Article

Radar-SR3: A Weather Radar Image Super-Resolution Generation Model Based on SR3

Zhanpeng Shi [1] , Huantong Geng [1,2,*], Fangli Wu [1], Liangchao Geng [3] and Xiaoran Zhuang [4]

1 School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China; 202212200009@nuist.edu.cn (Z.S.); 202212210021@nuist.edu.cn (F.W.)
2 China Meteorological Administration Radar Meteorology Key Laboratory, Nanjing 210023, China
3 School of Atmospheric Science, Nanjing University of Information Science & Technology, Nanjing 210044, China; 202311010012@nuist.edu.cn
4 Jiangsu Meteorological Observatory, Nanjing 210008, China; zxrxz3212009@163.com
* Correspondence: htgeng@nuist.edu.cn

Abstract: To solve the problems of the current deep learning radar extrapolation model consuming many resources and the final prediction result lacking details, a weather radar image super-resolution weather model based on SR3 (super-resolution via image restoration and recognition) for radar images is proposed. This model uses a diffusion model to super-resolve weather radar images to generate high-definition images and optimizes the performance of the U-Net denoising network on the basis of SR3 to further improve image quality. The model receives high-resolution images with Gaussian noise added and performs channel splicing with low-resolution images for conditional generation. The experimental results showed that the introduction of the diffusion model significantly improved the spatial resolution of weather radar images, providing new technical means for applications in related fields; when the amplification factor was 8, Radar-SR3, compared with the image super-resolution model based on the generative adversarial network (SRGAN) and the bicubic interpolation algorithm, the peak signal-to-noise ratio (*PSNR*) increased by 146% and 52% on average. According to this model, it is possible to train radar extrapolation models with limited computing resources with high-resolution images.

Keywords: weather radar image; super-resolution; diffusion model; attention mechanism; residual connection; S-band radar

1. Introduction

Weather radar is one of the key tools in the field of meteorology and natural disaster warning and is widely used to monitor and track rain, storms, lightning, and other weather phenomena in the atmosphere [1]. In meteorology, the role of weather radar cannot be ignored; it not only provides a wide range of meteorological data but also has the characteristics of high spatial and temporal resolution of weather radar detection data, which is of great significance for meteorological operations, such as short-term forecasting as well as small- and medium-scale weather monitoring, which are also the main tool for nowcasting. The timely monitoring of weather changes and the effective management of disaster risks are crucial for short-term weather forecasting. The performance and image quality of meteorological radar directly impact the accuracy of predictions. Current radar echo extrapolation deep learning models, such as PredRNN [2] and MotionRNN [3], tend to produce blurry images to achieve higher mean squared error (*MSE*) scores, thereby affecting the final prediction results. GAN-LSTM [4] and GAN-rcLSTM [1] have been incorporated as a generative adversarial network (GAN) module based on recurrent neural networks to generate radar images with clearer details and more accurate predictions. Currently, extrapolation models take high-resolution radar images as input in the hope of obtaining better extrapolation results. However, using high-resolution radar

images as input inevitably increases the model's parameters, thereby affecting training and inference efficiency.

Image super-resolution (SR) is a classic problem in computer vision and image processing. Its goal is to reconstruct a high-resolution (HR) image from a given low-resolution (LR) input [5], aiming to enhance the quality and details of the image. Super-resolution methods include traditional methods and deep learning methods. Among traditional methods, interpolation-based methods have simple algorithms and fast processing speeds but may result in a loss of detail in areas with complex textures [6], such as aliasing or color blocks. Methods based on deep learning can be categorized into those using convolutional neural networks (CNNs), generative adversarial networks (GANs), and diffusion models for super-resolution. The CNN-based super-resolution deep learning network SRCNN [7] was the first to introduce CNNs to the task of image super-resolution. In subsequent models, such as VDSR [8], residual networks were incorporated into deep learning models to enhance the model's receptive field while preserving shallow-layer information. With the success of generative adversarial networks (GANs) in image generation, SRGAN [9] was the first to introduce GAN to the field of super-resolution. SRGAN, compared with traditional CNNs, focuses more on the semantic differences between the original and reconstructed images rather than the pixel-wise color and brightness differences. In recent years, there have been studies focusing on super-resolution generation of weather radar images, all of which use CNNs such as U-Net [10] or RABPN [11]. Although super-resolution generation based on CNN is fast, the details of radar images are still missing compared with real observation images. Super-resolution generation methods have been widely studied and applied in downscaling in recent years. ESRGAN [12] used GAN to achieve $4\times$ resolution downscaling of surface wind speed. The study of [13] also used GAN to downscale time-evolving atmospheric fields.

In recent years, diffusion models have gained widespread attention in the field of image generation [14–16]. These models use a U-Net architecture in deep learning for denoising and offer the advantages of lower training complexity and simpler model architecture compared with adversarial generative networks [17]. Diffusion models have also been successfully applied to image super-resolution tasks. The SR3 model [18] has achieved significant accomplishments in various tasks [19]. In contrast to traditional diffusion probability models, SR3 not only focuses on image reconstruction but also integrates object recognition information into the reconstruction process, resulting in better performance in image super-resolution tasks.

Diffusion models have demonstrated significant potential in various domains. For instance, the Imagen [20] model applies a conditional diffusion model for image super-resolution, consisting of a text encoder and a series of diffusion models for image super-resolution [9]. The Dreamix [21] video generation model utilizes a video diffusion model to combine low-resolution spatiotemporal information from the original video with newly synthesized high-resolution information during inference, ensuring consistency with guiding textual prompts. In the meteorological domain, SwinRDM [22] builds upon SwinRNN [23] by incorporating a diffusion model module to predict high-spatial-resolution and high-quality atmospheric details.

Despite the widespread success of diffusion model-based image super-resolution models in various fields, their application in weather meteorological radar images remains a relatively unexplored area, requiring in-depth research. Therefore, this paper leverages an improved SR3 model named Radar-SR3 to generate high-quality and high-resolution radar echo images.

