

Article

LSTMAtU-Net: A Precipitation Nowcasting Model Based on ECSA Module

Huantong Geng ^{1,2}, Xiaoyan Ge ^{2,*} , Boyang Xie ², Jinzhong Min ³ and Xiaoran Zhuang ⁴

¹ Jiangsu Open University, Nanjing 210036, China; htgeng@nuist.edu.cn

² School of Computer, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211221043@nuist.edu.cn

³ School of Meteorology, Nanjing University of Information Science and Technology, Nanjing 210044, China; jzmin@nuist.edu.cn

⁴ Jiangsu Meteorological Observatory, Nanjing 210008, China; zxrxz3212009@163.com

* Correspondence: 20211249485@nuist.edu.cn

Abstract: Precipitation nowcasting refers to the use of specific meteorological elements to predict precipitation in the next 0–2 h. Existing methods use radar echo maps and the Z–R relationship to directly predict future rainfall rates through deep learning methods, which are not physically constrained, but suffer from severe loss of predicted image details. This paper proposes a new model framework to effectively solve this problem, namely LSTMAtU-Net. It is based on the U-Net architecture, equipped with a Convolutional LSTM (ConvLSTM) unit with the vertical flow direction and depthwise-separable convolution, and we propose a new component, the Efficient Channel and Space Attention (ECSA) module. The ConvLSTM unit with the vertical flow direction memorizes temporal changes by extracting features from different levels of the convolutional layers, while the ECSA module innovatively integrates different structural information of each layer of U-Net into the channelwise attention mechanism to learn channel and spatial information, thereby enhancing attention to the details of precipitation images. The experimental results showed that the performance of the model on the test dataset was better than other examined models and improved the accuracy of medium- and high-intensity precipitation nowcasting.



Citation: Geng, H.; Ge, X.; Xie, B.; Min, J.; Zhuang, X. LSTMAtU-Net: A Precipitation Nowcasting Model Based on ECSA Module. *Sensors* **2023**, *23*, 5785. <https://doi.org/10.3390/s23135785>

Academic Editor: Ram M. Narayanan

Received: 6 May 2023

Revised: 12 June 2023

Accepted: 15 June 2023

Published: 21 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Severe convective weather occurs frequently in China, and short-term heavy precipitation causes other disaster chain reactions, which cause serious economic losses and casualties. Precipitation nowcasting can make an early warning plan for this weather and reserve a time window for disaster prevention and mitigation [1,2]. It can also ensure the correct operation of emergency services, energy management, maritime services, and other industries. Numerical Weather Prediction (NWP) [3] is a commonly used precipitation forecasting method that uses numerical calculations to solve physical equations to predict precipitation in a certain period of time in the future. However, the nowcasting performance of NWP is poor because the forecast period of nowcasting is less than the time required for the model to reach steady-state simulation [4]. It relies on many computational data, which requires high-performance computers. In order to solve these problems, radar-based precipitation nowcasting has become an important development direction.

Since Ligda [5] proposed the precipitation nowcasting based on radar elements, this method has been widely recognized and deeply studied [6–11], so it has become the mainstream forecasting technology for heavy precipitation nowcasting. It first uses the radar echo extrapolation model to predict the future radar echo and then uses the Z–R relationship [12] to obtain precipitation data. However, predicting in these two steps will

lead to cumulative error, while precipitation nowcasting based on precipitation elements or fusion elements [13–19] will reduce this error. It directly uses several precipitation frames or fusion feature frames to predict future precipitation maps. However, the precipitation data used in this method have the following problems: First, due to the very few to no rain samples in the precipitation data, there is the problem of an uneven distribution of data samples and a deviation of the data grade distribution in the precipitation dataset, as shown in Figure 1. Second, precipitation data have the characteristics of high-order nonlinearity and an extremely complex spatial and temporal distribution [20]. This makes it difficult to model with a specific inflexible target.

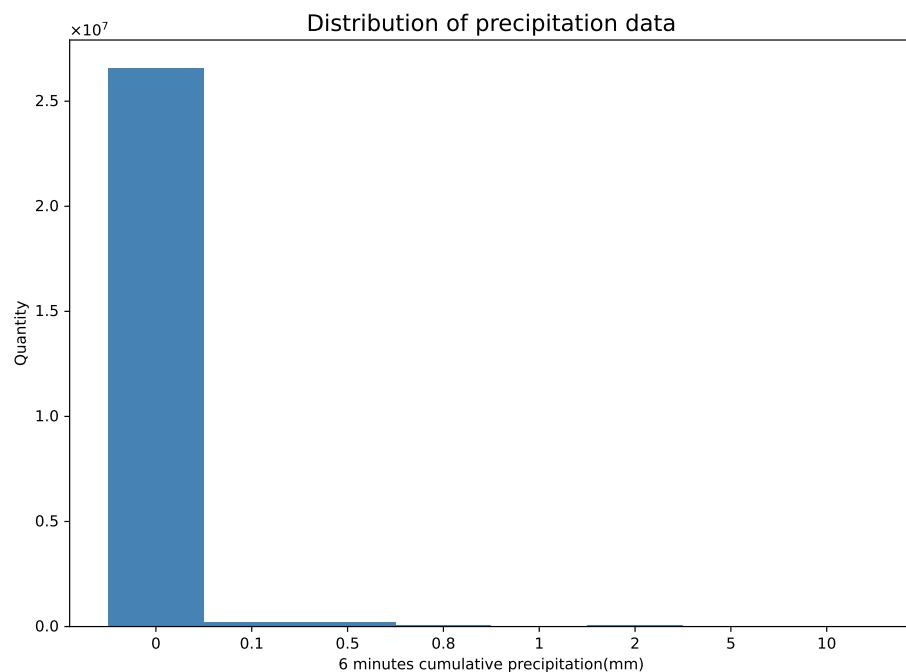


Figure 1. Precipitation data distribution statistics. The data were derived from several random sequences of the precipitation dataset in the 2022 Jiangsu Meteorological AI Algorithm Challenge. The spatial resolution of the single-time data is 480×560 ; the time period is 6 min; the total sample number is 101.

Deep learning methods based on precipitation features' or fusion features' extrapolation include Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). RNN-based models have strong modeling capabilities for time series data, which can be used to mine the spatiotemporal evolution of precipitation sequences [13,14], but they cannot solve the problem through parallel computing, prolong the prediction time, and have the problem of gradient explosion. CNN-based models not only have a strong feature extraction ability, but also can shorten prediction time by parallel computing. Among them, the authors in [21] introduced the famous U-Net model, which was originally used for medical image segmentation, but also has had success in precipitation nowcasting [15–19]. However, it has limited long-term dependency capabilities due to the inherent local nature of the convolutional operation. In addition, the models based on the RNN and CNN frameworks both have the problem of gradually inaccurate details in a long delivery cycle.

Motivated by this, this paper proposes LSTMAtU-Net, a deep learning network based on precipitation features' or fusion features' extrapolation. It is based on the U-Net architecture, equipped with a ConvLSTM [10] unit with the vertical flow direction, the ECSA module, and DSC [22]. The framework significantly improves the accuracy of precipitation nowcasting. Here, we used the dataset of the 2022 Jiangsu Meteorological AI Algorithm Challenge to evaluate the framework, which predicts the structured data of

precipitation elements in the next two hours based on the weather radar images and the structured data of meteorological elements that have occurred in the past.

In summary, the main contributions of this paper are as follows:

1. The LSTMAtU-Net model combines the advantages of RNNs and CNNs. We propose a model that combines the RNN and CNN structures, namely the LSTMAtU-Net model. It has a strong feature extraction capability and improved long-term dependency capabilities, and the core convolutional structure greatly reduces the risk of gradient explosion.
2. A new component, the ECSA module, is proposed, which uses mean pooling at different scales as a way to achieve the weighting of the channel and spatial information of convolutional features based on the ECA module [23]; thus, focusing on image details alleviates, to a certain extent, the problem of the RNN's and CNN's predictions showing the gradual inaccuracy of the image details.
3. We propose a completely new loss function. It first optimizes multiple prediction objectives and then designs the corresponding weighted loss function to solve the serious imbalance of precipitation data.

