A Heterogeneous Spatiotemporal Attention Fusion Prediction Network for Precipitation Nowcasting

Dan Niu[1,3†], Hongshu Che[1], Chunlei Shi[1], Zengliang Zang[4], Hongbin Wang[2], Xunlai Chen[5], and Qunbo Huang[6]

***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 work, a heterogeneous spatiotemporal attention fusion prediction network (HST-AFP) is proposed for radar echo extrapolatio(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 (NWP) data. With the help of the proposed attention-based ST diffusion module (ASTD), the multi-encoder 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 scale-wise attention fusion module (SWAF) and deep residual spatial and temporal attention mechanism (DRSTA). It can adaptively blend multi-source ST features and rescale the multi-scale temporal-wise and spatial-wise features for better prediction. Experiments in a real-world dataset of South China show that compared with the ingenious RNN-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.**

***Index Terms*—Precipitation nowcasting, heterogeneous ST data, ST diffusion, deep residual ST attention**

I. Introduction

PRECIPTATION 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]–[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]–[9].

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 zero to two hours lead time. Thus the faster and more accurate extrapolation based methods are widely adopted [14]–[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 are explored as a way to fill this gap, which can capture complex nonlinear spatiotemporal patterns and combine heterogeneous data sources for prediction [20]–[25]. In principle, it can weaken 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 a spatiotemporal sequence forecasting problem, where previous radar map sequence are the input and the sequence of a fixed number of future radar maps are output [10]. Some progress has been made by utilizing deep learning techniques for precipitation nowcasting [18], [26]. Firstly, the pioneering recurrent neural network (RNN) and long short-term memory (LSTM) encoder-decoder framework proposed in [27]–[29] provide a general framework for tackling this problem. Klein

This work was supported by National Natural Science Foundation of China (No. 42005120), the National key R&D Program of China (No. 2019YFE0110100, 2018YFC1506905), Natural Science Foundation of Jiangsu Province of China (No. BK20202006), Key R&D Program of Jiangsu Province (No. BE2019052).

† Corresponding author(email:danniu1@163.com).

[1]School of Automation, Southeast University, Nanjing 210096, China

[2]Key Laboratory of Transportation Meteorology of China Meteorological Administration, Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210041, China (email: wanghb@cma.gov.cn)

[3]Key Laboratory of Measurement and Control of CSE, Ministry of Education, Nanjing 210096, China

[4]Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 210096, China

[5]Shenzhen Key Laboratory of Severe Weather in South China, and Shenzhen Meteorological Bureau, Shenzhen 518040, China

[6]93110 Troops, PLA, Beijing 100843, China

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. presented a predictive recurrent neural network (PredRNN) [31] in the light of the idea that spatiotemporal predictive learning should memorize both spatial appearances and temporal variations in a unified memory pool. To alleviate the gradient propagation difficulties in 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 the long-term dependencies, and outperform the previous models (ConvLSTM and TrajGRU) in some real datasets. However, gradients 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 spatiotemporal prediction [6], [33]–[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 U-Net based models are proposed for precipitation nowcasting based on the radar echo sequences [34]–[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], a SE-ResUNet network was proposed to predict rainfall dynamics for the city of Beijing China. It combined the strengths of U-Net, ResNet, Sqeeze-and-Excitation attention, and enabled significant performance improvement.

Besides, the above-mentioned 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, ground or satellite measurements [40], [41]. It is clear that the NWP data can supply important prior prediction knowledge for meteorological parameters [42]–[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]–[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 work, 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 spatiotemporal resolution in this work) to further improve the precipitation nowcasting. This architecture is flexible enough to add relevant multi-source 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 spatial-wise and temporal-wise fusion features from multiple heterogeneous spatiotemporal data sources (NWP forecasts and radar echo observations, and more if supplied). Experiment 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 work are summarized as follows:

1) To achieve precipitation nowcasting by accurate radar echo extrapolation, we propose a heterogeneous spatiotemporal attention fusion prediction network to mine knowledge from multiple heterogeneous ST data sources, where multi-scale residual learnings are integrated to construct a very deep ST fusion prediction network.
2) An attention-based spatiotemporal diffusion module (ASTD) is proposed to convert sparse spatiotemporal tensor into a spatial-wise and temporal-wise 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, scale-wise attention fusion module (SWAF) and deep residual ST attention mechanism (DRSTA) are proposed to adaptively rescale and blend the multi-source and multi-scale 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 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