This article makes the following contributions:

(1) It explores the application of the SR3 super-resolution model on weather meteorological radar images and conducts a comparative analysis of the proposed super-resolution model with commonly used super-resolution models using the Jiangsu radar dataset.

(2) It proposes a new ResNet Block with Attention (RA) module to replace the convolutional modules in the U-Net model, thereby improving the denoising network of

U-Net. The Radar-SR3 super-resolution model, incorporating this novel RA module, achieved an optimal peak signal-to-noise ratio and structural similarity index on a weather radar dataset.

2. Problem Description and Materials

2.1. Problem Description

Recently, the application of radar echo extrapolation deep learning models in weather forecasting has made significant progress [1–3,22,23]. Since the RNN model has memory units, researchers prefer to use RNN-based models for radar echo extrapolation. However, the final extrapolation results of the models often suffer from issues of blurring and distortion. Under complex meteorological conditions, blurred predictions may lead to misjudgments of potential extreme weather events. As shown in Table 1, with the same batch size and time length, the size of radar images can significantly impact the parameters of Recurrent Neural Network (RNN)-based extrapolation models. When the width and height of radar images increase by a factor of 8, the average growth in model parameters is 15-times. This increase can pose a bottleneck in computational resources, affecting both training and inference speed. In extreme cases, it may even render training on some high-resolution weather radar datasets practically infeasible.

Table 1. Number of parameters of radar extrapolation model under different input sizes.

Method	Parameters	
	Input Size: 1, 20, 1, 128, 128 *	Input Size: 1, 20, 1, 16, 16 *
PredRNN	131,714,048	7,851,008
MIM [24]	249,720,192	14,380,416
ConvLSTM [25]	96,379,969	5,547,073
MotionRNN	132,511,361	8,648,321

* From left to right, the input sizes represent: batch size, time length, number of image channels, image height, image width.

2.2. Materials

The meteorological radar dataset consists of time-series data of radar echo, with its physical interpretation being the basic reflectivity factor at the 3 km elevation. A higher water droplet content in the atmosphere results in higher radar reflectivity [26,27]. The dataset is compiled using a network of multiple S-band meteorological radars in Jiangsu Province, covering the period from April to September in the years 2019 to 2021, and the data used are 3 km cappi (constant altitude plan position indicator). The radar echo data underwent quality control processes such as clutter suppression and discrete noise filtering. Additionally, data with a low proportion of radar echoes were manually excluded, covering the entire area of Jiangsu Province.

The data values range from 0 to 70 dBZ, with a horizontal resolution of 1 km $\times$ 1 km, a time resolution of 6 min, and a grid size of 480×560 pixels for single-time data. To facilitate the training of deep learning models while preserving image information, padding on both sides and center cropping were applied, resulting in images of 512×512 pixels.

To balance the resources, time, training effectiveness, and recognition performance required for deep learning model training, a total of 31,122 samples were selected. These samples were split into training, validation, and test sets in a ratio of 7:2:1, respectively.

Considering training time, the 512×512-pixel images are initially downsampled to images of size 128×128 and 16×16. The 128×128 images are defined as HR images, while the 16×16 images are defined as LR images. An example of data visualization is shown in Figure 1.

(a)

(b)

Figure 1. Sample of data: (**a**) 128 × 128-pixel HR image; (**b**) 16 × 16-pixel LR image.

3. Radar Echo Image Super-Resolution Model Based on Improved SR3

3.1. Denoising Diffusion Probability Model

The denoising diffusion probability model (DDPM) is inspired by non-equilibrium thermodynamics [28]. If you add noise to pixels in a high-dimensional image space, like ink spreading in water, and then reverse the process, this can generate images from the noise, resulting in unexpected combinations of images. The denoising diffusion probability model includes a deep learning denoising network and a diffusion process. The diffusion process includes a forward diffusion process and a reverse diffusion process.

3.1.1. Denoising Network Based on U-Net Model

The U-Net model [29] was initially designed to address the segmentation of medical images. It introduces an encoder–decoder [30] architecture, utilizing a U-shaped network structure to capture contextual information. The encoder part of U-Net is responsible for extracting image features from low-resolution inputs. Since the goal of the SR3 model is to reconstruct low-resolution images into high-resolution images, the decoder part of U-Net in the SR3 model works to gradually increase the resolution of the feature maps through deconvolution and up-sampling operations. To preserve image details and structural information, the U-Net in the SR3 model incorporates skip connections. These connections link the feature maps between the encoder and decoder, allowing information to be transmitted across different scales, facilitating feature fusion, and ultimately generating high-resolution images. The framework of the U-Net model is shown in Figure 2.
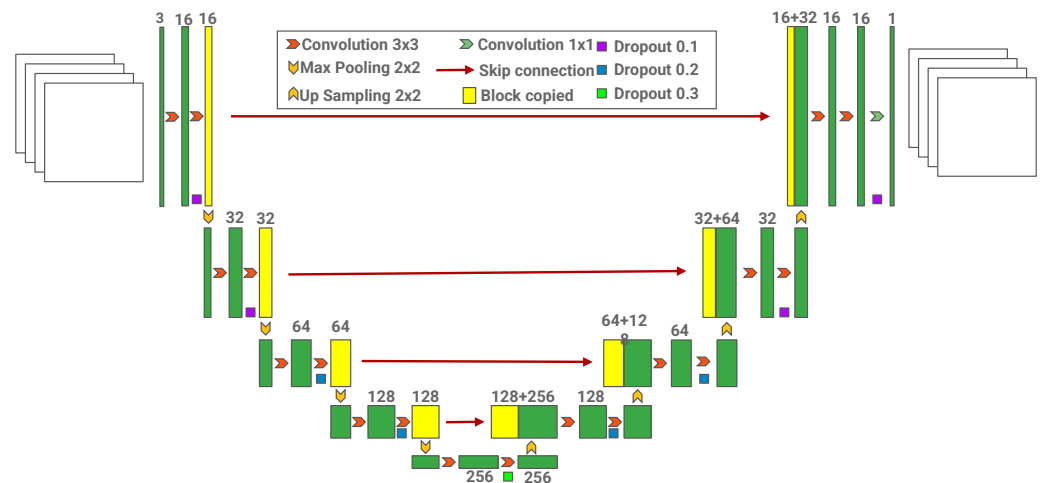

Figure 2. Structure of U-Net.

