remote sensing

MDPI

Article

Two-Stage UA-GAN for Precipitation Nowcasting

Liujia Xu [1], Dan Niu [1,2,*], Tianbao Zhang [1], Pengju Chen [1], Xunlai Chen [3] and Yinghao Li [1]

[1] School of Automation, Southeast University, Nanjing 210096, China
[2] Key Laboratory of Measurement and Control of CSE, Ministry of Education Research Laboratory, Nanjing 210096, China
[3] Shenzhen Key Laboratory of Severe Weather in South China, Shenzhen Meteorological Bureau, Shenzhen 518040, China
* Correspondence: 101011786@seu.edu.cn

Abstract: Short-term rainfall prediction by radar echo map extrapolation has been a very hot area of research in recent years, which is also an area worth studying owing to its importance for precipitation disaster prevention. Existing methods have some shortcomings. In terms of image indicators, the predicted images are not clear enough and lack small-scale details, while in terms of precipitation accuracy indicators, the prediction is not accurate enough. In this paper, we proposed a two-stage model (two-stage UA-GAN) to achieve more accurate prediction echo images with more details. For the first stage, we used the Trajectory Gated Recurrent Unit (TrajGRU) model to carry out a pre-prediction, which proved to have a good ability to capture spatiotemporal movement of rain field. In the second stage, we proposed a spatiotemporal attention enhanced Generative Adversarial Networks (GAN) model with a U-Net structure and a new deep residual attention module in order to carry out the refinement and improvement of the first-stage prediction. Experimental results showed that our model outperforms the optical-flow based method Real-Time Optical Flow by Variational Methods for Echoes of Radar (ROVER), and some well-known Recurrent Neural Network (RNN)-based models (TrajGRU, PredRNN++, ConvGRU, Convolutional Long Short-Term Memory (ConvLSTM)) in terms of both image detail indexes and precipitation accuracy indexes, and is visible to the naked eye to have better accuracy and more details.

Keywords: precipitation nowcasting; GAN; attention mechanism; spatiotemporal prediction

Citation: Xu, L.; Niu, D.; Zhang, T.; Chen, P.; Chen, X.; Li, Y. Two-Stage UA-GAN for Precipitation Nowcasting. *Remote Sens.* **2022**, *14*, 5948. https://doi.org/10.3390/rs14235948

Academic Editor: Ka Lok Chan

Received: 15 October 2022
Accepted: 17 November 2022
Published: 24 November 2022

1. Introduction

Precipitation nowcasting, which refers to forecast rainfall in a very short term (often in 0–6 h) at the minute level, kilometer level, or street level [1,2], has the following characteristics: stronger timeliness, higher resolution, and higher accuracy. Precipitation nowcasting mainly includes nowcasting from a variety of observations as well as different kinds of numerical weather models. It is of great significance to protect people's lives and property. From 17 July 2021, the torrential rain disaster in Henan Province has caused huge economic losses in Zhengzhou, Xinxiang, Hebi, Zhoukou, and other cities. The yields of crops such as corn, peanuts, and soybeans were also seriously and negatively affected [3]. Precipitation nowcasting can forecast heavy rain for hours in advance. Therefore, it plays an important role in the early prevention and control of rainstorm disasters and minimizing damage and losses.

Typically, the traditional weather forecasting method, also known as Numerical Weather Prediction (NWP), is a prediction scheme based on fluid dynamics and thermodynamic equations. Recent studies have sought to assimilate radar and lighting observations into NWP models using various data assimilation (DA) techniques with the aim of solving problems related to spinup in high-resolution NWP, thus improving short-term severe weather prediction [4]. Although recent advances in numerical weather prediction have enabled us to predict atmospheric dynamics at very high resolutions, the computational

cost of nowcasting with frequent update cycle requirements is sometimes too high [5]. In addition, even small disturbances in initial conditions, boundary conditions, and rounding errors can cause the descending accuracy of the NWP method [6,7].

With the development of observation technology and the increase in the number of observation satellites and radars, prediction methods based on the extrapolation of radar echo maps to infer short-term rainfall have shown good prediction performance. The radar echo maps have a high correlation to the rainfall rate through the intensity of the radar echo. By extrapolating the radar echo maps, the short-term precipitation intensity within the area can be predicted. To convert radar echo maps into rainfall intensity maps, some methods can be used [8]. An example of radar echo image is given in Figure 1. It is a Constant Altitude Plan Position Indicator (CAPPI) radar echo map captured by a Doppler radar at an altitude of 2 km, covering an area of 512 km × 512 km around Hong Kong. The radar echo map was originally a two-dimensional array. By grouping different echo intensities according to different thresholds and assigning different colors to different groups, a colored radar echo image is generated, as shown in Figure 1.

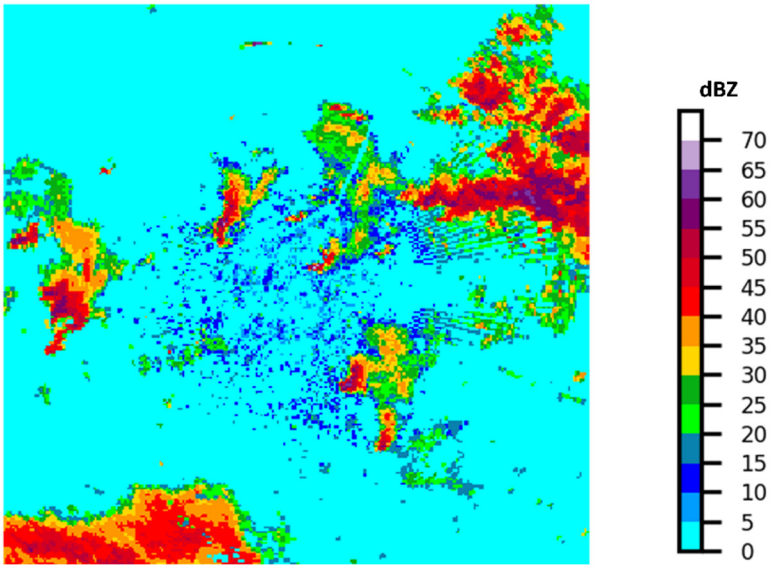

Figure 1. An example of radar echo image centered in Hong Kong City at 24 May 2015, 02:48:00 (UTC).

There are five general categories [9] that current radar echo extrapolation methods belong to: object-based methods, area-based methods, statistical methods, probabilistic methods, which are all traditional methods, and artificial intelligence (AI) methods, which are modern methods becoming popular in recent years. Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) is an object-based methodology featuring simplicity [10] that can forecast storm movement depending on a weighted linear fit to the previous storm, with inability to capture nonlinear patterns. Conventional area-based nowcasting methods including tracking of radar echoes by correlation (TREC) [11] and the McGill Algorithm for Precipitation Nowcasting by Lagrangian Extrapolation (MAPLE) [12], provided by Germann and Zawadzki, tend to calculate the motion tendency of the rain fields in certain areas. TREC predicts the radar motions by obtaining the quantitative and qualitative maximum correlations between two neighboring arrays of radar reflectivity factor data, while MAPLE uses a modified semi-Lagrangian advection scheme to estimate the motion field of precipitation. Since TREC usually has a poor forecasting effect on rapidly changing convective precipitation and has discontinuities in the resulting vector field, continuity of TREC vectors (COTREC) was proposed by Li to improve the consistency [13]. Classical statistical methods include spectral prognosis (S-PROG) [14], a nowcasting model for rain fields that have both spatial and dynamic scaling properties, by using a notch filter

to carry out spectral decomposition and scale-dependent temporal evolution. Probability forecasts [15], another part of the McGill Algorithm, improve the accuracy of nowcasting by adding probabilistic information, belonging to the category of probabilistic methods. In order to improve the radar-echo extrapolation prediction technology, many scholars have introduced the widely-used optical flow method in the field of computer vision to track the echo motion of the radar-echo. The ROVER algorithm [16] is a typical optical flow model. One of the known weaknesses in the above-mentioned traditional methods is that those models cannot take abundant advantage of the historical sequential information. They also cannot provide an end-to-end prediction of an entire sequence.

To solve these two problems, also benefiting from the development of the computation power, machine learning methods are widely applied in the field of computer vision as well as in extrapolation of radar echo maps, which is substantially a video prediction problem. Klein et al. [17] added a dynamic convolution layer to the traditional Convolutional Neural Networks (CNNs) structure to generate two prediction probability vectors to predict precipitation echoes. Shi E et al. added RNN to the dynamic convolutional layer proposed by Klein to construct Recurrent Dynamic Convolutional Neural Networks (RDCNN) [18], which has achieved good results in both forecast accuracy and forecast timeliness. Since the traditional LSTM cannot realize the extraction of spatial features, Shi et al. changed the input-to-state and state-to-state transform to convolution, and ConvLSTM network was proposed [19]. In order to adapt to the non-temporal and spatial invariance of most motions in practical situations, Shi et al. improved the model and introduced a trajectory GRU model (TrajGRU) [20] with learnable convolutions. Singh [21] also realized the prediction based on radar echo image sequence by adding a convolution structure on the basis of recurrent neural network to adapt to the spatiotemporal dependence of radar echo images. Feng et al. [22] conducted long-term and case-by-case tests on radar nowcasting using neural networks and related cross-over algorithms. Results showed that the accuracy of neural networks on moderate rainfall intensity test items was significantly improved. Wang et al. [23] proposed an improved ST-LSTM cell (Spatiotemporal LSTM) based on LSTM cells and applied it to a new end-to-end model PredRNN. Compared with PredRNN, a gradient unit module was added to construct PredRNN++ [24] to further capture the long-term memory, as also presented by Wang et al. in 2019. Agrawal et al. [25] regarded precipitation forecasting as a conversion problem from pictures to pictures, and used a convolutional neural network with a U-net structure to achieve the forecasting purpose. The most obvious disadvantage of models based on RNN is that errors will accumulate over time. Moreover, due to the use of the MSE or MAE loss function, the models tend to reduce the average error of the entire radar echo map, resulting in a blurry predicted picture. Based on the above two reasons, some scholars proposed to apply GAN to generate more realistic radar echo extrapolation prediction [26–29]. Usually, GAN consists of two parts: a generator to generate 'fake' images, and a discriminator whose task is to identify 'fake' images from the generator. Liu and Lee introduced a Meteorological Predictive Learning GAN model (MPL-GAN), which is based on conditional GAN, dealing with the uncertainty and sharpening the prediction [27]. Ravuri provided deep generative models (DGMs) [28] using two discriminators to correct the output of the generator and enhance details. A Conditional Generative Adversarial 3-D Convolutional Neural Network (denoted as ExtGAN) [29] was presented by Wang et al., using a combination of CNN model and GAN model to refine the prediction.