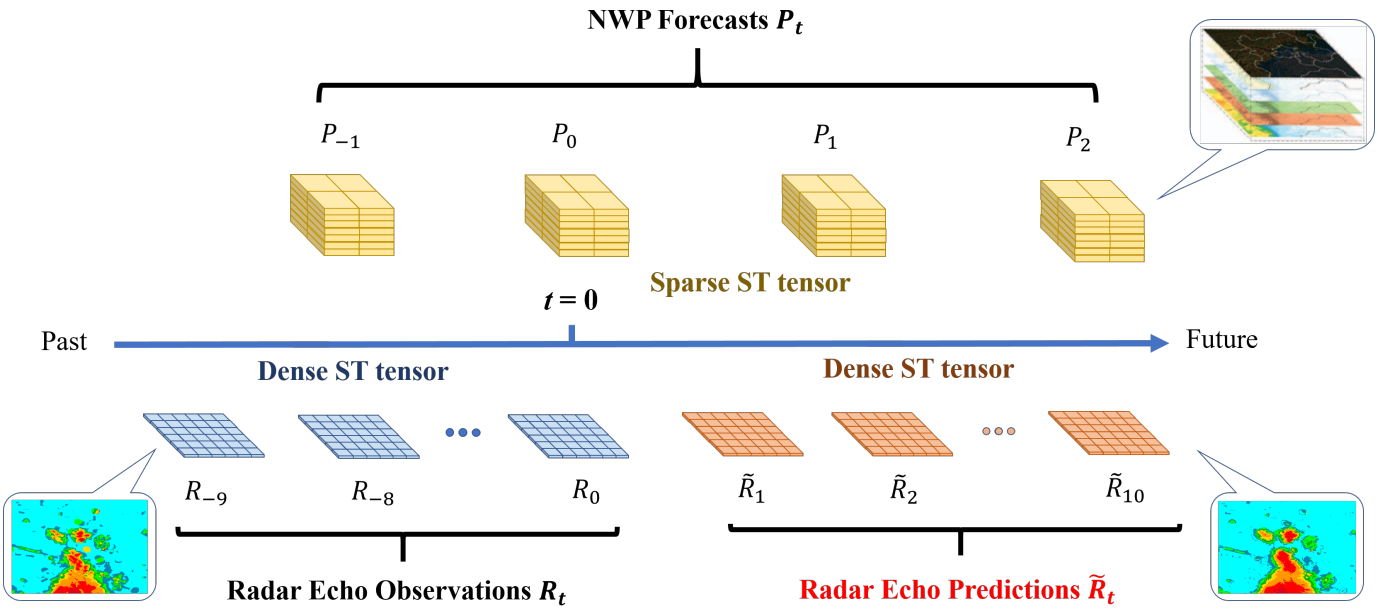

Fig. 1: The heterogeneous spatiotemporal structure of data in our task. On one hand, the radar echo history observation data and the NWP forecast data describe weather situations for different periods (the past and the future). On the other hand, they have different spatial and temporal resolutions. We seek to extract information from these heterogeneous data and produce a precipitation nowcasting.

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 work, the heterogeneous input data are first introduced, and then the problem to be solved is defined.

A. Heterogeneous Spatiotemporal Data

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

NWP Forecast Data (P).The NWP forecast product data with 3km spatial resolution and 1h temporal resolution are from Global/Regional Assimilation and Prediction System (GRAPES), which is a new-generation numerical weather prediction system developed by the China Meteorological Administration (CMA). In order to reduce the memory cost, the region covering 300km×300km 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 corresponding to a scope of 3km×3km in the real world. Each data grid carries simulation results of different meteorological parameters. The GRAPES performs hourly numerical simulation at 0:00, 12:00 (UTC) per day and each simulation covers the next 24 hours. Considering compute resources and GPU memory costs of model training, some parameters that are closely related to precipitation are selected: Cr, rain, rh with five height channels from 600hPa to 1000hPa. We concatenate the seven parameter channels at time t and form a comprehensive NWP forecast data P_t.

Radar Echo Observation Data (R). The radar echo dataset used in this work is a subset of the three-year weather radar intensities provided by Guangdong Meteorological Bureau from 2017 to 2019. The radar CAPPI reflectivity images, which have resolution of 300 × 300 pixels and also cover a 300km×300km area. It is obtained from some S-band radars, which are located at Guangzhou, Shenzhen, Shaoguan, etc. The spatial range is same with that of NWP forecast data. However, the radar echo observation interval is 12 minutes and the spatial resolution is 1km. It is clear that the two type data have different spatiotemporal resolution (12 minutes versus 1 hour, 1km versus 3km) 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 minutes and nowcasting is usually done for the following 2 hours, i.e., to predict the 10 frames ahead.