3.1.2. Diffusion Process

The diffusion process includes a forward diffusion process and a reverse diffusion process, using a parameterized Markov chain trained by variational inference to generate samples that match the data after a limited time [22].

In the forward diffusion process, a set of data $x_0 \sim q(x)$ obtained by sampling from the real data distribution, a series of noise-added samples $x_1, x_2, \ldots x_{t-1}, x_t, x_{t+1} \ldots, x_T,$

are obtained by superimposing Gaussian noise on the samples in T steps. The recursive formula from the origin HR image x_0 to an HR image after adding Gaussian noise at time step t x_t is:

$$x_t = \sqrt{\alpha_t \alpha_{t-1} \ldots \alpha_1} x_0 + \sqrt{1 - \alpha_t \alpha_{t-1} \ldots \alpha_1} z \tag{1}$$

z is the noise that conforms to the standard normal distribution, and α_t is the weight, which decreases as T increases. Let $\alpha_t \alpha_{t-1} \ldots \alpha_2 \alpha_1 = \overline{\alpha_t}$; then, for any T time step:

$$x_T = \sqrt{\overline{\alpha_T}} x_0 + \sqrt{1 - \overline{\alpha_T}} z \tag{2}$$

The final x_T is a noise that conforms to the standard normal distribution, also because z is the noise that conforms to the standard normal distribution. According to Formula (2):

$$\overline{\alpha_T} \approx 0 \tag{3}$$

Because $\alpha_t \alpha_{t-1} \ldots \alpha_2 \alpha_1 = \overline{\alpha_t}$, then $\overline{\alpha_t}$ can be controlled by α_t and time step T.

The reverse diffusion process uses a U-Net model to learn image noise to achieve denoising. That is, for the noise $\tilde{z}$ and time step t and the image x_t at time step t, we obtain:

$$\tilde{z} = UNet(x_t, t) \tag{4}$$

In the reconstruction stage, x_{t-1} is found under the premise of knowing x_t, which can be obtained by $q(x_{t-1}|x_t)$. According to Formula (1) and the properties of normal distribution:

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} z_t \sim N(\sqrt{\alpha_t} x_{t-1}, 1 - \alpha_t) \tag{5}$$

Then, according to the conditional probability formula:

$$q(x_{t-1}|x_t) = \frac{q(x_t, x_{t-1})}{q(x_t)} \sim N\left(\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1 - \overline{\alpha_t}}} \tilde{z}\right), \beta_t \cdot \frac{1 - \overline{\alpha_{t-1}}}{1 - \overline{\alpha_t}}\right) \tag{6}$$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1 - \overline{\alpha_t}}} \tilde{z}\right) + \sqrt{\beta_t \cdot \frac{1 - \overline{\alpha_{t-1}}}{1 - \overline{\alpha_t}}} z \tag{7}$$

$\tilde{z} = UNet(x_t, t), z \sim N(0, 1), 1 - \alpha_t = \beta_t$, according to Formula (7). The image x_{t-1} at time step $t-1$ can be obtained from the image x_t at time t and the noise $\tilde{z}_t$.

3.2. SimAM Attention

The Attention mechanism was first proposed by John K. Tsotsos in 1995 [31] for the field of visual images. In 2014, Google Mind applied the Attention mechanism to image classification in Recurrent Convolutional Neural Network (RNN) models [32]. The Attention mechanism generates weight vectors for each input element, determining which parts significantly impact the model output by calculating the weights of each feature map.

Traditional attention mechanisms include spatial and channel attention mechanisms [33]. The spatial attention mechanism focuses on the importance of different spatial features in the image, while the channel Attention mechanism emphasizes the significance of features between different channels. Adding Attention mechanisms to deep learning models can often lead to improved performance. However, increased parameters increase the complexity of deep learning models, resulting in longer training and inference times.

SimAM [34], on the other hand, is an Attention mechanism based on mature neuroscience theories. It simultaneously infers spatial and channel weights from the current neurons, achieving performance improvements without affecting model complexity. The structure of the SimAM Attention mechanism is shown in Figure 3.
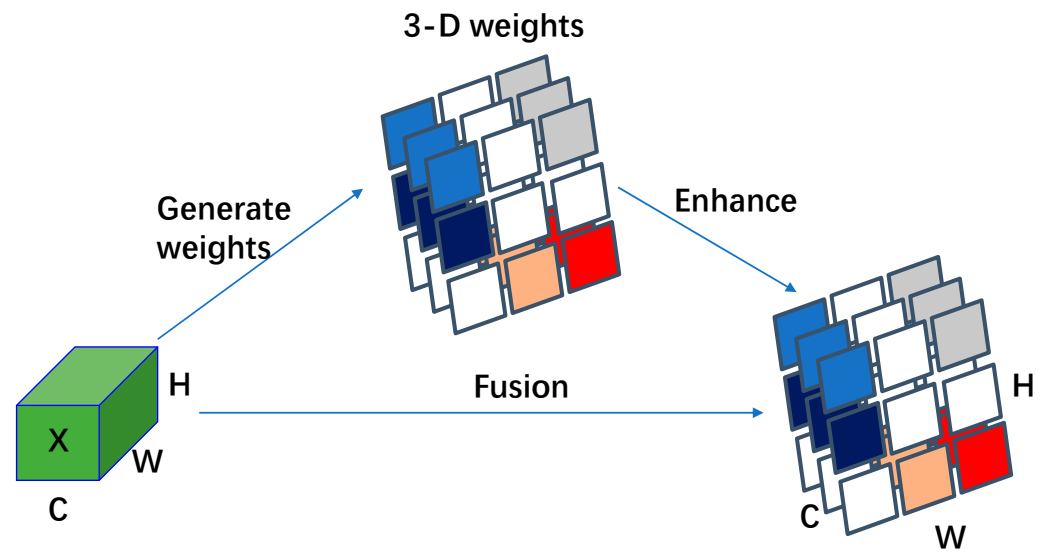
Figure 3. SimAM Attention mechanism. Different colors represent different weights.

