

## Article

# MAFormer: A New Method for Radar Reflectivity Reconstructing Using Satellite Data

Kuoyin Wang <sup>1,†</sup>, Yan Huang <sup>1,†</sup>, Tingzhao Yu <sup>1,\*</sup> , Yu Chen <sup>1</sup>, Zhimin Li <sup>1</sup> and Qiuming Kuang <sup>2</sup><sup>1</sup> Public Meteorological Service Center, China Meteorology Administration, Beijing 100081, China<sup>2</sup> Geovis Environment Technology Co., Ltd., Beijing 101399, China

\* Correspondence: tsingzao@hotmail.com

† These authors contributed equally to this work.

**Abstract:** Radar reflectivity plays a crucial role in detecting heavy rainfall and is an important tool for meteorological analysis. However, the coverage of a single radar is limited, leading to the use of satellite data as a complementary source. Consequently, how to bridge the gap between radar and satellite data has become a growing research focus. In this paper, we present MAFormer, a novel model for reconstructing radar reflectivity using satellite data within the Transformer framework. MAFormer consists of two modules: the Axial Local Attention Module and the Mixup Global Attention Module, which extract both local saliency and global similarity. Quantitative and qualitative experiments demonstrate the effectiveness of our proposed method. Specifically, the MAFormer model exhibits notable advancements when compared to state-of-the-art deep learning techniques. It demonstrates an improvement ranging from 0.01 to 0.05 in terms of the Heidke skill score, indicating its superior performance. Additionally, MAFormer effectively mitigates false alarm rates by approximately 0.016 to 0.04, which further highlights its enhanced accuracy and reliability.

**Keywords:** radar reconstruction; satellite; transformer; axial local attention; mixup global attention



**Citation:** Wang, K.; Huang, Y.; Yu, T.; Chen, Y.; Li, Z.; Kuang, Q. MAFormer: A New Method for Radar Reflectivity Reconstructing Using Satellite Data. *Atmosphere* **2023**, *14*, 1723. <https://doi.org/10.3390/atmos14121723>

Academic Editor: Tomeu Rigo

Received: 22 October 2023

Revised: 18 November 2023

Accepted: 21 November 2023

Published: 23 November 2023



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

## 1. Introduction

Radar reflectivity is a critical factor in weather nowcasting that serves as a reliable indicator for severe weather conditions. With its ability to offer detailed descriptions of local areas at a resolution of approximately 1 km, radar reflectivity plays a significant role in assessing and predicting weather patterns.

However, the effectiveness of radar is hindered by its limited coverage and susceptibility to obstruction by mountains and other physical barriers. This limitation poses a challenge in obtaining comprehensive radar reflectivity data across larger regions. To overcome this issue, satellite data can be employed as a supplementary tool to radar observations. Satellites provide a broader view of the Earth's atmosphere, allowing for a global assessment of weather conditions but at a coarser resolution of around 4 km.

In recent years, deep learning techniques have gained significant attention in the field of satellite data processing and meteorology, offering innovative solutions to complex weather prediction and analysis tasks. Convolutional neural networks (CNNs) have been widely employed in weather forecasting tasks, primarily for the analysis of meteorological images and satellite data. CNNs excel at capturing spatial dependencies in data, making them suitable for tasks such as meteorological forecasting [1], spatial downscaling [2,3], weather classification [4,5], and cloud classification [6]. Han et al. [7] transform meteorological nowcasting into two stages, i.e., precipitation level classification and accurate precipitation regression. Moreover, cross-channel 3D convolution is employed to fuse raw 3D Doppler radar data and to extract effective multi-source information automatically. Shi et al. exploit ConvLSTM [8] and TrajGRU [9] for precipitation nowcasting, which has been the baseline among many spatial-temporal related tasks. Analogously, Yu et al. [10]

propose ATMConvGRU, a cascaded network that distinguishes from the previous parallel architecture, with more deep nonlinear feature extraction. Though effective, training feasible forecasting models is relatively difficult due to their recurrent mechanism and their inner architecture. To solve this key problem, Ayzel et al. propose Dozhdya.Net [11], which is an all convolutional neural network for radar-based precipitation nowcasting. Training such a network is efficient and experimental results show that it is crucial for early warning of hazardous events. Agrawal et al. [12] treat forecasting as an image-to-image translation problem and leverage the power of the ubiquitous UNet [13] architecture. Quantities of experiments demonstrate that this method outperforms commonly used models such as optical flow model, persistence model, and NOAA numerical prediction. Similar contributions can also be found in [14–16]. Klocek et al. [17] achieve 6 h precipitation nowcasting under the encoder-forecaster LSTM framework with radar mosaic sequences as input. The recently proposed MetNet [18] has also shown dramatic superiority compared with numerical weather prediction. Basically, MetNet provides a framework and a promising direction up to 7 to 8 h forecasting. One main advantage of MetNet originates from its integrating of multi-source information such as satellite data, radar data, elevation, longitude, latitude, and time. Compared with typical ConvLSTM, MetNet adds an extra feature extraction module to explore more abstract spatial–temporal representation. The advanced version MetNet-2 [19] promotes forecasting time range from 8 h to 12 h with a larger receptive field. The former introduced deep learning methods use radar to directly predict future rain rates, free of physical constraints, while they accurately predict low-intensity rainfall, their operational utility is limited because of their lacking of constraints producing blurry nowcasts at longer lead times, yielding poor performance on rarer medium-to-heavy rain events. To address these challenges, Ravuri et al. [20] present a deep generative model for the probabilistic nowcasting of precipitation. On the other hand, Kuang et al. [21] impose the meteorological motion equation into TransUNet [22] for temperature forecasting, which is supposed to be a further integration of meteorological prior and machine learning methods. Benefiting from these achievements, it can be concluded that another promising direction for meteorological downscaling should be data-driven machine learning techniques.

When addressing the issue of radar reflectivity and satellite data gaps in satellite data-based radar reconstruction, Duan et al. [23] extended the UNet method for reconstructing radar reflectivity from Himawari-8. It is a weather satellite developed by the Japan Aerospace Exploration Agency (JAXA) and manufactured by Mitsubishi Electric Corporation. The satellite was launched from the Tanegashima Space Center in Kagoshima Prefecture, Japan, and is operated from the Japan Meteorological Agency (JMA) headquarters in Tokyo, Japan. They made adjustments, such as employing one convolution operation instead of two in each convolution block to reduce the risk of overfitting. Additionally, they removed skip connections as their findings indicated that the high-resolution spatial information they provided was not always beneficial. In a similar vein, Zhu et al. [24] aimed to extract deep network representations by reconstructing radar reflectivity data from Numerical Weather Prediction (NWP) simulations and satellite observations. They subsequently examined the relationship between these representations and physical quantities like NWP variables and satellite images. Their research highlighted the potential of data-driven deep learning models in bridging representation gaps across multiple scales and data sources. They also utilized Himawari-8 for radar reconstruction. Conversely, Yang et al. [25] proposed a novel deep learning technology based on the attention mechanism to reconstruct radar reflectivity using observations from China's new-generation geostationary meteorological satellite, FengYun-4A. To account for complex surface effects, they incorporated topography data into their model.