Suppose the current moment is $t = 0$. Given the NWP forecast $P = [P_t]_{t=-q}^{m}$ (real four-dimensional sparse tensor) from previous q hours to future hours, the radar echo observation $R = [R_t]_{t=-l+1}^{0}$ (three-dimensional 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 $\tilde{R} = [\tilde{R}_t]_{t=1}^{k}$ (three-dimensional dense tensor), where P, R and $\tilde{R}$ share the same x-y scope. Specifically, our goal is to find a mapping such that

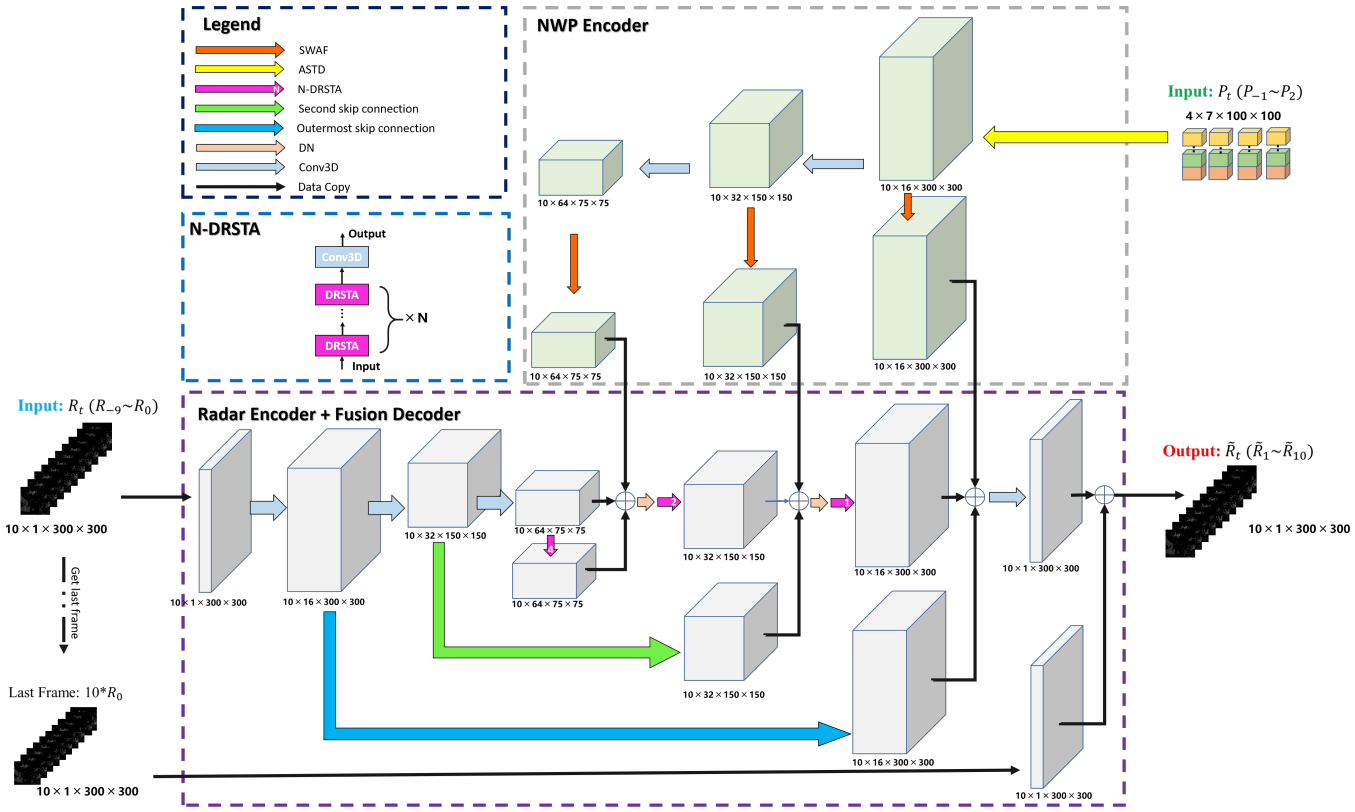

Fig. 2: The architecture of multi-source deep residual attention fusion prediction network (MS-DRAFP). $\oplus$ denotes the element-wise addition.

$$\min_{f} loss(\tilde{R}_t]_{t=1}^{k}, R_t]_{t=1}^{k}) \tag{1}$$
$$s.t.\ [\tilde{R}_t]_{t=1}^{k} = f([P_t]_{t=-q}^{m}, [R_t]_{t=-l+1}^{0})$$

Here we should emphasize the heterogeneous spatiotemporal structure of our data, as illustrated in Fig. 1. On 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, 3km & 1h) distinguishes it from the dense radar input and predicted radar output tensors (R *and* $\tilde{R}$, 1km & 12min). We seek to mining knowledge from these heterogeneous data.