Although the existing GAN model improves the clarity of the predicted echo maps and enhances the performance of details, the capability of improving prediction accuracy performance is not satisfactory. The prediction accuracy, especially for light and moderate precipitation, needs to be further improved [27,29]. In ref. [30], a two-stage network was proposed, wherein the second-stage network was used to further refine the raw prediction by the first stage network. Inspired by ref. [30], we proposed a two-stage GAN-based precipitation nowcasting model in this paper. To merge the advantage of RNN and GAN models, in the first stage, as a pre-prediction process, TrajGRU model was used to capture

the spatiotemporal transformation to give the first-stage prediction. In the second stage, a deep residual attention-enhanced GAN model was proposed to refine the first-stage prediction, which further enhances the prediction accuracy and sharpen the predicted maps with more details. A real-world dataset of radar echo maps (public benchmark dataset HKO-7), covering Hong Kong and surrounding areas, was used in this paper to train and evaluate the proposed model fairly. It is known that both prediction accuracy and image refinement with more details are important [28,29]. Therefore, we evaluated the prediction performance using two categories of indicators. The first focuses on echo image refinement index and then Root Mean Square Error (RMSE), Structural Similarity Index Measure (SSIM), and Sharpness were employed [29]. The second is the widely-used accuracy indicators for precipitation prediction [19,20,29,30], including critical success index (CSI), Heidke Skill Score (HSS), and False Alarm Rate (FAR). Our goal was to further enhance the precipitation prediction accuracy, while handling blurry extrapolation and increasing the small-scale details of predicted radar echo maps. Experiments showed that our model outperforms traditional methods (ROVER) and some deep-learning models (ConvLSTM, ConvGRU, TrajGRU, and PredRNN++) in terms of both image evaluation metrics and prediction accuracy metrics, which is demonstrated in Section 3.3.

This paper is developed in the following order. The analysis and definition of the extrapolation problem of the radar echo map, as well as the detailed information of our proposed model are demonstrated in Section 2. In Section 3, experimental comparisons of the models are presented. Section 4 presents the conclusions.

2. Methods

To obtain the short-term rainfall intensity prediction using the radar echo map extrapolation method, first we need to produce a sequence of predicted radar echo maps, and then the predicted radar echo maps can be converted into predicted rainfall intensity maps [8]. In this paper, we focused on generating a very close prediction to the real future radar echo image sequence by the previous sequence extrapolation. Since the RNN and CNN model mentioned before often bring blurry prediction results, and the prediction accuracy of some GAN models is unsatisfactory [31], we now propose a two-stage GAN-based prediction and refinement model. We applied TrajGRU in the first stage to obtain a rough prediction. For the second stage, we proposed a GAN model to further enhance prediction accuracy and, more importantly, achieve more small-scale detailed prediction. To let the second-stage network have better generation ability, three innovative ideas were applied to our model:

- A U-Net structure was used to build the generator, so that feature maps at different scales can be preserved and fused through up-sampling, down-sampling, and integrating global and local skip connections in the decoder part.
- A stacked residual attention module was designed and added into the decoder part of our generator to extract and adaptively rescale the multiscale sequence- and spatial-wise features. Experiments showed that the attention mechanism in the decoder enhanced the generator to obtain better prediction accuracy.
- In most of the existing two-stage models, the input of the second stage contains only the output of the first stage so that the second stage is only a refinement of the first stage. However, the data input to our second-stage generator consisted of one image generated by the first stage, and four history truth images in order to let our generator obtain more historical spatiotemporal information. By reviewing a certain amount of historical radar echo information, the generator becomes not only a refinement module, but also an optimization module with predictive ability.

In this section, we first state the problem in formulation, and then introduce our two-stage prediction model in detail.

2.1. Formulation of Prediction Problem

In general, the radar echo data are recorded and saved as a grayscale image every 6 min [19,20], which means 10 frames will be saved in a single hour. We can predict the precipitation in the next hour through the radar echo maps of the past hours. In this work, we actually usde the previous 10 frames of radar echo map to predict the next 10 frames of the radar echo map. Then the predicted radar echo intensity value can be converted to precipitation intensity value [8]. Essentially, using radar echo map extrapolation to do precipitation nowcasting is a spatiotemporal sequence prediction [19], which receives radar echo images in the past as input and outputs possible estimation of future radar echo images. Using a tensor $X_t \in R^{C\times W\times H}$ to represent the observed radar echo maps at time t, with C, W, H, representing the channel number of the radar echo image, and the width and height of the images. We can describe the radar echo map extrapolation problem as (1):

$$\hat{X}_{t+1}, \ldots, \hat{X}_{t+O} = \underset{X_{t+1},\ldots,X_{t+O}}{\operatorname{argmax}} \; p(X_{t+1}, \ldots, X_{t+O} | \widetilde{X}_{t-I+1}, \widetilde{X}_{t-I+2}, \ldots, \widetilde{X}_t) \tag{1}$$

Here, p is the conditional probability. $\hat{X}_{t+1}$ is referred to as the prediction at time $t+1$, and $\widetilde{X}_t$ means the real radar echo image at time t. With function (1), we expect to find the length-O prediction map sequence closest to the real radar echo maps: $t+1 \sim t+O$, through the real radar echo maps in the past from time $t-I+1$ to time t. In other words, I represents the number of images input to the model, and O represents the number of prediction outputs.

2.2. Network Structure

Figure 2 illustrates the structural framework of the proposed model. The model is composed of two stages: the pre-prediction stage on the left with a gray background and the refinement stage on the right with a yellow background. The first stage employs the TrajGRU model. Unlike the original TrajGRU model [20], which takes 5 frame inputs and outputs 20 frames, our input here was 10 frame inputs, and the output was also 10 frames. By increasing the number of input echoes, we allow the model to obtain more historical radar echo information and more spatiotemporal transformation can be extracted.

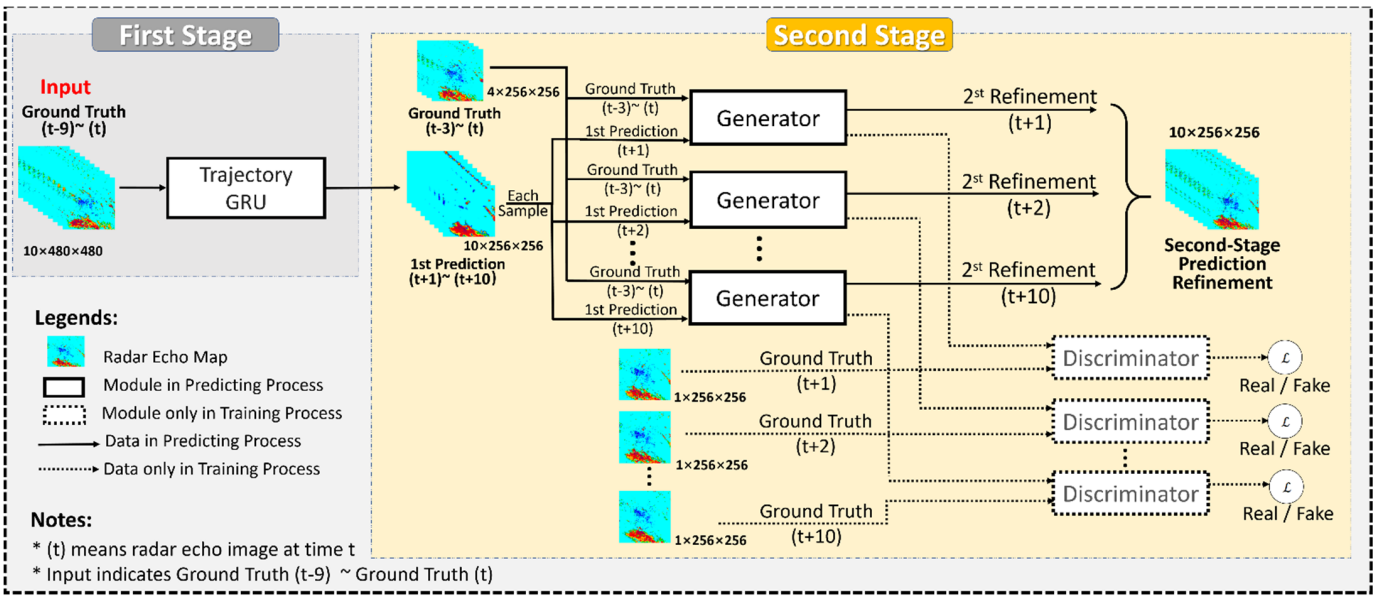

Figure 2. The structural framework and process details of the proposed model, consisting of a pre-prediction stage, the first stage on the left with gray background, and a refinement stage, the second stage with the yellow background on the right. Detailed structure parameters can be found in Tables 1 and 2.