While significant progress has been made in using existing methods for radar reconstruction, most of these methods, such as UNet, were originally developed for natural or medical image segmentation and may not be applicable for radar reconstruction. Addition-

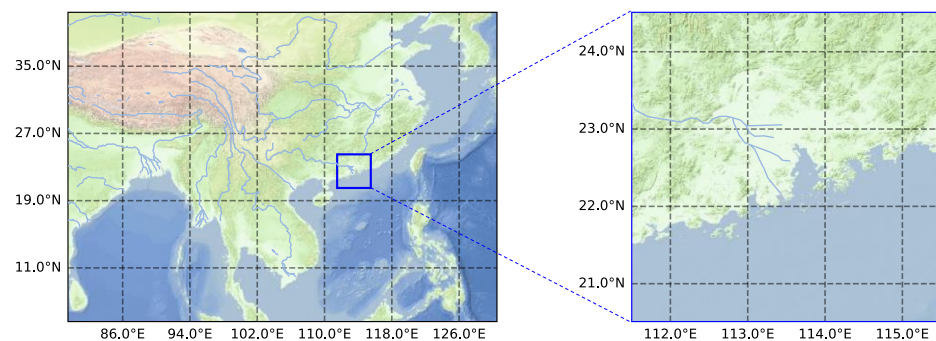
ally, the unique properties of the atmosphere pose significant challenges for satellite-based radar reconstruction, further complicating the issue.

Considering the property of atmosphere, two key challenges remain in radar reconstruction. Firstly, the atmosphere is an integral, highly-correlated system, with a singular change affecting all aspects globally. Secondly, extreme local weather events require specialized handling. However, existing methods frequently overlook these challenges. To overcome these challenges and meet the need for improving radar reflectivity reconstruction, we introduce a novel satellite-based method. The contributions of this paper can be summarized as follows:

1. This paper proposes a new transformer network, called MAFormer (Mixup-global with Axial-local attention Transformer), which includes two modules: the Mixup Global Attention Module (MGAM) and the Axial Local Attention Module (ALAM).
2. The MGAM extracts large-scale global-similarity features, while the ALAM is designed for small-scale local-singularity feature extraction. MAFormer, when combined with the vanilla transformer model, can accurately reconstruct radar reflectivity from single satellite data.
3. Quantitative and qualitative experiments were conducted, comparing the MAFormer model against state-of-the-art methods. The results of these experiments demonstrate the superiority of the proposed approach. Overall, this method offers a promising solution to the challenges of radar reconstruction and holds significant potential for further advancements in satellite-based data processing.

## 2. Materials

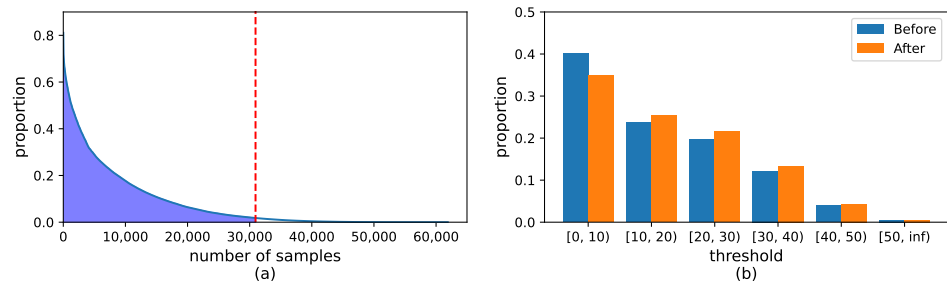
The study area has good coverage of 20.51° N–24.50° N, 111.50° E–115.49° E (see Figure 1). Specifically, the radar and satellite data are collected from 8 July 2022 to 20 September 2023 for research purposes. The satellite data has a high spatial temporal frequency with a resolution of 1 km and a time interval of 10 min, while the radar data has a resolution of 1 km and a time interval of 6 min.



**Figure 1.** An overview of the study area, denoted by the blue box. (The dashed lines represent the corresponding latitude and longitude).

However, we should note that it is not the case that all of these collected data useful. For example, there are many situations where the occupied region with radar reflectivity is rather small, which is useless and somewhat detrimental for model training. As a result, data selection is essential. In this paper, we sort all collected samples in descending order according to the proportion of the study area occupied by regions with radar reflectivity greater than 10 dbz. Then, we select the top 50% samples (the left part of Figure 2a) in this order for modeling. See the following Figure 2a for details.

Casting on the former illustrations, we select data according to the occupied region portion instead of a hard reflectivity threshold. Consequently, there is no such a deterministic reflectivity range. Additionally, an intuitive comparison between the data distribution before and after data selection is presented in Figure 2b. There is merely little difference between them. As a result, this selection strategy almost has no effects on the results.



**Figure 2.** Explanation about data selection. (a) Data selection strategy. The read dashed line represents 50% number of samples. (b) Distributions before and after data selection.

During the preprocessing, the satellite and radar data are aligned temporally to account for the temporal difference between them, resulting in numerous satellite–radar pairs. The data from 8 July 2022 to 30 June 2023 are used for model training and validation, while the data from 1 July 2023 to 20 September 2023 are reserved for model evaluation.

The satellite data in this study follows the approach described in [26], using Himawari-8. However, starting from 13 December 2022, Himawari-8 was replaced by Himawari-9. Therefore, Himawari-9 data are also collected in this study. Considering the high similarity between these two satellites, it is reasonable to integrate their data. This combination of Himawari-8 and Himawari-9 also helps to validate the generalization of the proposed model.

Following [26], out of the 16 channels available, only 7 channels (C8, C9, C10, C11, C13, C14, and C16) are selected for analysis. The rationale behind the selection of these channels has been addressed in [26]. As for radar, the quality-controlled radar reflectivity data used in this study are provided by the China Meteorological Administration.

### 3. Methods

#### 3.1. Preliminary

The transformer [27] architecture has emerged as a revolutionary approach in various fields of natural language processing and computer vision, and it has also shown promising results among meteorological applications. It was first designed to overcome the limitations of recurrent neural networks in handling long-range dependencies effectively. Specifically, transformer relies heavily on a self-attention mechanism and operates by processing all inputs simultaneously, which enables parallelization and efficient computation.

At the core of the transformer architecture lies the self-attention mechanism. It allows the model to weigh the importance of different positions or elements within the input feature space. The self-attention mechanism computes a weighted sum of values based on their relevance to each other. It attends to different positions in the input sequence by computing attention scores, which are ultimately used to obtain context-aware representations. The self-attention is performed by computing akin of correlation factors among three tokens, i.e., query, value and key. Specifically, based on a given query, as well as keys and values, the output of the attention module is calculated by a weighted sum of all values, where the weight assigned to each value is determined by the correlation between corresponding key and the query. Formally, a set of queries, keys, and values are packed together into matrices  $Q$ ,  $K$ , and  $V$  to compute the attention weights according to the following equation:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where  $Q, K \in \mathcal{R}^{N \times d_k}$ ,  $V \in \mathcal{R}^{N \times d_v}$  are all vectors that denote queries, keys, and values, respectively,  $N$  represents the number of features,  $d_k$  and  $d_v$  are the feature dimensions of each individual query/key and value.