3.3. SR3 Model

The SR3 model describes the super-resolution problem as a conditional generation problem. Compared with DDPM, which predicts the noise in the image each time through the U-Net network to generate a denoised image, the SR3 model first passes the original low-resolution image through up-sampling, which interpolates low-resolution images into high-resolution images and adds them to the training process. At the same time, the noise-added image and the interpolated high-resolution image are input. That is, the number of input channels changes from three channels in DDPM to six channels. The denoising network U-Net can perform conditional denoising based on the high-resolution image after low-resolution interpolation. Therefore, compared with DDPM, random denoising becomes a conditional generative model controlled based on the low-resolution image. At the same time, the denoising network U-Net in SR3 no longer obtains noise based on time step *t* but directly accepts the noise at the current time, thereby achieving faster inference speed.

Due to the particularity of weather radar images, the original U-Net denoising network in SR3 struggles to handle the global structure of the image and cross-channel dependencies, so the U-Net denoising network in SR3 is needed to improve the ability to capture global information, thus achieving a better denoising effect. This paper introduces an Attention mechanism to capture global information and channel information based on the original U-Net denoising network. It introduces residual connections to increase the number of model parameters and enhance the denoising ability of the U-Net model.

3.4. Radar-SR3 Model

Radar-SR3 replaces the original U-Net denoising network with an improved denoising U-Net network based on SR3. The overall process is shown in Figure 4.
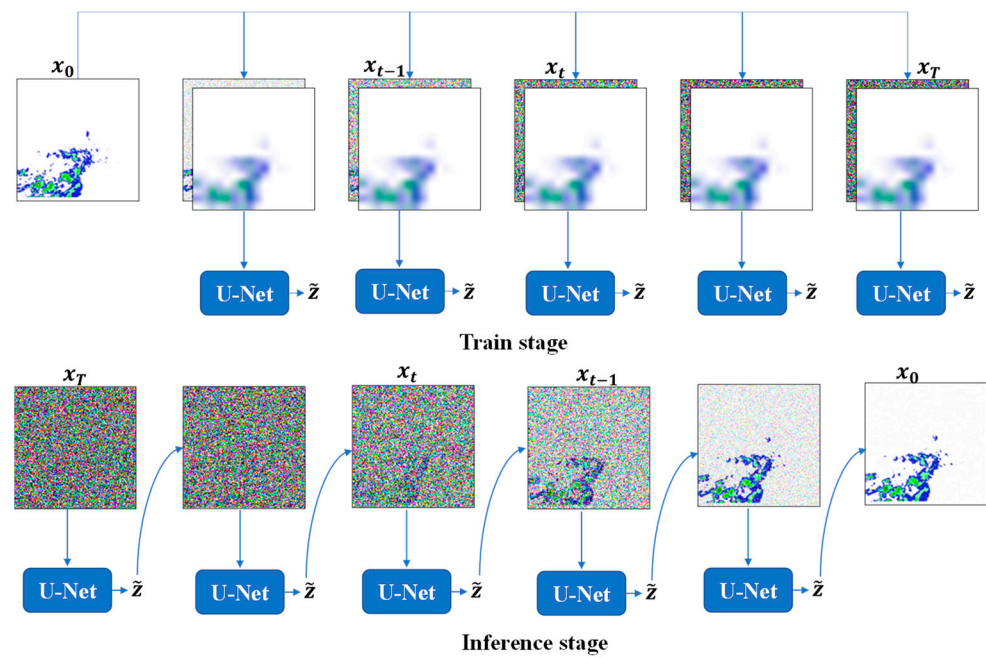
Figure 4. Overall process of Radar-SR3, where $\tilde{z}$ means noise.

3.4.1. Residual Connection with Attention Mechanism

Residual connection was first proposed in ResNet [35] in 2016 and won first place in the ImageNet Image Recognition Challenge in 2015. Residual connections are implemented by adding the input to the result of activating a nonlinear activation function. This method can reduce the problem of a network's gradient vanish and improve model expression capabilities. In the residual connection, the input x is mapped to a function $f(x)$, which is then added to the original input to output $y = x + f(x)$. This can reduce the vanishing gradient problem because deeper network parameters have less impact on the model output, thus ensuring stability and convergence speed during the training process. This article uses the Swish activation function [36] as the activation function in the residual block. Compared with the Relu activation function, Swish is a smooth and non-monotonic function, and its performance on multiple-depth models is better than the Relu function.

Based on the residual connection, this article adds different Attention mechanism modules for the U-Net denoising model to capture noise and features on each channel and space, named the RA (ResNet Block with Attention) module. The added attention mechanisms include the self-attention mechanism [37], the CBAM Attention mechanism [38], and the SimAM Attention mechanism [28]. The image quality and structural similarity indicators were compared, and the one with the highest index was selected as an integral part of the RA module. The RA module is shown in Figure 5.
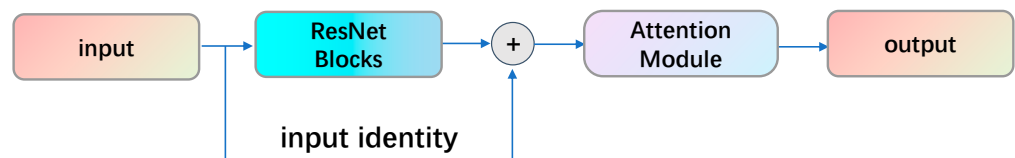

Figure 5. Sample of RA Module.