Table 1. Encoder part in the first-stage prediction model.

Name	Kernel	Stride	Padding	L	Channels Input/Output
Conv1	7×7	5×5	1×1	-	1/8
TrajGRU1	3×3	1×1	1×1	13	8/64
Conv2	5×5	3×3	1×1	-	64/192
TrajGRU2	3×3	1×1	1×1	13	192/192
Conv3	3×3	2×2	1×1	-	192/192
TrajGRU3	3×3	1×1	1×1	9	192/192

Table 2. Forecaster part in the first stage prediction model.

Name	Kernel	Stride	Padding	L	Channels Input/Output
TrajGRU1	3×3	1×1	1×1	13	192/192
DeConv1	4×4	2×2	1×1	-	192/192
TrajGRU2	3×3	1×1	1×1	13	192/192
DeConv2	5×5	3×3	1×1	-	192/64
TrajGRU3	3×3	1×1	1×1	9	64/64
DeConv3	7×7	5×5	1×1	-	64/8
Conv4	1×1	1×1	0×0	-	8/1

The second stage is a deep residual attention-enhanced GAN model. Note that the input to our GAN is composed of 5 frames of radar echo maps, one predicted echo map by the first stage and the last 4 frames from the past 10 real echo maps. The output was refined 10 echo maps, in pursuit of higher prediction accuracy and more small-scale details. Details of the two stages are stated in the following subsection.

2.2.1. First Stage: Spatiotemporal Prediction Net

The purpose of the first stage is to use many and past true radar echo images to find the spatial-temporal variation regularity between the past radar echo sequences, and to make a rough prediction through this spatial-temporal regularity. The TrajGRU model [20] was employed in the first stage prediction. It can actively learn the location-variant structure for recurrent connections, since the local correlation structure of consecutive echo maps will be different for different spatial locations and timestamps, especially for motion patterns such as rotation and scaling. The main formula of TrajGRU is shown in Equation (2).

$$\begin{aligned} \mathcal{U}_t, \mathcal{V}_t &= \gamma(\mathcal{X}_t, \mathcal{H}_{t-1}) \\ Z_t &= \sigma\left(\mathcal{W}_{xz} * \mathcal{X}_t + \sum_{l=1}^{L} \mathcal{W}_{hz}^{l} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})\right) \\ \mathcal{R}_t &= \sigma\left(\mathcal{W}_{xr} * \mathcal{X}_t + \sum_{l=1}^{L} \mathcal{W}_{hr}^{l} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})\right) \\ \mathcal{H}'_t &= f\left(\mathcal{W}_{xh} * \mathcal{X}_t + \mathcal{R}_t \circ \left(\sum_{l=1}^{L} \mathcal{W}_{hh}^{l} * \text{warp}(\mathcal{H}_{t-1}, \mathcal{U}_{t,l}, \mathcal{V}_{t,l})\right)\right) \\ \mathcal{H}_t &= (1 - Z_t) \circ \mathcal{H}'_t + z_t \circ \mathcal{H}_{t-1} \end{aligned} \quad , \tag{2}$$

In this formula, L refers to the number of permitted links. $\mathcal{U}_t, \mathcal{V}_t \in \mathbb{R}^{L \times H \times W}$ are the flow fields used to store local connections, whose generating network is γ. $\mathcal{W}_{hi}^{l}$, $\mathcal{W}_{hf}^{l}$, $\mathcal{W}_{hc}^{l}$, $\mathcal{W}_{ho}^{l}$ indicate the weights for projecting the channels. '*' stands for convolution and '$\circ$' stands for the Hadamard product. Warp function generates location from $\mathcal{U}_t, \mathcal{V}_t$ through the bilinear sampling method [32,33]. Define $\mathcal{M} = \text{warp}(\mathcal{I}, \mathbf{U}, \mathbf{V})$ where $\mathcal{M}, \mathcal{I} \in \mathbb{R}^{C \times H \times W}$ and $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{H \times W}$, then:

$$\mathcal{M}_{c,i,j} = \sum_{m=1}^{H} \sum_{n=1}^{W} \mathcal{J}_{c,m,n} max\left(0,\ 1 - \left|i + \mathbf{V}_{i,j} - m\right|\right) max\left(0,\ 1 - \left|j + \mathbf{U}_{i,j} - n\right|\right) \tag{3}$$

It determines that TrajGRU is efficient in capturing the spatiotemporal correlations and achieved good performance for precipitation nowcasting. It is very suitable as the first stage of our model. The first stage prediction model is an encoder–forecaster structure. The detailed parameters of the encoder and forecaster are shown in Tables 1 and 2.

In the two tables, the two dimensions in kernel size, stride size, and padding size represent height and width. L stands for the number of state-to-state transition links. The number of input channel and output channel is also given in the tables. Since no major changes are made to the original TrajGRU but the number of input and output frames and parameters are modified, the first-stage pre-prediction model will not be introduced too much here. We focus on the deep residual attention enhanced GAN model in the second refinement stage.

2.2.2. Second Stage: Detail Refinement Stage

Although TrajGRU has achieved a high prediction accuracy [20,24,29], it can be seen from the naked eye that the predicted radar echo image is very blurred. As shown in Figure 3 and discussed in refs. [28,29], this is a common problem that appears in some ML-based models. Such results are usually related to two factors. (1) The loss function usually employs the MSE/MAE of the predicted images, which makes the model prediction results tend to be smooth and mean prediction [34,35]. Moreover, the predictability of radar echo is related to the echo scale, and small-scale high frequency echo details usually have low predictability [12,13,36–38]. To deal with this "blurry" problem, a U-Net shaped GAN with deep residual Attention block (UA-GAN) model was proposed in the second stage, the detail-refinement stage.

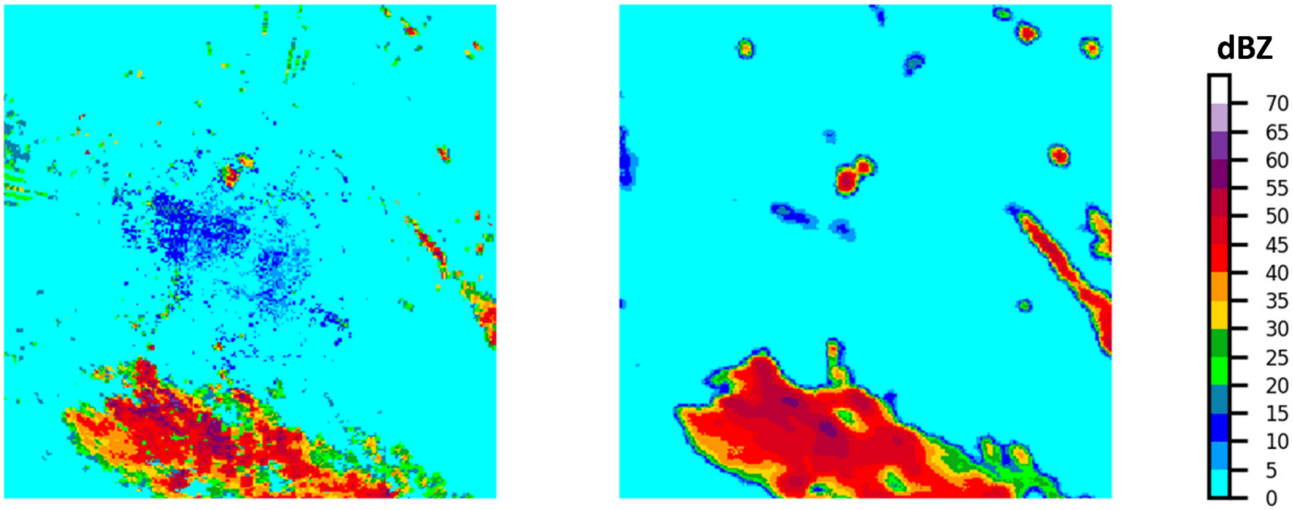

Figure 3. An example of TrajGRU prediction image (the first stage) and the corresponding ground truth image: (**a**) ground truth; (**b**) result by using the first stage prediction model.

General description. As shown in Figure 2, UA-GAN consists of a generator and a discriminator, and the discriminator is only used during model training. The generator employs an encoder–decoder model, which are described in detail in the following paragraphs starting with "Encoder of Generator" and "Decoder of Generator". The "2DSRAB" module is an important module we proposed and used in the "Decoder of Generator" part, so this module is illustrated in a separate paragraph starting with "2DSRAB in Decoder" before the description of the decoder part.

Generator. The structure of the proposed generator is illustrated in Figure 4, in which different colored rectangles represent different processing modules. We adopted the structure of U-Net to build the generator. It employs a encoder–decoder with global multi-scale skip connections to retain the multi-scale spatial-temporal feature maps in the

encoder, and integrates local residual learning in our proposed attention model (2DSRAB) of decoder part to adaptively rescale multi-scale spatiotemporal features for guiding a high-to-low level residual prediction generation. The input form of the generator is also the key for the better performance of UA-GAN network. As for the input in Figure 4, it is made up of five frames, including one frame which is the output of the first stage, and four ground truth frames, which are also part of the input sequence of the first stage. One frame is the first stage's pre-prediction result which our generator will take in to carry out the optimization and refinement, and these four ground truth images are always the last four frames input to the first stage. In other words, in the second stage, each frame is actually optimized and refined individually by the generator. For example, the input of the first stage is the real image from time point $t-9$ to t, and the output is from time point $t+1$ to $t+10$. In the second stage, when we want to predict the echo image at time point $t+x$, the first frame input to the generator is the predicted image by the first stage at time point $t+x$, and the last four frames are real images in the past time point $t-3$ to t. By this input policy, the generator can not only enjoy the relatively accurate prediction results generated by TrajGRU in the first stage and focus on improving the small-scale details of the prediction maps, but also can take advantage of the original historical radar echo maps, from which the required spatial and temporal features can be captured to further improve the prediction accuracy. These five frames of radar echo maps will be sent to the generator as five different channels, that is, when the width and height of the input radar echo maps are 256, the actual size of the tensor fed to the generator is 5 * 256 * 256. Thus, in the generator we actually utilized 2D convolution instead of 3D convolution, which reduces the overall parameter amount of the model.