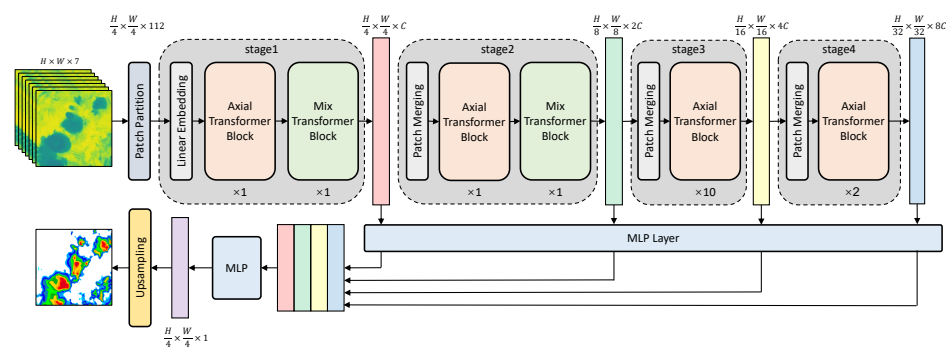
Apart from self attention, another four main components of transformer are followings:

- (a) Encoder–Decoder: The transformer architecture is composed of an encoder and a decoder. The encoder processes the input feature, generating a series of representations capturing the contextual information of each element. The decoder takes the encoder’s output and generates the reconstructed output.
- (b) Multi-Head Attention: The self-attention mechanism is further enhanced by using multiple heads. Each head learns different relationships between positions in the input sequence, allowing the model to capture various forms of dependencies. The outputs of multiple attention heads are concatenated and linearly transformed to obtain the final representations.
- (c) Feed-Forward Neural Networks: Transformer models include feed-forward neural networks (FFNs) to process the attention-based representations. FFNs consist of multiple layers of fully connected networks, enabling the model to model complex non-linear relationships.
- (d) Residual Connections and Layer Normalization: To mitigate the vanishing gradient problem and improve gradient flow, residual connections are employed. Residual connections provide skip connections, allowing the model to bypass certain layers and retain valuable information. Layer normalization is applied after each sub-layer to stabilize the training process.

### 3.2. Overview

The proposed MAFormer model consists of two main parts: a Swin-transformer [28] based multi-level encoder and a simple convolution-based decoder. The pipeline operates by extracting multi-scale features from the available satellite data, denoted as  $\mathcal{X} \in \mathcal{R}^{H \times W \times C}$ . Here,  $H$  and  $W$  denote the height and width of the study area, where  $H$  and  $W$  correspond to the latitude and the longitude, respectively.  $C$  represents the number of satellite channels for modeling. These features are then integrated using a multi-layer perceptron network for radar reconstruction. To restore the original spatial size, an efficient bilinear upsampling strategy is employed. More details on the feature size of multiple scales can be found in Figure 3.

MAFormer builds upon SwinTransformer by introducing two key components that differentiate it from other deep learning-based radar reconstruction methods. Firstly, to capture the local properties of weather processes, an axial attention mechanism is embedded into the typical transformer block. Secondly, to extract global long-range spatial similarity, the mixup attention module is introduced. This allows MAFormer to incorporate global information through the Mixup Global Attention Module (MGAM) while also extracting local characteristics using the Axial Local Attention Module (ALAM). These modules enable satellite-based radar reflectivity reconstruction under the transformer framework, thus giving rise to the name “MAFormer”.



**Figure 3.** An overview of the proposed MAFormer for satellite based radar reflectivity reconstruction. Basically, MAFormer consists of a transformer-based encoder for satellite feature extraction and a MLP-based decoder for radar reconstruction. Different modules are distinguished by colors with description, while the boxes without description denotes middle level features of MAFormer.



### 3.3. Mixup Global Attention Module

As illustrated in Section 3.1, one key component of transformer is the multi-head self-attention, which extracts long-range spatial dependency. In order to obtain multi-scale features in a cascaded manner as convolutional networks, SwinTransformer introduces the concept of Shift Window. This operation in turns hinders model from utilizing *global* information. Nevertheless, the weather system is an integrated union that the global effects must be taken into consideration. To resolve this problem, the Mixup Global Attention Module is introduced.

The mixup operation is first introduced in SegFormer [29], which focuses on semantic segmentation. Here, we generalize it to global feature extraction for satellite data. Specifically, mixup is embedded to the feedforward module using a  $3 \times 3$  depth-wise convolution, and it can be formulated mathematically as

$$X_{out} = MLP(GELU(Conv_{3 \times 3}(MLP(X_{in})))) + X_{in} \quad (2)$$

where  $X_{in}$  is the feature from efficient self-attention. Here, MLP, GELU, and Conv denote multi-layer perception, GaussianError, Linear Units and convolutional layer, respectively.

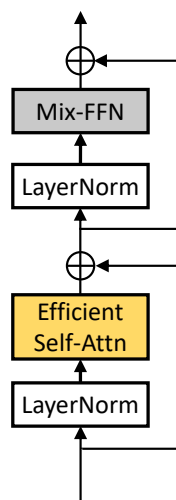
After that, the Mixup Global Attention Module (MGAM) incorporates efficient self-attention and layer normalization via residual connection as

$$\begin{aligned} X_{mid} &= ESA(LN(X_{in})) + X_{in} \\ X_{out} &= MixupFFN(LN(X_{mid})) + X_{mid} \end{aligned} \quad (3)$$

where ESA and LN represent Efficient Self Attention and Layer Normalization, respectively.

Through the former illustration, Mixup attention first extracts the high-level features using convolution and then mix these features globally via MLP. Consequently, it can depict global features which is needed for meteorological applications. Then, we call it the Mixup Global Attention Module. Consequently, a detailed illustration can be found in Figure 4.

### Mixup Global Attention Module



**Figure 4.** Illustration of the Mixup Global Attention Module (MGAM). MGAM consists of an efficient self-attention and a depth-wise convolution based feedforward operation. Both of these two parts are preprocessed through a layernorm. The cross sign in circle denotes element-wise addition.

### 3.4. Axial Local Attention Module

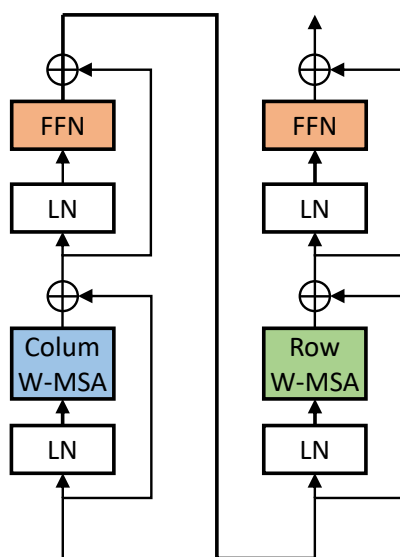
MGAM exploits the global relevance while neglects local specification. Nevertheless, one key difference between meteorological and other researches lies in the locality. To explore this property, this subsection proposes an Axial Local Attention Module (ALAM).

