 min). In this article, not simple convolution-based upsample but ASTD module is proposed first to deal with ST

resolution distinction

$$\hat{P}_t = \text{ASTD}(P_t). \tag{2}$$

More details about the ASTD module are explained in Section III-D. Then, the feature extraction module is designed to extract multiscale ST representations by stacked downscaling blocks, which will be shown in detail.

2) Radar Echo Encoder: The radar echo observation encoder is in charge of extracting the information based on the previous l observations including the current one: $[R_t]_{t=-l+1}^0 = R_{-9}, R_{-8}, \ldots, R_0$. Since the ST resolution of the input radar echo map is the same as the predicted output radar echo sequence, no ST diffusion module and but only feature extraction module is required to obtain multiscale feature tensors. Here, the feature extraction module shares the similar architecture with that of the NWP encoder.

3) Feature Extraction Module: It is employed to extract multiscale ST features for the encoder. First, a 3-D convolution (Conv3D) operation is used to extract the low-level (small-scale) features. Then, two downscaling layers are designed to extract middle- and large-scale ST feature maps for further blending in the decoder part

$$[STE_{R,L}, STE_{P,L}] = \text{Conv3D}\left(R_t, \hat{P}_t\right) \tag{3}$$

where $STE_{R,L}$ is the extracted low-level (small-scale) ST features from the input past radar echo sequence R_t. $STE_{P,L}$ is the low-level ST features extracted from the NWP ST diffusion data $\hat{P}_t$. As shown in Fig. 2, $STE_{R,L}$ and $STE_{P,L}$ features are utilized for further encoding and also directly connected to the decoder part by the outermost skip connection.

Subsequently, the encoder is composed of two consecutive downscaling blocks DSL

$$[STE_{R,M}, STE_{P,M}] = DSL(STE_{R,L}, STE_{P,L}) \tag{4}$$

$$[STE_{R,H}, STE_{P,H}] = DSL(STE_{R,M}, STE_{P,M}). \tag{5}$$

In (4) and (5), DSL consists of a Conv3D with stride 2 followed by a rectifier linear unit (ReLU) and a batch normalization (BN) unit. The Conv3D layer achieves bilinear downsampling, which halves the spatial size of input feature maps and doubles the number of feature maps. The ReLU activation function is used to model nonlinear relations. BN is a widely adopted technique that enables faster and more stable training of a network. $STE_{R,M}$ and $STE_{P,M}$ features are the middle-level (middle-scale) ST features extracted from the low-level features $STE_{R,L}$ and $STE_{P,L}$, respectively. They are also the input for further extracting the high-level (large-scale) ST features $STE_{R,H}$ and $STE_{P,H}$. The middle-scale feature maps or receptive fields $STE_{R,M}$ and $STE_{P,M}$ and the large-scale feature maps $STE_{R,H}$ and $STE_{P,H}$ will also connected to the corresponding decoder layer via other two global skip connections (second and innermost skip connections), respectively. To ease the flow of multisource and multiscale ST feature information and make the decoder to fuse residual information, both the global/local skip connection and deep residual attention learning is integrated in this article to achieve better prediction. The local

skip connection and deep residual attention learning will be shown in the following subsections.

B. Fusion Decoder

In the fusion decoder part, multiscale ST feature information extracted from the NWP forecast data and radar echo map sequence in the encoder part will be adaptively rescaled and fused to achieve a high-to-low level residual prediction. First, 4-stacked DRSTA block is proposed to construct the large-scale residual prediction module RP_L, which can extract and concentrate the discriminative high-level ST prediction features from the deep feature maps of encoder $STE_{R,H}$

$$STD_{R,H} = RP_L(STE_{R,H}). \tag{6}$$

The DRSTA block can achieve quite large depth and provide very large receptive field size, which will be presented detailed in Section III-C. For high-level ST encoder features $STE_{P,H}$ from the NWP forecast data, the SWAF module is assigned to obtain the high-level refined prediction feature $STD_{P,H}$

$$STD_{P,H} = \text{SWAF}(STE_{P,H}) \tag{7}$$

where the SWAF module provides an adaptive fusion method to improve the prediction accuracy without increasing much weight parameters. It consists of Conv3D layers, in which a specific set of weights is trained to distinguish and fuse different scale variation trend information from the NWP forecast data.

