atmosphere

MDPI

Article

A Novel Multi-Input Multi-Output Recurrent Neural Network Based on Multimodal Fusion and Spatiotemporal Prediction for 0–4 Hour Precipitation Nowcasting

Fuhan Zhang [1], Xiaodong Wang [1] and Jiping Guan [2,*]

[1] School of Computer, National University of Defense Technology, Changsha 410000, China; zhangfuhan19@nudt.edu.cn (F.Z.); xdwang@nudt.edu.cn (X.W.)
[2] School of Meteorology and Oceanography, National University of Defense Technology, Changsha 410000, China
* Correspondence: guanjiping@nudt.edu.cn

Abstract: Multi-source meteorological data can reflect the development process of single meteorological elements from different angles. Making full use of multi-source meteorological data is an effective method to improve the performance of weather nowcasting. For precipitation nowcasting, this paper proposes a novel multi-input multi-output recurrent neural network model based on multimodal fusion and spatiotemporal prediction, named MFSP-Net. It uses precipitation grid data, radar echo data, and reanalysis data as input data and simultaneously realizes 0–4 h precipitation amount nowcasting and precipitation intensity nowcasting. MFSP-Net can perform the spatiotemporal-scale fusion of the three sources of input data while retaining the spatiotemporal information flow of them. The multi-task learning strategy is used to train the network. We conduct experiments on the dataset of Southeast China, and the results show that MFSP-Net comprehensively improves the performance of the nowcasting of precipitation amounts. For precipitation intensity nowcasting, MFSP-Net has obvious advantages in heavy precipitation nowcasting and the middle and late stages of nowcasting.

Keywords: radar echo data; reanalysis data; precipitation amount grid data; deep learning; spatiotemporal prediction; multimodal fusion; multi-task learning; RNN; precipitation nowcasting

Citation: Zhang, F.; Wang, X.; Guan, J. A Novel Multiple-Input Multiple-Output Recurrent Neural Network Based on Multimodal Fusion and Spatiotemporal Prediction for 0–4 h Precipitation Nowcasting. *Atmosphere* **2021**, *12*, 1596. https://doi.org/10.3390/atmos12121596

Academic Editors: Amin Talei and Miodrag Rancic

Received: 20 October 2021
Accepted: 22 November 2021
Published: 29 November 2021

1. Introduction

Precipitation nowcasting can forecast the distribution and development of precipitation in nearby periods with high temporal and spatial resolution. Accurate precipitation nowcasting can not only provide convenience for people's daily life [1,2] but also help in disaster prevention and mitigation [3,4]. The current common operational system for precipitation forecasts is the numerical weather prediction (NWP) [5] model, but it cannot provide accurate nowcasting due to spin-up [6]. In order to improve accuracy, deep learning has become an important development direction of precipitation nowcasting. Many scholars have conducted research [7–17] in this direction. Precipitation nowcasting based on deep learning includes precipitation intensity nowcasting and precipitation amount nowcasting, which predict the instantaneous value and the cumulative value of precipitation.

Precipitation intensity nowcasting based on deep learning is generally realized by radar echo extrapolation. First, the future radar echo data are predicted by the past radar echo data, and then the Z–R relationship is used to convert the predicted radar echo data into precipitation intensity data to realize precipitation intensity nowcasting. Various types of deep learning neural networks are used in radar echo extrapolation, including convolutional neural networks (CNN), recurrent neural networks (RNN), and generative adversarial networks (GAN). RNN uses the hidden state to save time-related features and continuously updates the hidden state according to new data as time advances. Compared with other deep learning neural networks, RNN has strong modeling capabilities for

sequence data. In 2015, Shi et al. [18] formulated radar echo extrapolation as a spatiotemporal prediction problem and used the ConvLSTM network, which applied a convolution structure to LSTM for prediction. Shi et al. [19] further proposed TrajGRU to improve the effect of precipitation intensity nowcasting. This approach uses the generated optical flow [20] to guide the connection structure in the network, and the point in the convolution structure is connected to points with a higher correlation instead of a fixed number of surrounding points. Many spatiotemporal prediction methods [21–23] have regarded radar echo extrapolation as one of the tasks to evaluate the spatiotemporal prediction ability of their methods. Wang et al. [24] proposed a spatiotemporal prediction method called PredRNN, which makes the spatial features of each layer of ConvLSTM interact in time series. PredRNN adds spatiotemporal memory units and connects them through a zigzag structure to make features spread in space and time. Bonnet et al. [10] used the spatiotemporal prediction method PredRNN++ [25] to achieve precipitation intensity nowcasting. PredRNN++ utilizes causal LSTM units to integrate temporal and spatial features and the gradient highway (GHU) to alleviate gradient disappearance. In order to improve the long-term extrapolation ability of the spatiotemporal prediction model, HPRNN [26] proposes a hierarchical prediction strategy which reduces the accumulation of prediction errors over time through a recurrent coarse-to-fine mechanism. Lin et al. [27] proposed self-attention memory (SAM) to memorize the features of long-distance dependence in terms of spatial and temporal domains. SAM can be embedded in most spatiotemporal prediction recurrent neural networks. Wu et al. [22] utilized the MotionRNN framework to capture complex changes in motion and adapt to the spatiotemporal changes in the scene, simultaneously modeling transient changes and motion trends through the MotionGRU unit.

There are few precipitation amount nowcasting methods based on machine learning and deep learning. Zhang et al. [28] used a multi-layer perceptron to forecast the precipitation amount data of 56 weather stations in China for the next 3 h. The forecast is derived from 13 physical factors related to precipitation in the surrounding area. RN-Net [9] regards precipitation amount nowcasting as a spatiotemporal prediction problem. It takes past precipitation amount grid data and radar echo data as input data, and then forecasts the precipitation amount grid data for the next 2 h.

Unlike most other computer vision problems, some data in the meteorological field have multiple sources. Multiple sensors from multiple angles observe a meteorological element to obtain data. The WRF model [29] uses a variety of data assimilation methods to fuse meteorological data. Most weather nowcasting methods based on deep learning do not take advantage of the multi-source data. They only use themselves as the basis for forecasting. Among the few weather nowcasting methods that use multi-source data, the multimodal fusion method is relatively simple. LightNet [30] uses WRF simulation data and observation data as the basis for 0–6 h lightning nowcasting, and RN-Net [9] uses radar echo data and observation data as the basis for 0–2 h precipitation amount nowcasting. Both use the late fusion method to fuse a variety of data. However, simply fusing the extracted spatiotemporal features cannot fully utilize the advantages of multi-source data.

In this paper, we make full use of multi-source meteorological data through a deep learning network design. For 0–4 h precipitation nowcasting, precipitation grid data, radar echo data, and reanalysis data are used as the basis for nowcasting. We deeply combine spatiotemporal prediction and multimodal fusion and adopt the multi-task learning [31] strategy. This paper proposes a novel precipitation nowcasting model, MFSP-Net, based on spatiotemporal scale fusion, which simultaneously realizes precipitation intensity nowcasting and precipitation amount nowcasting. MFSP-Net can take meteorological data with different temporal and spatial resolutions as its input. The dual-input dual-output MFSP-LSTM unit in the model retains the spatiotemporal information flow of the two sources of input data while fusing them at the hidden state level. The global spatiotemporal receptive field of the MFSP-LSTM unit is expanded by introducing the SAM unit. Multi-task learning strategy is used in training, and three kinds of nowcasting are learned in parallel. The dataset used in the experiment includes 20-month precipitation amount grid data, radar

echo data, and reanalysis data, and two months of precipitation amount forecast data of the WRF model for comparison. The experimental results show that MFSP-Net is better than RN-Net [9] for precipitation amount nowcasting. Compared with other precipitation intensity nowcasting models, MFSP-Net has apparent advantages in heavy precipitation nowcasting.

The rest of this paper is organized as follows: Section 2 introduces the data used in the paper. Section 3 introduces common networks composed of multimodal fusion and spatiotemporal prediction and details the proposed MFSP-Net. The results and analysis of the experiment and the ablation study are in Section 4. Discussions and conclusions are given in Sections 5 and 6, respectively.

2. Data

2.1. Dataset

The dataset used in this paper includes precipitation amount grid data, radar echo data, and reanalysis data, which are related to precipitation. Below, we introduce the three types of data.

Precipitation amount grid data are precipitation fusion products that combine automatic weather station precipitation amount data and satellite retrieval precipitation products. The automatic weather stations' precipitation amount data contain the precipitation amount data of more than 30,000 automatic weather stations in China. The satellite retrieval precipitation products are the precipitation products derived from real-time satellites (CMORPH) [32] developed by the Climate Prediction Center of the National Centers for Environmental Prediction (NCEP). The first probability density function matching and the optimal interpolation are used for fusion. Its overall error is within 10%, which can be approximated as the truth of precipitation amount data.

We use Doppler radar mosaic data as the radar echo data in this paper. The radar echo data contain many echo noises, such as interference echoes, non-meteorological echoes, etc., which affect the nowcasting. In the experiment, we construct a singular point filter and a bilateral filter to filter the value domain and the spatial domain, effectively eliminating the pulsation and clutter while retaining the echo characteristics. In addition, a high-pass filter is constructed to remove data below 15 dBZ, and only data related to precipitation are retained.

The reanalysis data comprise precipitation-related data from the ERA5 dataset [33] of the European Centre for Medium-Range Weather Forecasts (ECMWF). The data contain temperature, relative humidity, geopotential, vorticity, and wind (divided into v direction and u direction) data at 500 hPa, 700 hPa, and 850 hPa. When using the reanalysis data, we concatenate them into a tensor with 18 channels.

The spatial range and time range of the three types of data are the same. Their spatial range is 21–33° N and 112–124° E. The approximate spatial range is shown in Figure 1. This area is located in southeastern China and has a subtropical monsoon climate. Their time range includes May–September 2017–2020. In the dataset, 465 days for training, 29 days for validation, and 57 days for testing are included. The data on some days were incomplete due to equipment failure or other reasons. The time and spatial resolution of different data are different, as shown in Table 1. In addition, due to the large gap in the numerical range of various data, we have normalized various data separately.