The idea behind axial attention [30] is to perform attention separately along each axis or dimension of the input feature. This means that attention is applied horizontally along one axis (e.g., the row or latitude axis in meteorological data) and vertically along another axis (e.g., the column or longitude). By decomposing the attention mechanism in this way, the computational complexity is reduced and the effectiveness within each axis. Similarly to a typical attention block, the layer normalization and residual connection are also employed. To be specific, ALAM first extracts the column-wise attention with feedforward operation, and then followed by a row-wise one. Before each operation, the widely used layer normalization is employed and a residual connection is added between every two layer normalization. Mathematically, ALAM can be defined as

$$\begin{aligned} X_{mid} &= CWA(LN(X_{in})) + X_{in} \\ X_{out} &= RWA(LN(X_{mid})) + X_{mid} \end{aligned} \tag{4}$$

where CWA and RWA denote Column-wise Window Attention and Row-wise Window Attention, respectively. A more intuitive illustration can also be found in Figure 5.

### Axial Local Attention Module



**Figure 5.** Illustration of the proposed Axial Local Attention Module (ALAM). ALAM consists of two cascaded axial attention module within each shift window.

#### 3.5. Method Analysis

In order to fully exploit the multi-scale features of meteorological satellite data, MAFormer incorporates both a MGAM module and a ALAM for global and local feature extraction. Different from most of the existing works [10], in which the axial attention is used for global feature extraction and convolution for local, MAFormer employs axial attention within each window, i.e., local areas, for local feature extraction, and mixup accompanied with MLP for global extraction. We should note that these two key differences distinguish our method from other contributions. And the experimental results in the following Section 4 also demonstrates the superiority of our MAFormer method.

### 4. Results

#### 4.1. Experiments Setting

For fair comparison, all of the experimental setting in this section are fixed. Specifically, the optimizer is set to be AdamW [31] with an initial learning rate  $10^{-4}$  and use a Poly learning rate scheduler with 1500 steps of linear warm-up start by  $10^{-6}$ . AdamW is a

widely used optimizer for model training. It improves Adam [32] with decoupled weight decay, leading to more generalized performance. Training stops after 80,000 iterations and the batch size is set to be 16.

4.2. Metrics

Following the setting of [23], this paper also employs two types of evaluation metrics, i.e., the classification score and the regression score. Specifically, the classification score includes false alarm rate (FAR), probability of detection (POD), critical success index (CSI), and Heidke skill score (HSS).

To be specific, suppose the confusion matrix for binary classification is defined as Table 1, then the former metrics can be defined as

$$\begin{aligned}
 FAR &= \frac{FP}{TP + FN} \\
 POD &= \frac{TP}{FP + TP} \\
 CSI &= \frac{TP}{TP + FP + FN} \\
 HSS &= \frac{2 * (TP * TN - FN * FP)}{FN * 2 + FP * 2 + 2 * TP * TN + (FN + FP) * (TP + TN)}
 \end{aligned}
 \tag{5}$$

Note that for obtaining the binary confusion matrix, a threshold must be predefined. Without specific illustration, the threshold is set to be 30 dBZ, which is a widely used threshold to distinguish heavy rain.

Table 1. Confusion matrix definition for classification.

		Ground Truth	
		1	0
Prediction	1	True Positive (TP)	False Positive (FP)
	0	False Negative (FN)	True Negative (TN)

While for regression score, the typical root mean squared error (RMSE), mean absolute error (MAE), and the widely used peak signal noise ratio (PSNR), structure similarity (SSIM), are measured as

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_p - Y_t)^2} \\
 MAE &= \frac{1}{N} \sum_{i=1}^N |Y_p - Y_t| \\
 PSNR &= 10 \log_{10} \left( \frac{MAX(Y_t)}{MSE} \right) \\
 SSIM &= \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}
 \end{aligned}
 \tag{6}$$

For FAR, RMAE, and MAE, the smaller the better, while for CSI, POD, HSS, PSNR, and SSIM, the larger the better.

4.3. Quantitative Results

This subsection first presents a quantitative comparison between the proposed MAFormer and other prevalent methods such as UNet [13], DeepLab [33], SegFormer [29], and Swin-Transformer [28]. The results are presented in Table 2. As the former illustrated, the threshold is set to be 30dBZ for classification metrics such as FAR, CSI, POD, and HSS.



Table 2 provides interesting insights into the performance of various methods used for meteorological analysis. The results show that transformer-based methods, such as SegFormer, SwinTransformer, and the proposed MAFormer, outperformed convolution-based methods like DeepLab and UNet. This is because transformer-based methods tend to focus more on global properties, which is a key indicator for large-scale weather processes.

**Table 2.** Comparison between the proposed MAFormer and other state-of-the-art methods. *cls.* and *reg.* denotes classification score and regression score, respectively. The bold values represent the best, while the blue values denote the second best.

Method	<i>cls.</i>				<i>reg.</i>			
	FAR	CSI	POD	HSS	RMSE	MAE	PSNR	SSIM
DeepLab	0.345	0.324	0.390	0.401	7.491	9.658	28.433	0.502
UNet	0.343	0.344	0.419	0.421	7.314	9.477	28.597	0.555
SegFormer	0.367	<b>0.371</b>	<b>0.473</b>	0.441	7.278	9.442	28.629	0.498
Swin	0.345	0.357	0.439	0.434	7.171	9.288	28.773	0.540
MAFormer (ours)	<b>0.327</b>	<b>0.369</b>	<b>0.450</b>	<b>0.451</b>	<b>7.110</b>	<b>9.231</b>	<b>28.826</b>	<b>0.604</b>

Furthermore, among the transformer-based methods, MAFormer stands out by performing better than SegFormer and SwinTransformer in most cases. This could be attributed to the fact that in addition to focusing on global indicators, MAFormer also concentrates on local small-scale characteristics. By incorporating both global and local features, MAFormer can potentially capture a wider range of weather phenomena, leading to more accurate predictions.

The success of transformer-based methods can be attributed to their ability to model long-range dependencies and capture global patterns. Such capabilities are especially useful in the field of weather forecasting where multiple factors and processes interact with one another over large spatial and temporal scales. The use of transformers enables the models to effectively incorporate such interactions and dependencies in their predictions.