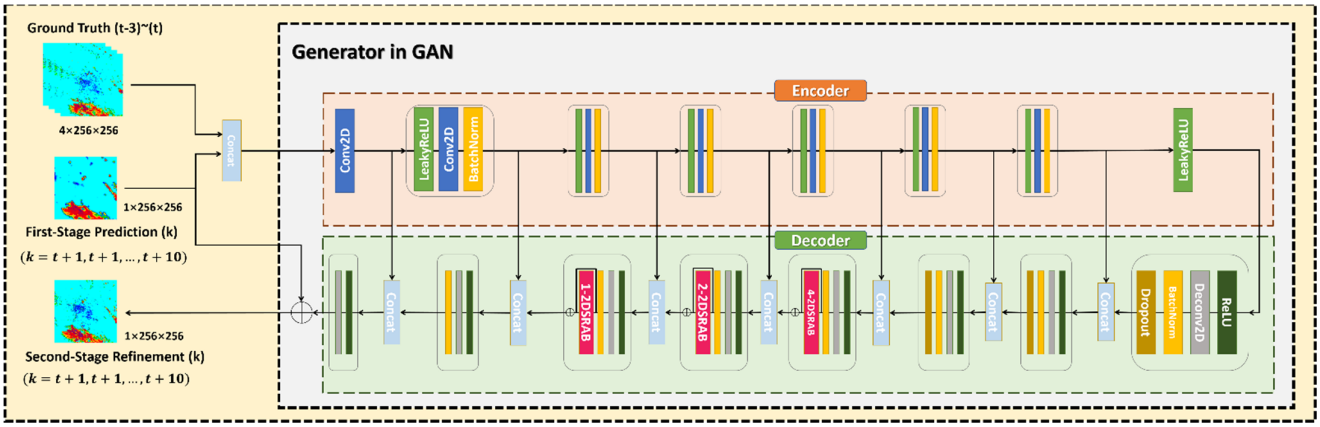

Figure 4. Structure of the generator in UA-GAN. It is a U-Net structure, also could be divided into two parts: encoder and decoder. Five echo frames including one predicted echo maps from first-stage prediction network, are input into the generator and one refined echo frame is output. Only one echo frame is optimized and refined at a time by the generator in UA-GAN.

Encoder of the Generator. In the encoder part, we continuously downsampled the input radar echo images using an eight-layer convolutional network, together with BatchNorm and LeakyReLU modules. In Figure 4, Conv2D modules are represented using dark blue rectangles, BatchNorm modules using light yellow rectangles, and LeakyReLU modules using grass green. The first layer has only one Conv2D module. The second to seventh layers have the same structure, which consists of LeakyReLU, Conv2D and BatchNorm in order, and the last layer has only one LeakyReLU module. The detail of parameters in the encoder part are listed in Table 3. The feature maps output by each layer were copied and fused by skip connections to the decoder side.

Table 3. Encoder part of generator used in UA-GAN (Stage 2).

Name	Kernel	Stride	Padding	Channels Input/Output
Conv1	4×4	2×2	1×1	5/64
Conv2	4×4	2×2	1×1	64/128
Conv3	4×4	2×2	1×1	128/256
Conv4	4×4	2×2	1×1	256/528
Conv5	4×4	2×2	1×1	528/528
Conv6	4×4	2×2	1×1	528/528
Conv7	4×4	2×2	1×1	528/528
Conv8	4×4	2×2	1×1	528/528

2DSRAB in the Decoder. In the decoder part, we designed a new spatiotemporal attention mechanism, two-dimensional stacked residual attention block (2DSRAB), and applied it to the decoder side of the UA-GAN network generator. Instead of treating all features equally, a spatiotemporal attention module was proposed for temporal-wise and spatial-wise weightings, which strengthens the discriminative learning ability and the representational power of deep networks. The structure of this spatiotemporal attention block is shown in Figure 5. On the left of Figure 5 is the basic attention block, in which the light purple rectangle represents the average pooling layer, and the light green rectangle represents the max pooling layer. The input of the basic attention block has the size of $B * C * W * H$, where B stands for batch size, C stands for channel, and W and H stand for width and height. Thus, in the basic attention block, we first calculated the average and maximum value on an image scale, in other words, on a spatial scale. Then, we compute the average and maximum values on a channel scale. Since five radar echo images were input into the generator as five different channels, the attention on the channel scale just means on the temporal scale. Max pooling filters out more recognizable features, while average pooling retains more common features. Both average and max-pooled features are simultaneously used to greatly improve the representation power of networks. The short-cut eases the flow of information. In this basic attention module, we initially realized the construction of a spatiotemporal attention mechanism.

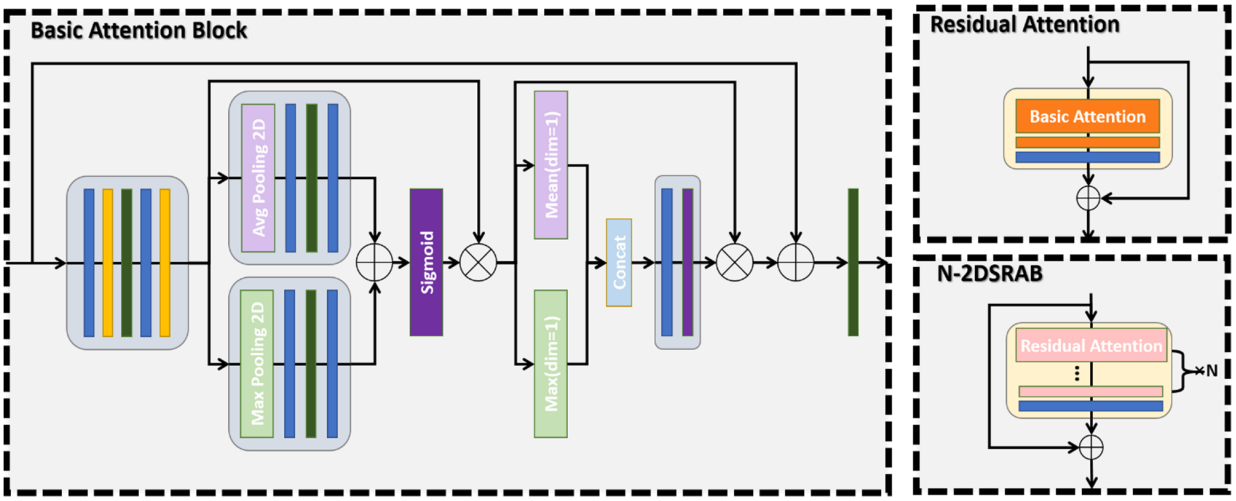

Figure 5. The structure of the proposed attention module 2DSRAB and the structure of its main components. On the left is the basic attention block, and on the upper right is the residual attention block. On the lower right is the 2DSRAB block consisting of N residual attention blocks.

Using this basic attention block, we could build a residual attention block by stacking two basic attention blocks, one Conv2D module, and local skip connection, displayed on the upper right of Figure 5, with the basic attention block colored orange. This is called

residual attention block. The lower right of the Figure 5 exhibits the 2DSRAB module, consisting of the N residual attention module (colored pink), one Conv2D module, and local skip connection. Local skip connection can stabilize the training of very deep networks and ease the flow of spatiotemporal information. The 2DSRAB module is quite beneficial to form very deep trainable prediction and refinement networks, which adaptively rescales and blends multi-scale spatiotemporal features.

Decoder of the Generator. The decoder also consists of an eight-layer convolutional network, together with a BatchNorm module and ReLU module. In Figure 4, the ReLU module is colored dark green, the Deconv2D module is colored gray, and the Dropout module colored light brown. The first to the seventh layers consist of ReLU, DeConv2D, and BatchNorm in turn. Additionally, the eighth layer contains only ReLU and DeConv2D. At the same time, in the first, second, and third layers, a dropout module was added to randomly discard some neurons to reduce the joint effect of feature extraction. This also improves the adaptive ability of individual feature extractors, achieving the goal of improving the generalization ability of the network. To build a better decoder, we added the 2DSRAB module in three places of the decoder, which were the fourth, fifth, and sixth layers respectively. In Figure 4, the 2DSRAB module is represented by a rose-red rectangle, and the N in N-2DSRAB represents the depth of this attention module. It is worth mentioning that, before the decoder actually outputs the final results, one of the input frames, the prediction echo frame provided by the first stage, was also added to the final output result, thereby retaining the important original information initially input into the generator. The details of the parameters in the decoder are shown in Table 4.

Table 4. Decoder part of the generator used in UA-GAN (Stage 2).