Then, the high-level radar prediction features $STD_{R,H}$ and the NWP prediction features $STD_{P,H}$ are incorporated with the original high-level radar encoder representations $STE_{R,H}$ by the innermost skip connection, and high-level residual attention prediction RAP_H is obtained

$$RAP_H = STD_{R,H} + STD_{P,H} + STE_{R,H}. \tag{8}$$

Similar to the multiencoding part, two consecutive upsample fusion prediction modules are designed, including middle-scale (middle-level) residual prediction module RP_M and small-scale (low-level) residual prediction module RP_S

$$RAP_M = RP_M(RAP_H) + STE_{R,M} + \text{SWAF}(STE_{P,M}) \tag{9}$$

$$RAP_L = RP_S(RAP_M) + STE_{R,L} + \text{SWAF}(STE_{P,L}) \tag{10}$$

where RAP_M and RAP_L are middle- and low-level residual attention prediction, respectively. $RP_M(RAP_H)$ and $RP_S(RAP_M)$ are extracted middle- and low-level fusion prediction features, respectively. $\text{SWAF}(STE_{P,M})$ and $\text{SWAF}(STE_{P,L})$ are the middle- and low-level NWP refined prediction features, respectively. As shown in (9) and (10), the fusion prediction features will be blended with the NWP refined prediction features and original radar encoder representations to produce lower level residual attention prediction by the second and outermost skip connections. RP_M and RP_S are the residual prediction module consisting of a upsample module USL and stacked-DRSTA. USL is composed of three operations: a DeConv3D, BN, and ReLU. The stacked DRSTA is employed to adaptively concentrate and rescale more useful ST
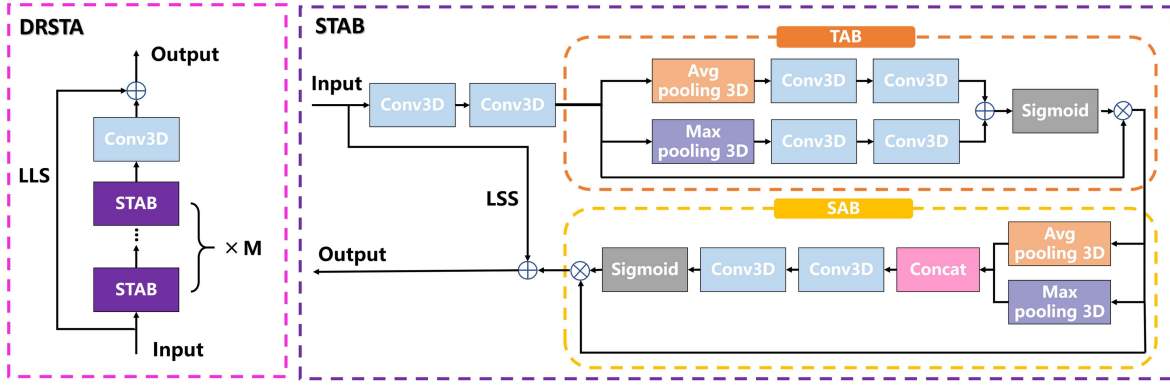
Fig. 3.   DRSTA block.

fusion features. The global skip connections in the decoder part blend the multisource and multiscale ST features to restore the lost information and generate the forecast radar echo sequence. Finally, we employ a $1\times1\times1$ 3-D convolution to generate the residual prediction for the last radar echo frame and output the predicted radar echo sequence $\tilde{R}_t$

$$\tilde{R}_t = \tilde{R}_1 \ldots \tilde{R}_{10} = \mathrm{Conv}_{1\times1\times1}(RAP_L + R_0). \quad (11)$$

It is verified that this operation can further enhance the whole model's forecasting ability for precipitation.

C. DRSTA Block

As shown in the decoder part, the stacked DRSTA block is proposed to form a very deep trainable prediction network, which adaptively rescales and discriminatively blends multisource and multiscale ST sequence features. Inspired by the Residual in Residual structure [50], local long and short skip connections are integrating to construct very deep trainable networks. As shown in Fig. 3, the DRSTA block consists of spatiotemporal attention block (STAB), a Conv3D, and local long skip (LLS) connection. The depth of representation is of crucial importance for feature extraction, but simply stacking residual blocks hardly obtains better improvements. LLS connection can stabilize the training of very deep network and ease the flow of ST information across STABs. The STAB further contains two Conv3D, 3-D temporal attention block (TAB), 3-D spatial attention block (SAB), and local short skip (LSS) connection. LSS further allows the main parts of the network to learn more informative residual information. Such a deep residual attention structure with LLS and LSS allows us to train a very deep network and blends deep multisource and multiscale ST features to generate finer features, which favors the high reconstruction and prediction performance.