The rest of the paper is organized as follows: Section 2 briefly introduces the related work on nowcasting. Section 3 describes our proposed model. Comprehensive experiments that verified the effectiveness of our model are given in Section 4. The conclusion is given in Section 5.

2. Related Works

In this section, we will introduce deep learning models for nowcasting. Since our model also includes an attention mechanism, we also introduce the related research work on attention mechanisms.

2.1. Nowcasting Models Based on ConvRNN

ConvLSTM [10] was the first precipitation nowcasting model based on the ConvRNN, which adds convolutional layers to FC-LSTM, giving the network the ability to extract spatial information on top of being able to extract temporal information very well. The authors of [24] considered [10] the uneven extraction of temporal and spatial information and proposed the PredRNN model, which adds a unit storing propagation spatial information in each LSTM module to enable the model to propagate longitudinally. Continuing the study of [24], the authors of [25] proposed the PredRNN++, which adds the Gradient Highway Unit (GHU) to the PredRNN to deal with the gradient disappearance problem. To solve the difficulty of predicting higher-order nonsmooth information in radar echo maps, the authors of [26] proposed the MIM mechanism, which can effectively extract nonsmooth information. RN-Net [14], which fuses the spatiotemporal characteristics of rainfall data and radar data, provides a more adequate basis for forecasting. Based on the research of [14], the same team [13] proposed the MFSP-Net model, which uses meteorological data with different spatiotemporal resolutions as the input and combines spatiotemporal prediction and multi-model fusion in depth with a multi-task learning strategy. GAN-rcLSTM [11] introduced the Generative Adversarial Network (GAN) technology. The discriminator consists of several 3D convolutions and a fully connected layer. The generator is the proposed residual LSTM module, which can reduce the gradient disappearance problem and avoid the weakness problem of deep LSTM extrapolation to a certain extent. In addition, it makes the extrapolation sequence clearer through the generator and discriminator.

2.2. Nowcasting Models Based on U-Net

U-Net [21] is a classical fully convolutional network, which consists of a contracting path and a symmetric expanding path. The contracting path is used to obtain context information, and the expanding path is used for accurate positioning. ResU-Net [19] introduced the attention mechanism to the ResNet of [21], where the attention mechanism module was incorporated to learn temporal information. Broad-U-Net [18] has asymmetric

parallel convolutions and an Atrous Spatial Pyramid Pooling (ASPP) module by which it combines multiscale features to learn more complex patterns. SmaAt-U-Net [15] reduces the network size significantly, but the prediction performance is comparable to other models. Based on [15], the authors of [16] proposed AA-TransU-Net, which introduces the Transformer mechanism into precipitation nowcasting with some improvement in forecast accuracy.

2.3. Attention Mechanism

The core idea of the attention mechanism is to find correlations among the original data and then highlight some of their important features, and some studies in precipitation nowcasting have introduced the attention mechanism to precipitation nowcasting. The Squeeze-And-Extraction Network (SENet) [27], which belongs to the family of channel attention mechanisms, is very simple to construct and does not require the introduction of new functions or layers. Besides, it has good properties in terms of model and computational complexity. The Convolutional Block Attention Module (CBAM) [28] is a fusion of spatial and channel attention mechanisms, which can learn the importance of the channel and space separately and also can be easily embedded into any known framework. The Efficient Channel Attention (ECA) module [23] solves the side effects caused by the dimensionality reduction in SENet and achieves good results using only a few parameters.

The ConvRNN-based nowcasting model, developed from ConvLSTM and GAN-rcLSTM, is dedicated to solving the effective transmission of spatiotemporal information, the problem of gradient disappearance, and the ambiguity of prediction. Although the accuracy of precipitation was improved to a certain extent, it still has the problem that it cannot be parallelized, which will prolong the prediction time and produce low-quality prediction results due to the stochastic dynamic environment of precipitation [29,30]. The U-Net-based nowcasting models developed from U-Net and AA-TransU-Net are all devoted to solving the problem of the limited ability of the convolutional model to model long-range dependencies [29]. Although the ability to learn the time factor gradually becomes stronger, the effect is still not good in a long delivery cycle. Therefore, we combined these two structures to improve the long-term dependence capability, shorten the prediction time, and capture the complex spatial information to produce better prediction results. In addition, in order to better focus on image details, we introduced an attention mechanism, using the ECSA module. This is an improved version of the ECA module [23]. The existing ECA module has the problem of losing much of the spatial structure information, because it contains a Global Average Pooling (GAP) module function and the feature map size is generally 7×7 [31], while the U-Net architecture contains feature maps of different sizes. Therefore, based on this motivation and the idea of [31], the ECSA module adapted to different scales was established.

3. Methodology

3.1. Problem Definition

The problem of precipitation nowcasting based on precipitation features or fusion features can be approximated as a sequence prediction problem. The input and predicted data in this task are both structured data sequences of precipitation elements. Consider X as the input data, which is a series of three-dimensional tensors in time period M , and \hat{Y} as the predicted outcome, which is a sequence of data in time period N . $X \in R^{H \times W \times C_{in}}$, $\hat{Y} \in R^{H \times W \times C_{out}}$, where H, W is the spatial resolution of the image, C_{in} is the number of input data channels multiplied by the time period M , and C_{out} is the number of predicted data channels multiplied by the time period N . This is due to the need to map the time dimension to the channel dimension. F is the proposed model LSTMAtU-Net, then the whole prediction process is defined as:

$$\hat{Y} = F(X) \quad (1)$$

3.2. LSTMAtU-Net

The model proposed in this paper is based on the U-Net architecture. The model first replaces the convolutional part of U-Net with DSC. Next, the ECSA module and the ConvLSTM unit with the vertical flow direction are placed in the skip-connection of the encoder and decoder. The importance weights of each feature of the generated image are derived by the former, and the latter extracts and memorizes the spatiotemporal representations to learn the importance of the temporal factor. Finally, the functions of the model's encoder and decoder are consistent with those of U-Net, with the encoder responsible for feature extraction and the decoder responsible for restoring the image to its original size, resulting in a pixel-by-pixel prediction. The model framework is shown in Figure 2. The settings of the DSC and ECSA modules reduce the number of parameters in order to have better performance. In Table 1, we compare the LSTMAtU-Net model with the ECSA module and SENet regarding the number of parameters of our model. Our model reduces the model parameters by about 41.86 % compared to the model without these two modules.

Table 1. Number of parameters of the compared models.

Model	Parameters
LSTMAtU-Net with SENet	32,023,892
LSTMAtU-Net with ECSA	31,948,797
LSTMAtU-Net with ECSA and DSC	18,619,421

Encoder: As seen in the left half of Figure 2, this starts with two convolutional layers to process the input image into a feature map suitable for the encoder. This is followed by four consecutive modules, which contain a Max pooling layer (red arrow in Figure 2) and a double convolutional layer (each convolutional layer includes a 3×3 DSC, Batch-Norm (BN) [32] layer, and Rectified Linear Unit (ReLU) [33]) (brown arrow in Figure 2). The featured image obtains a larger field of view and more essential features as the network layers go deeper.

Decoder: As seen in the right half of Figure 2, this starts with the output of the ConvLSTM unit, followed by 4 consecutive modules, differing from the encoder's module in that it contains a bilinear upsampling layer (green arrow in Figure 2) and connects the upsampled feature mapping of doubled the size to the feature mapping output from the previous ECSA module. Finally, the predicted image is output with a 1×1 convolutional layer (purple arrow in Figure 2).