3.4.2. Improved U-Net Denoising Network

This paper modifies the original U-Net model by replacing its convolutional modules with residual connection modules, enhancing its effectiveness and depth. In the downsampling and up-sampling layers, the residual connection blocks are replaced with RA modules better to capture noise and features from the original image. The architecture

of the U-Net network is illustrated in Figure 6, consisting of an encoding segment and a decoding segment. The encoding segment employs three layers of residual blocks to extract shallow semantic information from the image. Two RA modules are utilized to capture deeper image correlations. The decoding component uses two RA modules to reconstruct the deep semantic features of the decoded image and three layers of residual blocks to restore shallow semantic information.
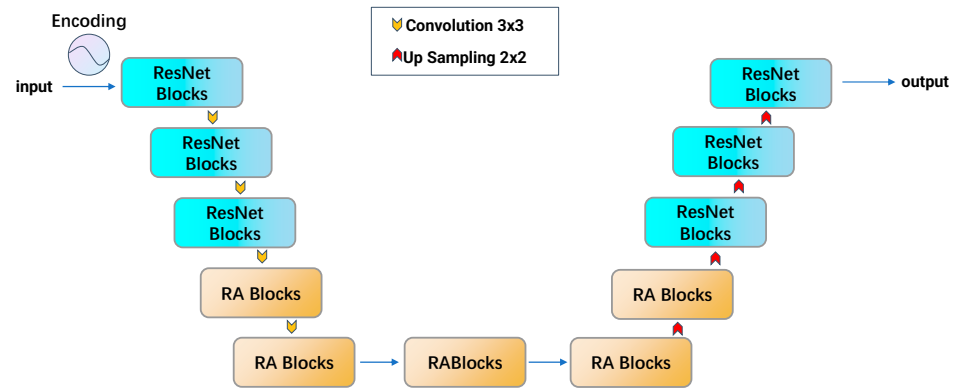

Figure 6. Improved U-Net denoising network.

As depicted in Figure 7, the residual block includes a Group Normalization, a Swish activation function, and a 3×3 convolutional kernel with a stride of 1 in two-dimensional convolution. The down-sampling layer involves a 3×3 convolutional kernel, while the up-sampling layer utilizes a 2×2 convolutional kernel.
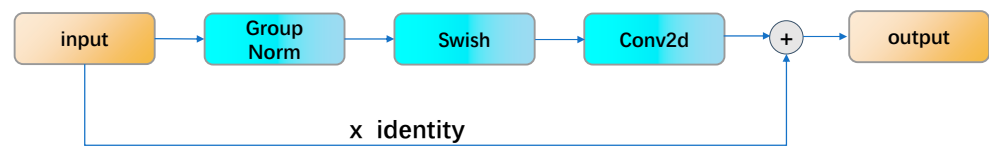

Figure 7. Residual block in U-Net.

4. Experiments and Results

4.1. Experimental Setup

The experimental environment is as follows: The CPU is an Intel® Core™ i9-13900K processor with a frequency of 5.0 GHz, and the memory is 32 GB. The GPU is an NVIDIA GeForce RTX 4090. The software environment includes PyTorch 2.0.1 and CUDA 11.8. The batch size is set to 24, and the Adam optimizer [39] is used for optimization with an initial learning rate of 1×10^{-4}. The L1 loss function is employed as the loss function.

4.2. Evaluation Metrics

This paper uses the peak signal-to-noise ratio (*PSNR*) and the structural similarity index measurement (*SSIM*) [40] as quantitative metrics to evaluate the super-resolution performance of the algorithm. *PSNR* is employed to assess the consistency between generated images and ground truth, while SSIM is used to evaluate the structural similarity between generated images and ground truth. The definitions of *SSIM* and *PSNR* are given in Formulas (9) and (10).

For an image x and y with a size of $m \times n$, the mean squared error (*MSE*) between x and y is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[x(i,j) - y(i,j) \right]^2 \tag{8}$$

$$PSNR = 20 \cdot log_{10}\left(\frac{MAX_x}{\sqrt{MSE}}\right) \tag{9}$$

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{10}$$

μ_x and μ_y represent the average of x and y. σ_x^2 and σ_y^2 represent the variance of x and y. σ_{xy} is the covariance of x and y.

$c_1 = \left(k_1L\right)^2, c_2 = \left(k_2L\right)^2, c_3 = c_2/2$, $k_1 = 0.01$, $k_2 = 0.03$, L is the range of pixel values. If the number of image channels is three, the range of L is 0~255. *SSIM* is a number between 0 and 1. The larger it is, the smaller the difference between the output image and the image-free image is; that is, the image quality is very good. When the two images are the same, *SSIM* = 1.

4.3. Results

4.3.1. Comparative Experiment

Figure 8 shows the super-resolution results of Radar-SR3 at different training stages. Figure 9 shows the relationship between the number of training epochs and PNSR. As the number of training epochs increases, the images generated by the Radar-SR3 model become closer to the actual value. Figure 10 selects five time steps to show the super-resolution process of Radar-SR3. Table 2 shows different models' *PSNR* and *SSIM* indicators when the amplification factor is 8. Compared with the SR3 model, Radar-SR3 improves *PSNR* by 0.44 while keeping *SSIM* unchanged.
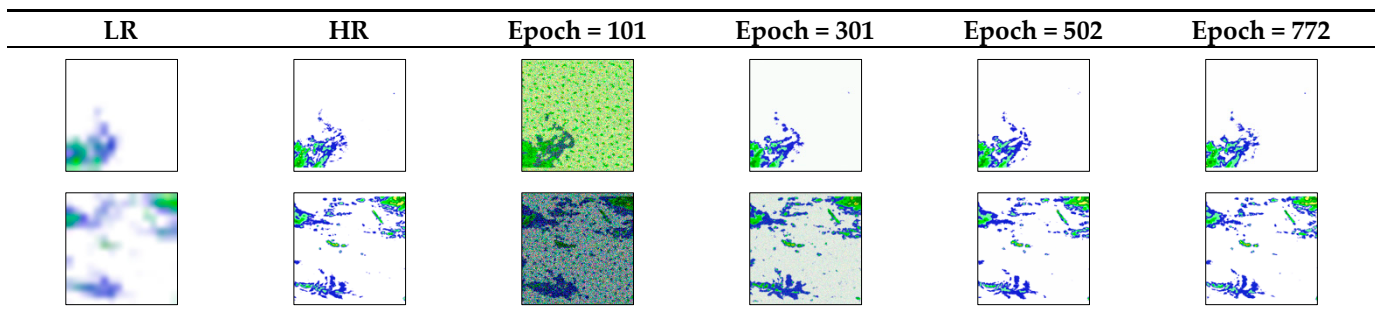