Instead of treating all the features equally, the TAB and the SAB are proposed for temporalwise and spatialwise weightings by modeling the interdependencies, which strengthen the discriminative learning ability and the representational power of deep networks. As shown in Fig. 3, the 3-D TAB contains two 3-D pooling descriptor branches for channelwise statistic, sigmoid, and a short cut. The two 3-D pooling descriptor branches further contain 3-D average pooling, two Conv3D and 3D max

pooling, and two Conv3D. Pooling descriptor gathers important clue about distinctive object features to infer finer channelwise attention. Both the average-pooled and max-pooled features are simultaneously used to greatly improve representation power of networks rather than using each independently [51]. The shortcut eases the flow of information and allow abundant low-frequency information to be bypassed. The simple gating mechanism with sigmoid is utilized to learn the nonlinear interaction between channels and capture dependencies from the aggregated information. Meanwhile, the 3-D SAB consists of two pooling branches (average and max) followed by concat operation, Conv3D, BN unit, sigmoid, and shortcut. With temporal and spatial attention, the residual component in the STAB is adaptively rescaled.

Such TAB and SAB mechanisms allow our proposed network to concentrate on more useful ST features, which adaptively rescale and blend discriminative features across temporal sequences and spatial regions. As a result, the radar echo and NWP forecast feature maps at different time steps and different spatial fields have different impacts for precipitation forecasting.

D. ASTD Module

In this article, the NWP forecast data P_t is a sparse tensor with low spatial resolution (3 km) and low temporal resolution (1 h). However, the prediction output for precipitation nowcasting is the radar echo map sequence with high spatial resolution (1 km) and high temporal resolution (12 min). In this article, not simple convolution-based upsample but ASTD module is proposed to deal with ST resolution distinction. As shown in Fig. 4, the ASTD contains a spatial diffusion module (SAM) and a temporal diffusion module (TAM). Considering that different weather parameters of the NWP forecast data at different time points and different spatial regions have different variation trend, the STAB is employed to design the SAM and the TAM. The SAM consists of a STAB, DeConv, and DSL, by which the spatial resolution is upscaled to 1 km. Subsequently, the TAM, which contains a DSL and two STAB, is designed to increase the time resolution and the time steps increase from 4 to 10.

By employing the attention-based nonlinear diffusion module ASTD, not by the linear or manually nonlinear interpolation, multilevel NWP ST features can be better extracted and fused
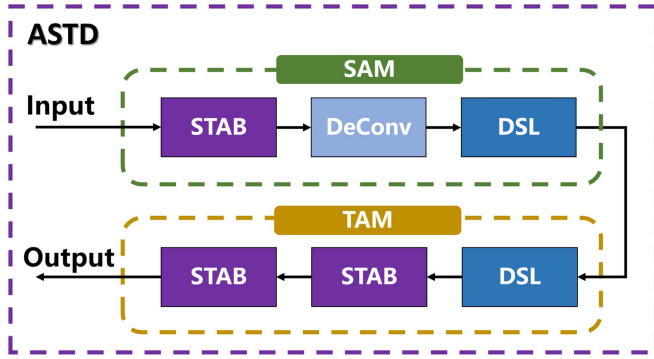
Fig. 4.   ASTD module.

to improve prediction performance, which can be verified from experiment results.

IV. EXPERIMENTS

A. Experimental Setup

The radar echo dataset and the NWP forecast dataset used in this article are the subset of the three-year weather radar intensities and GRAPES _3km meteorological products provided by Guangdong Meteorological Bureau from 2017 to 2019. Since not every day is rainy and our nowcasting target is precipitation, we select 356 rainy events to form our dataset. In order to reduce the memory cost, the region covering 300 km × 300 km of the Pearl River Delta is selected, covering longitude ranges from 112° E to 115°E and latitude from 22°N to 25°N.

As shown in Fig. 2, our output is to predict the most likely length-10 radar echo sequence in the future. The model inputs comprise two types of heterogeneous ST data. The radar echo input $[R_T]_{t=-9}^0$ employs the previous 2-h radar echo observations including the current one, and the input dimension is $10 \times 1 \times 300 \times 300$, due to 1-km spatial resolution and 12-min temporal resolution. The initial channel number of feature maps is 1. The width and height of the initial input tensor are 300. Another heterogeneous input is the NWP forecast data $[P_T]_{t=-Q}^k$ with seven parameter channels at each time slot from past 1 h to future 2 h, and the input dimension is $4 \times 7 \times 100 \times 100$ (sparse ST tensor) due to 3-km spatial resolution and 1-h temporal resolution. For preprocessing, the radar intensities are linearly transformed to pixel values and are clipped to be between 0 and 255. Moreover, to alleviate the noise impact in training and evaluation, the pixel values of some noisy regions are further removed by applying K-means clustering to the monthly pixel average [10]. Then, the radar echo sequence instances are sliced using a 20-frame-wide sliding window, and each sequence is 20 frames long [10 for the input, and 10 for forecasting (2 h)]. The total 356 precipitation events are split into a training set of 254 samples and a test set of 102 samples.