Skip-connection section: This includes the ECSA module and ConvLSTM unit with the vertical flow direction.

The channel attention mechanism generally uses GAP to aggregate global information in each channel, while the U-Net architecture outputs different sizes of feature maps in each layer, which leads to a large amount of lost spatial information. To solve this problem, we propose the ECSA module, which is a local cross-channel interaction module without dimensionality reduction, with the structure shown in Figure 3, where the convolutional kernel size is experimentally shown to be the best with a size of 3.

Given a feature map $X \in R^{C \times H \times W}$ and denoting the adaptive average pooling layer as $P(\cdot, \cdot)$, $C1D_3(\cdot)$ as a 1×3 one-dimensional convolution, and $\sigma(\cdot)$ as the Sigmoid activation function, then the output \tilde{X} of the ECSA module can be represented as shown in (2).

$$\tilde{X} = \sigma(C1D_3(P(X, R))) \otimes X \quad (2)$$

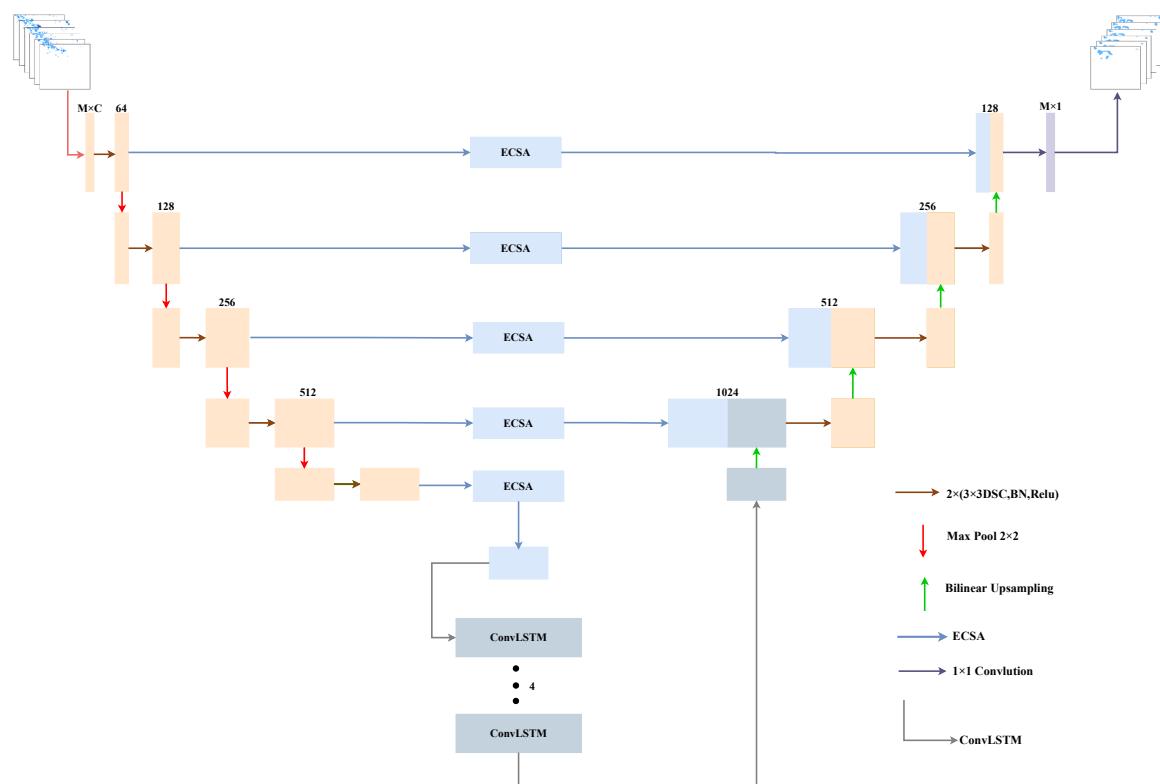


Figure 2. LSTMAtU-Net model framework. It contains the encoder–decoder structure. Each bar in the figure represents a multi-channel feature map, and the number represents the number of channels. The input is an example of precipitation, and the output is the corresponding prediction map generated by the LSTMAtU-Net model. On the right side of the diagram is the description of each arrow.

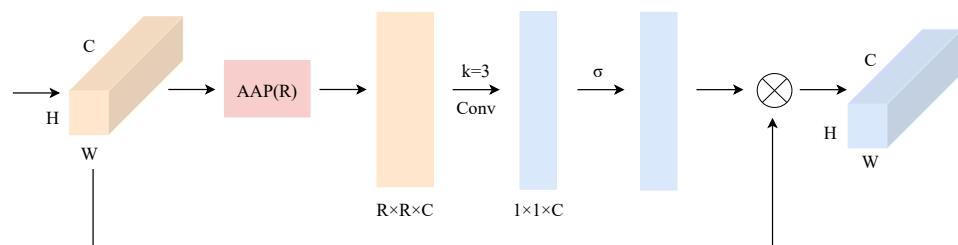


Figure 3. ECSA module architecture diagram. R denotes for each layer different feature map sizes adapted to the output of different scales; for example, R in the first layer of the ECSA module is taken as 16. AAP is the Adaptive Average Pooling layer.

It first adaptively outputs features of different scales R (16×16 for the first layer, 8×8 for the second layer, 4×4 for the third layer, 2×2 for the fourth layer, and 1×1 for the fifth layer) according to the different feature map dimensions of each layer to obtain the spatial structure information of $R \times R \times C$. Then, the feature compression is performed along the spatial dimension, and each feature channel is transformed into $R \times R$ real numbers; then, this information is convolved to learn the importance of different channel features, and at this time, a feature map of dimension $1 \times 1 \times C$ is output. Finally, the output feature map is multiplied by the original input feature map channel, so that the spatial structure information and channel information can be integrated simultaneously to improve the attention to the precipitation area.

The detailed structure of the ConvLSTM unit with the vertical flow direction is shown in Figure 4. It is built with convolution to combine the CNN layer with the LSTM cell in order to grasp both the spatial and temporal information. This structure is equivalent to

rotating the original ConvLSTM unit 90 degrees clockwise so that the information flow is delivered in the vertical direction, so as to realize the combination with the U-Net structure. Its input is the output \tilde{X} of the ECSA attention module, and M^l is the spatial-temporal memory described. It passes vertically from the $l - 1$ th layer to the current node in the same time step, and the final hidden state H^l also depends on the spatial memory at this time. Its forgetting gate f , memory gates i and g , and output gate o are consistent with the gate structure of the ConvLSTM unit. It has the following update formula, where $*$ is the convolution operation and \otimes denotes the Hadamard product.

$$\begin{aligned} g &= \tanh(W_{xg} * \tilde{X} + W_{hg} * H^{l-1} + W_{mg} * M^{l-1} + b_g) \\ f &= \sigma(W_{xf} * \tilde{X} + W_{hf} * H^{l-1} + W_{mf} * M^{l-1} + b_f) \\ i &= \sigma(W_{xi} * \tilde{X} + W_{hi} * H^{l-1} + W_{mi} * M^{l-1} + b_i) \\ M^l &= f \otimes M^{l-1} + i \otimes g \\ o &= \sigma(W_{xo} * \tilde{X} + W_{ho} * H^{l-1} + W_{mo} * M^l + b_o) \\ H^l &= o \otimes \tanh(M^l) \end{aligned} \quad (3)$$

All states of the ConvLSTM were initialized to 0 before the first spatiotemporal state appeared.