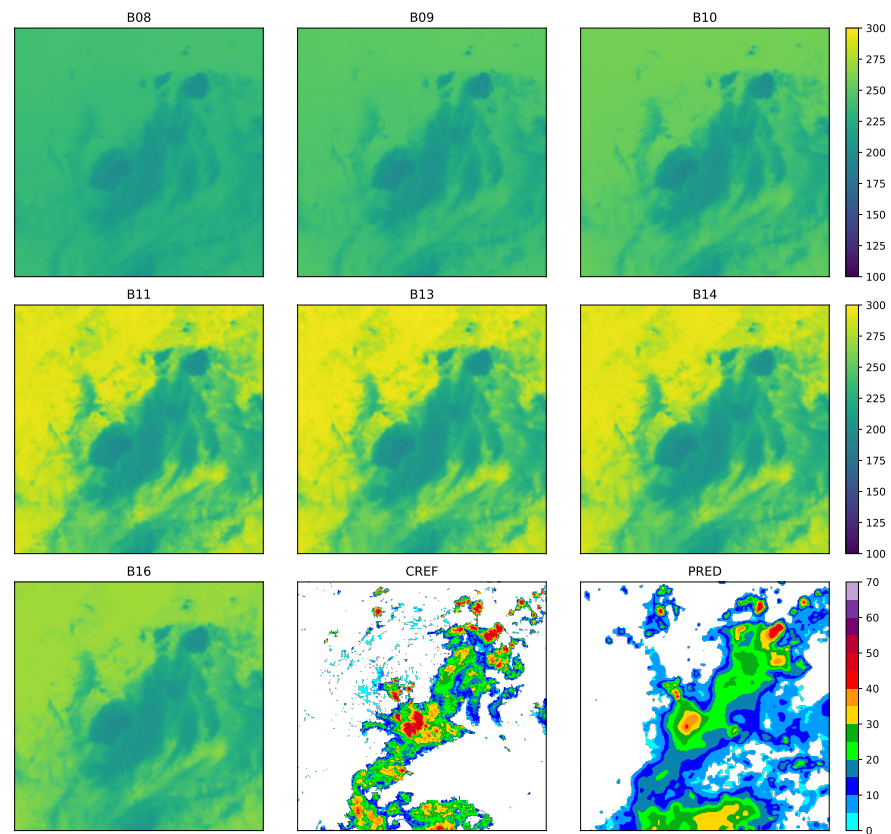
It is important to note that the success of MAFormer is not solely due to its focus on both global and local features. The model architecture and design also play a crucial role in achieving superior performance. For example, MAFormer uses attention mechanisms that allow the model to dynamically assign different weights to different parts of the input sequence, boosting the model's ability to capture important features.

#### 4.4. Qualitative Results

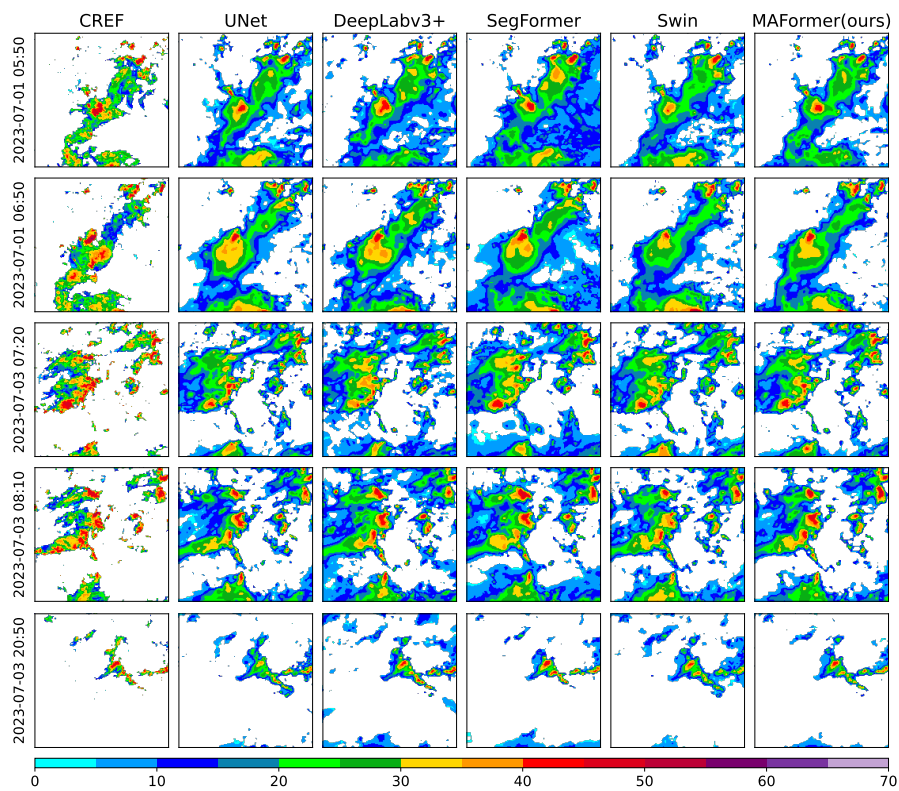
Except for quantitative comparison, this subsection also presents qualitative results for intuitive comparison.

Figure 6 first presents the inputs satellite data for different channels and also the groundtruth/reconstructed radar reflectivity. According to the physical property and display format, we split the input satellite channels into three parts. The first part includes B08, B09, and B10; the second part includes B11, B13, and B14; and the third part includes only B16. Here,  $B_k$  denotes the  $k$ -th channel of satellite. For example, B08 is the 8-th channel, B11 represents the 11-th channel, and so on. CREF stands for the groundtruth composite radar reflectivity. And PRED indicates the model predicted or reconstructed radar reflectivity. From Figure 6, the reconstructed data can roughly depicts radar reflectivity especially larger than 20 dBZ.

Furthermore, Figure 7 shows the reconstructed results of different methods. From this figure, most of the methods can reconstruct radar reflectivity at the large scale. For SegFormer, it tends to predict small values (the blue parts).



**Figure 6.** Illustration about the input satellite data (B08–B16), the groundtruth radar reflectivity (CREF), and the reconstructed data (PRED).