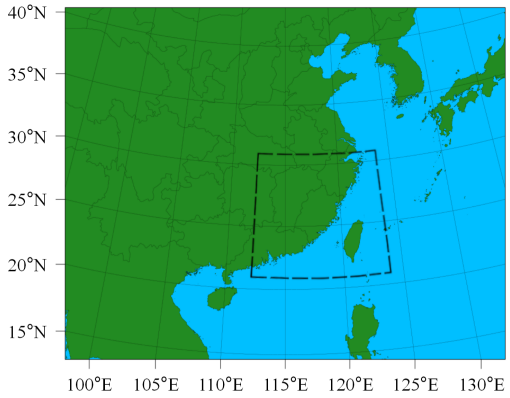

Figure 1. Schematic diagram of the location of the dataset. The area within the dotted line is the spatial range of the dataset.

Table 1. The parameter information of the three types of data. The original time resolution of the radar echo data is 6 min.

	Time Resolution	Spatial Resolution	Size
Precipitation amount grid data	1 h	10 km	120 × 120
Radar echo data	12 min (6 min)	5 km	240 × 240
Reanalysis data	1 h	25 km	48 × 48

2.2. WRF Model

The WRF model is used to compare the 0–4 h precipitation amount nowcasting effect with MFSP-Net. The WRF model [34] is configured with a one-domain nested grid system. The horizontal resolution of the domain is 10 km, with 120 × 120 grid points. The domain has 35 vertical layers, with the model top at 50 hPa. The boundary conditions are updated every 6 h from the 0.25° × 0.25° National Centers for Environmental Prediction (NCEP) Final Operational Model Global Tropospheric Analysis. The primary physical parameterization schemes are shown in Table 2. The model is integrated every 6 h, the forecast time is 12 h, and the results are output every 1 h.

Table 2. Physical parameterization schemes.

Name	Scheme
Microphysics	Thompson scheme
Cumulus parameterization	Kain–Fritsch (new Eta) scheme
Planetary boundary layer	Mellor–Yamada–Janjic TKE scheme
Surface layer	Revised MM5 Monin–Obukhov scheme
Longwave radiation	Rapid radiative transfer model for GCMs
Shortwave radiation	Rapid radiative transfer model for GCMs

3. Model

The deep learning neural networks used in this paper combine multimodal fusion and spatiotemporal prediction. In Section 3.1, we review the framework of the spatiotemporal prediction method and introduce the architecture of the ConvLSTM network. In Section 3.2, we introduce two spatiotemporal prediction models that use the simple multimodal fusion method and use them as comparison networks in the experiment. We introduce the architecture of MFSP-Net in Section 3.3. In Section 3.4, we detail the structure of the MFSP-LSTM unit.

3.1. Spatiotemporal Prediction Network

Spatiotemporal prediction is performed for spatiotemporal sequence data, such as video frames. The spatiotemporal sequence data can be regarded as a series of 3D tensors $X_1, X_2 \cdots X_t$ in the time period t. The size of each 3D tensor is $C \times M \times N$, where C, M, and N represent the number of channels, length, and width, respectively. The spatiotemporal prediction predicts the most probable length-n sequence in the future given the previous length-q sequence including the current observation.

The ConvLSTM network is an RNN with an encoding–decoding structure. It encodes the past data to extract the spatiotemporal features and then decodes them to make predictions. In the ConvLSTM unit, cell outputs C_t, hidden states H_t, and gates i_t, f_t, g_t, and o_t are 3D tensors similar to input X_t. To better understand these, we can imagine them as vectors standing on a spatial grid. ConvLSTM determines the future state of each cell in the grid by the inputs and past states of its local neighbors. This can easily be achieved by using convolution operators in the state-to-state and input-to-state transitions. The key equations of ConvLSTM unit are shown as follows:

$$\begin{aligned} g_t &= tanh(W_{xg} * X_t + W_{hg} * H_{t-1} + b_g) \\ i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \\ C_t &= f_t \odot C_{t-1} + i_t \odot g_t \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \\ H_t &= o_t \odot tanh(C_t) \end{aligned} \tag{1}$$

where σ, $*$, and $\odot$ denote the sigmoid activation function, the convolution operator, and the Hadamard product, respectively. The use of the input gate i_t, forget gate f_t, output gate o_t, and input-modulation gate g_t controls information flow across the memory cell C_t.

3.2. Multimodal Fusion and Spatiotemporal Prediction

Most spatiotemporal prediction methods use a single-input scheme, and they cannot directly input multiple data. In order to realize the multi-input scheme, the most convenient method is to combine the spatiotemporal prediction method with the primary multimodal fusion method, such as early fusion and late fusion [35]. Here, we choose ConvLSTM as the primary network.

The early fusion method is the simplest way to extend ConvLSTM to the multi-input scheme. It first fuses multiple data by the fusion convolutional layer and passes it to ConvLSTM. From the temporal standpoint, one can view this as a type of early fusion. The architecture of the dual-input EF-ConvLSTM is shown in Figure 2. Its structure is similar to ConvLSTM and consists of an encoder and a decoder. The fusion convolutional layer is added to the input port of the encoder.

The LF-ConvLSTM can be regarded as a multi-encoder ConvLSTM network, and the architecture of the dual-input LF-ConvLSTM is shown in Figure 3. Each type of input data has an encoder. The spatiotemporal feature of each type of data is extracted through the encoder and stored in cell outputs and hidden states. Then, the cell outputs and hidden states of the various kinds of data are respectively fused in the fusion module and used as the input of the decoder.

Both the early fusion method and the late fusion method can realize the multi-input scheme. Similar models have achieved significant effects, such as RN-Net and LightNet. However, the two did not make full use of multi-source data. The early fusion method fuses the spatiotemporal information flow of multiple data on the time scale. The late fusion method fuses the spatiotemporal information flow of multiple data on the spatial scale.

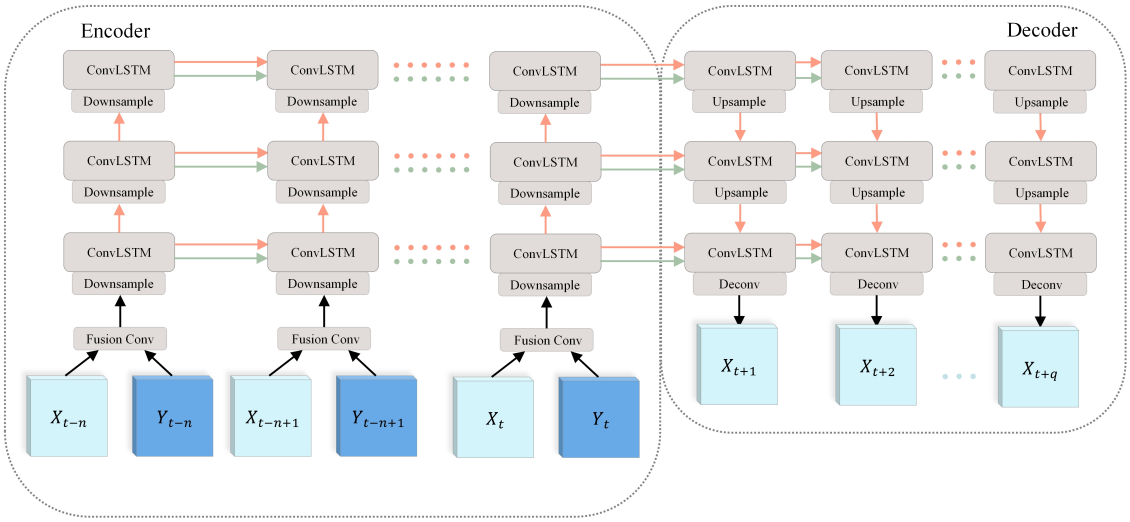

Figure 2. The dual-input early fusion ConvLSTM (EF-ConvLSTM) architecture. The two inputs are 3D tensors, defined as X_t and Y_t. EF-ConvLSTM forecasts the $X_{t+1} \cdots X_{t+q}$ in the future given the previous $X_{t-n} \cdots X_t$ and $Y_{t-n} \cdots Y_t$ including the current data.

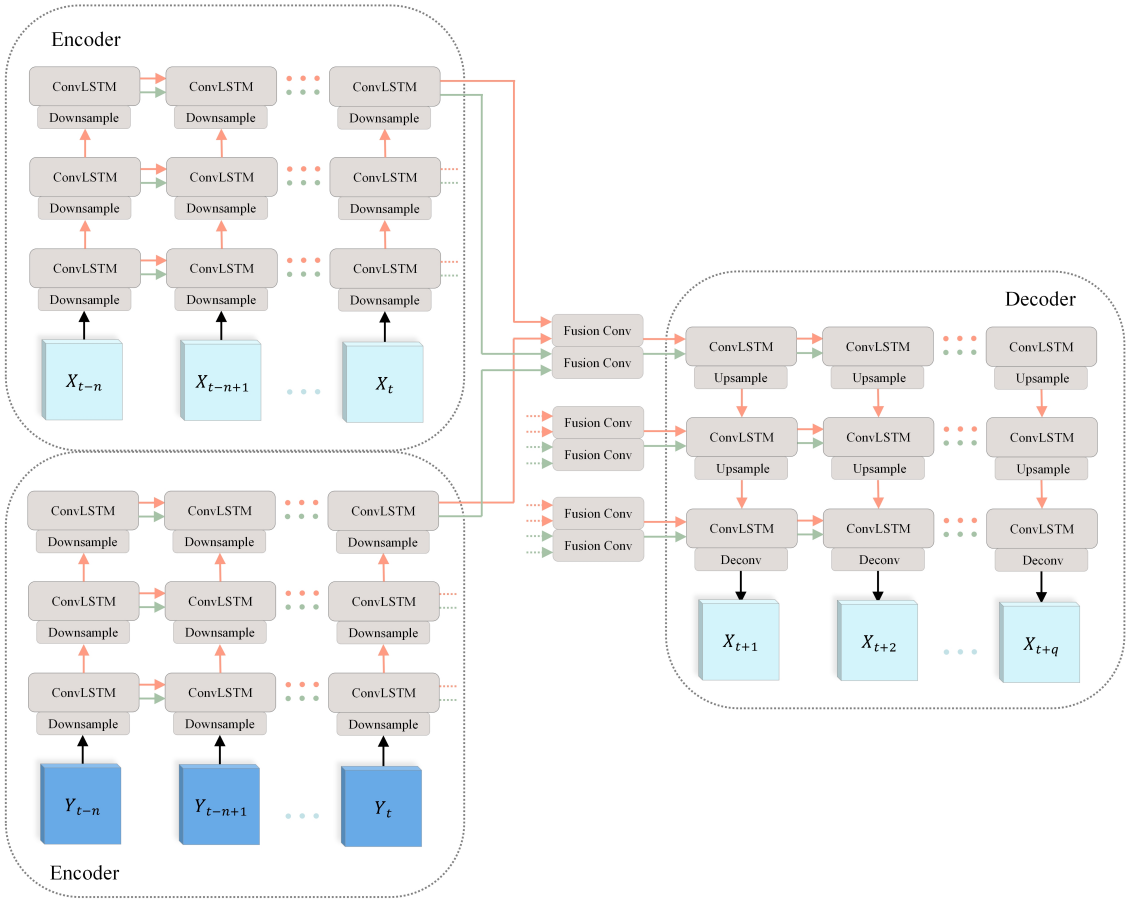

Figure 3. The dual-input late fusion ConvLSTM (LF-ConvLSTM) architecture. It contains two encoders, one decoder, and the fusion module. The two inputs are 3D tensors, defined as X_t and Y_t. LF-ConvLSTM forecasts the $X_{t+1} \cdots X_{t+q}$ in the future given the previous $X_{t-n} \cdots X_t$ and $Y_{t-n} \cdots Y_t$ including the current data.