Name	Kernel	Stride	Padding	Depth	Input/Output Channels
DeConv1	4×4	2×2	1×1	-	528/528
DeConv2	4×4	2×2	1×1	-	1024/528
DeConv3	4×4	2×2	1×1	-	1024/528
DeConv4	4×4	2×2	1×1	-	1024/528
2DSRAB1	1×1	1×1	0×0	4	528/528
DeConv5	4×4	2×2	1×1	-	1024/256
2DSRAB2	1×1	1×1	0×0	2	256/256
DeConv6	4×4	2×2	1×1	-	528/128
2DSRAB3	1×1	1×1	0×0	1	128/128
DeConv7	4×4	2×2	1×1	-	256/64
DeConv8	4×4	2×2	1×1	-	128/1

Discriminator. The training goal of the discriminator of the UA-GAN model was used to judge the real radar echo map as true and the prediction map generated by the generator as false. By adopting a five-layer convolutional structure, the discriminator could obtain good discriminative performance. The structure of the discriminator is presented in Figure 6, and the module color is the same as in Figures 4 and 5.

The first layer of the discriminator was composed of Conv2D and LeakyReLU, and the last layer was composed of Conv2D and Sigmoid. The remaining layers all consisted of Conv2D, BatchNorm, and LeakyReLU. Both the second-stage refinement map and ground-truth image were fed into the discriminator. The detailed parameters of the discriminator are listed in Table 5.

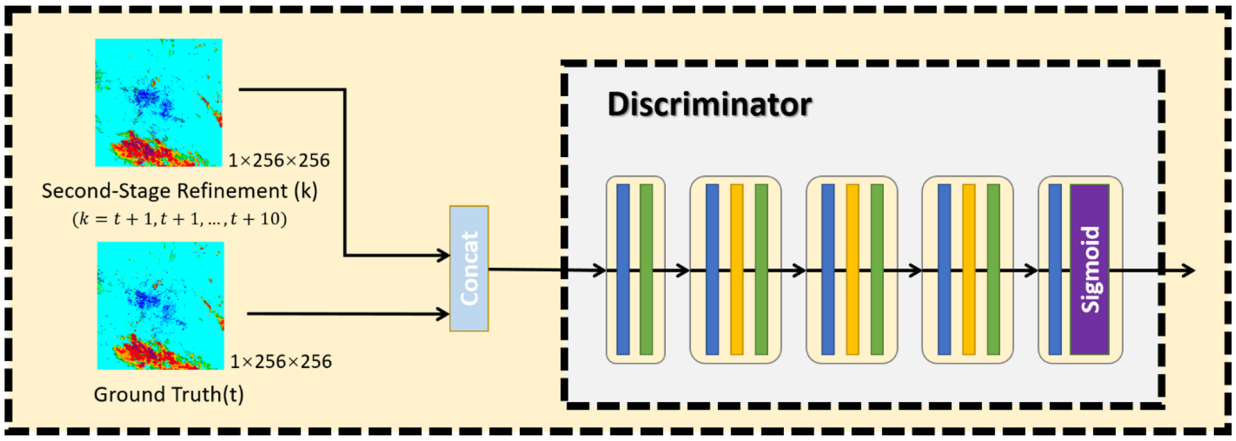

Figure 6. Structure of the discriminator.

Table 5. The structure of the discriminator used in UA-GAN (Stage 2).

Name	Kernel	Stride	Padding	Input/Output Channels
Conv1	4×4	2×2	1×1	2/64
Conv2	4×4	2×2	1×1	64/128
Conv3	4×4	2×2	1×1	128/256
Conv4	4×4	1×1	1×1	256/528
Conv5	4×4	1×1	1×1	528/1

2.3. Loss Function

Two different loss functions were adopted respectively for the first-stage and second-stage models.

2.3.1. Loss Function of First Stage

The first stage mainly employed the TrajGRU model. Shi et al. [20] designed a very effective loss function, weighted MSE, and weighted MAE for the TrajGRU model, which can improve the prediction accuracy of heavy rain. In this paper, we used a similar loss function but changed the threshold and the value of each weight. By assigning different weights to the pixels of different precipitation intensities, the greater the precipitation intensity, the greater the weight to enhance the sensitivity of the model for heavy rain prediction [20,29,30]. The weights are defined in (4), where dBZ refers to the radar echo value.

$$w(z) = \begin{cases} 1, & z < 20 \text{ dBZ} \\ 2, & 20 \text{ dBZ} \le z < 35 \text{ dBZ} \\ 6, & 35 \text{ dBZ} \le z < 45 \text{ dBZ} \\ 60, & z \ge 45 \text{ dBZ} \end{cases}, \tag{4}$$

The loss function of the first stage is stated as (5):

$$Loss_{stage1} = \frac{1}{T} \sum_{t=1}^{T} \sum_{i,j} \left(w_{t,i,j} \left(\left(Y_{t,i,j} - \hat{X}_{t,i,j} \right)^2 + \left| Y_{t,i,j} - \hat{X}_{t,i,j} \right| \right) \right), \tag{5}$$

where T stands for the number of radar echo images in the sequence and $w_{t,i,j}$ is the weights of $(i,\ j)$ pixel on the t-th image. $Y_{t,i,j}$ and $\hat{X}_{t,i,j}$ are the values of $(i,\ j)$ pixels on the t-th ground truth and predictive images. In this way, we used the sum of the weighted MSE and weighted MAE together as the loss function of the first stage.

2.3.2. Loss Function of Second Stage

In a GAN model, the generator and the discriminator have different tasks. The generator's job is to generate realistic images to fool the discriminator, while the discriminator's job is to tell if the image is from the generator or a real image. They achieve better performance by competing with each other. Thus, the goal of the conditional GAN can be described as (6):

$$\mathcal{L}_{CGAN}(G, D) = \mathbb{E}_{x,z}[\log D(z, x)] + \mathbb{E}_z[\log(1 - D(z, G(z)))], \tag{6}$$

In which the generator (G) intends to minimize, while the discriminator (D) tries to maximize. In Formula (6), x represents the ground-truth echo images and z represents other conditional information to generate a specific prediction image. Of course, only using the previous loss function of conditional GAN will only allow the generator to imitate the small-scale features of real images, but not fine-tune the generator on the pixel scale. Therefore, on the basis of the basic loss function, we added MSE and MAE functions to enable the generator to have better echo intensity prediction accuracy at the pixel scale, which is shown in the following Equation (7).

$$\begin{aligned} \mathcal{L}oss_{stage2} = \omega_1(\mathbb{E}_{x,z}[\log D(z, x)] - \mathbb{E}_z[\log(1 - D(z, G(z)))] \\ +\omega_2(x - G(z))^2 + \omega_3|x - G(z)| \end{aligned} \tag{7}$$

The designed loss function consisted of three parts: basic GAN loss, MSE, and MAE, where $\omega_1, \omega_2, \omega_3$ denote the weights of the three parts, respectively. To make the three loss parts have similar orders of magnitude, here we set $\omega_1 = 1$, $\omega_2 = \omega_3 = 100$.

3. Experiments

3.1. Radar Echo Image Dataset

HKO-7 dataset. The radar echo maps dataset used in this paper were from the famous and public dataset HKO-7 [20], provided by Hong Kong Observatory (HKO), which contains 7 years of radar echo data from 2009 to 2015. The original size of each radar CAPPI reflectivity image was 480 × 480 pixels, taken from an altitude of 2 km every 6 min, which means it can collect 240 frames of radar echo images in a single day. The radar echo map in this dataset covers Hong Kong and its surrounding region with an area of about 512 km × 512 km. With the formula $pixel = 255 \times \frac{dBZ}{70}$, we can convert the radar echo intensity to pixel value and then limit the range within (0, 255). An example of radar echo image is shown in Figure 1.

Dataset filtering and segmentation. Since precipitation does not occur every day, the radar echo image at the time of no precipitation is not meaningful for developing the network, so before dividing the training set and the test set, it is necessary to filter out some data with precipitation and discard the part without precipitation. To preserve the integrity of the precipitation process, it is necessary to retain the whole process of precipitation generation and disappearance. Finally, we divided the filtered precipitation data into training set and test set. There were 812 days in the training set and 131 days in the test set. In the second stage, in order to shorten the training time, we randomly selected a part from the training and test set of the first stage as the training and test set of the second stage. In the second stage, 80,000 echo frames were selected for the training set, with 724 days involved and 8000 echo frames for the test set, 113 days. The details of the two datasets used in the two stages are demonstrated in Tables 6 and 7.

Table 6. The dataset used in the first stage.

	Training Set	**Test Set**
Years	2009–2014	2015
Days	812	131
Frames	192,168	31,350

Table 7. The dataset used in the second stage.

	Training Set	Test Set
Years	2009–2014	2015
Days	724	113
Frames	80,000	8000

3.2. Evaluation

For achieving "no blurry" and high-accuracy echo prediction, we employed both image quality evaluation and prediction accuracy indexes to evaluate the performance of the proposed network at the same time.

Image quality evaluation index. To compare echo images generated by our model with the real images considering image structure and clarity, three widely-used metrics, RMSE, SSIM, and sharpness, were employed [27,29,30,39,40]. Their definitions are given in Equations (8)–(10).