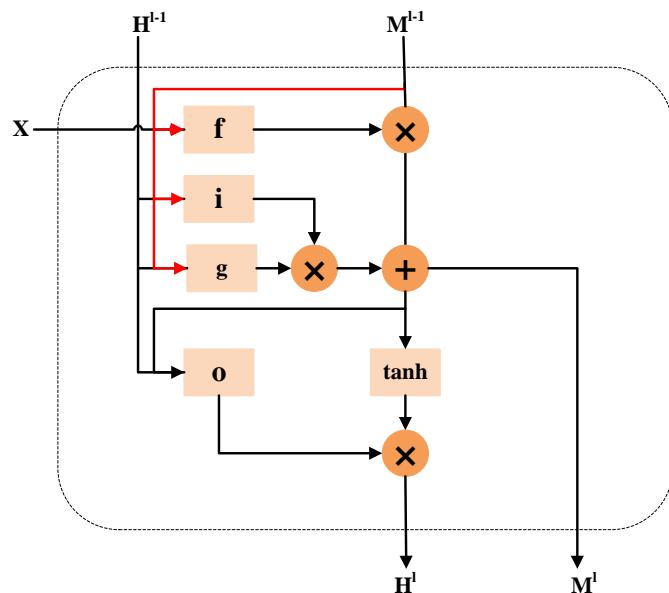


Figure 4. The structure of the ConvLSTM unit with the vertical flow direction, where f is the forgetting gate, i and g are the memory gates, and o is the output gate.

3.3. Loss Function

The previous loss function of precipitation nowcasting was the Mean-Squared Error (MSE), and it can accelerate the regression of forecast data; however, it is obviously not suitable for optimizing a single measurement criterion because precipitation is a strong random and chaotic event. Therefore, we made several improvements on the basis of the MSE.

Firstly, the weighted MSE loss function in [19] was used because of the imbalance of the precipitation data samples (the first formula in (5)).

Secondly, the Critical Success Index (CSI) [34] is generally used to evaluate precipitation in the meteorological field. The case where all precipitation was greater than the set threshold was considered as 1, and the others were set as 0. Setting the TPs (prediction = 1, truth = 1), FPs (prediction = 1, truth = 0), TNs (prediction = 0, truth = 0), and FNs (prediction = 0, truth = 1), the CSI score was calculated as shown in (4). This

shows that the CSI score is a categorical indicator and the MSE is a regression loss function, so optimizing the MSE does not result in a significant improvement in the CSI score. Therefore, the binary cross-entropy loss function [35] was introduced and slightly modified into a loss function suitable for regression problems. A target value y_i greater than the threshold was set to 1 and for the rest of the cases to 0 to obtain y'_i , and the difference between the predicted value and the threshold was passed through the Sigmoid function to obtain the probability of whether the predicted value is greater than the threshold to calculate the loss function value.

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (4)$$

Our final loss function, TLoss, is the last equation in (5), and it was experimentally proven that the improved binary cross-entropy loss function with the weight value set to 0.5 achieved the best results.

$$\begin{aligned} \text{WMSE} &= \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 * (e^{y_i * 0.6} - 0.8) \\ \hat{y}_p &= \sigma(\hat{y} - p) \\ \text{BCE}_p &= -\frac{1}{n} \sum_{i=1}^n (y'_i) * \log \hat{y}_p + (1 - y'_i) * \log(1 - \hat{y}_p) \\ \text{TLoss} &= \text{WMSE} + 0.5 * (\text{BCE}_{p1} + \text{BCE}_{p2} + \text{BCE}_{p3}) \end{aligned} \quad (5)$$

In this loss function, BCE_p denotes the binary cross-entropy loss function for each frame greater than p mm, $p1$ denotes light rainfall, $p2$ denotes moderate rainfall, and $p3$ denotes heavy rainfall.

4. Experiments

4.1. Dataset

In this paper, we used the dataset from the 2022 Jiangsu Meteorological AI Algorithm Challenge. Selecting this dataset for the experiments can make our model more widely applicable in other engineering applications.

The 2022 Jiangsu Weather AI Algorithm Challenge dataset, which has precipitation feature maps produced by Jiangsu Meteorological Station, is elemental data for the period of April–September 2019–2021. It is stored as a grayscale map, obtained by interpolating data from automatic meteorological observation stations in Jiangsu and surrounding areas to a uniform network. Interpolation objectively reflects the spatial distribution of meteorological elements and matches other meteorological data. The precipitation element is the cumulative 6 min precipitation at the automatic station, i.e., the cumulative value of precipitation observations over a 6 min period up to the current moment. The resolution of a frame is 480×560 pixels, ranging from 0–10 mm. In addition, the competition also provides meteorological radar data; these data were obtained from the quality control and mosaic of multiple S-band meteorological radars in Jiangsu Province. Quality control is the premise of ensuring the correct and credible meteorological observation data entering the application system such as weather forecast. The mosaic splices the fragmented regional radar monitoring map into a complete regional radar map of Jiangsu Province. Radar echo data cover the whole Jiangsu region with a horizontal resolution of 0.01° (about 1 km). It has the same spatial and temporal resolution as the precipitation element. The data range is 0–70 dBZ.

In terms of data preprocessing, first of all, because of the strong correlation between radar echo elements and precipitation rainfall magnitude and distribution, it can be seen from Table 2 that the performance of the fused data was usually stronger than that of a single precipitation factor, so the two elements were stacked at the channel level to form multi-source fused data for precipitation nowcasting. Secondly, in order to facilitate the research and due to the limitation of computing power and training cost, we downsampled the data again to 120×140 pixels by bilinear interpolation. Finally, in order to predict the

precipitation in the next two hours, the multi-source fusion data of the past two hours were used to predict the precipitation in the next two hours, a sequence of 40 times in total. The data were divided into 27,625 groups and divided into training set and test sets with a ratio of 8:2.

Table 2. The quantitative evaluation results of the single factor and the fusion features are the average evaluation results of the 10-frame precipitation nowcasting in the previous 1 h. “ $r \geq \gamma$ mm” refers to the skill score under the γ mm rainfall threshold within 6 min. “ \uparrow ” means the higher the score, the better, while “ \downarrow ” means the lower the score, the better. And the bold font indicates better performance.

Method	$r \geq 0.05$ mm CSI \uparrow	$r \geq 0.2$ mm CSI \uparrow	$r \geq 0.5$ mm CSI \uparrow
U-Net (single factor)	0.3993	0.2432	0.0978
U-Net (fusion features)	0.3878	0.2565	0.1098
ConvLSTM (single factor)	0.4132	0.2660	0.1328
ConvLSTM (fusion features)	0.4204	0.2984	0.1487

4.2. Performance Metrics

The evaluation metrics used were the CSI [34], the Probability Of Detection (POD), and the False Alarm Rate (FAR). The CSI is the World Meteorological Organization’s scoring standard for the accuracy of quantitative precipitation forecast. The POD refers to the proportion of the predicted actual precipitation area in all actual precipitation areas. The FAR refers to the proportion of the actual non-precipitation area in the forecast precipitation area in relation to the total forecast precipitation area. The calculation of the CSI is given in Section 3. The POD and FAR scores were calculated as follows. We set the thresholds to 0.05 mm, 0.2 mm, and 0.5 mm to calculate these metrics, corresponding to light rain, light to moderate rain, and moderate to heavy rain [13]. In addition, in order to show the impact of the next 2 h rainfall nowcasting, we evaluated it in two time periods of 1 h and 2 h. The TPs, FNs, FPs, and TNs are also given in Section 3.