3.3. MFSP-Net

As shown in Table 1, the time resolution and spatial resolution of the three types of data in the dataset are different. To make full use of these multi-source data, we hope that the model can input multi-source data with different spatiotemporal resolutions and fuse them on time and space scales. For this purpose, MFSP-Net deeply fuses multi-source data in each MFSP-LSTM unit, and its special network structure makes it adapt to input data with different spatiotemporal resolutions. The MFSP-LSTM unit has a major input, minor input, major output, and minor output. The data corresponding to the major input and major output dominate the development of the spatiotemporal information flow of this layer, while the data corresponding to the minor input and minor output play an auxiliary role. Since the precipitation amount grid data reflect the most natural precipitation development process, we take the features and status of the precipitation amount grid data as the major input. The radar echo data and reanalysis data reflect the rough precipitation development process and contain some incorrect information. We use the features and status of radar echo data and reanalysis data as the minor input. In addition, to better guide network training, we use the multi-task learning strategy. In order to obtain three kinds of accurate nowcasting simultaneously, MFSP-Net needs to deeply fuse three types of data while retaining their respective spatiotemporal information flows. MFSP-Net is an RNN with an encoding–decoding structure, and its architecture is shown in Figure 4. Below, we introduce the encoder and decoder of MFSP-Net separately.

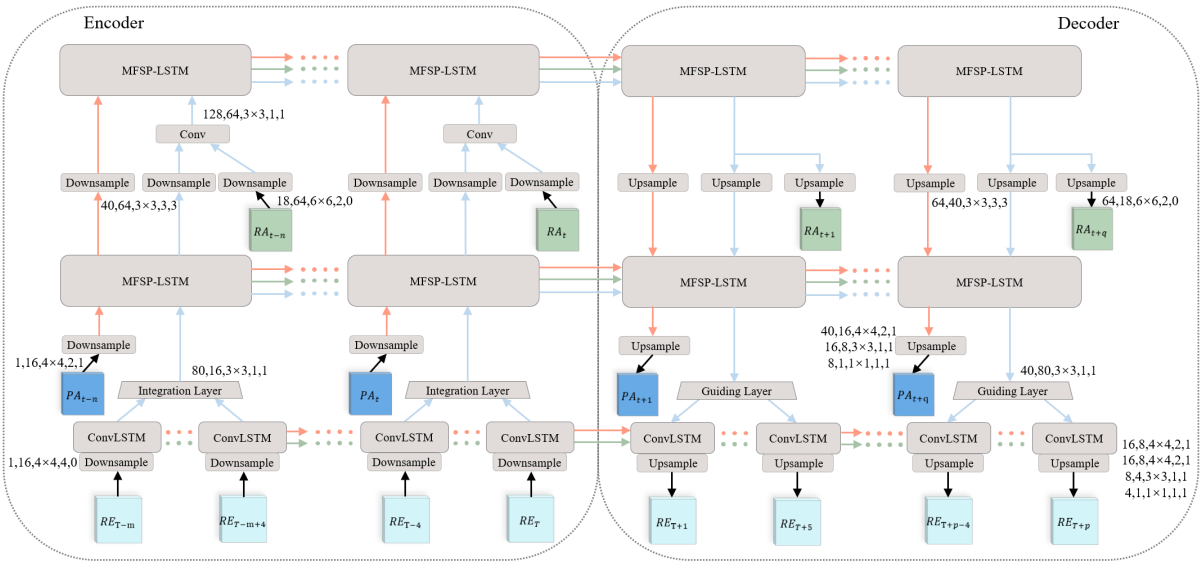

Figure 4. The three-input three-output MFSP-Net architecture. The three types of data are radar echo data RE, precipitation amount grid data PA, and reanalysis data RA. MFSP-Net forecasts the $PA_{t+1} \cdots PA_{t+q}$, $RA_{t+1} \cdots RA_{t+q}$ and $RE_{T+1} \cdots RE_{T+p}$ in the future given the previous $PA_{t-n} \cdots PA_t$, $RA_{t-n} \cdots RA_t$ and $RE_{T-m} \cdots RE_T$ including the current data.

Encoder: As the spatial resolutions of RE, PA, and RA are 5 km, 10 km, and 25 km, respectively, it is impossible to fuse the three types of data in one layer. In the encoder of MFSP-Net, each layer inputs one type of data and inputs the data in descending order of the spatial resolution of the data. The RNN units of the first layer (Bottom-up) are ConvLSTM units and input radar echo data. The RNN units of the second and third layers are dual-input dual-output MFSP-LSTM units. The major input of the second layer is precipitation amount grid data, and the minor input is the spatiotemporal features extracted from the previous layer. The major input of the third layer is the major output of the previous layer, and the minor input is the fusion feature of the reanalysis data and the minor output of the previous layer. This network structure can enhance the network's applicability to the spatiotemporal resolution of data and enhance the scalability of the network input scheme. In addition, the input data and the spatiotemporal features of the previous layer

need to pass through a downsample unit before inputting the RNN unit to transform the two into the same size as the next layer and obtain higher-level features. Since the time resolutions of radar echo data, precipitation amount data, and reanalysis data are 12 min, 1 h, and 1 h, respectively, the time resolutions of the first and second layers are different. Meanwhile, the radar echo data are instantaneous, and the precipitation amount grid data are cumulative. Therefore, we correspond 1 frame of precipitation amount grid data to 5 frames of radar echo data in the cumulative period in the network. Drawing on the idea of HPRNN [26], we added an integration layer between the first and second layers to unify the time resolution.

Decoder: The decoder architecture is similar to the encoder, but their data flow direction is opposite. The structural difference can be considered as three points: first, the downsample unit is replaced by the upsample unit; second, the two types of input data of the third layer MFSP-LSTM units are zero tensors. The minor output data of the third layer are used as the minor input of the second layer and also used as the basis for the reanalysis data forecast; third, the minor output of the second layer passes through the guiding layer to obtain the spatiotemporal features of five frames of radar echo data.

We set special parameters for the upsample layer and downsample layer between each layer. The specific parameters are shown in Figure 4. The kernel size of all RNN units' convolution is 3, and the step size is 1. In addition, MFSP-Net also supports dual-input schemes. When the input data are precipitation amount grid data and radar echo data, it becomes PA-RE-MFSP-Net. PA-RE-MFSP-Net is still a three-layer structure, but the minor input of the third layer of the encoder and the minor output of the third layer of the decoder do not contain reanalysis data. When the input data are precipitation amount grid data and reanalysis data, it becomes PA-RA-MFSP-Net. Its structure is like MFSP-Net without the second layer. The input data of RE-RA-MFSP-Net are radar echo data and reanalysis data. The RNN units of the first and second layers are ConvLSTM units, and the RNN units of the third layer are MFSP-LSTM units. We use these dual-input schemes as part of the ablation study to explore the respective roles of the three kinds of input data.

3.4. MFSP-LSTM

In order to deeply fuse multiple data while retaining their respective spatiotemporal information flow, we propose a new dual-input dual-output RNN unit: MFSP-LSTM. The detailed structure of MFSP-LSTM is shown in Figure 5. MFSP-LSTM retains the spatiotemporal transformation information flow of the major input data X_t and the minor input data Y_t, respectively, and fuses the two types of data from the hidden state level to achieve spatiotemporal-scale fusion. Update equations of the MFSP-LSTM unit can be presented as follows:

$$
\begin{aligned}
g_{x\_t} &= tanh(W_{xg} * X_t + W_{hg\_x} * H_{t-1} + b_{g\_x}) \\
i_{x\_t} &= \sigma(W_{xi} * X_t + W_{hi\_x} * H_{t-1} + b_{i\_x}) \\
f_{x\_t} &= \sigma(W_{xf} * X_t + W_{hf\_x} * H_{t-1} + b_{f\_x}) \\
C_t &= f_{x\_t} \odot C_{t-1} + i_{x\_t} \odot g_{x\_t} \\
g_{y\_t} &= tanh(W_{yg} * Y_t + W_{hg\_y} * H_{t-1} + b_{g\_y}) \\
i_{y\_t} &= \sigma(W_{yi} * Y_t + W_{hi\_y} * H_{t-1} + b_{i\_y}) \\
f_{y\_t} &= \sigma(W_{yf} * Y_t + W_{hf\_y} * H_{t-1} + b_{f\_y}) \\
N'_t &= f_{y\_t} \odot N_{t-1} + i_{y\_t} \odot g_{y\_t} \\
o_{x\_t} &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} * C_t + W_{no\_x} * N'_t + b_{o\_x}) \\
H'_t &= o_{x\_t} \odot tanh(W_{1\times1} * [C_t, N'_t]) \\
o_{y\_t} &= \sigma(W_{yo} * y_t + W_{no\_y} * N'_t + b_{o\_y}) \\
N''_t &= o_{y\_t} \odot tanh(N'_t) \\
[H_t, N_t], M_t &= SAM([H'_t, N''_t], M_{t-1})
\end{aligned}
\tag{2}
$$

The MFSP-LSTM unit contains the temporal memory C_t of the major input, the temporal memory N'_t of the minor input, the hidden state H_t of the major input, and the hidden state N_t of the minor input, where t represents the time step. N'_t, N''_t, and H'_t are intermediate variables and will not be saved or transmitted. We use the features of the precipitation amount grid data or the major output of the previous layer as the major input and the features of other data or the minor output of the previous layer as the minor input. The current temporal memory C_t of the major input depends on the past state C_{t-1} and is controlled through a forget gate f_{x_t}, an input gate i_{x_t}, and a modulation gate g_{x_t}. The current temporal memory N'_t of the minor input depends on the past state N_{t-1} and is controlled through a forget gate f_{y_t}, an input gate i_y_t, and a modulation gate g_{y_t}. The calculation of the hidden state of the major input includes the time memory of the major input and the minor input to realize the fusion of the two input data. In order to make the hidden state and time memory have the same dimensionality, we concatenate these time memories together and then apply a 1×1 convolution layer for dimension reduction. The hidden state of the minor input only depends on its time memory, which is conducive to preserving the spatiotemporal transformation information of the minor input data. In addition, to improve the MFSP-LSTM unit's adaptability to long-distance dependence, we followed Lin's work [27] and added the SAM to the MFSP-LSTM unit. We expanded the global spatiotemporal receptive field of H_t and N_t by the memory Mt.