RMSE is a widely-used intuitive error evaluation index [19,20,41]. The smaller the RMSE is, the smaller the error is between the two figures. In Equation (8), x stands for the ground-truth echo image while $\widetilde{x}$ stands for the predictive echo image. W and H are the width and height of the image, respectively.

$$\mathrm{RMSE}(x,\widetilde{x}) = \sqrt{\frac{\sum_i \sum_j \left(x_{i,j} - \widetilde{x}_{i,j}\right)^2}{HW}} \tag{8}$$

SSIM (structural similarity) [42], with its formulation given by Equation (9), is also a famous image quality evaluation metric [30,40]. It estimates the structural similarity of two images by calculating brightness, variance, and covariance. In Equation (9), μ is the mean value and σ is the covariance. C_1, C_2 are constants used to avoid division by 0. The larger SSIM is, the more similar the two images are.

$$\mathrm{SSIM}(x,\widetilde{x}) = \frac{(2\mu_x\mu_{\widetilde{x}} + C_1)(2\sigma_{x\widetilde{x}} + C_2)}{\left(\mu_x^2 + \mu_{\widetilde{x}}^2 + C_1\right)\left(\sigma_x^2 + \sigma_{\widetilde{x}}^2 + C_2\right)} \tag{9}$$

Sharpness [39] of an image can be estimated by calculating the gradient difference between the two images. Higher sharpness means a sharper image with more details. Equation (10) defines sharpness, where $\nabla_i x = \left|x_{i,j} - x_{i-1,j}\right|$, $\nabla_j x = \left|x_{i,j} - x_{i,j-1}\right|$.

$$\mathrm{Sharpness}(x,\widetilde{x}) = 10\log_{10}\frac{max_{\widetilde{x}}^2}{\frac{1}{WH}\left(\sum_i \sum_j \left|\left(\nabla_i x + \nabla_j x\right) - \left(\nabla_i \widetilde{x} + \nabla_j \widetilde{x}\right)\right|\right)} \tag{10}$$

Apart from image quality indexes for evaluating the details in predicted echo maps, employing some metrics for focusing on prediction accuracy is also quite important.

Precipitation nowcasting accuracy index. As shown in refs. [19,20,29,30], three commonly used precipitation nowcasting metrics, namely, critical success index (CSI), Heidke Skill Score (HSS), and false alarm rate (FAR) were also employed to evaluate the prediction accuracy in this work. To calculate them, we first binarized the image 0–1 using a specific threshold. This means that pixel value larger than threshold was set to 1, while the smaller was set to 0. As shown in Table 8, TP, TN, FP, and FN represent different results.

Table 8. Confusion matrix used in calculating CSI, HSS, and FAR.

	Prediction = 1	**Prediction = 0**
Truth = 1	TP	FN
Truth = 0	FP	TN

Then, we can calculate CSI, HSS, and FAR score using the following Equation (11):

$$\begin{aligned} \text{CSI} &= \frac{\text{TP}}{\text{TP+FP+FN}} \\ \text{HSS} &= \frac{2\times(\text{TP}\times\text{TN}-\text{FP}\times\text{FN})}{(\text{TP+FN})(\text{FN+TN})+(\text{TP+FP})(\text{FP+FN})} \\ \text{FAR} &= \frac{\text{FP}}{\text{TP+FP}} \end{aligned}, \tag{11}$$

3.3. Results

For the sake of discussion, we used UA-GAN to refer to our entire two-stage model. In the experiments, we compared our two-stage model with one optical-flow based model: ROVER [16], four well-known deep-learning models: ConvLSTM [19], ConvGRU, TrajGRU [20], which is also the first-stage result of our model, and PredRNN++ [24] on the image quality evaluation indexes RMSE, SSIM, sharpness, and on the forecasting evaluation indexes CSI, HSS, and FAR. Meanwhile, to demonstrate the importance of our proposed attention mechanism, we also conducted an ablation experiment to delete all the 2DSRAB in the second stage, which is be referred to as UA-GAN (without attention).

Experiment analysis. In our experiments, we used the past 10 frames of radar echo images to predict the 10 future echo frames, that is, to predict the future one-hour radar echo images using the past one hour.

First, three image quality evaluation indexes with different methods were compared as shown in Table 9. In the table, '↓' means the smaller the better, and '↑' means the larger the better. We mark the best result within a specific metric with bold face.

Table 9. Image quality comparisons of radar echo prediction.

Model	RMSE ↓	SSIM ↑	Sharpness ↑
TrajGRU (first stage) [20]	0.132	0.570	33.674
PredRNN++ [24]	0.142	0.556	39.517
ConvGRU [20]	0.137	0.568	26.877
ConvLSTM [19]	0.139	0.565	30.464
ROVER [16]	0.152	0.404	22.173
UA-GAN (without attention)	0.128	0.514	**72.483**
UA-GAN (Ours)	**0.099**	**0.585**	63.395

From Table 9, all the deep-learning models outperformed the optical flow-based ROVER algorithm [16]. Among the deep-learning models, our UA-GAN model achieved the best results in RSME and SSIM. It was second only to UA-GAN without an attention module in sharpness. It improved the RMSE score of UA-GAN (without attention) (second best) from 0.128 to 0.099 (decrease of 22.6%), the SSIM score of TrajGRU (first stage, second best) from 0.570 to 0.585 (increase of 2.63%), and sharpness score of PredRNN++ (second best) from 39.517 to 63.395 (increase of 60.4%). It is clear that the proposed UA-GAN model is beneficial to generate sharper echo map prediction with more small-scale details.

In addition, to provide an all-round evaluation of the algorithms' prediction accuracy performance, we also present the forecasting evaluation scores for multiple thresholds (25 dBZ, 35 dBZ, 40 dBZ, 45 dBZ, and 50 dBZ) that correspond to different rainfall levels [19,20,29,30]. The test results are shown in Tables 10–12. The best results are also marked with bold face.

Table 10. Scores of **CSI(↑)** at echo thresholds = 25, 35, 40, 45, and 50 dBZ.

Model	25 dBZ	35 dBZ	40 dBZ	45 dBZ	50 dBZ
TrajGRU (first stage) [20]	0.331	0.261	0.208	0.158	0.121
PredRNN++ [24]	0.342	0.271	0.217	0.161	0.114
ConvGRU [20]	0.326	0.257	0.207	0.160	0.111
ConvLSTM [19]	0.323	0.255	0.203	0.155	0.113
ROVER [16]	0.309	0.231	0.169	0.115	0.078
UA-GAN (without attention)	0.328	0.257	0.198	0.141	0.097
UA-GAN (Ours)	**0.381**	**0.313**	**0.256**	**0.193**	**0.138**

Table 11. Scores of **HSS(↑)** at echo thresholds = 25, 35, 40, 45, and 50 dBZ.

Model	25 dBZ	35 dBZ	40 dBZ	45 dBZ	50 dBZ
TrajGRU (first stage) [20]	0.406	0.345	0.290	0.240	0.176
PredRNN++ [24]	0.406	0.346	0.291	0.237	0.189
ConvGRU [20]	0.401	0.341	0.290	0.239	0.172
ConvLSTM [19]	0.396	0.337	0.284	0.231	0.174
ROVER [16]	0.370	0.297	0.225	0.158	0.107
UA-GAN (without attention)	0.404	0.340	0.277	0.208	0.150
UA-GAN (Ours)	**0.473**	**0.409**	**0.349**	**0.277**	**0.205**

Table 12. Scores of **FAR(↓)** at echo thresholds = 25, 35, 40, 45, and 50 dBZ.

Model	25 dBZ	35 dBZ	40 dBZ	45 dBZ	50 dBZ
TrajGRU (first stage) [20]	0.564	0.636	0.679	0.693	0.531
PredRNN++ [24]	0.604	0.640	0.677	0.705	0.631
ConvGRU [20]	0.587	0.622	0.654	0.664	**0.511**
ConvLSTM [19]	0.581	0.628	0.658	0.677	0.512
ROVER [16]	0.585	0.612	0.749	0.774	0.745
UA-GAN (without attention)	0.566	0.632	0.704	0.778	0.728
UA-GAN (Ours)	**0.466**	**0.514**	**0.562**	**0.605**	0.627

Judging from the performance of UA-GAN in important CSI and HSS scores, our method achieved a significant accuracy improvement compared with optical flow methods and other deep-learning methods, especially at high echo intensity thresholdds (heavy rainfall). Compared with the deep-learning models, the optical flow-based ROVER method had a relatively poor prediction performance, and presented a gap in the evaluation indices. In deep-learning approaches, it can be seen that the proposed UA-GAN achieved better nowcasting scores than other four methods for almost all three prediction accuracy metrics, and, particularly importantly, had an obvious improvement at the 45 dBZ and 50 dBZ (heavy rainfall) thresholds. At the 50 dBZ threshold, the CSI under our UA-GAN was over 0.017 higher than that under the TrajGRU method (increase of about 14%), and also 0.024 higher than that under the PredRNN++ model (increase of nearly 21%). Moreover, the HSS was also much improved by about 16.5% relative to that under the TrajGRU method, of over 8.5% relative to that under the PredRNN++ method. It is clear that our proposed method had better prediction accuracy even for heavy rainfall, which usually is a more difficult task. In addition, to verify the effectiveness of the proposed attention mechanism (2DSRAB) in the generator, an ablation experiment was conducted and the test results under UA-GAN (without attention) were also supplied. It was demonstrated that the proposed model with the 2DSRAB attention mechanism could remarkably enhance the three accuracy metrics relative to that without the attention mechanism.

Figure 7 further shows the prediction accuracy scores for the 6 to 60 min lead times. They are the average scores for the whole test dataset. The important CSI and HSS scores had similar trends, and they all decreased with time. From this figure, even for different lead times and different rainfall intensities, the proposed UA-GAN also outperformed other models whose nowcasting performances degraded faster as the lead time increased.