In this article, the proposed HST-AFP network is compared with a typical optical-flow-based method (ROVER [17]), two ingenious RNN-based methods (TrajGRU [1] and PredRNN++[32]), and two well-known CNN-based methods (SmaAt-UNet [37] and SE-ResUNet [36]). Since the input of

TABLE I
SKILL SCORES (≥ 20 DBZ, LIGHT RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↑
OpticalFlow[19]	0.490	0.563	0.628	0.322
TrajGRU[1]	0.564	0.644	0.695	0.260
PredRNN++[32]	0.554	0.628	0.724	0.306
SmaAt-UNet[37]	0.541	0.621	0.677	0.279
SE-ResUNet[36]	0.536	0.619	0.646	**0.251**
HST-AFP_Radar	**0.571**	0.645	**0.757**	0.306
HST-AFP	0.570	**0.647**	0.718	0.274

TABLE II
SKILL SCORES (≥ 35 DBZ, MODERATE RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↑
OpticalFlow[19]	0.371	0.441	0.455	0.519
TrajGRU[1]	0.357	0.491	0.495	**0.459**
PredRNN++[32]	0.346	0.474	0.571	0.550
SmaAt-UNet[37]	0.321	0.446	0.456	0.499
SE-ResUNet[36]	0.348	0.480	0.535	0.519
HST-AFP_Radar	0.364	0.495	**0.692**	0.574
HST-AFP	**0.370**	**0.503**	0.624	0.537

other models contains only the radar echo maps, the proposed HST-AFP with only radar echo input (HST-AFP _ Radar) is also tested for fair comparison. The models are optimized by the Adam optimizer [52] by setting $\beta1 = 0.5$ and $\beta2 = 0.999$. The minibatch size is 8. The learning rate is initialized as 1×10^{-4} and decreased by 0.7 at every five epochs. The frequencies of different rainfall levels are highly imbalanced. Thus, the weighted loss function B-MSE + B-MAE is designed [1], [33]. We train these models with early stopping on the sum of B-MSE and B-MAE.

Four commonly used precipitation nowcasting metrics, including critical success index (CSI), Heidke skill score (HSS), probability of detection (POD), and false alarm rate (FAR), are employed to evaluate the prediction accuracy [1]. Moreover, to give an all-round performance evaluation, we calculate the skill scores for three radar reflectivity thresholds (20, 35, 45 dBZ) that correspond to different rainfall levels. For the skill scores at a specific threshold τ, the pixel values in forecasting and ground truth are first converted to 0/1 by thresholding with τ. Then, the TP (prediction = 1, truth = 1), FN (prediction = 0, truth = 1), FP (prediction = 1, truth = 0), and TN (prediction = 0, truth = 0) are calculated. The four nowcasting metrics are defined as follows:

$$CSI = \frac{TP}{TP + FP + FN}$$

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

$$POD = \frac{TP}{TP + FN}$$

$$FAR = \frac{FP}{TP + FP}. \tag{12}$$

TABLE III
SKILL SCORES (≥ 45 DBZ, HEAVY RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↑
OpticalFlow[19]	0.129	0.212	0.207	0.772
TrajGRU[1]	0.154	0.252	0.209	**0.651**
PredRNN++[32]	0.154	0.253	0.322	0.783
SmaAt-UNet[37]	0.119	0.196	0.169	0.720
SE-ResUNet[36]	0.147	0.241	0.249	0.750
HST-AFP_Radar	0.162	0.262	**0.372**	0.786
HST-AFP	**0.166**	**0.267**	0.307	0.748

TABLE IV
GPU MEMORY USAGE IN TRAINING STEP (BATCH SIZE = 1) AND THE
FORECAST TIME SPENT

Models	GPU memory usage (MB)	Forecast time spent (s)
TrajGRU[1]	4174	0.5413
PredRNN++[32]	12966	0.4188
HST-AFP_Radar	3391	0.029
HST-AFP	5129	0.144

TABLE V
SKILL SCORES (≥20 DBZ, LIGHT RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↓
HST-AFP	**0.570**	**0.647**	**0.718**	0.274
HST-AFP_NASTD	0.566	0.645	0.711	**0.272**
HST-AFP_NSWAF	0.569	**0.647**	0.700	0.275
HST-AFP_NDRSTA	0.459	0.519	0.635	0.389