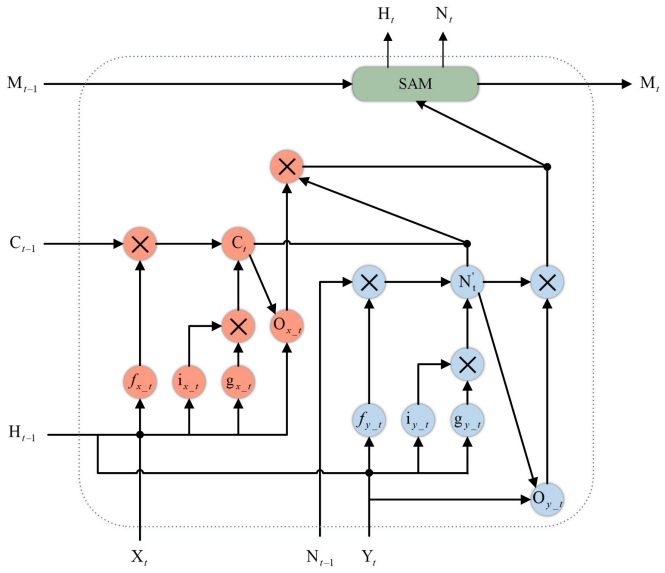

Figure 5. The dual-input early fusion ConvLSTM (EF-ConvLSTM) architecture. The two inputs are 3D tensors, defined as X_t and Y_t. EF-ConvLSTM forecasts the $X_{t+1} \cdots X_{t+q}$ in the future given the previous $X_{t-n} \cdots X_t$ and $Y_{t-n} \cdots Y_t$ including the current data.

4. Experiment

In the experiment, our task can be defined as nowcasting the precipitation amount data and the precipitation intensity data (radar echo data) of the next 4 h, based on the precipitation amount grid data, reanalysis data, and the radar echo data of the past 4 h. In Section 4.1, we introduce the implementation details of the experiment. We introduce the performance metric of precipitation amount nowcasting and precipitation intensity nowcasting in Section 4.2. In Section 4.3, for precipitation amount nowcasting, we compare MFSP-Net with the WRF model and deep learning models, including different multimodal fusion methods and benchmark models. For precipitation intensity nowcasting, we compare MFSP-Net with different benchmark models. In Section 4.4, we conduct ablation studies to verify the effectiveness of the network structure and explore the role of multi-source data. We visualize three representative examples for further analysis in Section 4.5.

4.1. Implementation Details

Our experimental platform used Ubuntu16.04 with 32 GB memory and two Nvidia RTX 2080 GPUs. The proposed neural networks were implemented with Pytorch [36] and trained end-to-end. All network parameters were initialized with a normal distribution. All models were trained using the Adam optimizer [37] with a starting learning rate of 10^{-4}. The training process was stopped after 60,000 iterations, and the batch size of each iteration was set to 4. All data normalized to the range of [0, 1] were used as network input data. Inspired by Tran's work [38], we combined MAE, MSE, and SSIM as the *Loss* function of the network to reduce image blurring. Since meteorological data are sparse, the result is close to 1 when calculating SSIM. For this reason, we increased the weight of SSIM. In addition, MFSP-Net makes predictions for three types of data. We calculated the *Loss* for the three types of data and added weight to the *Loss* as part of the *Multi-Loss* function. The value of precipitation amount data is small, which causes the *Loss* of precipitation amount data to be much smaller than other data. For this reason, we increased the weight of the *Loss* of precipitation amount data. The calculation formula of the *Multi-Loss* is as follows:

$$\begin{aligned} Loss(Z, \widehat{Z}) &= a * MAE(Z, \widehat{Z}) + b * MSE(Z, \widehat{Z}) + c * SSIM(Z, \widehat{Z}) \\ Multi\text{-}Loss &= A * Loss(PA, \widehat{PA}) + B * Loss(RE, \widehat{RE}) + C * Loss(RA, \widehat{RA}) \end{aligned} \tag{3}$$

where Z, PA, RE, and RA are the truth data, and $\widehat{Z}$, $\widehat{PA}$, $\widehat{RE}$, and $\widehat{RA}$ are the forecast data. In the experiment, a, b, and c are set to 1, 1, and 2500, respectively. A, B, and C are set to 50, 1, and 1, respectively.

4.2. Performance Metric

The nowcasting result of deep learning methods in the experiment is regarded as multi-frame sequence data. The precipitation amount nowcasting result contains four frames of data, which are the cumulative precipitation data for every 1 h in the next 4 h. The precipitation intensity nowcasting result contains 20 frames of radar echo data, and the time resolution is 12 min. The nowcasting effect is evaluated by comparing these results with truth radar echo data or precipitation amount data. Commonly used metrics for precipitation nowcasting in the meteorological field include the Critical Success Index (CSI), probability of detection (POD), and false alarm rate (FAR). We use the thresholds of 0.5 mm, 2 mm, and 5 mm to calculate these metrics for precipitation amount and use the thresholds of 0.5 mm/h, 2 mm/h, and 5 mm/h to calculate these metrics for precipitation intensity. The dBZ is the unit of radar echo data, which can be converted to mm/h by the Z–R relationship. These threshold settings refer to the precipitation level, and the corresponding relationship is shown in Table 3.

Table 3. Correspondence between threshold and precipitation level.

Precipitation Amount per Hour (mm)	Precipitation Level
$r < 0.5$	No or hardly noticeable
$0.5 \leq r < 2.0$	Light
$2.0 \leq r < 5.0$	Light to moderate
$5.0 \leq r$	Moderate or greater

In addition, the nowcasting result in the experiment can be regarded as image data. Therefore, we have introduced MSE in computer vision as part of the performance metric. MSE calculates the L2 distance between the truth data and the forecast data. The calculation formulas for the above four evaluation indicators are as follows:

$$MSE = [\sum_{x=1}^{w} \sum_{y=1}^{h} (\widehat{Z}_{xy} - Z_{xy})^2] / (w * h) \tag{4}$$

$$CSI = NA/(NA + NB + NC) \tag{5}$$

$$POD = NA/(NA + NC) \tag{6}$$

$$FAR = NB/(NA + NB) \tag{7}$$

where w and h are the width and height of the data. $\widehat{Z}_{xy}$ and Z_{xy} are the values of the forecast data $\widehat{Z}$ and truth data Z in the coordinates (x, y). NA, NB, NC, and ND represent the number of true-positive, false-positive, false-negative, and true-negative grid points.

Finally, the performance metric of precipitation amount nowcasting is calculated based on the precipitation amount grid data, including four metrics of the first frame within 1 h, the first two frames within 2 h, and four frames within 4 h. The performance metric of precipitation intensity nowcasting is calculated based on the radar echo data and includes 4 metrics of the first 5 frames within 1 h, the first 10 frames within 2 h, and 20 frames within 4 h.

4.3. Experimental Results and Analysis

The experiment was divided into two parts: precipitation amount nowcasting and precipitation intensity nowcasting. In the experiment of precipitation amount nowcasting, we researched different multimodal fusion methods and compared MFSP-Net with the WRF model and the benchmark models. For precipitation intensity nowcasting, we compared the MFSP-Net with the benchmark models.

4.3.1. Precipitation Amount Nowcasting

In the precipitation amount nowcasting, the multimodal fusion method includes fusion on the spatial scale, fusion on the time scale, and simultaneous fusion on the time scale and space scale, respectively corresponding to the three-input EF-ConvLSTM, the three-input LF-ConvLSTM, and MFSP-Net. In addition, the three-input late fusion TrajGRU can be regarded as the three-input RN-Net [9]. We use common spatiotemporal prediction models as the benchmark models for precipitation amount nowcasting, including ConvLSTM, TrajGRU, and PredRNN. They are similar to the radar echo extrapolation, which forecasts future development using past precipitation amount data.

In addition, to compare with the traditional method, the WRF model was run to obtain the precipitation amount forecasts. Its spatial range is the same as the dataset, and its time range covers the testing set of the dataset. The WRF model is integrated every 6 h and forecasts the next 12 h with a time resolution of 1 h of precipitation amount. The deep learning methods forecast the precipitation amount for the next 4 h every 1 h. To compare the two types of methods, we used the comparison method proposed by Zhang et al. [9], as shown in Figure 6. First, we extract 4 h of data every 1 h from the 12 h WRF model precipitation amount forecast, with a total of 9 sets of data. Then, we compare each set of data with the corresponding actual precipitation amount. However, the WRF model has a spin-up period whose duration cannot be determined, and the forecast in this period is usually not used. To avoid the spin-up period, the best evaluation results among the nine sets of data are used as the WRF model evaluation result within 12 h. Meanwhile, we also compared our precipitation amount nowcasting methods based on the deep learning of these nine time periods with the corresponding actual precipitation amount. The average of the nine evaluation results is used as the deep learning method evaluation result within 12 h. We compare all the 12 h WRF model forecasts integrated every 6 h in the testing set with the deep learning methods forecasts through the above method. This comparison method solves the problem of the different forecasting frequencies of the two methods and avoids the spin-up period of the WRF model.