III. HETEROGENEOUS SPATIOTEMPORAL ATTENTION FUSION PREDICTION NETWORK

Considering spatiotemporal heterogeneity and the high dimensionality of the multi-source spatiotemporal sequences especially for multi-step predictions, such prediction problem is nontrivial, unless the spatiotemporal features of the multi-source heterogeneous data are captured and fused well. This section presents the architecture of the proposed heterogeneous spatiotemporal attention fusion prediction network (HST-AFP), as illustrated in Fig. 2. The main network architecture is the multi-encoder and fusion-decoder structure. The past radar echo maps and the NWP forecast data are concatenated as the inputs of model, respectively. 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 spatiotemporal representations by stacked downscaling block, which usually halves the image size and doubles the number of feature maps. In the NWP forecast Encoder, attention based spatiotemporal diffusion module (ASTD) is firstly proposed to deal with spatiotemporal resolution distinction. In the decoder part, scale-wise attention fusion module (SWAF), and deep residual spatial and temporal attention module (DRSTA) are proposed to allow our decoder network to fuse discriminatively multi-level 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 multi-source and multi-scale spatiotemporal features, which guide the update of the lower level features and sequence prediction process with the multi-scale deep residual learning.

A. Multi-Encoder

Multi-source 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.

NWP Forecast Encoder. It is responsible for encoding the NWP forecast data. We stack the cr, rain, 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 two hour prediction but also the data at the previous one hour and the current timestamp are selected to supply more weather and variation trend information. P_t is a sparse tensor due to low spatial resolution (3km) and low temporal resolution (1h). However, the prediction output is the radar echo map sequence with high spatial resolution (1km) and high temporal resolution (12min). In this work, not simple convolution-based upsample but attention based spatiotemporal diffusion module (ASTD) is proposed firstly to deal with spatiotemporal resolution distinction.

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

More details about ASTD module is explained in subsec. D. Then feature extraction module is designed to extract multi-scale spatiotemporal representations by stacked downscaling blocks, which will be shown in detailed.

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: $[R_t]_{t=-l+1}^{0} = R_{-9}, R_{-8}, ..., R_0$. Since the spatiotemporal resolution of the input radar echo map is same with the predicted output radar echo sequence, no spatiotemporal diffusion module and only feature extraction module is required to obtain multi-scale feature tensors. Here the feature extraction module share the similar architecture with that of NWP encoder.

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

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

where $STE_{R,L}$ is the extracted low-level (small-scale) spatiotemporal features from the input past radar echo sequence R_t. $STE_{P,L}$ is the low-level spatiotemporal features extracted from the NWP spatiotemporal diffusion data $\hat{P}_t$. As shown in Fig. 2, $STE_{R,L}$, $STE_{P,L}$ 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 Eqs. (4)-(5), DSL consists of a 3-dimensional convolution (Conv3D) with stride 2 followed by a rectifier linear unit (ReLU) and a batch normalization unit. The Conv3D layer achieves bilinear downsampling, which halves the spatial size of input feature maps and doubles the number of feature maps. ReLU activation function is used to model nonlinear relations. Batch Normalization (BN) is a widely adopted technique that enables faster and more stable training of network. $STE_{R,M}$ and $STE_{P,M}$ features are the middle-level (middle-scale) spatiotemporal 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) spatiotemporal 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 multi-source and multi-scale spatiotemporal 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 work 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, multi-scale spatiotemporal 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. Firstly, 4-stacked deep residual spatiotemporal attention block (4-DRSTA) is proposed to construct the large-scale residual prediction module RP_L, which can extract and concentrate the discriminative high-level spatiotemporal prediction features from the deep feature maps of encoder $STE_{R,H}$.

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

The deep residual spatiotemporal attention block (DRSTA) can achieve quite large depth and provide very large receptive field size, which will be presented detailed in following subsection C. For high-level spatiotemporal encoder features $STE_{P,H}$ from the NWP forecast data, scale-wise attention fusion module (SWAF) is assigned to obtain the high-level refined prediction feature $STD_{P,H}$.

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

where scale-wise attention fusion module (SWAF) 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 with the multi-encoding part, two consecutive upsample fusion prediction module are designed, including middle-scale (middle-level) residual prediction module RP_M and small-scale (low-level) residual prediction module RP_S.