**Figure 7.** An intuitive comparison between the proposed MAFormer and other methods.

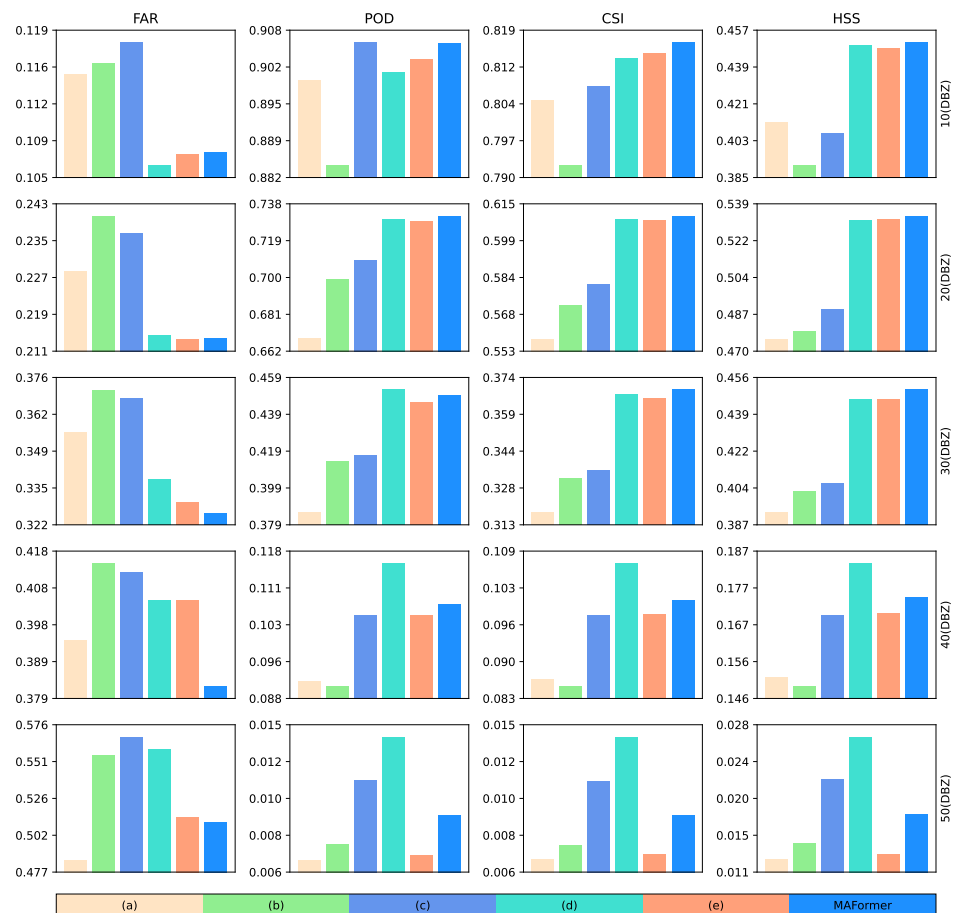
### 5. Discussion

#### 5.1. Analysis of Satellite Channel Importance

As Section 4.4 described, the satellite channels can be divided roughly into three parts. Consequently, this subsection conducts experiments considering channel importance. Table 3 and Figure 8 present the results. Specifically, we use (a)–(e) to denote the six cases by removing certain channels, and the detailed configuration can be found in Table 3.

**Table 3.** Channel importance considering regression metrics (the check mark  $\checkmark$  indicates a channel is selected).

No.	Channel							RMSE	MAE	PSNR	SSIM
	C8	C9	C10	C11	C13	C14	C16				
(a)				$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	9.916	7.712	28.204	0.591
(b)	$\checkmark$	$\checkmark$	$\checkmark$					10.123	7.826	28.024	0.529
(c)		$\checkmark$			$\checkmark$			9.776	7.590	28.328	0.566
(d)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$	9.314	7.171	28.748	0.578
(e)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	<b>9.299</b>	<b>7.161</b>	<b>28.762</b>	<b>0.620</b>



**Figure 8.** Channel importance considering classification metrics.

Comparing the results of different experiments, namely (a) and (b), it is evident that channels C11, C13, and C14 play a more crucial role in MAFormer reconstruction. Additionally, the results obtained from (a), (b), and (e) highlight that channels C8, C9, and C10 are also important. It should be noted that the comparison between (c) and (a)/(b) is not entirely accurate as (c) only removes one channel, creating an imbalance. To rectify

this, experiment (d) also removes one channel. Moreover, the findings suggest that channel C16 holds less significance compared to the other channels. Furthermore, experiment (c) selects three specific channels, and when combined with the results from experiment (e), it indicates that incorporating more channels can benefit radar reconstruction.

5.2. Effectiveness of the Proposed Modules

Basically, the proposed MAFormer for satellite based radar reconstruction consists of two new modules, namely MGAM and ALAM. To evaluate the effectiveness of the proposed two modules, this subsection also conduct quantitative and qualitative experiments.

Tables 4 and 5 first present the quantitative results considering both of classification and regression metrics. The baseline method is set to be SwinTransformer.

**Table 4.** Results of the proposed two modules considering regression metrics. The ↓ and ↑ indicate the metrics are decreasing and increasing, respectively.

	RMSE	PSNR	MAE	SSIM
Swin	9.288	28.773	7.171	0.540
+ ASM	9.273 ↓	28.786 ↑	7.148 ↓	0.586 ↑
+ MSM	9.239 ↓	28.818 ↑	7.131 ↓	0.588 ↑
MA	<b>9.231</b> ↓	<b>28.826</b> ↑	<b>7.110</b> ↓	<b>0.604</b> ↑

**Table 5.** Results of the proposed two modules considering classification metrics. The ↓ and ↑ indicate the metrics are decreasing and increasing, respectively.

	FAR					POD				
	10	20	30	40	50	10	20	30	40	50
Swin	0.110	0.218	0.345	0.419	0.601	<b>0.912</b>	0.731	0.439	0.090	0.006
+ ASM	0.109 ↓	0.215 ↓	0.349	0.403 ↓	0.596 ↓	0.907	0.732 ↑	<b>0.467</b> ↑	0.092 ↑	0.004
+ MSM	0.109 ↓	0.216 ↓	0.342 ↓	0.409 ↓	0.547 ↓	0.909	<b>0.734</b> ↑	0.446 ↑	0.103 ↑	0.008 ↑
MA	<b>0.108</b> ↓	<b>0.214</b> ↓	<b>0.327</b> ↓	<b>0.382</b> ↓	<b>0.510</b> ↓	0.906	0.731	0.450 ↑	<b>0.107</b> ↑	<b>0.009</b> ↑
	CSI					HSS				
	10	20	30	40	50	10	20	30	40	50
Swin	<b>0.819</b>	0.607	0.357	0.085	0.006	0.448	0.528	0.434	0.148	0.011
+ ASM	0.816	0.610 ↑	<b>0.373</b> ↑	0.086 ↑	0.004	0.448	0.532 ↑	0.449 ↑	0.151 ↑	0.007
+ MSM	0.818	<b>0.611</b> ↑	0.362 ↑	0.096 ↑	0.007 ↑	0.450 ↑	0.533 ↑	0.440 ↑	0.167 ↑	0.015 ↑
MA	0.816	0.610 ↑	0.369 ↑	<b>0.101</b> ↑	<b>0.009</b> ↑	<b>0.451</b> ↑	<b>0.533</b> ↑	<b>0.451</b> ↑	<b>0.174</b> ↑	<b>0.018</b> ↑

Starting with regression metrics, which provide insights into the model’s ability to accurately predict continuous numerical values, we found that both modules consistently improved the model’s performance. However, the combination of these two modules showcased a more substantial improvement compared to using each module separately. This highlights the synergistic effect of incorporating multiple modules in enhancing the model’s regression capabilities.