TABLE VI
SKILL SCORES (≥35 DBZ, MODERATE RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↓
HST-AFP	**0.370**	**0.503**	**0.624**	0.537
HST-AFP_NASTD	0.368	0.500	0.617	0.537
HST-AFP_NSWAF	0.366	0.499	0.594	**0.526**
HST-AFP_NDRSTA	0.244	0.346	0.389	0.625

TABLE VII
SKILL SCORES (≥45 DBZ, HEAVY RAINFALL)

Models	CSI↑	HSS↑	POD↑	FAR↓
HST-AFP	**0.166**	**0.267**	**0.307**	**0.748**
HST-AFP_NASTD	0.153	0.249	0.276	0.762
HST-AFP_NSWAF	0.160	0.258	0.294	0.758
HST-AFP_NDRSTA	0.080	0.134	0.157	0.871

B. Quantification Results

Tables I–III show the precipitation nowcasting metric results for the 2-h prediction. $R \geq \tau$ denotes the skill score at the τ dBZ echo reflectivity threshold. In these tables, "↑" means the higher value is better, and "↓" means the lower value is better. The best result is also marked with bold face. It is clear that among the typical models, the deep learning models outperform the optical-flow-based ROVER model [17], and there is a gap in evaluation indices. The nonlinear and convolutional structure of the deep learning network is able to learn some complex ST patterns in the dataset. However, it is difficult to update the future flow fields reasonably in the optical-flow-based methods. Among the deep learning models, the proposed HST-AFP network mines knowledge from multiple heterogeneous ST data sources and performs the best at the nearly four metrics over the two newly proposed UNet-based methods and also two RNN-based methods and especially has an obvious improvement at the 35- and 45-dBZ thresholds. Note that even the HST-AFP _ Radar method can also achieve better prediction performance compared with other models. At the 45-dBZ threshold, the CSI of the proposed HST-AFP is over 0.019 higher than that of the SE-ResUNet method (increase by about 12.9%) and also 0.012 higher than that of the PredRNN++ model (increase by about 7.8%). Also, the HSS is much improved, about by 10.8% than that of the SE-ResUNet method, over by 5.5% than that of the PredRNN++ method. It is shown that the proposed method has better prediction performance for heavy rainfall, which is usually a difficult task. On the other hand, comparing HST-AFP with HST-AFP _ Radar, it is clear that the prediction accuracy can be further enhanced by designing reasonable modules to effectively extract and fuse the NWP forecast data (even with limited useful information due to low ST resolution). At the 45-dBZ thresholds, the important CSI can be further increased by about 2.5%, and the HSS is also increased by about 2%.

In addition, the time spent during the forecast and GPU memory usage for the model training with batch size 1 are also compared in Table IV. It is clear that the proposed CNN-based method occupies less GPU memory and spends remarkably shorter forecast time than two RNN-based methods. The forecast time of the proposed HST-AFP _ Radar can be reduced by more than 90% compared with that of the PredRNN++ method. The time is also about two times less even the NWP forecast data are added as the model input (HST-AFP).

C. Effect of ASTD, SAF, and DRSTA

We study the effects of the ASTD module, the SWAF module, and the DRSTA block in this part.

1) ASTD Module: To demonstrate the effect of the proposed ASTD, we replace it with a commonly used convolution-based upsample module, and the test results are shown in Tables V–VII. It can be seen that the prediction accuracy of the HST-AFP _ NASTD (non-ASTD version) decreases obviously, especially at 45-dBZ threshold. This indicates that converting sparse ST NWP tensor into a spatialwise and temporalwise dense form is quite effective, and simply upsample is not applicable to effectively mine the ST information from the heterogeneous sparse NWP tensor.

2) SWAF Module: We also show the effect of the SWAF. When comparing the test results of HST-AFP and HST-AFP _ NSWAF (non-SWAF version), we find that prediction networks with SWAF would perform better than that without SWAF, which provides an adaptive fuse method to improve the prediction accuracy without increasing much weight parameter.