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

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}}$$
(6)

4.3. Implementation Details

LSTMAtU-Net was implemented in Pytorch. We first normalized the precipitation data M . The normalization procedure was as follows, and the logarithmic transformation can make the precipitation data distribution as close to a Gaussian distribution as possible, thus enhancing the model fitting effect with the MSE as the loss function. We set 0.001 as the learning rate, used the Adam optimizer, and set the batch size to 20. The LSTMAtU-Net model uses the TLoss as the verification loss function. In the loss function, p1 takes 0.05, p2 takes 0.5, and p3 takes 1. Other models use the MSE loss function. All experiments were trained and tested on an NVIDIA Tesla A100 GPU.

$$M = \log(1 + M) \quad (7)$$

4.4. Quantitative Analysis on the 2022 Jiangsu Weather AI Algorithm Challenge Dataset

We compared the LSTMAtU-Net model with other models on the 2022 Jiangsu Meteorological AI Algorithm Challenge dataset. We calculated 6 min of cumulative rainfall to evaluate the performance of these models. The average evaluation results of the 10-frame precipitation forecast in the previous hour are shown in Table 3, and the average evaluation results of the 20-frame precipitation forecast in two hours are shown in Table 4. The best performance is indicated by bolded font.

The CSI is the main basis for comparing the performance of precipitation nowcasting methods. From the results of Tables 3 and 4, the LSTMAtU-Net model achieved the best results on the CSI evaluation metric, and the evaluation metrics of medium and high intensity precipitation were at least 4% higher than other SOTA models. In addition, we also provide a plot of the evaluation metrics for different prediction steps in Figure 5. The red dotted lines show the prediction results of our model, and it can be seen from the figure that this model had the best prediction results in most of the prediction intervals with medium and high thresholds. This means that the LSTMAtU-Net model we propose not only improved the accuracy of precipitation nowcasting to a certain extent, but also had fair performance in long-term dependence. However, it did not achieve good performance on the FAR index, and the false alarm rate needs to be reduced. Next, we compared the evaluation results of different methods.

Comparison of the models based on the ConvRNN architecture and U-Net architecture: From Tables 3 and 4, it can be seen that the overall CSI of the model based on the ConvRNN architecture [10,11,24–26] was relatively high. Compared with the overall average score of the model based on the U-Net architecture [15,16,21], the overall average CSI score of the model based on the ConvRNN architecture increased by at least 2% in the evaluation results within the first hour and second hour. The precipitation nowcasting accuracy of [15,16,21] was low, and the FAR scores were high; however, their POD scores were relatively high. The POD scores with thresholds of 0.05 mm, 0.2 mm, and 0.5 mm in the evaluation results within 2 h were at least 0.68, 0.44, and 0.16, respectively, while the POD scores in [10,11,24–26] were basically below these values. We believe that the U-Net architecture model can capture complex spatial changes, so their POD scores were higher. The information transmitted decayed faster with time than the model based on the ConvRNN architecture. It can be clearly seen from the third picture in Figure 5 that the gray line representing the SmaAt-U-Net model had a higher score than most of the models based on the ConvRNN architecture in the first two time steps, but the latter's score decreased faster, so their CSI scores were relatively low. The precipitation forecast accuracy of [10,11,24–26] was higher, and the FAR scores were lower; however, their POD scores were relatively low. This result was because the models based on the ConvRNN architecture have obvious advantages in the processing time dimension, but their ability to capture features is not strong, so their CSI scores were higher and POD scores relatively lower.

LSTMAtU-Net is a model architecture based on the U-Net architecture, so its POD score was close to or even superior to that in [15,16,21]. In the evaluation results within 2 h, the POD scores of 0.2 mm and 0.5 mm reached greater than 0.5 and 0.3, respectively. It is also equipped with a ConvLSTM unit with the vertical flow direction, which enhanced the ability to process in the time dimension. It can be seen that the LSTMAtU-Net's score was the highest in most of the prediction intervals with medium and high thresholds in Figure 5, so it achieved the best results in the CSI evaluation metric.

A visual example of random selection is provided in Figure 6 to further visualize the predictive power of the model. The second behavior is the prediction results of our model. Firstly, it can be seen that the image details predicted by the LSTMAtU-Net model at three time nodes were more accurate than the other models, as it alleviated the problem of the RNN and CNN prediction image details being gradually inaccurate. Secondly, our model was superior to the other models in the area and intensity of medium and high precipitation, which improved the prediction accuracy of medium and high precipitation to a certain extent. However, neither our model nor the other models were accurate in portraying larger value precipitation areas, and the images became more and more blurred over time; these shortcomings need to be explored in future work.

Table 3. Comparison of average evaluation results of 10 frames of precipitation nowcasting in the first hour of different models. “ $r \geq \gamma \text{ mm}$ ” refers to the skill score under the $\gamma \text{ mm}$ rainfall threshold within 6 min. “ \uparrow ” means the higher the score, the better, while “ \downarrow ” means the lower the score, the better. The best performance is indicated by bolded font.

Model	CSI \uparrow	$r \geq 0.05 \text{ mm}$		$r \geq 0.2 \text{ mm}$		$r \geq 0.5 \text{ mm}$	
		POD \uparrow	FAR \downarrow	POD \uparrow	FAR \downarrow	POD \uparrow	FAR \downarrow
U-Net [21]	0.3878	0.7607	0.5227	0.2565	0.6308	0.6313	0.1098
SmaAt-U-Net [15]	0.4019	0.7795	0.5183	0.2877	0.6042	0.5914	0.1360
AA-TransU-Net [16]	0.3364	0.7351	0.5741	0.2464	0.5752	0.6104	0.1152
ConvLSTM [10]	0.4204	0.7139	0.5059	0.2984	0.4837	0.5198	0.1487
PredRNN [24]	0.4427	0.6353	0.4244	0.2972	0.4297	0.4977	0.1454
PredRNN++ [25]	0.4021	0.6591	0.5059	0.2872	0.4502	0.4819	0.1515
MIM [26]	0.4094	0.6290	0.4722	0.2716	0.3925	0.4385	0.1377
GAN-rcLSTM [11]	0.4286	0.5917	0.4216	0.2886	0.4035	0.4648	0.1314
LSTMAtU-Net	0.4455	0.7766	0.4876	0.3267	0.6662	0.5859	0.1908
							0.4318

Table 4. Comparison of the average evaluation results of 20 frames of precipitation nowcasting in two hours of different models. “ $r \geq \gamma \text{ mm}$ ” refers to the skill score under the $\gamma \text{ mm}$ rainfall threshold within 6 min. “ \uparrow ” means the higher the score, the better, while “ \downarrow ” means the lower the score, the better. The best performance is indicated by bolded font.

Model	CSI \uparrow	$r \geq 0.05 \text{ mm}$		$r \geq 0.2 \text{ mm}$		$r \geq 0.5 \text{ mm}$	
		POD \uparrow	FAR \downarrow	POD \uparrow	FAR \downarrow	POD \uparrow	FAR \downarrow
U-Net [21]	0.3285	0.7220	0.5950	0.1904	0.4783	0.6827	0.0601
SmaAt-U-Net [15]	0.3459	0.7414	0.5814	0.2193	0.4711	0.6152	0.0802
AA-TransU-Net [16]	0.2994	0.6836	0.6144	0.1883	0.4459	0.6508	0.0700
ConvLSTM [10]	0.3696	0.6873	0.5652	0.2311	0.3932	0.5557	0.0986
PredRNN [24]	0.3749	0.5600	0.4933	0.2221	0.3298	0.5570	0.0985
PredRNN++ [25]	0.3570	0.6414	0.5670	0.2260	0.3801	0.5124	0.1067
MIM [26]	0.3579	0.5686	0.5234	0.2095	0.3131	0.4628	0.0963
GAN-rcLSTM [11]	0.3662	0.5334	0.4922	0.2162	0.3189	0.5449	0.0888
LSTMAtU-Net	0.3813	0.7249	0.5542	0.2564	0.5364	0.6425	0.1350
							0.3025