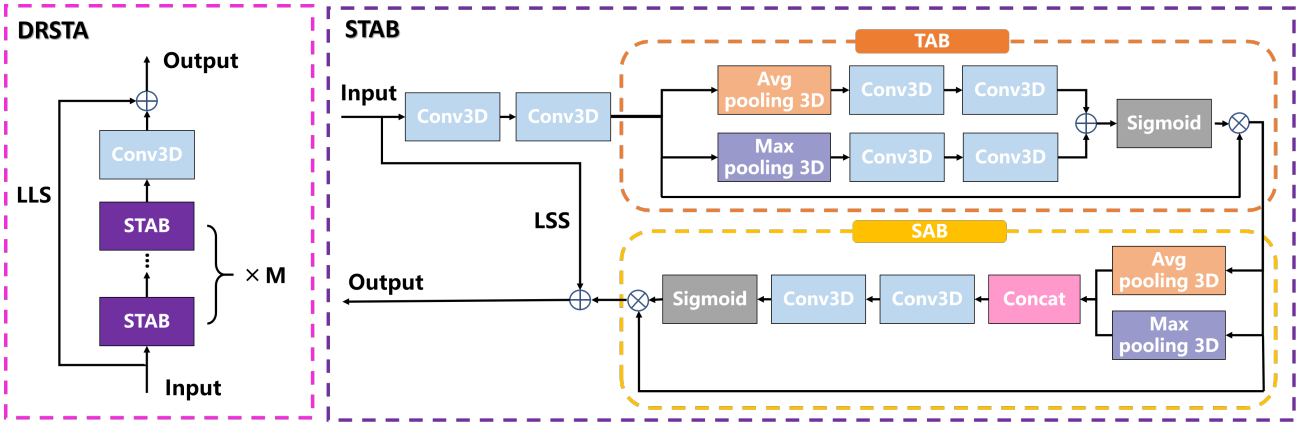

Fig. 3: Deep residual spatial and temporal attention block (DRSTA).

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

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

where RAP_M and RAP_L are middle-level and low-level residual attention prediction. $RP_M(RAP_H)$ and $RP_S(RAP_M)$ are extracted middle-level and low-level fusion prediction features. $SWAF(STE_{P.M})$ and $SWAF(STE_{P.L})$ are the middle-level and low-level NWP refined prediction features. As shown in Eq. (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, Batch Normalization (BN) and ReLU. Stacked DRSTA are employed to adaptively concentrate and rescale more useful spatiotemporal fusion features. The global skip-connections in the decoder part blend the multi-source and multi-scale spatiotemporal features to restore the lost information and generate the forecast radar echo sequence. Finally, we employ a 1×1×1 3D convolution to generate the residual prediction for the last radar echo frame and output the predicted radar echo sequence $\tilde{R}_t$.

$$\tilde{R}_t = \tilde{R}_1 ... \tilde{R}_{10} = Conv_{1\times1\times1}(RAP_L + R_0) \tag{11}$$

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

C. Deep Residual Spatiotemporal Attention Block

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

Instead of treating all features equally, temporal and spatial attention modules (TAB and SAB) are proposed for temporal-wise and spatial-wise weightings by modeling the interdependencies, which strengthen the discriminative learning ability and the representational power of deep networks. As shown in Fig. 3, the 3D temporal attention block (TAB) contains two 3D pooling descriptor branches for channel-wise statistic, sigmoid and a short cut. The two 3D pooling descriptor branches further contain 3D average pooling, two Conv3D and 3D max pooling, two Conv3D, respectively. Pooling descriptor gathers important clue about distinctive object features to infer finer channel-wise attention. Both average-pooled and max-pooled features are simultaneously used to greatly improve representation power of networks rather than using each independently [51]. The short-cut 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 3D spatial attention block (SAB) consists of two pooling branches (average and max) followed by concat operation, Conv3D, batch normalization unit, sigmoid and short cut. With temporal and spatial attention, the residual component in the STAB is adaptively rescaled.

Such TAB and SAB mechanisms allow our proposed net-

work to concentrate on more useful spatiotemporal 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. Attention based Spatiotemporal Diffusion Module

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

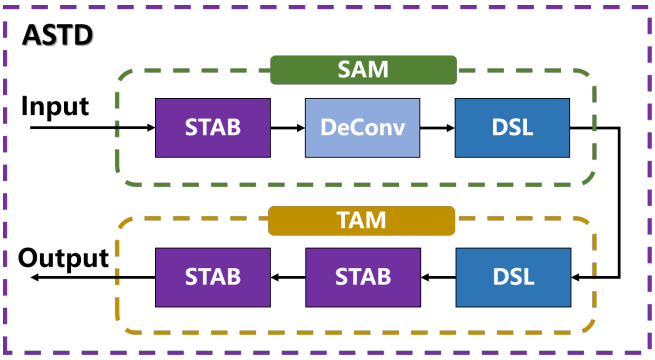

Fig. 4: Attention based spatiotemporal diffusion module (ASTD).

By employing the attention based nonlinear diffusion module ASTD, not by the linear or manually nonlinear interpolation, multi-level NWP spatiotemporal features can be better extracted and fused to improve prediction performance, which can be verified from experiment results.

IV. Experiments

A. Experiments Setup

The radar echo dataset and NWP forecast dataset used in this paper 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 300km×300km of the Pearl River Delta is selected, covering longitude ranges from 112° to 115°E and latitude from 22° 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 spatiotemporal data. The radar echo input $[R_T]_{t=-9}^{0}$ employs the previous two hour radar echo observations including the current one, and the input dimension is $10 \times 1 \times 300 \times 300$, due to 1km spatial resolution and 12min temporal resolution. The initial channel number of feature maps is 1. The width and height of the initial input tensor is 300. Another heterogeneous input is the NWP forecast data $[P_T]_{t=-Q}^{k}$ with seven parameter channels at each time slot from past one hour to future two hours, and the input dimension is $4 \times 7 \times 100 \times 100$ (sparse ST tensor) due to 3km spatial resolution and 1h 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 hours)). The total 356 precipitation events are split into a training set of 254 samples and a test set of 102 samples.