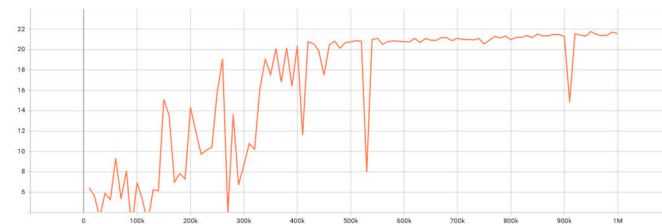**Figure 8.** Super-resolution results of Radar-SR3 at different training stages.

Figure 9. *PSNR* changes with training iterations. The abscissa represents the number of iterations, and the ordinate represents the value of *PSNR*.
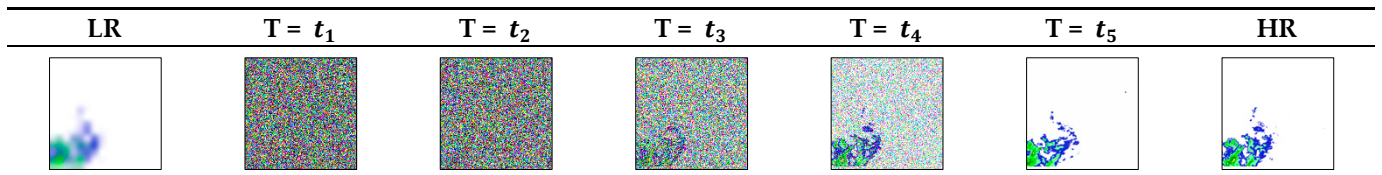

Figure 10. Process of super resolution, T represents time step, select five time steps for visualization.

Table 2. Comparison of super-resolution performance of different models.

Method	$16 \times 16 \rightarrow 128 \times 128$	
	PSNR/dB ↑	*SSIM* ↑
SR3	21.33	0.885
Bicubic	14.28	0.582
SRGAN	8.82	0.063
Radar-SR3	21.77	0.885

Figure 11 selects an example. First, the LR image is interpolated into an HR image through the Bicubic algorithm, from 16×16 to 128×128. Then, the interpolated image is compared with SR3, SRGAN, and Radar-SR3 for details. Compared with SR3, the super-resolution reconstruction effect of Radar-SR3 in high-echo areas is closer to the real value, and the details in some discontinuous echo areas are richer. A comparison can be obtained, and Radar-SR3 super-resolution model imaging is more precise and more detailed than Bicubic algorithm imaging; although the SRGAN model can restore high-echo areas well, it generates severely gridded images during the generation process, and overall imaging is not available. But the Radar-SR3 model can restore the details more clearly based on generating clear images, generating smoother and continuous effects, and the generated images are closer to the authentic images.
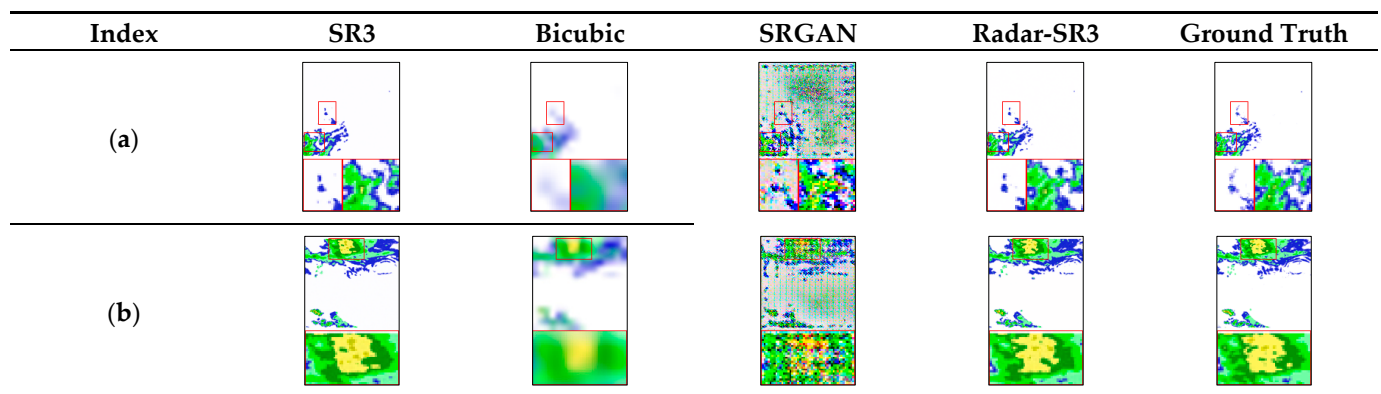

Figure 11. Comparison of details. Use the red frame to enlarge image details and compare them.