3) DRSTA Block: We further demonstrate the effect of DRSTA by comparing the skill scores of HST-AFP and HST-AFP _ NDRSTA (non-DRSTA version). It is clear that the skill score performance will be significantly decreased when the DRSTA block is removed. It is vital to form a very deep trainable prediction network and adaptively rescale and blend
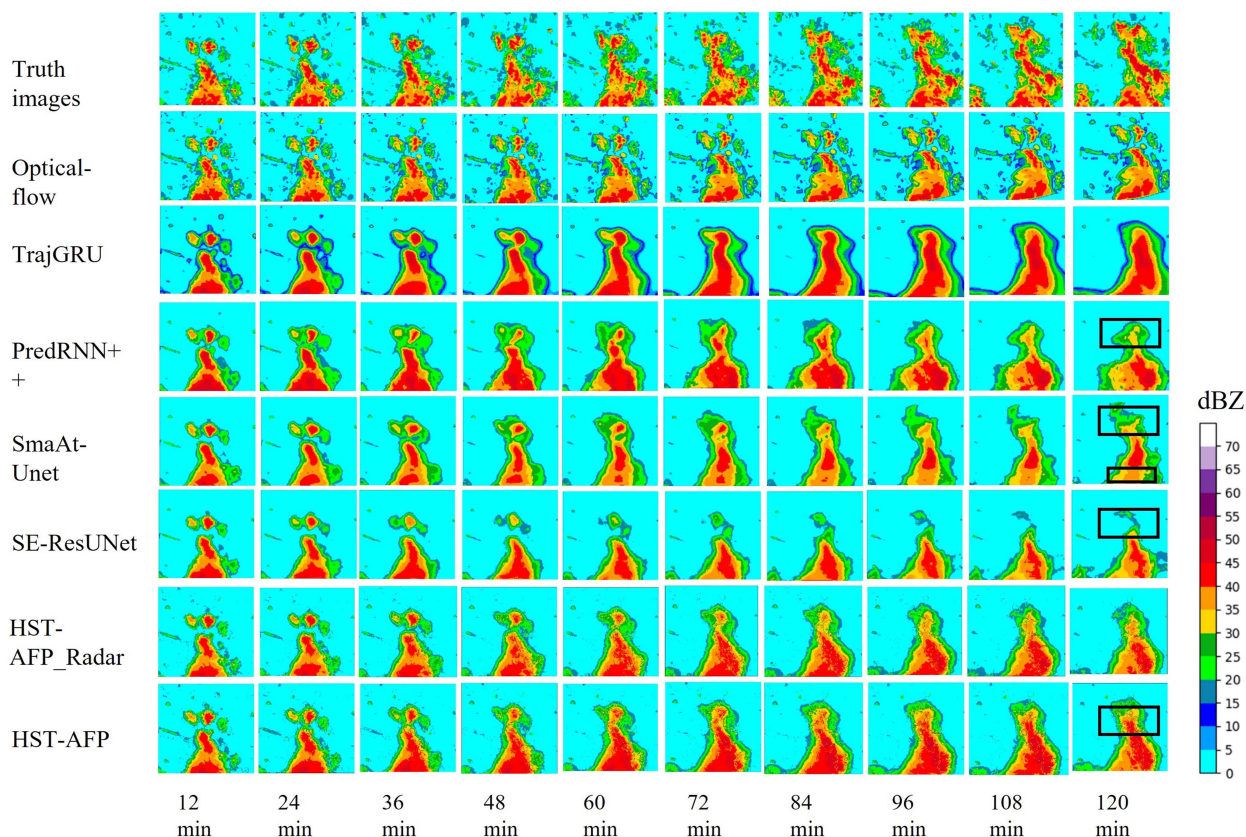
Fig. 5. DRSTA block.

multisource and multiscale ST features to achieve a high-to-low-level residual fusion prediction.

In addition, visualization of the comparisons among the evaluated methods is shown in Fig. 5. Nearly, all the methods predict accurately at the first moment. However, significant differences can be observed as the lead time increased. In the optical-flow-based ROVER method, forecasting scale is gradually reduced, and echo intensity change is nearly ignored. In deep learning methods, the small-scale details are gradually lost, and the boundaries become smooth. The blurring effect may be caused by the inherent uncertainties of the task. Although the optical-flow-based method can give sharper predictions than deep learning methods, more false alarms will be triggered and less prediction precise is obtained in general. As time goes by, TrajGRU tends to exaggerate the forecasting scale, and the major echo region's intensity tends to be overestimated. In SmaAt-Unet and SE-ResUNet, the echo scale cannot be effectively predicted, and some echo regions are lost. Compared with other deep learning methods, the proposed HST-AFP method shows the best performance. As time goes by, the intensity and the position are closer to the truth images.

sources. An ASTD module was proposed in the multiencoder part to convert sparse NWP ST tensor into a spatialwise and temporalwise dense form, which can be effectively extracted by the ST encoder. In the fusion decoder part, the SWAF module was designed to adaptively blend multisource ST features. Moreover, the DRSTA mechanism was proposed to achieve a very deep trainable residual fusion prediction network and discriminatively rescale the multiscale temporalwise and spatialwise fusion features for guiding a high-to-low-level residual fusion prediction. We showed that the proposed HST-AFP has noteworthy shorter forecast time but has better prediction performance than two ingenious RNN-based and the newly proposed two UNet-based precipitation nowcasting methods. The challenges are that sharp and accurate predictions of the whole radar maps in longer term predictions are quite difficulty. For future work, we will employ this CNN-based fusion prediction network framework to blend more heterogeneous data sources (e.g., station OBS) to further improve prediction performance as well as enhance the predicted details of the radar echo images. We will also try to build an operational nowcasting system with the Guangdong Meteorological Bureau.