In this work, the proposed HST-AFP network is compared with typical optical flow based method (ROVER [17]), two ingenious RNN-based methods (TrajGRU [1], PredRNN++ [32]), and two well-known CNN-based methods (SmaAt-Unet [37] and SE-ResUNet [36]). Since the input of 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 $\beta 1 = 0.5$, $\beta 2 = 0.999$. The minibatch size is 8. The learning rate is initialized as 1×10^{-4} and decreased by 0.7 at every 5 epoch. 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), 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 (20dBZ, 35dBZ, 45dBZ) 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) is calculated. The four nowcasting

metrics are defined as follows:

$$
\begin{aligned}
CSI &= \frac{TP}{TP+FP+FN} \\
HSS &= \frac{2\times(TP\times TN - FP\times FN)}{(TP+FN)(FN+TN)+(TP+FP)(FP+FN)} \\
POD &= \frac{TP}{TP+FN} \\
FAR &= \frac{FP}{TP+FP}
\end{aligned}
\tag{12}
$$

B. Quantification Results

Tables I to III show the precipitation nowcasting metric results for the two-hour 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 spatiotemporal 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 35dBZ and 45dBZ thresholds. Note that, even HST-AFP_Radar method can also achieve better prediction performance compared with other models. At the 45dBZ threshold, the CSI of the proposed HST-AFP is over 0.019 higher than that of SE-ResUNet method (increase by about 12.9%), and also 0.012 higher than that of PredRNN++ model (increase by about 7.8%). Also, the HSS is much improved, about by 10.8% than that of SE-ResUNet method, over by 5.5% than that of 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 spatiotemporal resolution). At the 45dBZ 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 PredRNN++ method. The time is also about two times less even the NWP forecast data are added as the model input (HST-AFP).

TABLE I: Skill scores($\geq$ 20dBZ, 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($\geq$ 35dBZ, 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

TABLE III: Skill scores($\geq$ 45dBZ, 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(batchsize=1) and the forcast 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

C. Effect of ASTD, SAF and DRSTA

We study the effects of the attention based spatiotemporal diffusion module (ASTD), scale-wise attention fusion module (SWAF) and deep residual spatial and temporal attention block (DRSTA) in this part.

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

Scale-wise attention fusion module (SWAF). 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