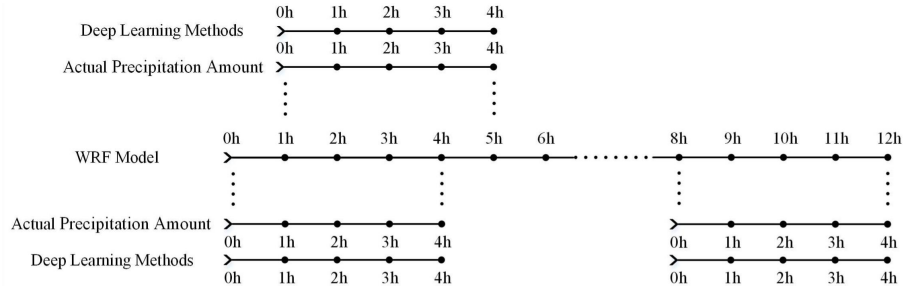

Figure 6. Schematic diagram of comparison method between deep learning methods and the WRF model.

The average values of precipitation amount nowcasting evaluation results within 1 h, 2 h, and 4 h are shown in the upper part of Tables 4–6, respectively. We compare the evaluation results from three aspects: between the deep learning models and the WRF model, between multimodal fusion methods and benchmark models, and between different multimodal fusion methods. In the comparison, we use CSI as the main evidence.

Table 4. Average evaluation results of one frame of precipitation amount nowcasting in the first hour. For MSE and CSI, the best performance is reported using red, and the second best is reported using blue. "↑" means that the higher the score, the better, while "↓" means that the lower the score, the better. "$r \geq \gamma$" means the skill score at the γ mm precipitation amount threshold in 1 h.

Method	MSE/Frame ↓	r ≥ 0.5 mm			r ≥ 2 mm			r ≥ 5 mm		
		CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓
WRF	2.4569	0.1573	0.3615	0.7223	0.1005	0.2662	0.7557	0.0653	0.1483	0.6260
ConvLSTM	1.4878	0.4120	0.6648	0.4799	0.3234	0.4335	0.4398	0.2250	0.2699	0.4246
TrajGRU	1.5063	0.4249	0.6747	0.4655	0.3333	0.4538	0.4435	0.2358	0.2948	0.4588
PredRNN	1.5262	0.3872	0.6412	0.5056	0.3096	0.4051	0.4321	0.2088	0.2424	0.3985
EF-ConvLSTM	1.1037	0.5206	0.6553	0.2830	0.4483	0.5717	0.3249	0.3605	0.4633	0.3807
LF-ConvLSTM	1.0400	0.5267	0.6517	0.2669	0.4604	0.5878	0.3200	0.3816	0.4944	0.3739
LF-TrajGRU (RN-Net)	1.0668	0.5318	0.6840	0.2949	0.4602	0.5994	0.3352	0.3824	0.5124	0.3988
MFSP-Net	**1.0904**	**0.5415**	**0.6685**	**0.2596**	**0.4753**	**0.6181**	**0.3270**	**0.3996**	**0.5439**	**0.3990**
MFSP-Net (without SAM)	1.1307	0.5326	0.6594	0.2651	0.4597	0.5815	0.3129	0.3737	0.4836	0.3781
MFSP-Net (without *Multi-Loss*)	1.0711	0.5353	0.6657	0.2679	0.4637	0.5707	0.2879	0.3812	0.4746	0.3402
PA-RA-MFSP-Net	1.4590	0.4351	0.5556	0.3326	0.3406	0.4302	0.3794	0.2537	0.3134	0.4285
PA-RE-MFSP-Net	1.1446	0.5265	0.6798	0.2998	0.4615	0.6175	0.3537	0.3815	0.5277	0.4205

Table 5. Average evaluation results of two frames of precipitation amount nowcasting in the first two hours

.

Method	MSE/Frame ↓	r ≥ 0.5 mm			r ≥ 2 mm			r ≥ 5 mm		
		CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓
WRF	2.8775	0.1567	0.3591	0.7405	0.0992	0.2634	0.7947	0.0636	0.1449	0.7403
ConvLSTM	1.6900	0.3560	0.6147	0.5447	0.2566	0.3368	0.4825	0.1689	0.1994	0.4758
TrajGRU	1.7172	0.3665	0.5939	0.5150	0.2636	0.3543	0.4975	0.1768	0.2182	0.5232
PredRNN	1.7064	0.3401	0.5808	0.5521	0.2500	0.2273	0.4770	0.1556	0.1789	0.4615
EF-ConvLSTM	1.4632	0.4267	0.5392	0.3367	0.3605	0.4633	0.3807	0.2340	0.2912	0.4757
LF-ConvLSTM	1.3740	0.4406	0.5590	0.3351	0.3608	0.4604	0.3858	0.2838	0.3617	0.4384
LF-TrajGRU (RN-Net)	1.3939	0.4467	0.5780	0.3438	0.3613	0.4638	0.3841	0.2838	0.3694	0.4471
MFSP-Net	**1.4261**	**0.4598**	**0.5795**	**0.3185**	**0.3854**	**0.5049**	**0.3904**	**0.3070**	**0.4136**	**0.4650**
MFSP-Net (without SAM)	1.4861	0.4478	0.5645	0.3250	0.3643	0.4668	0.3903	0.2751	0.3556	0.4684
MFSP-Net (without *Multi-Loss*)	1.4162	0.4467	0.5581	0.3159	0.3654	0.4496	0.3498	0.2817	0.3478	0.4154
PA-RA-MFSP-Net	1.7002	0.3641	0.4627	0.3732	0.2695	0.3372	0.4328	0.1882	0.2296	0.4963
PA-RE-MFSP-Net	1.4766	0.4481	0.5929	0.3606	0.3730	0.5035	0.4201	0.2918	0.3998	0.4901

Table 6. Average evaluation results of four frames of precipitation amount nowcasting in the four hours.

Method	MSE/Frame ↓	r ≥ 0.5 mm			r ≥ 2 mm			r ≥ 5 mm		
		CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓
WRF	3.5216	0.1542	0.3609	0.7690	0.0968	0.2625	0.8353	0.0616	0.1459	0.8236
ConvLSTM	1.8897	0.2860	0.5020	0.6083	0.1799	0.2294	0.5386	0.1076	0.1250	0.5552
TrajGRU	1.9453	0.2901	0.4684	0.5767	0.1863	0.2466	0.5819	0.1138	0.1391	0.6375
PredRNN	1.8929	0.2795	0.4761	0.6021	0.1800	0.2273	0.5382	0.0991	0.1127	0.5646
EF-ConvLSTM	1.8053	0.3120	0.3893	0.4038	0.2340	0.2912	0.4757	0.1641	0.2043	0.5659
LF-ConvLSTM	1.7260	0.3259	0.4106	0.4063	0.2471	0.3094	0.4629	0.1808	0.2255	0.5287
LF-TrajGRU (RN-Net)	1.7312	0.3289	0.4167	0.3958	0.2464	0.3072	0.4343	0.1793	0.2269	0.5232
MFSP-Net	**1.7813**	**0.3503**	**0.4377**	**0.3770**	**0.2741**	**0.3530**	**0.4644**	**0.2000**	**0.2629**	**0.5564**
MFSP-Net (without SAM)	1.8418	0.3411	0.4310	0.4008	0.2555	0.3261	0.4889	0.1768	0.2251	0.5758
MFSP-Net (without *Multi-Loss*)	1.7697	0.3298	0.4059	0.3718	0.2484	0.3005	0.4230	0.1720	0.2086	0.5133
PA-RA-MFSP-Net	1.9390	0.2783	0.3505	0.4369	0.1953	0.2429	0.5195	0.1254	0.1518	0.6035
PA-RE-MFSP-Net	1.8115	0.3444	0.4574	0.4365	0.2657	0.3538	0.5029	0.1906	0.2545	0.5780

As shown in the upper part of Tables 4–6, compared with the deep learning models, all metrics of the WRF model are inferior. Two factors cause this result [9]. First, the WRF model does not use the latest truth data but completely depends on WRF simulations. WRF simulations usually have deviations in the time domain and geographical area. Second, the parameterization scheme in the WRF model is manually designed by meteorological experts, which is different from the law reflected in historical meteorological data. For this reason, we need to develop precipitation nowcasting based on deep learning vigorously.

Compared with the multimodal fusion models, the precipitation amount nowcasting effect of the benchmark models is poor. The precipitation amount grid data are sparse, containing few meteorological spatiotemporal features, and cannot support its prediction.

Among multiple multimodal fusion methods, the early fusion method has the worst nowcasting effect, and MFSP-Net has the best nowcasting effect. Compared with other multimodal fusion methods, as time progresses, the improvement of MFSP-Net's nowcasting effect increases. This shows that MFSP-Net can evolve the spatiotemporal fusion features over a longer distance and accurately. MFSP-Net is better than other models in heavy precipitation nowcasting, which shows that it can better fuse multi-source data and use them to guide network training. In addition, MFSP-Net's POD and FAR are higher than other methods. This increases the range of correct forecasts while also increasing the range of incorrect forecasts.

4.3.2. Precipitation Intensity Nowcasting

The average values of precipitation intensity nowcasting evaluation results within 1 h, 2 h, and 4 h are shown in the upper part of Tables 7–9, respectively. In the field of precipitation intensity nowcasting, the radar echo extrapolation method has developed rapidly. For this reason, we use five radar echo extrapolation methods as the benchmark models for precipitation intensity nowcasting, including ConvLSTM, PredRNN, PredRNN++, SA-ConvLSTM, and Motion-PredRNN.

The nowcasting effect of the benchmark models for precipitation intensity nowcasting is far better than that for precipitation amount nowcasting. The radar echo data contain abundant meteorological spatiotemporal features, which can support the extrapolation process. Compared with the radar echo extrapolation method, MFSP-Net uses precipitation amount grid data and reanalysis data for additional guidance. The precipitation grid amount data reflect the most natural precipitation process and strongly correlate with the high-value area in the radar echo data. The reanalyzed data reflect the atmospheric development process, in which data such as temperature, wind direction, and relative humidity are closely related to the precipitation development process. Under the guidance

of these two types of data, the nowcasting effect of MFSP-Net on heavy precipitation is far better than the benchmark models for precipitation intensity nowcasting.