Figure 12 shows the super-resolution effects of different models. The pictures generated by the Radar-SR3 model are closer to the real values.
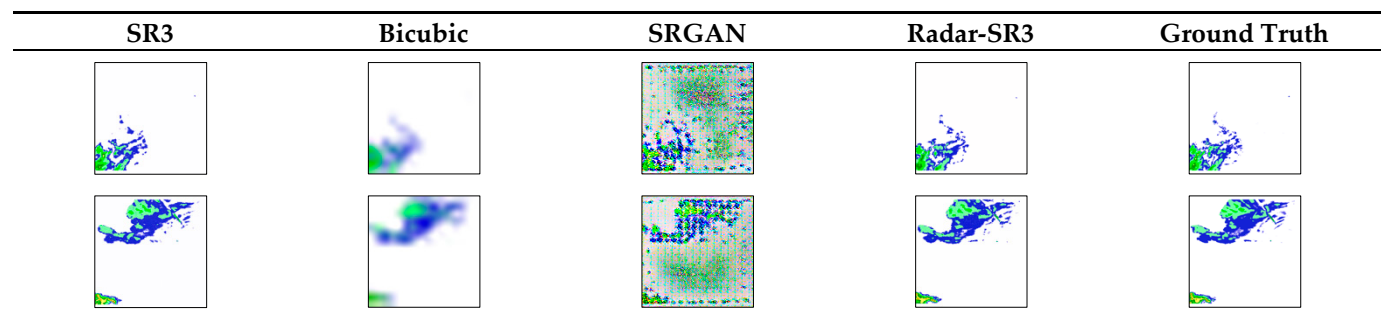

Figure 12. Comparison of generated images by different models.

4.3.2. Module Selection

To assess the denoising capability of the U-Net network combined with different Attention mechanisms, an 8-fold magnification module selection experiment was conducted on the Jiangsu radar dataset for the self-attention mechanism, CBAM Attention mechanism,

and SimAM Attention mechanism. Initially, the U-Net denoising network was used as the baseline model without adding any attention mechanism. Subsequently, the self-attention, CBAM, and SimAM Attention mechanisms were individually incorporated into the U-Net residual block. A comparative analysis was performed to examine the effects of each module on the U-Net denoising network. Evaluation metrics such as *PSNR* and *SSIM* were employed. The experimental results are presented in Table 3.

Table 3. Comparison of indicators and parameter quantities of different models.

Method	PSNR ↑	SSIM ↑	Parameters
Baseline	21.33	0.885	91,506,819
Self-Attention	21.38	**0.892**	97,807,491
CBAM	20.75	0.878	91,704,015
SimAM	**21.77**	0.885	**91,506,819**
SimAM + CBAM	21.04	0.880	91,704,015

Bold values represent optimal indicators.

Table 3 shows that the denoising capability, as measured by the *PSNR* indicator, is superior when the RA module combines the SimAM attention block with the residual block compared to other combinations. The *SSIM* is only 0.007 lower than using the self-attention module. Moreover, SimAM can enhance model performance without increasing training time or complexity due to its parameter-free nature. While the self-attention mechanism has the highest *SSIM* score, the improvement in *PSNR* compared to the baseline is not significant, and it has the highest number of parameters, meaning that it is not the optimal choice. Although the CBAM Attention mechanism can capture features from different channels and spatial dimensions, experimental results and the *PSNR/SSIM* metrics indicate suboptimal performance when applied to the U-Net denoising network. Attempting to concatenate the CBAM and SimAM Attention results in a slight improvement in both *PSNR* and *SSIM* compared to using CBAM alone. Consequently, the SimAM Attention is selected as the Attention module within the RA module.

5. Discussion

Weather radar is significant for nowcasting, and radar image super-resolution technology based on the conditional generation diffusion model can significantly alleviate the problem of poor imaging of the extrapolation model caused by various factors. Based on the SR3 super-resolution model, this paper first explores the feasibility of the SR3 model in weather radar super-resolution. Secondly, by improving the U-Net denoising network, the convolution block is replaced by a residual connection, and for the problem of difficulty in fusing multi-dimensional features, a residual module incorporating an attention mechanism is proposed, which includes a SimAM Attention module and multiple residual blocks. Using radar observation data in Jiangsu in the past three years and comparing experimental results, it was found that Radar-SR3 using the improved U-Net denoising model has better image generation capabilities than the SR3 model and is in the same dataset as commonly used image super-resolution algorithms; comparisons were made and excellent results were obtained. But Radar-SR3 still has flaws: the training time is too long. In the experiments in this paper, 1 epoch takes about 30 min, but it takes 500 epochs to achieve stable super-resolution generation results. If Denoising Diffusion Implicit Models (DDIMs) are used, the inference time can be reduced.

6. Conclusions

In a follow-up work, without introducing a new radar echo extrapolation model, radar echo prediction can be carried out through low-resolution radar echo images, and then the super-resolution generation of the extrapolated images can be performed through the Radar-SR3 model, obtaining a radar echo extrapolation image with clearer and richer details. Because

the radar echo extrapolation model accepts low-resolution sequence data, model training on high-resolution datasets becomes possible under limited computing resources.