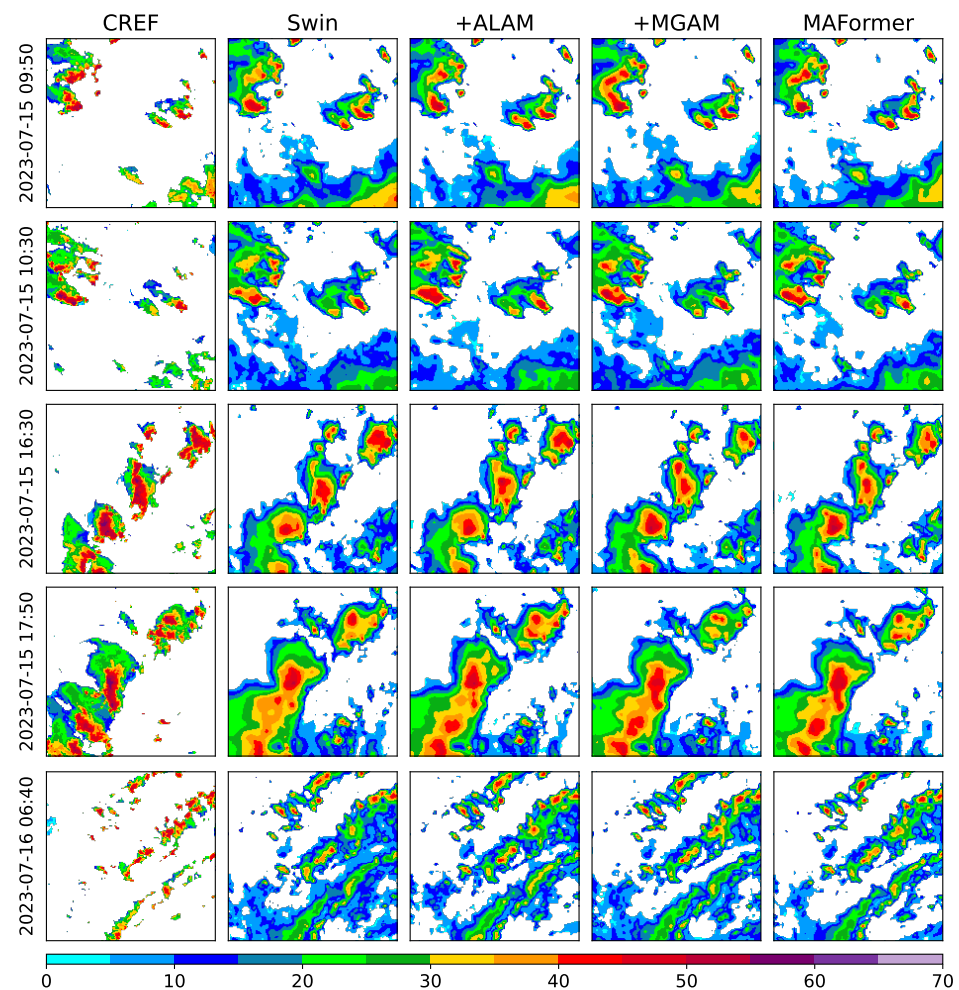
Moving on to classification metrics, which assess the model’s accuracy in classifying weather phenomena into different categories, we conducted evaluations using multiple threshold values, including 10 dBZ, 20 dBZ, 30 dBZ, 40 dBZ, and 50 dBZ. The results obtained were generally consistent with the regression metrics. Interestingly, we observed that the addition of a single module, whether it was ALAM or MGAM, consistently boosted the model’s performance across all thresholds. This indicates the stability and effectiveness of these modules in improving the model’s classification abilities regardless of the chosen threshold.

The improvement in model performance achieved through the addition of these modules can be attributed to several factors. Firstly, the ALAM module leverages attention mechanisms to dynamically allocate weights to different regions of interest, enabling the model to focus on important features and suppress noise or less relevant information. This

helps improve the model's ability to extract meaningful representations from the input data, leading to more accurate predictions.

Additionally, the MGAM module introduces a multi-granular attention mechanism that captures information at different spatial scales. By variably attending to local and global features, the model gains a more comprehensive understanding of the input data, which is particularly valuable in weather forecasting where both local and global characteristics play significant roles. The combination of these mechanisms contributes to the stability and consistent performance improvements observed across different thresholds.

Besides the former quantitative results, Figure 9 also presents qualitative comparison. Combined with the results presented in Tables 4 and 5, we conclude that the newly proposed two modules benefit for satellite based radar reconstruction.



**Figure 9.** An intuitive illustration of the proposed two modules, i.e., ALAM and MGAM.

## 6. Conclusions

The field of weather forecasting and analysis has always been an essential area of scientific research due to its significant impact on human activities and welfare. Radar reflectivity plays a crucial role in observing and understanding the atmospheric conditions, specifically in predicting and tracking severe weather phenomena like thunderstorms, hailstorms, and tornadoes. However, obtaining accurate and reliable radar reflectivity data is often challenging, as it requires complex and expensive instruments and equipment. Hence, researchers have developed new techniques to reconstruct radar reflectivity data from other sources, such as satellite data.

In this paper, a novel approach called the MAFormer is introduced that significantly improves the accuracy and reliability of radar reflectivity data reconstruction using satellite



data. This approach utilizes the transformer framework, which is a widely used neural network architecture that has been proven to be highly effective in various fields, including natural language processing, speech recognition, and image analysis. By leveraging the transformer's architecture, MAFormer can effectively capture and utilize the complex relationships and interactions among different radar and satellite channels, leading to more accurate and reliable results.

Specifically, MAFormer incorporates two newly proposed modules, the Mixup Global Attention Module (MGAM) and the Axial Local Attention Module (ALAM), which further enhance the model's performance. The MGAM explores global similarities between radar and satellite data, while the ALAM focuses on extracting local saliency patterns. Together, these modules enable MAFormer to reconstruct radar reflectivity data from satellite data more effectively and efficiently. Experimental results demonstrate the superiority of the proposed method.

In summary, the MAFormer approach represents a significant advancement in the field of radar reflectivity reconstruction, enabling more accurate and widespread predictions of severe weather phenomena. The utilization of the transformer framework and the two newly proposed modules enhances the model's ability to capture and utilize complex relationships and interactions among different radar and satellite channels. These advancements have the potential to revolutionize weather forecasting and analysis, leading to better preparation and mitigation measures against severe weather phenomena.

It is important to note that further investigations and evaluations are required to fully assess the performance and generalizability of MAFormer. Future research directions include exploring additional data fusion techniques, refining the attention mechanisms, and conducting extensive experiments on various real-world datasets. These advancements will significantly contribute to the field of radar reflectivity reconstruction, enabling more accurate and widespread predictions of severe weather phenomena.

**Author Contributions:** Conceptualization, T.Y.; methodology, K.W.; software, Z.L.; validation, K.W. and Y.C.; formal analysis, T.Y.; investigation, K.W. and Q.K.; resources, Y.C.; data curation, K.W.; writing—original draft preparation, T.Y. and Y.H.; visualization, K.W. and Y.H.; supervision, T.Y.; project administration, T.Y.; funding acquisition, T.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Science Foundation of China No. 62106270. and the Joint Research Project for Meteorological Capacity Improvement No. 23NLTSQ014.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The satellite data and radar data are available at <http://idata.cma/cmadaas/> (accessed on 8 September 2023), China Meteorological Administration Data as a Service (CMADaaS).