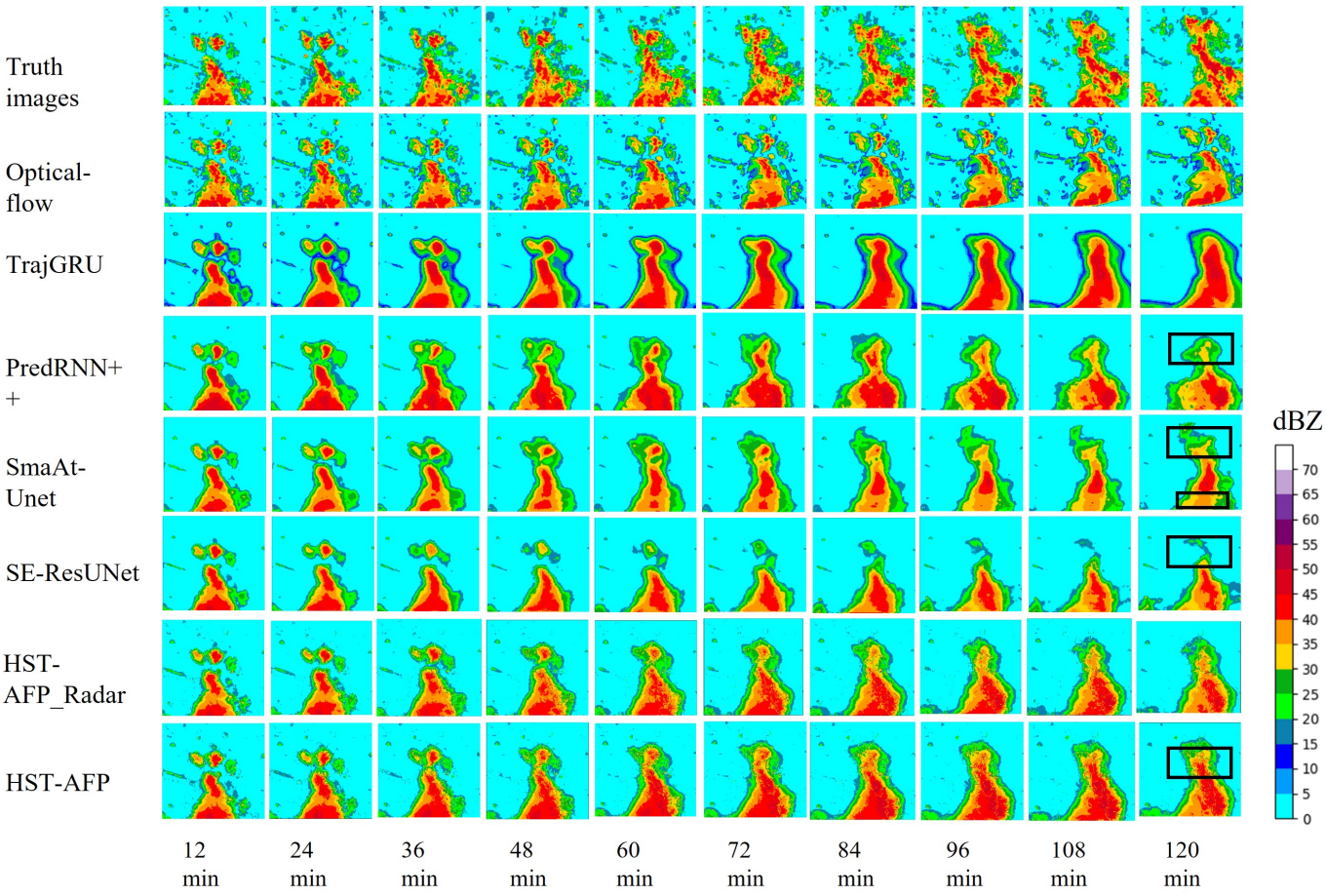

Fig. 5: Deep residual spatial and temporal attention block (DRSTA).

method to improve the prediction accuracy without increasing much weight parameter.

Deep Residual Spatial and Temporal Attention Block (DRSTA). 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 very deep trainable prediction network, and adaptively rescale and blend multi-source and multi-scale spatiotemporal features to achieve a high-to-low level residual fusion prediction.

TABLE V: Skill scores(≥20dBZ, 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

In addition, visualizing the comparisons among the evaluated methods are shown in Fig. 5. Nearly all methods predict accurately at the first moment. However, significant differences can be observed as the lead time increased. In the opticalflow-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

TABLE VI: Skill scores(≥35dBZ, 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(≥45dBZ, 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

boundaries become smooth. The blurring effect may be caused by the inherent uncertainties of the task. Although opticalflow-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 can not be effectively predicted and some echo regions are lost. Comparing with other deep learning methods, the proposed HST-AFP method shows the best performance. As time goes by, the intensity and the

position is closer to the truth images.

V. Conclusion

In this paper, we investigate to make the precipitation nowcasting via a CNN-based fusion prediction network framework extracting spatiotemporal information from multiple heterogeneous ST data sources. An attention-based spatiotemporal diffusion module (ASTD) has been proposed in the multi-encoder part to convert sparse NWP ST tensor into a spatial-wise and temporal-wise dense form, which can be effectively extracted by the ST encoder. In the fusion decoder part, scale-wise attention fusion module (SWAF) is designed to adaptively blend multi-source ST features. Moreover, deep residual spatial and temporal attention mechanism (DRSTA) is proposed to achieve very deep trainable residual fusion prediction network and discriminatively rescale the multi-scale temporal-wise and spatial-wise fusion features for guiding a high-to-low level residual fusion prediction. We have shown 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.

References

[1] X. Shi, Z. Gao, L. Lausen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[2] J. Ritvanen, B. Harnist, M. Aldana, T. M akinen, and S. Pulkkinen, "Advection-free convolutional neural network for convective rainfall nowcasting," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[3] X. Dong, Z. Zhao, Y. Wang, J. Wang, and C. Hu, "Motion-guided global–local aggregation transformer network for precipitation nowcasting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[4] B. N. Jones, "Machine learning tapped to improve climate forecasts.the approach helps to identify atmospheric processes and rank climate models by quality," 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 Sensing*, vol. 13, no. 21, p. 4285, 2021.

[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 Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–21, 2022.

[8] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, 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. Tr omel, "Nationwide radar-based precipitation nowcasting—a localization filtering approach and its application for germany," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 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," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[11] J. Sun, M. Xue, J. W. Wilson, I. Zawadzki, S. P. Ballard, J. Onvlee-Hooimeyer, P. Joe, D. M. Barker, P.-W. Li, B. Golding, et al., "Use of nwp for nowcasting convective precipitation: Recent progress and challenges," *Bulletin of the American Meteorological Society*, vol. 95, no. 3, pp. 409–426, 2014.

[12] G. Marchuk, *Numerical methods in weather prediction*. Elsevier, 2012.

[13] M. Tolstykh and A. Frolov, "Some current problems in numerical weather prediction," *Izvestiya Atmospheric and Oceanic Physics*, 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," *arXiv preprint arXiv:1912.12132*, 2019.