V. CONCLUSION

In this article, we investigated to make the precipitation nowcasting via a CNN-based fusion prediction network framework extracting ST information from multiple heterogeneous ST data

REFERENCES

[1] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5622–5632.

[2] J. Ritvanen, B. Harnist, M. Aldana, T. Mäkinen, and S. Pulkkinen, "Advection-free convolutional neural network for convective rainfall nowcasting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1654–1667, 2023.

[3] X. Dong, Z. Zhao, Y. Wang, J. Wang, and C. Hu, "Motion-guided global-local aggregation transformer network for precipitation nowcasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5119816.

[4] B. N. Jones, "How machine learning could help to improve climate forecasts," *Nature*, vol. 548, no. 7668, pp. 379–380, 2017.

[5] 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.

[6] 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, 2022, Art. no. 1004705.

[7] 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, 2022, Art. no. 4706021.

[8] S. Ravuri et al., "Skilful precipitation NowCasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, 2021.

[9] R. Reinoso-Rondinel, M. Rempel, M. Schultze, and S. Trömel, "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] 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. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

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

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

[13] M. Tolstykh and A. Frolov, "Some current problems in numerical weather prediction," *Izvestiya Atmos. Ocean. Phys.*, vol. 41, no. 3, pp. 285–295, 2005.

[14] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine learning for precipitation nowcasting from radar images," 2019, *arXiv:1912.12132*.

[15] C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst.*, 2020, pp. 92–97.

[16] M. G. Schultz et al., "Can deep learning beat numerical weather prediction?," *Philos. Trans. Roy. Soc. A*, vol. 379, no. 2194, 2021, Art. no. 20200097.

[17] W. Woo and W. Wong, "Application of optical flow techniques to rainfall nowcasting," in *Proc. 27th Conf. Severe Local Storms*, 2014, poster 35, [Online]. Available: https://ams.confex.com/ams/27SLS/webprogram/Paper254084.html

[18] R. Prudden et al., "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," 2020, *arXiv:2005.04988*.

[19] 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.

[20] M. Chantry, H. Christensen, P. Dueben, and T. Palmer, "Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI," *Philos. Trans. Roy. Soc. A*, vol. 379, no. 2194, 2021, Art. no. 20200083.

[21] L. Han, J. Sun, and W. Zhang, "Convolutional neural network for convective storm nowcasting using 3-D doppler weather radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1487–1495, Feb. 2020.

[22] J. Xia et al., "Machine learning-based weather support for the 2022 Winter Olympics," *Adv. Atmos. Sci.*, vol. 37, pp. 927–932, 2020.

[23] X. Yu and Y. Zheng, "Advances in severe convection research and operation in China," *J. Meteorol. Res.*, vol. 34, no. 2, pp. 189–217, 2020.

[24] H. Chen et al., "A landslide extraction method of channel attention mechanism U-Net network based on sentinel-2A remote sensing images," *Int. J. Digit. Earth*, vol. 16, no. 1, pp. 552–577, 2023.

[25] B. Yu, Y. Li, Q. Sun, and J. Shi, "Calculation and analysis of multi-scale earth gravity field parameters based on self-developed EIGEN-5C model software," *J. Geovis. Spatial Anal.*, vol. 6, no. 2, 2022, Art. no. 32.

[26] P. Dou, H. Shen, Z. Li, and X. Guan, "Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 103, no. 8, 2021, Art. no. 102477.

[27] D. Niu, L. Diao, L. Xu, Z. Zang, X. Chen, and S. Liang, "Precipitation forecast based on multi-channel ConvLSTM and 3D-CNN," in *Proc. Int. Conf. Unmanned Aircr. Syst.*, 2020, pp. 367–371.

[28] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single layer multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting," *Procedia Comput. Sci.*, vol. 135, pp. 89–98, 2018.

[29] S. Yao et al., "A ConvLSTM neural network model for spatiotemporal prediction of mining area surface deformation based on SBAS-InSAR monitoring data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5201722.

[30] B. Klein, L. Wolf, and Y. Afek, "A dynamic convolutional layer for short range weather prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4840–4848.

[31] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 879–888.

[32] 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. Int. Conf. Mach. Learn.*, 2018, pp. 5123–5132.

[33] 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, 2021, Art. no. 4103508.