Table 7. Average evaluation results of five frames of precipitation intensity nowcasting in the first hour. For MSE and CSI, the best performance is reported using red, and the second best is reported using blue. "↑" means that the higher the score, the better, while "↓" means that the lower the score, the better. "$r \geq \gamma$" means the skill score at the γ mm/h rainfall threshold.

Method	MSE/Frame ↓	r ≥ 0.5 mm/h			r ≥ 2 mm/h			r ≥ 5 mm/h		
		CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR↓
ConvLSTM	163.29	0.6081	0.7364	0.2249	0.4966	0.6108	0.2764	0.3936	0.4606	0.2742
PredRNN++	162.04	0.6145	0.7473	0.2263	0.5028	0.6265	0.2824	0.4024	0.4799	0.2852
PredRNN	152.84	0.6289	0.7572	0.2148	0.5235	0.6474	0.2702	0.4239	0.5008	0.2683
SA-ConvLSTM	145.58	0.6334	0.7782	0.2288	0.5253	0.6627	0.2839	0.4283	0.5126	0.2768
Motion-PredRNN	150.72	0.6291	0.7416	0.1973	0.5246	0.6372	0.2563	0.4232	0.4936	0.2596
MFSP-Net	**158.00**	**0.6279**	**0.7792**	**0.2380**	**0.5308**	**0.7014**	**0.3159**	**0.4605**	**0.5939**	**0.3308**
RE-RA-MFSP-Net	163.92	0.6205	0.7613	0.2317	0.5091	0.6787	0.3310	0.4268	0.5526	0.3496
PA-RE-MFSP-Net	158.16	0.6241	0.7464	0.2102	0.5268	0.6745	0.2959	0.4460	0.5545	0.3085
MFSP-Net (without SAM)	159.45	0.6209	0.7443	0.2125	0.5206	0.6513	0.2798	0.4327	0.5245	0.2905
MFSP-Net (without *Multi-Loss*)	315.35	0.5857	0.8601	0.3538	0.4446	0.8765	0.5263	0.3742	0.8598	0.6022

Table 8. Average evaluation results of 10 frames of precipitation intensity nowcasting in the first two hours.

Method	MSE/Frame ↓	r ≥ 0.5 mm/h			r ≥ 2 mm/h			r ≥ 5 mm/h		
		CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓
ConvLSTM	205.49	0.5435	0.6710	0.2650	0.4251	0.5224	0.3104	0.3164	0.3673	0.3105
PredRNN++	203.65	0.5462	0.6693	0.2567	0.4232	0.5181	0.3012	0.3108	0.3621	0.3005
PredRNN	196.22	0.5580	0.6806	0.2499	0.4421	0.5419	0.2975	0.3335	0.3878	0.2949
SA-ConvLSTM	188.54	0.5672	0.7091	0.2660	0.4491	0.5604	0.3082	0.3388	0.3977	0.2994
Motion-PredRNN	194.16	0.5599	0.6754	0.2412	0.4468	0.5453	0.2960	0.3377	0.3913	0.2972
MFSP-Net	**208.53**	**0.5640**	**0.7184**	**0.2812**	**0.4661**	**0.6237**	**0.3569**	**0.3897**	**0.5041**	**0.3751**
RE-RA-MFSP-Net	216.19	0.5492	0.6793	0.2635	0.4387	0.5780	0.3564	0.3465	0.4386	0.3752
PA-RE-MFSP-Net	208.66	0.5597	0.6925	0.2613	0.4604	0.6004	0.3429	0.3737	0.4674	0.3580
MFSP-Net (without SAM)	208.28	0.5564	0.6832	0.2558	0.4528	0.5717	0.3207	0.3593	0.4364	0.3380
MFSP-Net (without *Multi-Loss*)	397.96	0.5300	0.8061	0.3960	0.3996	0.8223	0.5648	0.3309	0.8010	0.6414

Table 9. Average evaluation results of 20 frames of precipitation intensity nowcasting in the four hours.

Method	MSE/Frame ↓	r ≥ 0.5 mm/h			r ≥ 2 mm/h			r ≥ 5 mm/h		
		CSI ↑	POD ↑	FAR↓	CSI ↑	POD ↑	FAR ↓	CSI ↑	POD ↑	FAR ↓
ConvLSTM	265.83	0.4545	0.5723	0.3254	0.3291	0.4012	0.3652	0.2247	0.2576	0.3716
PredRNN++	262.74	0.4510	0.5560	0.3066	0.3145	0.3763	0.3357	0.2048	0.2335	0.3327
PredRNN	256.57	0.4575	0.5592	0.2955	0.3339	0.4015	0.3352	0.2319	0.2647	0.3362
SA-ConvLSTM	251.45	0.4701	0.5967	0.3244	0.3396	0.4167	0.3576	0.2309	0.2662	0.3547
Motion-PredRNN	253.65	0.4686	0.5780	0.3027	0.3485	0.4226	0.3473	0.2410	0.2752	0.3432
MFSP-Net	**279.43**	**0.4731**	**0.6125**	**0.3366**	**0.3730**	**0.4957**	**0.4084**	**0.2934**	**0.3738**	**0.4332**
RE-RA-MFSP-Net	291.23	0.4485	0.5558	0.3126	0.3386	0.4360	0.3987	0.2465	0.3029	0.4154
PA-RE-MFSP-Net	280.89	0.4705	0.6031	0.3332	0.3680	0.4833	0.4090	0.2820	0.3514	0.4316
MMFSP-Net (without SAM)	277.99	0.4679	0.5885	0.3184	0.3611	0.4580	0.3845	0.2717	0.3293	0.4111
MFSP-Net (without *Multi-Loss*)	492.47	0.4522	0.7077	0.4523	0.3393	0.7139	0.6121	0.2772	0.6834	0.6859

4.4. Ablation Study

We conducted ablation studies on MFSP-Net, verified the effects of SAM and *Multi-Loss*, and explored the impact of different input data on precipitation nowcasting. The MFSP-LSTM unit in MFSP-Net-without-SAM removes the SAM. MFSP-Net-without-SAM can simultaneously complete precipitation amount nowcasting and precipitation intensity nowcasting. MFSP-Net-without-*Multi-Loss* loses the ability to perform multi-task learning. It uses precipitation amount grid data as label data to achieve precipitation amount nowcasting and radar echo data as label data to achieve precipitation intensity nowcasting. We designed three dual input schemes for the three input data: PA-RE, PA-RA, and PE-RA. PA-RE-MFSP-Net can complete two nowcastings simultaneously, PA-RA-MFSP-Net can only complete precipitation amount nowcasting, and RE-RA-MFSP-Net can only complete precipitation intensity nowcasting. The evaluation results of these ablation studies are shown in the bottom part of Tables 4–9.

Comparing the evaluation results of MFSP-Net and MFSP-Net-without-SAM, it can be found that SAM improves the nowcasting ability of the network in all periods, and the improvement increases with the increase of the threshold. SAM helps the network to better understand the long-distance dependence in the temporal domain and enables the network to extract higher-level meteorological spatiotemporal features in the spatial domain.

Comparing the evaluation results of MFSP-Net and MFSP-Net-without-*Multi-Loss*, it can be found that multi-task learning strategy has a comprehensive improvement on the two nowcasting effects, especially in 2–4 h. For precipitation amount nowcasting, MFSP-Net-without-*Multi-Loss* uses precipitation amount grid data as label data. As the major input data of MFSP-LSTM, the spatiotemporal information flow of the precipitation amount grid data has more of an interaction with the hidden state. Therefore, using the precipitation amount grid data alone to train the network still resulted in an excellent nowcast effect. For precipitation intensity nowcasting, MFSP-Net-without-*Multi-Loss* uses radar echo data as label data. The radar echo data are the minor input data of MFSP-LSTM. The spatiotemporal information flow interacts less with the hidden state, which cannot guide the network training thoroughly. The FAR and POD in the precipitation intensity nowcasting are abnormal, and the nowcasting effect is worse than that of the benchmark model based on radar echo extrapolation.

Comparing MFSP-Net with different input schemes, it can be found that PA-RE-MFSP-Net's precipitation amount nowcasting effect is better than PA-RA-MFSP-Net, and its precipitation intensity nowcasting effect is better than RE-RA-MFSP-Net. There are some errors in the reanalyzed data. When the meteorological spatiotemporal features are few, the network cannot correct these errors. The radar echo data are sufficient to support its nowcasting. The precipitation amount grid data strongly correlate with the high-value radar echo area, and their addition improves the nowcasting effect of heavy precipitation. The reanalyzed data reflect the atmospheric development process and result in less improvement than the radar echo data and precipitation amount grid data.

4.5. Visualization Results

Figure 7 visualizes two representative cases of precipitation amount nowcasting by MFSP-Net, MFSP-Net-without-SAM, MFSP-Net-without-*Multi-Loss*, PA-RE-MFSP-Net, PA-RA-MFSP-Net, RN-Net, EF-ConvLSTM, and the WRF model. Figure 8 visualizes a representative case of precipitation intensity nowcasting by MFSP-Net, MFSP-Net-without-SAM, MFSP-Net-without-*Multi-Loss*, PA-RE-MFSP-Net, RE-RA-MFSP-Net, and MotionPredRNN. Figure 8 and the right half of Figure 7 have the same nowcasting period. From Figure 7, we observe that the nowcasting effects of all deep learning models are better than WRF models. Comparing MFSP-Net with four networks of ablation study, the rules are similar to the conclusions of the ablation study. From Figures 7 and 8, it can be found that *Multi-Loss* strengthens the weak precipitation nowcasting ability of MFSP-Net, and SAM reduces the FAR of MFSP-Net. Comparing different dual-input schemes, it can be found that the reanalysis data played an important role. The data strengthen networks'

heavy precipitation nowcasting ability and improve nowcasting accuracy in the middle and later periods. In the precipitation intensity nowcasting, compared with the radar echo extrapolation network Motion-PredRNN, MFSP-Net's heavy precipitation nowcasting and the middle and late nowcasting are more accurate.