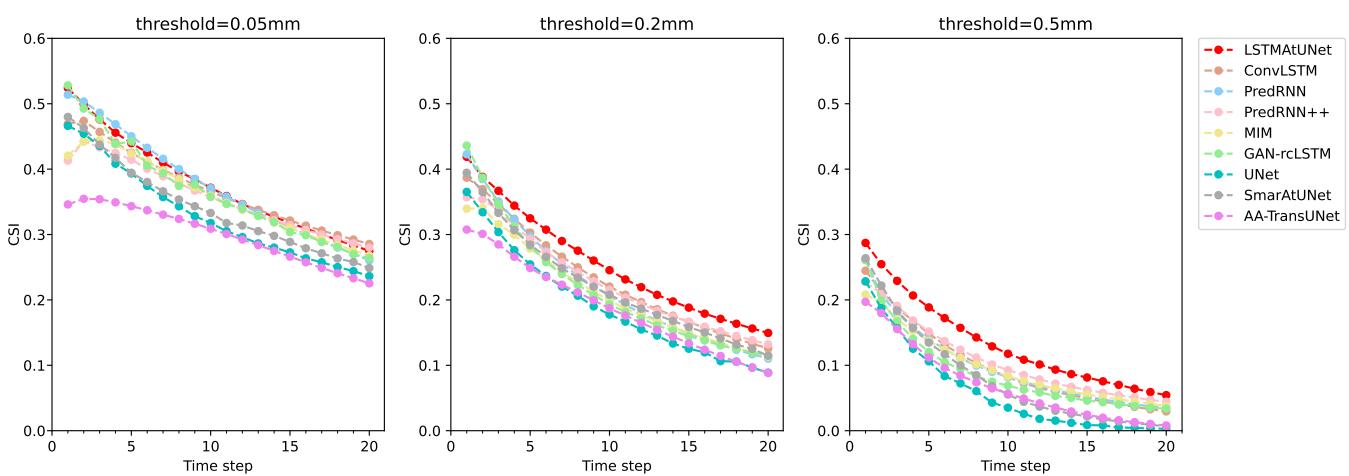


Figure 5. Prediction comparison of different time steps in two hours; 1 time step is 6 min.

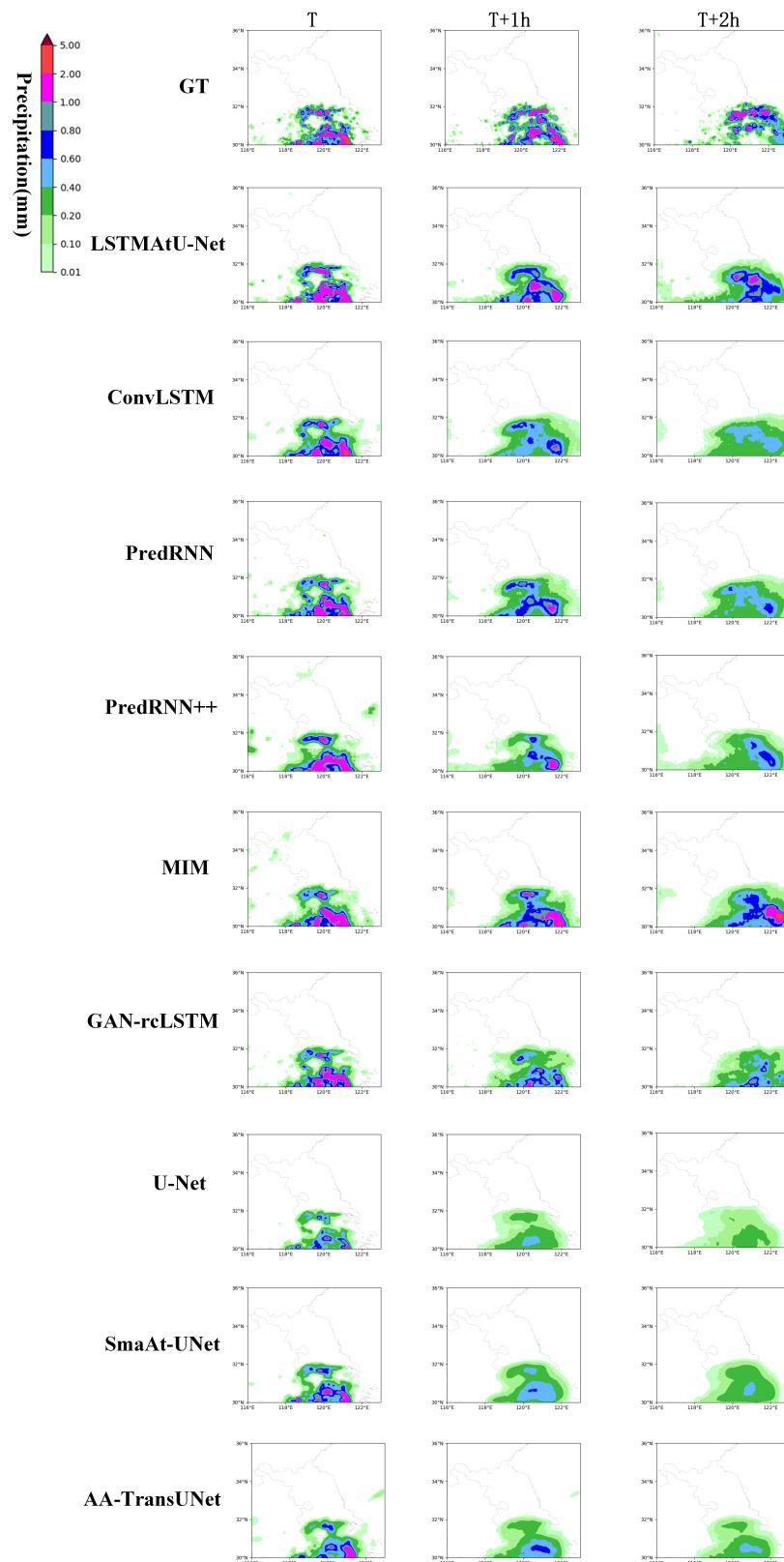


Figure 6. Visualization comparison with other models.

4.5. Ablation Experiments and Analyses

In order to fully explore the performance of the model, we used ablation experiments to explore the performance of the model architecture and the ECSA module and TLoss in

the model. Since our model was based on the U-Net architecture, we used the U-Net model as the baseline model.

Firstly, the precipitation forecast accuracy and visual images of the model with the ECSA module and the model without the ECSA module were compared to evaluate the influence of the ECSA module on the details of the forecast images. As shown in the third, fourth, sixth, and seventh lines of Tables 5 and 6 and Figure 7, it can be seen that the ECSA module improved the accuracy of the precipitation forecast to a certain extent, and the average precipitation results for the CSI metric in 2 h were the best. In addition, in the visual image, the model with the ECSA module was more accurate than the model without the ECSA module to predict the details of the image. Therefore, it can be shown that the ECSA module can alleviate the problem of the inaccurate prediction of image details to a certain extent.

Secondly, we compared the same model using different loss functions to evaluate the impact of the TLoss on the precipitation forecast accuracy. As shown in the fourth, fifth, seventh, and eighth lines of Tables 5 and 6 and Figure 7, it can be seen that the model using the TLoss was higher than the model using the MSE in the CSI metric, especially for the medium and high precipitation. In addition, in the image, the area and intensity of the prediction image using the TLoss function were more accurate at the medium and high thresholds. Therefore, it can be shown that the TLoss function improved the prediction accuracy of the medium and high precipitation to a certain extent.

Finally, the timing performance of the model architecture was compared. As shown in the first and fourth lines of Tables 5 and 6 and Figure 7, it can be seen that the model with the ConvLSTM unit achieved the best CSI metric. In addition, in the image, the two-hour prediction image of the model with the ConvLSTM unit was more accurate than the U-Net model. Therefore, it can be shown that the introduction of the ConvLSTM unit improved the long-term dependence.