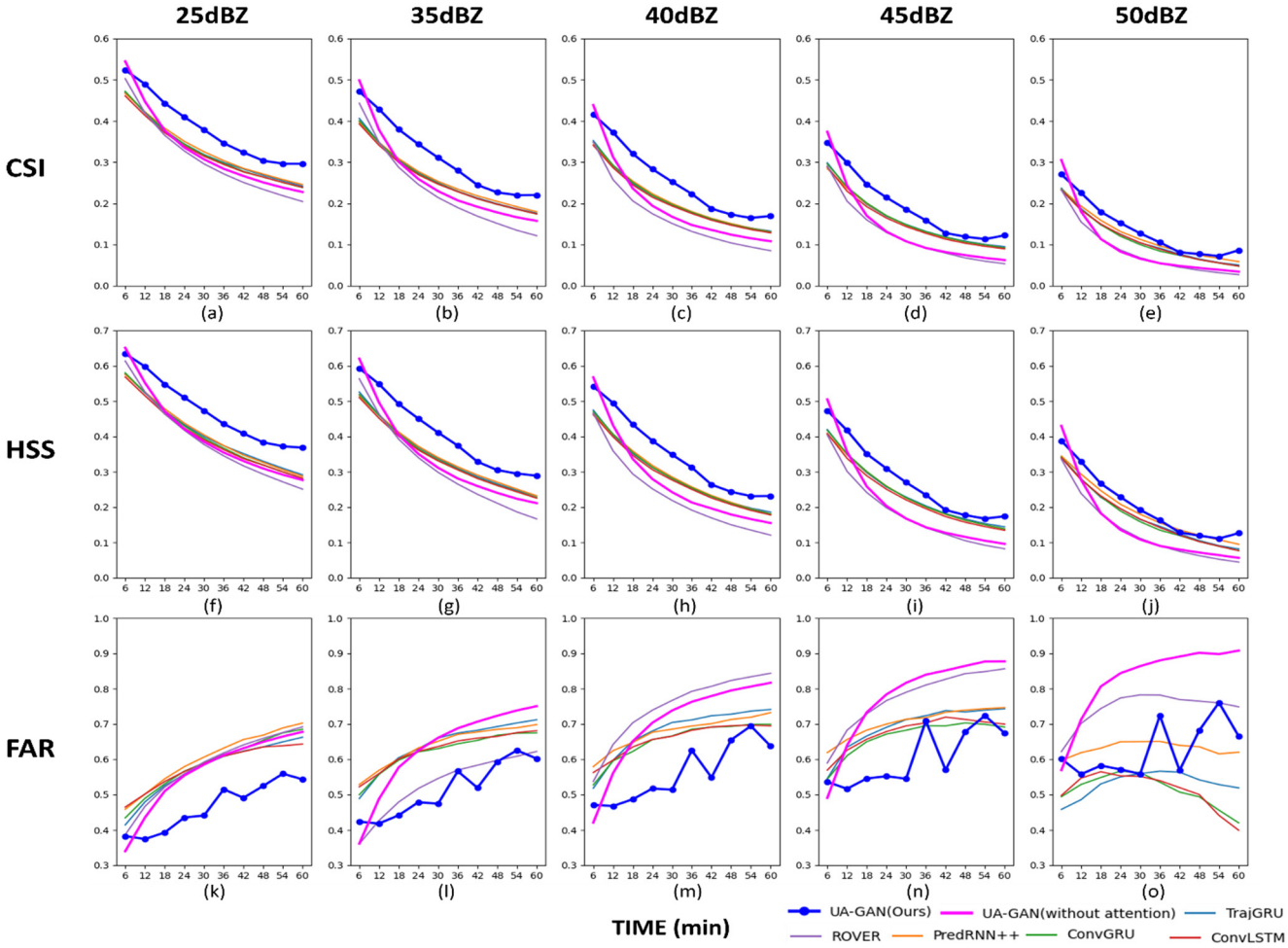

Figure 7. Subfigures (**a**–**o**) present the precipitation nowcasting evaluation scores (CSI, HSS, and FAR) for different lead times at different thresholds, respectively.

A visualization example of the radar echo maps predicted by different models is shown in Figure 8. The first line is the 10 radar echo frames in the past, and the second line is the ground truth of the future 10 echo frames. There were three major echo regions at bottom left, middle, and top right of the echo images, respectively. From the bottom left and top right corners of the images, it is clear that the major echo regions moved towards to the east. Moreover, the intensity of middle echo region was gradually weakening.

At the first moment, nearly all models predicted accurately. Then, significant differences were observed as the lead time increased. The forecasting echo scale in ROVER (optical flow) was gradually reduced and the echo intensity change was ignored. As time goes by, PredRNN++ and TrajGRU tended to exaggerate the forecasting scale and each region stuck to each other. The major echo region's intensity by the RNN-based methods (PredRNN++, TrajGRU, ConvGRU, and ConvLSTM) tended to be overestimated. Importantly, the small-scale details were gradually lost in extrapolations. The intensity distribution inside the echo map could not be forecasted correctly and the boundaries became smooth. By comparison, our UA-GAN provided the best performance, where the predicted echo frames were more realistic, and the details and distribution of each part were also better preserved. As the lead time increased, our prediction quality did not experience a significant change, while the echo intensity and position were more consistent with the real images than other methods.

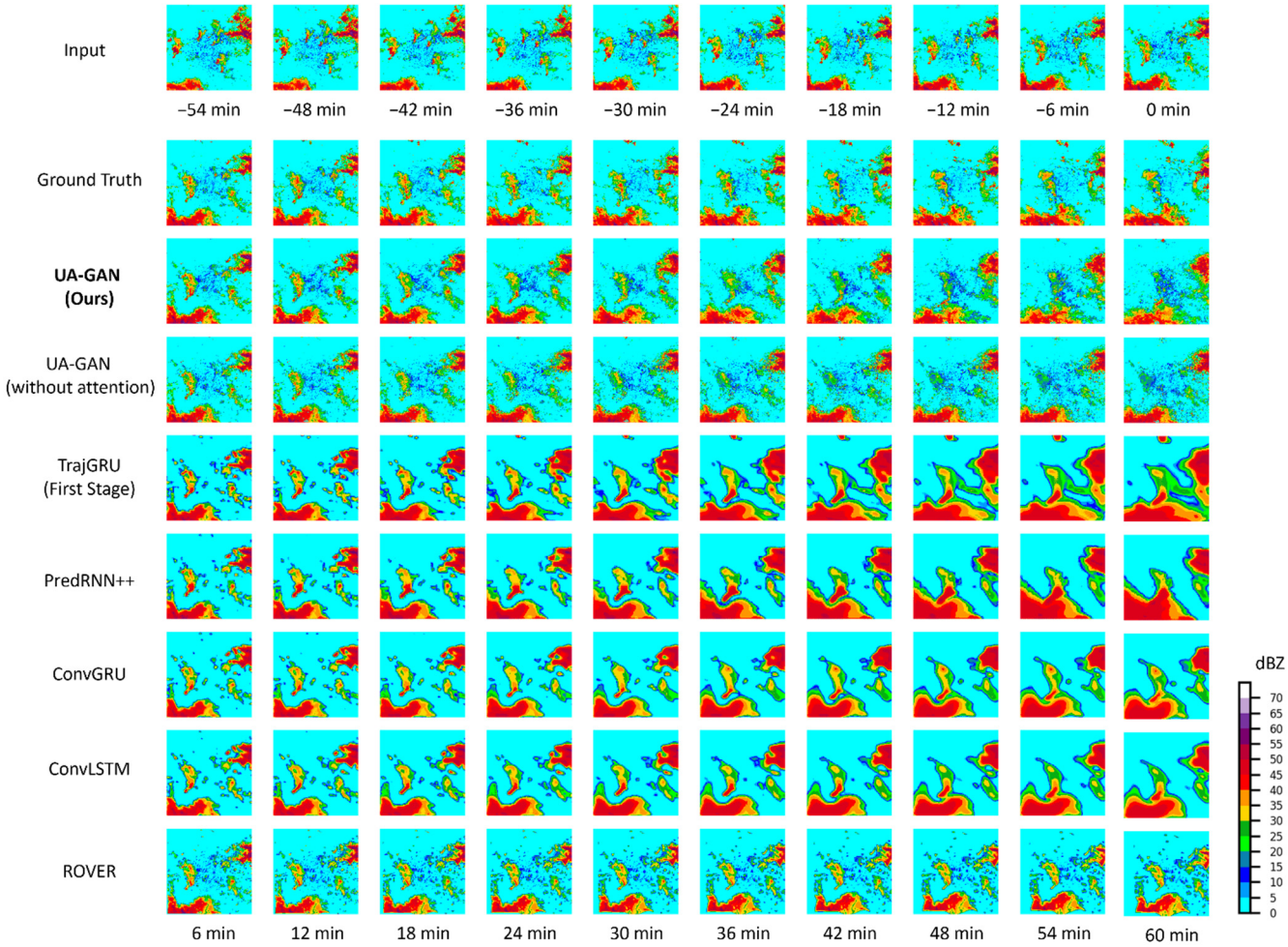

Figure 8. Visualization example of the predicted echo maps using different methods. The input echo sequence is 24 May 2015, from 02:48 to 03:48, and the predicted sequence is from 03:54 to 04:54.

Considering the attention mechanism, comparing with the proposed method without attention, it can be seen that our UA-GAN had more precise small-scale details at the bottom-left corners and could better predict echo weakened in the middle regions of the predicted echo maps

4. Conclusions

In this paper, a two-stage network was proposed to achieve the goal of radar echo extrapolation, whose first stage was a well-trained 10-in–10-out TrajGRU model, and the second stage was UA-GAN, a deep residual attention enhanced GAN model. Using TrajGRU to obtain spatiotemporal movement information of rain field, the first stage was able to produce a preliminary forecast. As for the second stage, we proposed a Unet with spatiotemporal attention generator in GAN. We input four past frames of echo images and one frame generated in the first stage into the generator, so that the generator could capture certain historical spatiotemporal features. Additionally, we designed a new attention block, 2DSRAB, in the decoder of the generator, which integrates the global residual learning and local deep residual spatiotemporal attention to adaptively rescale the multiscale features, and enables the generator to produce more accurate and more detailed prediction images. Experiments showed that our network outperformed traditional optical flow method and some well-known deep-learning methods in both image quality (RMSE, SSIM, and sharpness) and prediction accuracy metrics (CSI, HSS, and FAR). From Tables 9–12 and Figure 7, it is clear that our model can provide more accurate prediction echo images with

more small-scale details. Moreover, the proposed attention mechanism in the generator also further improved the prediction accuracy.