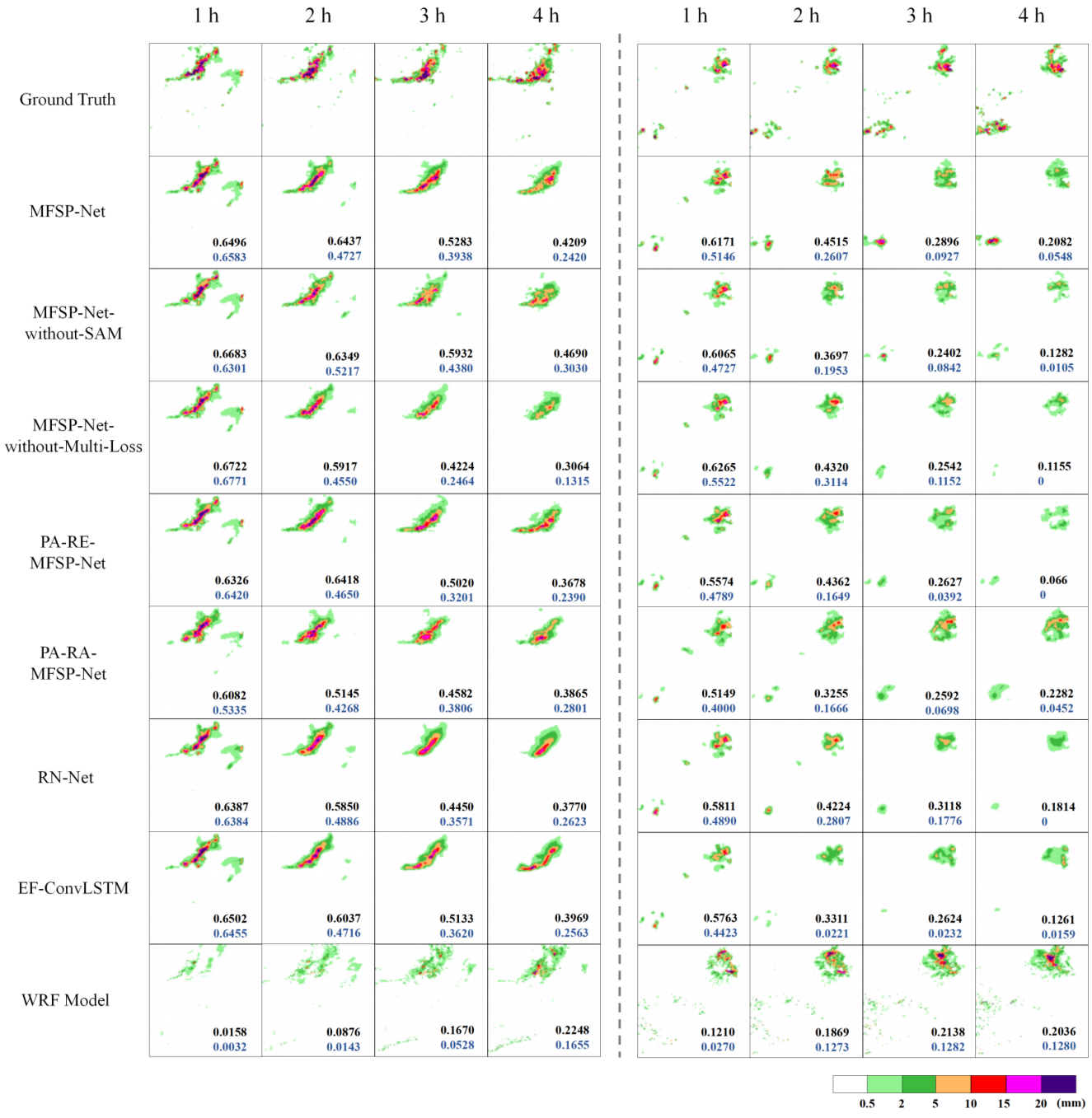

Figure 7. Visualization of two representative precipitation amount nowcasting cases. From top to bottom are the four frames' truth precipitation amount grid data of the next four hours and precipitation amount nowcasting conducted by MFSP-Net, MFSP-Net-without-SAM, MFSP-Net-without-*Multi-Loss*, PA-RE-MFSP-Net, PA-RA-MFSP-Net, RN-Net, EF-ConvLSTM, and WRF models. The black value and blue value in each nowcasting frame are the CSI with 2 mm and 5 mm as the threshold for this frame.

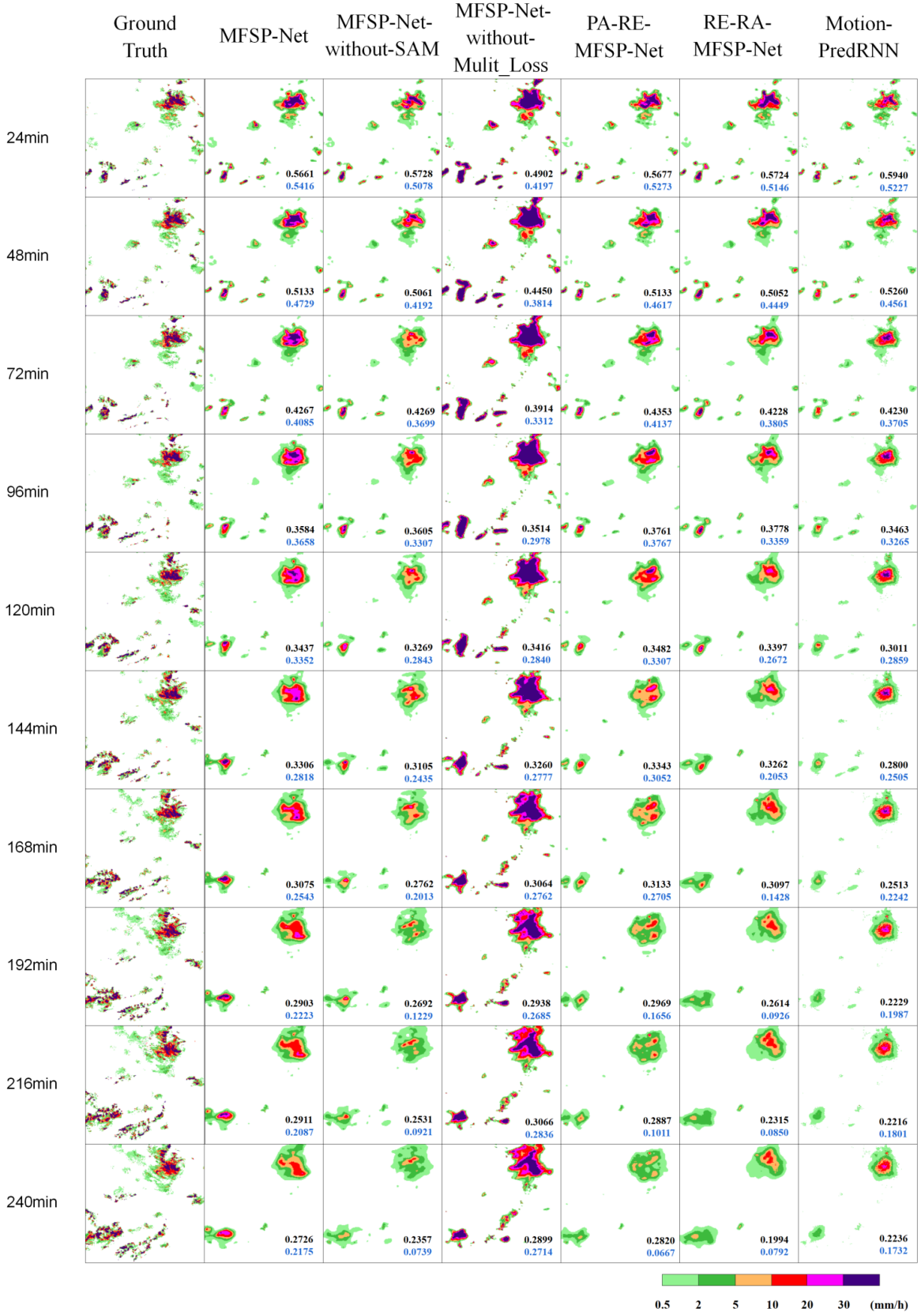

Figure 8. Visualization of one representative precipitation intensity nowcasting case. From left to right are the 10 frames' truth radar echo data of the next four hours and precipitation intensity nowcasting conducted by MFSP-Net, MFSP-Net-without-SAM, MFSP-Net-without-*Multi-Loss*, PA-RE-MFSP-Net, RE-RA-MFSP-Net, and Motion-PredRNN. The black value and blue value in each nowcasting frame are the CSI with 2 mm/h and 5 mm/h as the threshold for this frame.

5. Discussion

Currently, most of the weather nowcasting methods based on deep learning use the single-input scheme. The multimodal fusion method using the few multi-input methods is too simple to utilize the advantages of multi-source data fully. For this reason, we explore the combination of spatiotemporal prediction and multimodal fusion in precipitation nowcasting. The MFSP-Net proposed in this paper is a multi-input multi-output RNN based on multi-task learning, simultaneously realizing precipitation amount nowcasting and precipitation intensity nowcasting. In the experiment, the forecast effect of the WRF model is the worst, the multi-input scheme is better than the single-input scheme, and MFSP-Net is better than other multimodal fusion methods. Observing the experimental results and visualization cases, we are able to present the following points:

1. Compared with other data, the reanalysis data contain higher-level meteorological spatiotemporal features and errors. The higher-level spatiotemporal features enhance the network's heavy precipitation nowcasting and the middle and late nowcasting effects. However, the disadvantages caused by errors need to be eliminated by using a larger-scale network or adding other meteorological spatiotemporal features.
2. The precipitation amount grid data and radar echo data (precipitation intensity) are complementary. The two respectively represent the cumulative value and instantaneous value of precipitation. The precipitation amount grid data are very sparse and have weak continuity, but high accuracy. The radar echo data contain noise, but their continuity is strong. In precipitation nowcasting, combining the two types of data can improve the nowcasting effect. In the experiment, the forecasting effect of PA-RE-MFSP-Net was found to be similar to that of MFSP-Net.
3. There are two difficulties in precipitation nowcasting. The first is the errors of the data, especially the excessive noise of the radar echo data. A larger network can increase its tolerance for data errors, but it requires a larger dataset to support it. Therefore, we hope to use more accurate radar data or introduce satellite data in the next step. Secondly, the loss function cannot accurately reflect the prediction effect of the network. It can be observed in Tables 4–9 that there is no correct correspondence between CSI and MSE. We hope that relevant research can be carried out in the next step.