[34] V. Bouget, D. Béréziat, 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.

[35] J. G. Fernández and S. Mehrkanoon, "Broad-UNet: Multi-scale feature learning for nowcasting tasks," *Neural Netw.*, vol. 144, pp. 419–427, 2021.

[36] K. Song et al., "Deep learning prediction of incoming rainfalls: An operational service for the city of Beijing China," in *Proc. Int. Conf. Data Mining Workshops*, 2019, pp. 180–185.

[37] K. Trebing, T. Staczyk, and S. Mehrkanoon, "SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture," *Pattern Recognit. Lett.*, vol. 145, pp. 178–186, 2021.

[38] P. Esser, E. Sutter, and B. Ommer, "A variational U-Net for conditional appearance and shape generation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8857–8866.

[39] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722.

[40] K. G. Ghosh, "Analysis of rainfall trends and its spatial patterns during the last century over the gangetic West Bengal, Eastern India," *J. Geovis. Spatial Anal.*, vol. 2, no. 2, 2018, Art. no. 15.

[41] 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.

[42] 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.

[43] B. Wang et al., "Deep uncertainty quantification: A machine learning approach for weather forecasting," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2087–2095.

[44] 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.

[45] C. Cheng, M. Chen, J. Wang, F. Gao, and H. Yang, "Short-term quantitative precipitation forecast experiments based on blending of nowcasting with numerical weather prediction," *Acta Meteor. Sinica*, vol. 71, no. 3, pp. 397–415, 2013.

[46] H. H. Lin et al., "Multi-weather evaluation of nowcasting methods including a new empirical blending scheme," *Atmosphere*, vol. 11, no. 11, 2020, Art. no. 1166.

[47] G. Wang, W.-K. Wong, Y. Hong, L. Liu, J. Dong, and M. Xue, "Improvement of forecast skill for severe weather by merging radar-based extrapolation and storm-scale NWP corrected forecast," *Atmos. Res.*, vol. 154, pp. 14–24, 2015.

[48] Y. A. Geng et al., "LightNet: A dual spatiotemporal encoder network model for lightning prediction," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2439–2447.

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

[50] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[51] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

Dan Niu was born in Jiangsu, China, on April 16, 1986. He received the Ph.D. degree in verification technologies for nonlinear circuits and systems from the Graduate School of Information, Production and Systems, Waseda University, Shinjuku, Japan, in 2013.

He is currently an Associate Professor with the School of Automation, Southeast University, Nanjing, China. He is also a member of Key Laboratory of Measurement and Control of CSE, Ministry of Education, Nanjing, China. His research interests include artificial intelligence (AI)-based weather forecasting method and AI for electronic design automation.

Hongshu Che was born in Harbin, China. He received the B.E. degree in measurement and control technology and instruments from Hangzhou Dianzi University, Hangzhou, China, in 2019, and the M.E. degree in pattern recognition and intelligent systems from Southeast University, Nanjing, China, in 2022.

He is currently an Algorithm Engineer with Alibaba Group, Hangzhou. His research interests include time-series analysis, spatiotemporal sequence forecasting, and data mining.

Chunlei Shi was born in Tongling, China. She received the B.E. degree in automation from Shandong University, Jinan, China, in 2018, and the M.E. degree in artificial intelligence from the Joint Graduate School of Southeast University and Monash University, Suzhou, China, in 2023. She is currently working toward the Ph.D. degree in control science and engineering from with the School of Automation, Southeast University, Nanjing, China.

Her current research interests include artificial intelligence for weather spatiotemporal sequence prediction and time-series analysis.

Zengliang Zang was born in Jiangsu, China, on February 20, 1977. He received the Ph.D. degree in Meteorology from College of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing, China, in 2005.

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

Hongbin Wang was born in Shanxi, China, on August 23, 1985. He received the Ph.D. degree in atmospheric physics and atmospheric environment from the College of Atmospheric Sciences, Lanzhou University, Lanzhou, China, in 2013.

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

Xunlai Chen was born in Henan, China, on April 20, 1979. He received the Ph.D. degree in meterology from the Department of Atmospheric Science, Sun Yat-sen University, Guangzhou, China, in 2007.

He is currently a Senior Engineer of Meteorology with Shenzhen Meteorological Bureau, Shenzhen, China. His research interests include artificial intelligence for nowcasting and numerical weather prediction.

Qunbo Huang was born in Anhui, China, on May 10, 1988. He received the Ph.D. degree in computer science and Technology from National University of Defense Technology, Changsha, China, in 2018.

He is currently an Engineer with the Unit of 93110, People's Liberation Army, Beijing, China. His research interests include numerical simulation and satellite data assimilation.