References

1. 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]
2. Wang, Y.; Wu, H.; Zhang, J.; Gao, Z.; Wang, J.; Yu, P.S.; Long, M. PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 2208–2225. [CrossRef] [PubMed]
3. Wu, H.; Yao, Z.; Wang, J.; Long, M. MotionRNN: A Flexible Model for Video Prediction with Spacetime-Varying Motions. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 15430–15439.
4. Xu, Z.; Du, J.; Wang, J.; Jiang, C.; Ren, Y. Satellite Image Prediction Relying on GAN and LSTM Neural Networks. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
5. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating More Pixels in Image Super-Resolution Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, 17–24 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 22367–22377.
6. Chen, B.; Lin, M.; Sheng, K.; Zhang, M.; Chen, P.; Li, K.; Cao, L.; Ji, R. ARM: Any-Time Super-Resolution Method. In *Computer Vision—ECCV 2022*; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Lecture Notes in Computer Science; Springer Nature: Cham, Switzerland, 2022; Volume 13679, pp. 254–270. ISBN 978-3-031-19799-4.
7. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 8692, pp. 184–199. ISBN 978-3-319-10592-5.
8. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1646–1654.
9. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 105–114.
10. Geiss, A.; Hardin, J.C. Radar Super Resolution Using a Deep Convolutional Neural Network. *J. Atmos. Ocean. Technol.* **2020**, *37*, 2197–2207. [CrossRef]
11. Yu, Q.; Zhu, M.; Zeng, Q.; Wang, H.; Chen, Q.; Fu, X.; Qing, Z. Weather Radar Super-Resolution Reconstruction Based on Residual Attention Back-Projection Network. *Remote Sens.* **2023**, *15*, 1999. [CrossRef]
12. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *Computer Vision—ECCV 2018 Workshops*; Leal-Taixé, L., Roth, S., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11133, pp. 63–79. ISBN 978-3-030-11020-8.
13. Leinonen, J.; Nerini, D.; Berne, A. Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 7211–7223. [CrossRef]

14. Sasaki, H.; Willcocks, C.G.; Breckon, T.P. UNIT-DDPM: UNpaired Image Translation with Denoising Diffusion Probabilistic Models. *arXiv* **2021**, arXiv:2104.05358.
15. Li, H.; Yang, Y.; Chang, M.; Chen, S.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models. *Neurocomputing* **2022**, *479*, 47–59. [CrossRef]
16. Wu, Q.; Yang, C.; Zhao, W.; He, Y.; Wipf, D.; Yan, J. DIFFormer: Scalable (Graph) Transformers Induced by Energy Constrained Diffusion. In Proceedings of the The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, 1–5 May 2023.
17. Dhariwal, P.; Nichol, A.Q. Diffusion Models Beat GANs on Image Synthesis. In Proceedings of the Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, Virtual, 6–14 December 2021; Ranzato, M., Beygelzimer, A., Dauphin, Y.N., Liang, P., Vaughan, J.W., Eds.; pp. 8780–8794.
18. Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D.J.; Norouzi, M. Image Super-Resolution Via Iterative Refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4713–4726. [CrossRef] [PubMed]
19. Tang, Y. Hybrid Improved Models Combined SR3 Module for Animal Recognition in Electric Car's Actual Vision. In Proceedings of the 2022 International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 20–22 January 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 758–761.
20. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.L.; Ghasemipour, S.K.S.; Lopes, R.G.; Ayan, B.K.; Salimans, T.; et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In Proceedings of the NeurIPS, New Orleans, LA, USA, 28 November 2022.
21. Molad, E.; Horwitz, E.; Valevski, D.; Rav-Acha, A.; Matias, Y.; Pritch, Y.; Leviathan, Y.; Hoshen, Y. Dreamix: Video Diffusion Models Are General Video Editors. *arXiv* **2023**, arXiv:2302.01329. [CrossRef]
22. Chen, L.; Du, F.; Hu, Y.; Wang, Z.; Wang, F. SwinRDM: Integrate SwinRNN with Diffusion Model towards High-Resolution and High-Quality Weather Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 322–330. [CrossRef]
23. Hu, Y.; Chen, L.; Wang, Z.; Li, H. SwinVRNN: A Data-Driven Ensemble Forecasting Model via Learned Distribution Perturbation. *J. Adv. Model Earth Syst.* **2023**, *15*, e2022MS003211. [CrossRef]
24. Wang, Y.; Zhang, J.; Zhu, H.; Long, M.; Wang, J.; Yu, P.S. Memory in Memory: A Predictive Neural Network for Learning Higher-Order Non-Stationarity From Spatiotemporal Dynamics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, 16–20 June 2019; Computer Vision Foundation/IEEE: Piscataway, NJ, USA, 2019; pp. 9154–9162.
25. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, Montreal, QC, Canada, 7–12 December 2015; Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., Eds.; pp. 802–810.
26. Valjarević, A.; Morar, C.; Živković, J.; Niemets, L.; Kićović, D.; Golijanin, J.; Gocić, M.; Bursać, N.M.; Stričević, L.; Žiberna, I.; et al. Long Term Monitoring and Connection between Topography and Cloud Cover Distribution in Serbia. *Atmosphere* **2021**, *12*, 964. [CrossRef]
27. Schulte To Bühne, H.; Pettorelli, N. Better Together: Integrating and Fusing Multispectral and Radar Satellite Imagery to Inform Biodiversity Monitoring, Ecological Research and Conservation Science. *Methods Ecol. Evol.* **2018**, *9*, 849–865. [CrossRef]
28. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 6840–6851.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015—18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III. Navab, N., Hornegger, J., III, Wells, W.M., Frangi, A.F., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
30. Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, Doha, Qatar, 25–29 October 2014; A Meeting of SIGDAT, a Special Interest Group of the ACL. Moschitti, A., Pang, B., Daelemans, W., Eds.; ACL: Kerrville, TX, USA, 2014; pp. 1724–1734.
31. Tsotsos, J.K.; Culhane, S.M.; Wai, W.Y.K.; Lai, Y.; Davis, N.; Nuflo, F. Modeling Visual Attention via Selective Tuning. *Artif. Intell.* **1995**, *78*, 507–545. [CrossRef]
32. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent Models of Visual Attention. In Proceedings of the Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–13 December 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., Eds.; pp. 2204–2212.
33. Cai, Z.; Qiao, X.; Zhang, J.; Feng, Y.; Hu, X.; Jiang, N. RepVGG-SimAM: An Efficient Bad Image Classification Method Based on RepVGG with Simple Parameter-Free Attention Module. *Appl. Sci.* **2023**, *13*, 11925. [CrossRef]
34. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.

36. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018. Workshop Track Proceedings.

37. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R., Eds.; pp. 5998–6008.

38. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 11211, pp. 3–19. ISBN 978-3-030-01233-5.

39. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. Conference Track Proceedings.

40. 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]