6. Conclusions

For precipitation nowcasting, to make full use of multi-source meteorological data, this paper proposes a novel multi-input multi-output recurrent neural network model based on multimodal fusion and spatiotemporal prediction, named MFSP-Net. It uses precipitation grid data, radar echo data, and reanalysis data as input data and simultaneously realizes 0–4 h precipitation amount nowcasting and precipitation intensity nowcasting. MFSP-Net can perform the spatiotemporal-scale fusion of the three types of input data while retaining the spatiotemporal information flow of them. In the training phase, we use the multi-task learning strategy, which improves the performance of two nowcastings simultaneously. We carry out experiments and evaluations on the dataset of Southeast China. In the experiment, MFSP-Net comprehensively enhances the performance of the precipitation amount nowcasting. For the precipitation intensity nowcasting, MFSP-Net has obvious advantages in the heavy precipitation nowcasting and the middle and late stages of the nowcasting.

In order to further improve the performance of precipitation nowcasting, we will extend our current work to three aspects. Firstly, we will add meteorological satellite data as input data to provide additional meteorological spatiotemporal features for nowcasting. Secondly, in view of the mismatch between the loss value and the evaluation result that often occurs in training, we hope to design a new loss function that is more suitable for precipitation nowcasting. Finally, we hope to combine the visual transformer in the future model design to realize the global interaction of meteorological features.

Author Contributions: Conceptualization, X.W., J.G. and F.Z.; methodology, F.Z.; software, F.Z.; validation, X.W., J.G. and F.Z.; formal analysis, F.Z.; investigation, F.Z.; resources, J.G. and F.Z.; data curation, F.Z.; writing—original draft preparation, F.Z.; writing—review and editing, F.Z.; visualization, F.Z.; supervision, F.Z.; project administration, F.Z.; funding acquisition, J.G. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Project No. 41975066).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The ERA5 reanalysis data used in this paper can be download from the website https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form, accessed on 21 November 2021 . Other data cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Muhandhis, I.; Susanto, H.; Asfari, U. Determining Salt Production Season Based on Rainfall Forecasting Using Weighted Fuzzy Time Series. *J. Appl. Comput. Sci. Math.* **2020**, *14*, 23–27. [CrossRef]
2. Zhou, J.; Zhang, Y.; Tian, S.; Lai, S. Forecasting Rainfall with Recurrent Neural Network for irrigation equipment. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *510*, 042040. [CrossRef]
3. Zhou, J.; Xiang, J.; Huang, S. Classification and Prediction of Typhoon Levels by Satellite Cloud Pictures through GC–LSTM Deep Learning Model. *Sensors* **2020**, *20*, 5132. [CrossRef] [PubMed]
4. Osanai, N.; Shimizu, T.; Kuramoto, K.; Kojima, S.; Noro, T. Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network. *Landslides* **2010**, *7*, 325–338. [CrossRef]
5. Knievel, J.C.; Ahijevych, D.A.; Manning, K.W. Using temporal modes of rainfall to evaluate the performance of a numerical weather prediction model. *Mon. Weather Rev.* **2004**, *132*, 2995–3009. [CrossRef]
6. Chu, Q.; Xu, Z.; Chen, Y.; Han, D. Evaluation of the ability of the Weather Research and Forecasting model to reproduce a sub-daily extreme rainfall event in Beijing, China using different domain configurations and spin-up times. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 3391. [CrossRef]
7. Bouget, V.; Béréziat, D.; Brajard, J.; Charantonis, A.; Filoche, A. Fusion of Rain Radar Images and Wind Forecasts in a Deep Learning Model Applied to Rain Nowcasting. *Remote Sens.* **2021**, *13*, 246. [CrossRef]
8. Kumar, A.; Islam, T.; Sekimoto, Y.; Mattmann, C.; Wilson, B. Convcast: An embedded convolutional LSTM based architecture for precipitation nowcasting using satellite data. *PLoS ONE* **2020**, *15*, e0230114. [CrossRef]
9. Zhang, F.; Wang, X.; Guan, J.; Wu, M.; Guo, L. RN-Net: A Deep Learning Approach to 0–2 h Rainfall Nowcasting Based on Radar and Automatic Weather Station Data. *Sensors* **2021**, *21*, 1981. [CrossRef]
10. Bonnet, S.M.; Evsukoff, A.; Morales Rodriguez, C.A. Precipitation Nowcasting with Weather Radar Images and Deep Learning in São Paulo, Brasil. *Atmosphere* **2020**, *11*, 1157. [CrossRef]
11. Moon, S.H.; Kim, Y.H.; Lee, Y.H.; Moon, B.R. Application of machine learning to an early warning system for very short-term heavy rainfall. *J. Hydrol.* **2019**, *568*, 1042–1054. [CrossRef]
12. Parmar, A.; Mistree, K.; Sompura, M. Machine learning techniques for rainfall prediction: A Review. In Proceedings of the International Conference on Innovations in information Embedded and Communication Systems, Coimbatore, India, 17–18 March 2017.
13. Ayzel, G.; Heistermann, M.; Sorokin, A.; Nikitin, O.; Lukyanova, O. All convolutional neural networks for radar-based precipitation nowcasting. *Procedia Comput. Sci.* **2019**, *150*, 186–192. [CrossRef]
14. Adewoyin, R.A.; Dueben, P.; Watson, P.; He, Y.; Dutta, R. TRU-NET: A deep learning approach to high resolution prediction of rainfall. *Mach. Learn.* **2021**, *110*, 2035–2062. [CrossRef]
15. Lebedev, V.; Ivashkin, V.; Rudenko, I.; Ganshin, A.; Molchanov, A.; Ovcharenko, S.; Grokhovetskiy, R.; Bushmarinov, I.; Solomentsev, D. Precipitation nowcasting with satellite imagery. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2680–2688.
16. Yan, Q.; Ji, F.; Miao, K.; Wu, Q.; Xia, Y.; Li, T. Convolutional Residual-Attention: A Deep Learning Approach for Precipitation Nowcasting. *Adv. Meteorol.* **2020**, *2020*, 6484812. [CrossRef]
17. Tian, L.; Li, X.; Ye, Y.; Xie, P.; Li, Y. A Generative Adversarial Gated Recurrent Unit Model for Precipitation Nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 601–605. [CrossRef]
18. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.

19. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.k.; Woo, W.C. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017*; MIT Press: Cambridge, MA, USA, 2017; pp. 5617–5627.
20. Woo, W.C.; Wong, W.K. Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere* **2017**, *8*, 48. [CrossRef]
21. Wang, Y.; Zhang, J.; Zhu, H.; Long, M.; Wang, J.; Yu, P.S. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9154–9162.
22. Wu, H.; Yao, Z.; Wang, J.; Long, M. MotionRNN: A Flexible Model for Video Prediction With Spacetime-Varying Motions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021 (Computer Vision Foundation/IEEE), Virtual, 19–25 June 2021; pp. 15435–15444.
23. Oprea, S.; Martinez-Gonzalez, P.; Garcia-Garcia, A.; Castro-Vargas, J.A.; Orts-Escolano, S.; Garcia-Rodriguez, J.; Argyros, A. A Review on Deep Learning Techniques for Video Prediction. *arXiv* **2020**, arXiv:2004.05214.
24. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Philip, S.Y. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In *Advances in Neural Information Processing Systems*; 2017; pp. 879–888. Available online: https://papers.nips.cc/paper/2017/hash/e5f6ad6ce374177eef023bf5d0c018b6-Abstract.html (accessed on 21 November 2021).
25. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Philip, S.Y. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132.
26. Jing, J.; Li, Q.; Peng, X.; Ma, Q.; Tang, S. HPRNN: A Hierarchical Sequence Prediction Model for Long-Term Weather Radar Echo Extrapolation. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 4142–4146.
27. Lin, Z.; Li, M.; Zheng, Z.; Cheng, Y.; Yuan, C. Self-attention convlstm for spatiotemporal prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11531–11538.
28. Zhang, P.; Jia, Y.; Gao, J.; Song, W.; Leung, H.K. Short-term rainfall forecasting using multi-layer perceptron. *IEEE Trans. Big Data* **2018**, *6*, 93–106. [CrossRef]
29. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.Y.; Wang, W.; Powers, J.G. *G.: A Description of the Advanced Research WRF Version 3*; NCAR Tech. Note NCAR/TN-475+ STR; University Corporation for Atmospheric Research: Boulder, CO, USA, 2008. [CrossRef]
30. Geng, Y.; Li, Q.; Lin, T.; Jiang, L.; Xu, L.; Zheng, D.; Yao, W.; Lyu, W.; Zhang, Y. Lightnet: A dual spatiotemporal encoder network model for lightning prediction. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2439–2447.
31. Zhang, Y.; Yang, Q. An overview of multi-task learning. *Natl. Sci. Rev.* **2018**, *5*, 30–43. [CrossRef]
32. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503. [CrossRef]
33. Hersbach, H. The ERA5 Atmospheric Reanalysis. In *AGU Fall Meeting Abstracts*; 2016; Volume 2016, p. NG33D–01. Available online: https://ui.adsabs.harvard.edu/abs/2016AGUFMNG33D..01H/abstract (accessed on 21 November 2021).
34. Skamarock, W.C.; Klemp, J.B.; Dudhia, J. Prototypes for the WRF (Weather Research and Forecasting) model. In *Proceedings of the Ninth Conference Mesoscale Processes*; American Meteorological Society: FL, USA, 2001; pp. J11–J15. Available online: https://opensky.ucar.edu/islandora/object/articles:21028 (accessed on 21 November 2021).
35. Narayanan, A.L.; Siravuru, A.; Dariush, B. Gated Recurrent Fusion to Learn Driving Behavior from Temporal Multimodal Data. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1287–1294. [CrossRef]
36. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*; 2019; pp. 8026–8037. Available online: https://papers.nips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html (accessed on 21 November 2021).
37. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In *Conference Track Proceedings, Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015*; Bengio, Y., LeCun, Y., Eds.; ACM: San Diego, CA, USA, 2015.
38. Tran, Q.K.; Song, S.K. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. *Atmosphere* **2019**, *10*, 244. [CrossRef]