In the future, we will continue to work on new models to improve the prediction accuracy as well as enhance the small-scale details of the radar echo images. Moreover, we hope to introduce environmental field information and satellite products into the extrapolation model to improve the prediction of radar echo and further increase the lead time of radar extrapolation. We will also try to build an operational nowcasting system using the proposed algorithm.

Author Contributions: Conceptualization, L.X.; data curation, X.C. and T.Z.; funding acquisition, X.C.; methodology, L.X. and D.N.; project administration, D.N. and X.C.; software, D.N., L.X. and Y.L.; validation, P.C. and Y.L.; visualization, L.X.; writing—original draft, L.X.; writing—review and editing, D.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2019YFE0110100, 2018YFC1506905), Natural Science Foundation of Jiangsu Province of China (No. BK20202006), Zhishan Youth Scholar Program of Southeast University, the Key R&D Program of Jiangsu Province (No. BE2019052, BE2017076).

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Yichao Cao, Zengliang Zang and Chao Chen for helpful discussions and suggestions.

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

References

1. Gneiting, T.; Raftery, A.E. Weather Forecasting with Ensemble Methods. *Science* **2005**, *310*, 248–249. [CrossRef] [PubMed]
2. Schmid, F.; Wang, Y.; Harou, A. Nowcasting Guidelines—A Summary. In *WMO—No. 1198*; World Meteorological Organization: Geneva, Switzerland, 2017; Chapter 5.
3. Zhang, Y.; Guo, T.; Li, S.; Zhu, L.; Liu, H.; Yan, P. Considerations on Reduction of Main Agricultural Natural Disasters in Henan Province, Taking July 20th Flood in Henan Province as a Case. *Manag. Agric. Sci. Technol.* **2021**, *40*, 10–13+68. [CrossRef]
4. Xiao, X.; Sun, J.; Qie, X.; Ying, Z.; Ji, L.; Chen, M.; Zhang, L. Lightning Data Assimilation Scheme in a 4DVAR System and Its Impact on Very Short-Term Convective Forecasting. *Mon. Weather Rev.* **2021**, *149*, 353–373. [CrossRef]
5. Ayzel, G.; Heistermann, M.; Winterrath, T. Optical Flow Models as an Open Benchmark for Radar-Based Precipitation Nowcasting (Rainymotion v0.1). *Geosci. Model Dev.* **2019**, *12*, 1387–1402. [CrossRef]
6. Tolstykh, M.; Frolov, A. Some Current Problems in Numerical Weather Prediction. *Izv. Atmos. Ocean. Phys.* **2005**, *41*, 285–295.
7. 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]
8. Marshall, J.S.; Palmer, W. The distribution of raindrops with size. *J. Meteor.* **1948**, *5*, 165–166. [CrossRef]
9. Radhakrishnan, C.; Chandrasekar, V. CASA Prediction System over Dallas–Fort Worth Urban Network: Blending of Nowcasting and High-Resolution Numerical Weather Prediction Model. *J. Atmos. Ocean. Technol.* **2019**, *37*, 211–228. [CrossRef]
10. Dixon, M.; Wiener, G. TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-Based Methodology. *J. Atmos. Ocean. Technol.* **1993**, *10*, 785–797. [CrossRef]
11. Rinehart, R.E.; Garvey, E.T. Three-Dimensional Storm Motion Detection by Conventional Weather Radar. *Nature* **1978**, *273*, 287–289. [CrossRef]
12. Germann, U.; Zawadzki, I. Scale-Dependence of the Predictability of Precipitation from Continental Radar Images. Part I: Description of the Methodology. *Mon. Weather Rev.* **2001**, *130*, 2859–2873. [CrossRef]
13. Li, L.; Schmid, W.; Joss, J. Nowcasting of Motion and Growth of Precipitation with Radar over a Complex Orography. *J. Appl. Meteorol.* **1995**, *34*, 1286–1300. [CrossRef]
14. Seed, A.W. A Dynamic and Spatial Scaling Approach to Advection Forecasting. *J. Appl. Meteorol.* **2003**, *42*, 381–388. [CrossRef]
15. Germann, U.; Zawadzki, I. Scale Dependence of the Predictability of Precipitation from Continental Radar Images. Part II: Probability Forecasts. *J. Appl. Meteorol.* **2004**, *43*, 74–89. [CrossRef]
16. Woo, W.C.; Wong, W.K. Operational Application of Optical Flow Techniques to Radar-Based Rainfall Nowcasting. *Atmosphere* **2017**, *8*, 48. [CrossRef]
17. Klein, B.; Wolf, L.; Afek, Y. A Dynamic Convolutional Layer for Short Rangeweather Prediction. In Proceedings of the Computer Vision & Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
18. Shi, E.; Qian, L.I.; Daquan, G.U.; Zhao, Z. Weather Radar Echo Extrapolation Method Based on Convolutional Neural Networks. *J. Comput. Appl.* **2018**, *38*, 661.

19. 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*; MIT Press: Cambridge, MA, USA, 2015. [CrossRef]
20. 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. *arXiv* **2017**, arXiv:1706.03458. [CrossRef]
21. Singh, S.; Sarkar, S.; Mitra, P. Leveraging Convolutions in Recurrent Neural Networks for Doppler Weather Radar Echo Prediction. In Proceedings of the International Symposium on Neural Networks, Muroran, Japan, 21–26 June 2017.
22. Feng, H.; Long, M.; Li, Y.; Feng, X.; Wang, J.; Center, N.M.; Software, S.O.; University, T. The Application of Recurrent Neural Network to Nowcasting. *J. Appl. Meteorol. Sci.* **2019**. [CrossRef]
23. 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 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 3–9 December 2017; pp. 879–888.
24. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Yu, P.S. PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning. In Proceedings of the 35th International Conference on Machine Learning, New Orleans, LA, USA, 2–7 February 2018; Volume PMLR 80, pp. 5123–5132.
25. Agrawal, S.; Barrington, L.; Bromberg, C.; Burge, J.; Gazen, C.; Hickey, J. Machine Learning for Precipitation Nowcasting from Radar Images. *arXiv* **2019**, arXiv:1912.12132. [CrossRef]
26. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
27. Liu, H.B.; Lee, I. MPL-GAN: Towards Realistic Meteorological Predictive Learning Using Conditional GAN. *IEEE Access* **2020**, *8*, 93179–93186. [CrossRef]
28. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S. Skillful Precipitation Nowcasting Using Deep Generative Models of Radar. *arXiv* **2021**, arXiv:2104.00954. [CrossRef]
29. Wang, C.; Wang, P.; Wang, P.; Xue, B.; Wang, D. Using Conditional Generative Adversarial 3D Convolutional Neural Network for Precise Radar Extrapolation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5735–5749. [CrossRef]
30. Niu, D.; Huang, J.; Zang, Z.; Xu, L.; Che, H.; Tang, Y. Two-Stage Spatiotemporal Context Refinement Network for Precipitation Nowcasting. *Remote Sens.* **2021**, *13*, 4285. [CrossRef]
31. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
32. Ilg, E.; Mayer, N.; Saikia, T.; Keuper, M.; Dosovitskiy, A.; Brox, T. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
33. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. *Spatial Transformer Networks*; MIT Press: Cambridge, MA, USA, 2015.
34. Lee, A.X.; Zhang, R.; Ebert, F.; Abbeel, P.; Finn, C.; Levine, S. Stochastic Adversarial Video Prediction. *arXiv* **2018**, arXiv:1804.01523. [CrossRef]
35. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Liu, G.; Catanzaro, B. Video-to-Video Synthesis. *arXiv* **2018**, arXiv:1808.06601.
36. Pulkkinen, S.; Nerini, D.; Pérez Hortal, A.A.; Velasco-Forero, C.; Seed, A.; Germann, U.; Foresti, L. Pysteps: An Open-Source Python Library for Probabilistic Precipitation Nowcasting (v1.0). *Geosci. Model Dev.* **2019**, *12*, 4185–4219. [CrossRef]
37. Radhakrishna, B.; Zawadzki, I.; Fabry, F. Predictability of Precipitation from Continental Radar Images. Part V: Growth and Decay. *J. Atmos. Sci.* **2012**, *69*, 3336–3349. [CrossRef]
38. Turner, B.J.; Zawadzki, I.; Germann, U. Predictability of Precipitation from Continental Radar Images. Part III: Operational Nowcasting Implementation (MAPLE). *J. Appl. Meteorol. Climatol.* **2011**, *43*, 231–248. [CrossRef]
39. Mathieu, M.; Couprie, C.; Lecun, Y. Deep Multi-Scale Video Prediction beyond Mean Square Error. *arXiv* **2016**, arXiv:1511.05440.
40. Che, H.; Niu, D.; Zang, Z.; Cao, Y.; Chen, X. ED-DRAP: Encoder–Decoder Deep Residual Attention Prediction Network for Radar Echoes. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
41. Pontius, R.G.; Thontteh, O.; Chen, H. Components of Information for Multiple Resolution Comparison between Maps That Share a Real Variable. *Environ. Ecol. Stat.* **2008**, *15*, 111–142. [CrossRef]
42. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]