Table 5. The average evaluation results of 10 frames of precipitation nowcasting in the first hour. “ $r \geq \gamma$ mm” refers to the skill score under the γ mm rainfall threshold within 6 min. “ \uparrow ” means the higher the score, the better, while “ \downarrow ” means the lower the score, the better. The best performance is indicated by bolded font.

Model	$r \geq 0.05$ mm			$r \geq 0.2$ mm			$r \geq 0.5$ mm		
	CSI \uparrow	POD \uparrow	FAR \downarrow	CSI \uparrow	POD \uparrow	FAR \downarrow	CSI \uparrow	POD \uparrow	FAR \downarrow
U-Net+MSE	0.3878	0.7607	0.5227	0.2565	0.6308	0.6313	0.1098	0.3155	0.6492
U-Net+ECSA+MSE	0.3983	0.7598	0.5095	0.2708	0.6645	0.6298	0.1265	0.3690	0.6764
U-Net+ECSA+TLoss	0.4449	0.6867	0.4385	0.3188	0.5828	0.5480	0.1716	0.3402	0.5820
LSTMAtU-Net without ECSA+MSE	0.4204	0.6552	0.4184	0.3043	0.4675	0.5460	0.1515	0.2270	0.5478
LSTMAtU-Net+MSE	0.4368	0.7785	0.4913	0.2997	0.6951	0.6260	0.1673	0.4673	0.7170
LSTMAtU-Net+TLoss	0.4455	0.7766	0.4876	0.3267	0.6662	0.5859	0.1908	0.4318	0.6419

Table 6. The average evaluation results of 20 frames of precipitation nowcasting in two hours. “ $r \geq \gamma$ mm” refers to the skill score under the γ mm rainfall threshold within 6 min. “ \uparrow ” means the higher the score, the better, while “ \downarrow ” means the lower the score, the better. The best performance is indicated by bolded font.

Model	$r \geq 0.05$ mm			$r \geq 0.2$ mm			$r \geq 0.5$ mm		
	CSI \uparrow	POD \uparrow	FAR \downarrow	CSI \uparrow	POD \uparrow	FAR \downarrow	CSI \uparrow	POD \uparrow	FAR \downarrow
U-Net+MSE	0.3285	0.7220	0.5950	0.1904	0.4783	0.6827	0.0601	0.1719	0.4659
U-Net+ECSA+MSE	0.3429	0.7388	0.5806	0.2109	0.5308	0.6742	0.0759	0.2191	0.5479
U-Net+ECSA+TLoss	0.3750	0.6522	0.5128	0.2407	0.4535	0.5864	0.1105	0.2165	0.5180
LSTMAtU-Net without ECSA+MSE	0.3654	0.5560	0.5042	0.2163	0.3356	0.6418	0.0967	0.1471	0.6038
LSTMAtU-Net+MSE	0.3743	0.7315	0.5582	0.2394	0.5786	0.6805	0.1177	0.3224	0.7032
LSTMAtU-Net+TLoss	0.3813	0.7249	0.5542	0.2564	0.5364	0.6425	0.1350	0.3025	0.6173

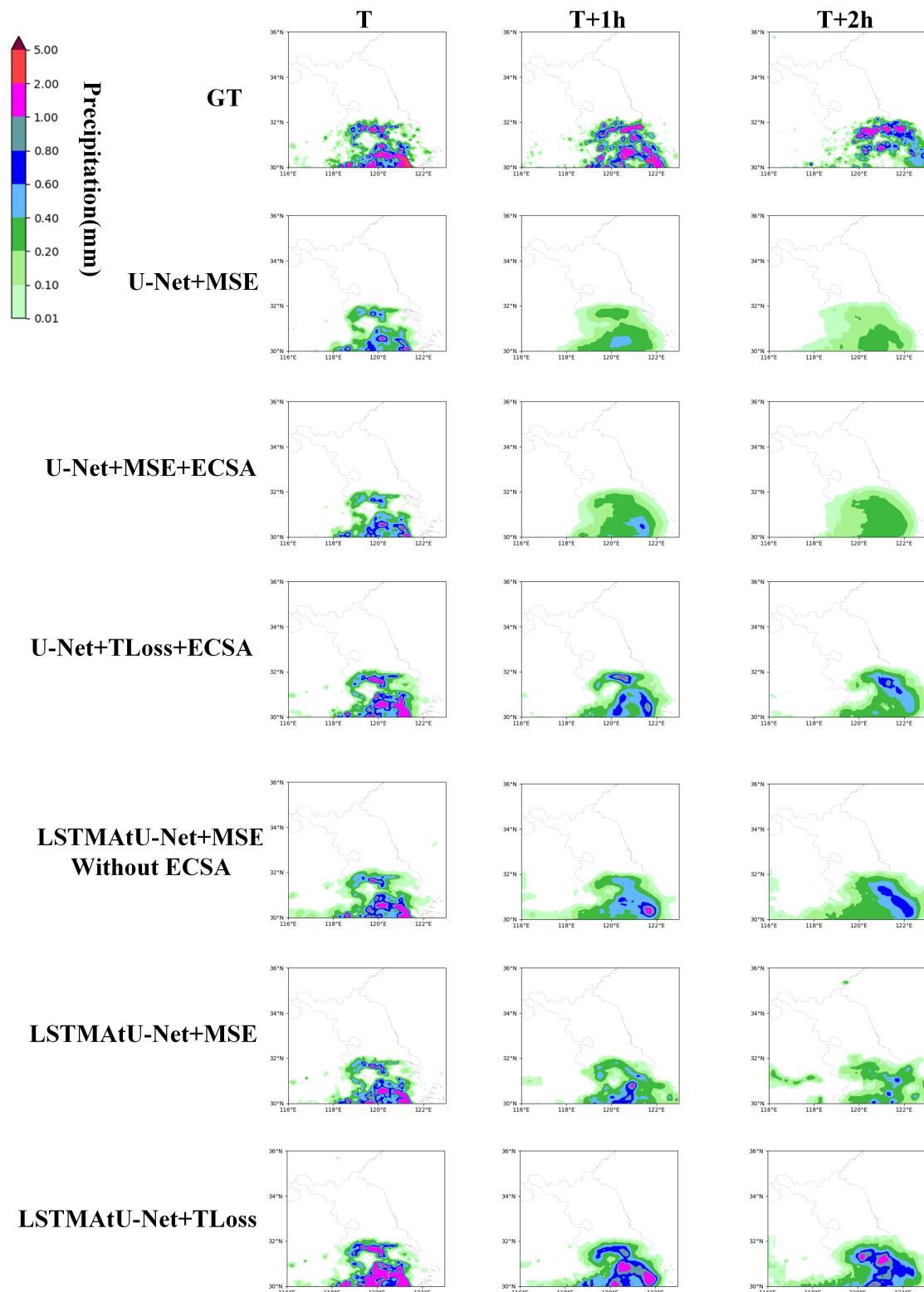


Figure 7. Comparison of the visualization results of the ablation experiments. The first behavior is the true value of precipitation. The second line, U-Net + MSE, means that the conventional U-Net model was used and the loss function was the MSE, and the other lines are similar. The first is the 6-minute cumulative precipitation from T to T + 6 min, the second the 6-minute cumulative precipitation from T + 60 min to T + 66 min, and the third the 6-minute cumulative precipitation from T + 114 min to T + 120 min.

5. Conclusions

In this paper, we proposed a new precipitation nowcasting model, LSTMAtU-Net. The model combines the CNN structure, the RNN structure, and an attention mechanism. It first performs convolutional and downsampling operations on the input image sequence to extract features from the images at various scales, followed by integrating the extracted feature information with channel information and spatial structure information through the ECSA module, grasping the temporal information through the ConvLSTM unit with the vertical flow direction, then fusing the information with the previous ones at the same scale for each upsampling, and finally, fusing them through a 1×1 convolution to output the predicted image. Here, we validated the feasibility and correctness of the proposed model using a competition dataset and a public dataset. It can be seen from the results that our proposed model outperformed the state-of-the-art precipitation nowcasting models.