**Conflicts of Interest:** Author Qiuming Kuang was employed by the company Geovis Environment Technology Co., Ltd., the company didn't participate the study. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The company didn't participate the study.

## Abbreviations

The following abbreviations are used in this manuscript:

ALAM	Axial Local Attention Module
MGAM	Mixup Global Attention Module
MAFormer	Transformer with MGAM and ALAM
FAR	False Alarm Rate
POD	Probability Of Detection
CSI	Critical Success Index
HSS	Heidke Skill Score



MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
PSNR	Peak Signal Noise Ratio
SSIM	Structure SIMilarity
CREF	Composite radar Reflectivity
PRED	Predicted reflectivity

## References

1. Yu, T.; Yang, R. Temporal Dynamic Network With Learnable Coupled Adjacent Matrix for Wind Forecasting. *IEEE Geosci. Remote. Sens. Lett.* **2023**, *20*, 1001605. [[CrossRef](#)]
2. Yu, T.; Kuang, Q.; Zheng, J.; Hu, J. Deep Precipitation Downscaling. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1001405. [[CrossRef](#)]
3. Yu, T.; Yang, R.; Huang, Y.; Gao, J.; Kuang, Q. Terrain-Guided Flatten Memory Network for Deep Spatial Wind Downscaling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2022**, *15*, 9468–9481. [[CrossRef](#)]
4. Yu, T.; Kuang, Q.; Hu, J.; Zheng, J.; Li, X. Global-Similarity Local-Saliency Network for Traffic Weather Recognition. *IEEE Access* **2021**, *9*, 4607–4615. [[CrossRef](#)]
5. Zhang, F.; Yu, T.; Li, Z.; Wang, K.; Chen, Y.; Huang, Y.; Kuang, Q. Deep Quantified Visibility Estimation for Traffic Image. *Atmosphere* **2022**, *14*, 61. [[CrossRef](#)]
6. Jena, K.K.; Bhoi, S.K.; Nayak, S.R.; Panigrahi, R.; Bhoi, A.K. Deep Convolutional Network Based Machine Intelligence Model for Satellite Cloud Image Classification. *Big Data Min. Anal.* **2023**, *6*, 32–43. [[CrossRef](#)]
7. Han, L.; Sun, J.; Zhang, W. Convolutional Neural Network for Convective Storm Nowcasting Using 3-D Doppler Weather Radar Data. *IEEE Trans. Geosci. Remote. Sens.* **2020**, *58*, 1487–1495. [[CrossRef](#)]
8. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.; Wong, W.; Woo, W. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 802–810.
9. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.; Wong, W.; Woo, W. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5617–5627.
10. Yu, T.; Kuang, Q.; Yang, R. ATMConvGRU for Weather Forecasting. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1003805. [[CrossRef](#)]
11. All convolutional neural networks for radar-based precipitation nowcasting. *Procedia Comput. Sci.* **2019**, *150*, 186–192. [[CrossRef](#)]
12. 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.
13. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect. Notes Comput. Sci.* **2015**, *9351*, 234–241.
14. Hernández, E.; Sánchez-Anguix, V.; Julián, V.; Cámara, J.P.; Duque, N.D. Rainfall Prediction: A Deep Learning Approach. *Lect. Notes Comput. Sci.* **2016**, *9648*, 151–162.
15. Lebedev, V.; Ivashkin, V.; Rudenko, I.; Ganshin, A.; Molchanov, A.; Ovcharenko, S.; Grokhovetskiy, R.; Bushmarinov, I.; Solomentsev, D. Precipitation Nowcasting with Satellite Imagery. In Proceedings of the International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2680–2688.
16. Qiu, M.; Zhao, P.; Zhang, K.; Huang, J.; Shi, X.; Wang, X.; Chu, W. A Short-Term Rainfall Prediction Model Using Multi-task Convolutional Neural Networks. In Proceedings of the International Conference on Data Mining, New Orleans, LA, USA, 18–21 November 2017; pp. 395–404.
17. Klocek, S.; Dong, H.; Dixon, M.; Kanengoni, P.; Kazmi, N.; Lufarenko, P.; Lv, Z.; Sharma, S.; Weyn, J.A.; Xiang, S. MS-nowcasting: Operational Precipitation Nowcasting with Convolutional LSTMs at Microsoft Weather. *arXiv* **2021**, arXiv:2111.09954.
18. Sønderby, C.K.; Espeholt, L.; Heek, J.; Deghani, M.; Oliver, A.; Salimans, T.; Agrawal, S.; Hickey, J.; Kalchbrenner, N. MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv* **2020**, arXiv:2003.12140.
19. Espeholt, L.; Agrawal, S.; Sønderby, C.K.; Kumar, M.; Heek, J.; Bromberg, C.; Gazen, C.; Hickey, J.; Bell, A.; Kalchbrenner, N. Skillful Twelve Hour Precipitation Forecasts using Large Context Neural Networks. *arXiv* **2021**, arXiv:2111.07470.
20. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. Skillful precipitation nowcasting using deep generative models of radar. *Nature* **2021**, *597*, 672–677. [[CrossRef](#)] [[PubMed](#)]
21. Kuang, Q.; Yu, T. MetPGNet: Meteorological Prior Guided Network for Temperature Forecasting. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1004305. [[CrossRef](#)]
22. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.
23. Duan, M.; Xia, J.; Yan, Z.; Han, L.; Zhang, L.; Xia, H.; Yu, S. Reconstruction of the Radar Reflectivity of Convective Storms Based on Deep Learning and Himawari-8 Observations. *Remote. Sens.* **2021**, *13*, 3330. [[CrossRef](#)]
24. Zhu, M.; Liao, Q.; Wu, L.; Zhang, S.; Wang, Z.; Pan, X.; Wu, Q.; Wang, Y.; Su, D. Multiscale Representation of Radar Echo Data Retrieved through Deep Learning from Numerical Model Simulations and Satellite Images. *Remote Sens.* **2023**, *15*, 3466. [[CrossRef](#)]

25. Yang, L.; Zhao, Q.; Xue, Y.; Sun, F.; Li, J.; Zhen, X.; Lu, T. Radar Composite Reflectivity Reconstruction Based on FY-4A Using Deep Learning. *Sensors* **2023**, *23*, 81. [[CrossRef](#)] [[PubMed](#)]
26. Lagerquist, R.; Stewart, J.Q.; Ebert-Uphoff, I.; Kumler, C. Using deep learning to nowcast the spatial coverage of convection from Himawari-8 satellite data. *Mon. Weather. Rev.* **2021**, *149*, 3897–3921.
27. 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, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008. [[CrossRef](#)]
28. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002. [[CrossRef](#)]
29. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Álvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. 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; pp. 12077–12090.
30. Ho, J.; Kalchbrenner, N.; Weissenborn, D.; Salimans, T. Axial Attention in Multidimensional Transformers. *arXiv* **2019**, arXiv:1912.12180.
31. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 6–9 May 2019.
32. 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.
33. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018—15th European Conference, Munich, Germany, 8–14 September 2018; Volume 11211, pp. 833–851. [[CrossRef](#)]

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