[15] C. Z. Basha, N. Bhavana, P. Bhavya, and V. Sowmya, "Rainfall prediction using machine learning & deep learning techniques," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 2020, pp. 92–97.

[16] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadtler, "Can deep learning beat numerical weather prediction?" *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200097, 2021.

[17] W. Woo and W. Wong, "Application of optical flow techniques to rainfall nowcasting," in *the 27th Conference on Severe Local Storms*, 2014.

[18] R. Prudden, S. Adams, D. Kangin, N. Robinson, S. Ravuri, S. Mohamed, and A. Arribas, "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," arXiv preprint arXiv:2005.04988, 2020.

[19] W. C. Woo and W. K. Wong, "Operational application of optical flow techniques to radar-based rainfall nowcasting," *Atmosphere*, vol. 8, no. 3, p. 48, 2017.

[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," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200083, 2021.

[21] L. Han, J. Sun, and W. Zhang, "Convolutional neural network for convective storm nowcasting using 3-d doppler weather radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1487–1495, 2019.

[22] J. Xia, H. Li, Y. Kang, C. Yu, L. Ji, L. Wu, X. Lou, G. Zhu, Z. Wang, Z. Yan, et al., "Machine learning-based weather support for the 2022 winter olympics," 2020.

[23] X. Yu and Y. Zheng, "Advances in severe convection research and operation in china," *Journal of Meteorological Research*, vol. 34, no. 2, pp. 189–217, 2020.

[24] H. Chen, Y. He, L. Zhang, S. Yao, W. Yang, Y. Fang, Y. Liu, and B. Gao, "A landslide extraction method of channel attention mechanism u-net network based on sentinel-2a remote sensing images," *International Journal of Digital 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," *Journal of Geovisualization and Spatial Analysis*, vol. 6, no. 2, p. 32, 2022.

[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," *International Journal of Applied Earth Observation and Geoinformation*, vol. 103, no. 8, p. 102477, 2021.

[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 *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 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 Computer Science*, vol. 135, pp. 89–98, 2018.

[29] S. Yao, Y. He, L. Zhang, W. Yang, Y. Chen, Q. Sun, Z. Zhao, and S. Cao, "A convlstm neural network model for spatiotemporal prediction of mining area surface deformation based on sbas-insar monitoring data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–22, 2023.

[30] B. Klein, L. Wolf, and Y. Afek, "A dynamic convolutional layer for short range weather prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 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," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[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 *International Conference on Machine Learning*. PMLR, 2018, pp. 5123–5132.

[33] L. Han, H. Liang, H. Chen, W. Zhang, and Y. Ge, "Convective precipitation nowcasting using u-net model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–8, 2021.

[34] V. Bouget, D. B er eziat, 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 Sensing*, vol. 13, no. 2, p. 246, 2021.

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

[36] K. Song, G. Yang, Q. Wang, C. Xu, J. Liu, W. Liu, C. Shi, Y. Wang, G. Zhang, X. Yu, et al., "Deep learning prediction of incoming rainfalls: An operational service for the city of beijing china," in *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2019, pp. 180–185.

[37] K. Trebing, T. Staczyk, and S. Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," *Pattern Recognition Letters*, 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8857–8866.

[39] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 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," *Journal of Geovisualization and Spatial Analysis*, vol. 2, no. 2, p. 15, 2018.

[41] C. Bai, D. Zhao, M. Zhang, and J. Zhang, "Multimodal information fusion for weather systems and clouds identification from satellite images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 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 Review*, vol. 148, no. 3, pp. 1099–1120, 2020.

[43] B. Wang, J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang, "Deep uncertainty quantification: A machine learning approach for weather forecasting," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & 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 Sensing*, vol. 11, no. 6, p. 642, 2019.

[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, C. C. Tsai, J. C. Liou, Y. C. Chen, C. Y. Lin, L. Y. Lin, and K. S. Chung, "Multi-weather evaluation of nowcasting methods including a new empirical blending scheme," *Atmosphere*, vol. 11, no. 11, p. 1166, 2020.

[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," *Atmospheric Research*, vol. 154, pp. 14–24, 2015.

[48] Y. A. Geng, Q. Li, T. Lin, L. Jiang, L. Xu, D. Zheng, W. Yao, W. Lyu, and Y. Zhang, "Lightnet: A dual spatiotemporal encoder network model for lightning prediction," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2439–2447.

[49] Y. A. Geng, Q. Li, T. Lin, W. Yao, L. Xu, D. Zheng, X. Zhou, L. Zheng, W. Lyu, and Y. Zhang, "A deep learning framework for lightning forecasting with multi-source spatiotemporal data," *Quarterly Journal of the Royal Meteorological Society*, 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.

[51] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

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