Although the proposed model showed some improvements, there were still some shortcomings. First, the forecast accuracy of medium- to high-intensity precipitation was improved, but was still low. Secondly, the accuracy of the forecast results and the clarity of the forecast images decreased with the increase of the forecast time.

The first problem was caused by the small percentage of medium- to high-intensity precipitation data and the averaging property of the convolutional operation; therefore, we need to think deeply about the setting of the weighted loss function in future studies. For the second problem, the Transformer architecture can be introduced and improved, which can effectively capture the temporal coherence of long time series. References [20,36–40] applied the Transformer architecture to time series forecasting with good results, while References [29,41–43] also used this architecture in their work on precipitation nowcasting to demonstrate the feasibility of this method. Therefore, our future research will focus on the Transformer architecture and the weighted loss function to improve the precipitation nowcasting accuracy.

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

Funding: This research was funded by the National Key Research Development Plan under Grant 2017YFC1502104 and the Beijing foundation of NJIAS under Grant BJJ202103.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

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

References

1. 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](#)]
2. 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* **2008**, *10*, 325–338.
3. Sun, J.; Xue, M.; Wilson, J.W.; Zawadzki, I.; Ballard, S.P.; Onvlee-Hooimeyer, J.; Joe, P.; Barker, D.M.; Li, P.W.; Golding, B.; et al. Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 409–426. [[CrossRef](#)]
4. 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*, 3391–3407. [[CrossRef](#)]
5. Ligda, M. Lightning Detection by Radar. *J. Atmos. Terr. Phys.* **1956**, *9*, 279–283. [[CrossRef](#)]
6. Woo, W.C.; Wong, W.K. Operational Application of Optical Flow Techniques to Radar-Based Rainfall Nowcasting. *Atmosphere* **2017**, *8*, 48. [[CrossRef](#)]

7. Sakaino H. Spatio-temporal image pattern prediction method based on a physical model with time-varying optical flow. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3023–3036. [[CrossRef](#)]
8. Germann, U.; Zawadzki, I. Scale-dependence of the predictability of precipitation from continental radar images Part I: Description of the methodology. *Mon. Weather. Rev.* **2002**, *130*, 2859–2873. [[CrossRef](#)]
9. Luo, C.; Li, X.; Ye, Y. PFST-LSTM: A SpatioTemporal LSTM Model with Pseudoflflow Prediction for Precipitation Nowcasting. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 843–857. [[CrossRef](#)]
10. 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. *arXiv* **2015**, arXiv:1506.04214.
11. Geng, H.; Wang, T.; Zhuang, X.; Xi, D.; Hu, Z.; Geng, L. GAN-rcLSTM: A Deep Learning Model for Radar Echo Extrapolation. *Atmosphere* **2022**, *13*, 684. [[CrossRef](#)]
12. Bowler, N.; Pierce, C.E.; Seed, A. Development of a precipitation nowcasting algorithm based upon optical flflow techniques. *J. Hydrol.* **2004**, *288*, 74–91. [[CrossRef](#)]
13. Zhang, F.; Wang, X.; Guan, J. A Novel Multi-Input Multi-Output Recurrent Neural Network Based on Multimodal Fusion and Spatiotemporal Prediction for 0–4 Hour Precipitation Nowcasting. *Atmosphere* **2021**, *12*, 1596. [[CrossRef](#)]
14. 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](#)]
15. Trebing, K.; Sta'nczyk, T.; Mehrkanon, S. SmaAt-U-Net: Precipitation nowcasting using a small attention-U-Net architecture. *Pattern Recognit. Lett.* **2021**, *145*, 178–186. [[CrossRef](#)]
16. Yang Y.; Mehrkanon S. Aa-transunet: Attention augmented transunet for nowcasting tasks. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 18–23 July 2022.
17. 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](#)]
18. Garcia Fernandez, J.; Mehrkanon, S. Broad-U-Net: Multi-scale feature learning for nowcasting tasks. *Neural Netw.* **2021**, *144*, 419–427. [[CrossRef](#)]
19. Song, K.; Yang, G.; Wang, Q.; Xu, C.; Liu, J.; Liu, W.; Shi, C.; Wang, Y.; Zhang, G.; Yu, X.; et al. Deep learning prediction of incoming rainfalls: An operational service for the city of Beijing China. In Proceedings of the 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 8–11 November 2019; pp. 180–185.
20. Wu, S.; Xiao, X.; Ding, Q.; Zhao, P.; Wei, Y.; Huang, J. Adversarial sparse transformer for time series forecasting. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17105–17115.
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
22. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
23. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vitural, 13–19 June 2020; pp. 13–19.
24. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; p. 10.
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, Stockholm Sweden, 10–15 July 2018; pp. 5123–5132.
26. Wang, Y.; Zhang, J.; Zhu, H.; Long, M.; Wang, J.; Yu, P.S. Memory in memory: A predictive neural network for learning higher-order nonstationarity from spatiotemporal dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 9154–9162.
27. Hu J., Shen L., Sun G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
28. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
29. Gao, Z.; Shi, X.; Wang, H.; Zhu, Y.; Wang, Y.B.; Li, M.; Yeung, D.Y. Earthformer: Exploring Space-Time Transformers for Earth System Forecasting. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 25390–25403.
30. Sun, F.; Bai, C.; Song, Y.; Zhang, J. MMINR: Multi-frame-to-Multi-frame Inference with Noise Resistance for Precipitation Nowcasting with Radar. In Proceedings of the 2022 26th International Conference on Pattern Recognition, Montreal, QC, Canada, 21–25 August 2022; pp. 97–103.
31. Guo, J.; Ma, X.; Sansom, A.; McGuire, M.; Kalaani, A.; Chen, Q.; Tang, S.; Yang, Q.; Fu, S. Spanet: Spatial pyramid attention network for enhanced image recognition. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo, London, UK, 6–10 July 2020; pp. 1–6.
32. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.

33. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
34. Jolliffe, I.T.; Stephenson, D.B. *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*; John Wiley and Son: Hoboken, NJ, USA, 2012.
35. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
36. Lim, B.; Arik, S.Ö.; Loeff, N.; Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [[CrossRef](#)]
37. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [[CrossRef](#)]
38. Liu, S.; Yu, H.; Liao, C.; Li, J.; Lin, W.; Liu, A.X.; Dustdar, S. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In Proceedings of the International Conference on Learning Representations, Virtual Event, Austria, 3–7 May 2021.
39. Nejad, A.S.; Alaiz-Rodríguez, R.; McCarthy, G.D.; Kelleher, B.; Grey, A.; Parnell, A. SERT: A Transfomer Based Model for Spatio-Temporal Sensor Data with Missing Values for Environmental Monitoring. *arXiv* **2023**, arXiv:2306.03042.
40. Sun, Y.; Fu, K.; Lu, C.T. DG-Trans: Dual-level Graph Transformer for Spatiotemporal Incident Impact Prediction on Traffic Networks. *arXiv* **2023**, arXiv:2303.12238.
41. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
42. Bai, C.; Sun, F.; Zhang, J.; Song, Y.; Chen, S. Rainformer: Features Extraction Balanced Network for Radar-Based Precipitation Nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
43. Franch, G.; Tomasi, E.; Poli, V.; Cardinali, C.; Cristoforetti, M.; Alberoni, P.P. Ensemble precipitation nowcasting by combination of generative and transformer deep learning models. In Proceedings of the Copernicus Meetings, Vienna, Austria, 23–28 April 2023.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.