Response to Reviewer #1

We sincerely thank Reviewer #1 for the careful evaluation of our manuscript and for the valuable comments and suggestions. We have carefully considered all of your comments and have revised the manuscript accordingly. Your feedback has been very helpful in improving the rigor, completeness, and clarity of our work, and we are grateful for your thoughtful suggestions.

In response to your review, we have prepared point-by-point responses in the same order as your comments. All of the reviewer's comments and concerns have been carefully considered and studied during the revision process, and these efforts are reflected in both our responses and the revised manuscript. For ease of reference, the revised manuscript has been prepared entirely in track changes mode so that the changes made can be identified clearly.

Our detailed responses are provided below.

Major Concerns

1.Methodological Clarity and Reproducibility.

The description of the Multidimensional Reconstruction Convolution module (Section 2.1) is detailed but lacks clarity in places. For instance, the derivation of the gating mechanism using γ and the thresholding operation is not fully explained. How is the threshold determined? Is it learned or fixed?

Response:

Thank you for pointing out this issue. We agree that the explanation in this part of the manuscript was not sufficiently detailed, especially regarding the setting of the threshold.

Here, the threshold is a fixed empirical hyperparameter rather than a learnable parameter optimized during training. In our experiments, this threshold was fixed at 0.5. The rationale for this choice is that, after being processed by the Sigmoid function, the correlation weights are mapped into a continuous probability interval of

(0, 1). Setting the threshold at 0.5, the midpoint of this interval, is an intuitive and effective choice, as it enables a balanced division of features into two categories, namely information-rich and redundant, for subsequent processing.

In addition, the Spatial Reconstruction Unit (SRU) does not simply discard low-weight features. Instead, it divides the features into two parts, W1 and W2, and then performs cross-reconstruction. In this way, both high-information and low-information features participate in the subsequent reconstruction process, rather than one part being completely removed. Under such a structure, if the threshold is set too low, for example 0.2, most features will fall into the "informative" branch, thereby weakening the separation effect. Conversely, if the threshold is set too high, for example 0.8, too many features will be assigned to the "non-informative" branch, which may lead to the loss of useful information. Therefore, the choice of 0.5 can provide sufficient separation while avoiding excessive sparsity.

We have added the corresponding explanation in Section 2.1 of the revised manuscript to make the presentation and justification clearer.

2.Dataset Description and Preprocessing

Section 3.2 introduces the radar echo dataset but lacks critical details: How were missing data or noise handled during quality control?

The Moving-MNIST dataset is described, but no details are given on how it was generated (e.g., number of digits per frame, velocity ranges). This is necessary for reproducibility.

Response:

The dataset used in this study was derived from meteorological radar and automatic weather station observational data collected in Jiangsu Province during April to September from 2019 to 2021, and was produced and provided by the Jiangsu Meteorological Observatory. This dataset consists of 3-km constant-altitude plan position indicator (CAPPI) base reflectivity products obtained after quality control and mosaicking of multiple S-band weather radars across Jiangsu Province. The data values range from 0 to 70 dBZ, with a horizontal resolution of approximately 1 km and a temporal resolution of 6 minutes. Since the radar echo products used in this work had already undergone quality control in the upstream processing pipeline by the data provider, missing data and obvious noise had already been handled and removed before the generation of PNG images and sequential CSV files. Therefore, in the training code of this study, only normalization was performed, and no additional quality control procedure was repeated.

The Moving-MNIST dataset was generated based on the original MNIST handwritten digit images. Specifically, two handwritten digits were first randomly selected from the MNIST test set and then randomly placed on a 64 × 64 grayscale canvas. A random initial velocity vector was assigned to each digit. The digits were then moved frame by frame according to their assigned velocities. When a digit reached the boundary of the canvas, it bounced back by reversing the corresponding component of its velocity vector. The two digits were rendered together in the same frame with overlapping allowed. This process was repeated for 20 frames to form a complete video sequence. Each frame contained exactly two digits, which is the standard setting of the dataset. In terms of motion speed, the initial velocity of each digit was uniformly sampled in both the horizontal and vertical directions from a range of approximately ±2 to ±6 pixels per frame, and the speed magnitude remained constant throughout the sequence, with only the direction reversed after boundary collisions.

The above information has been incorporated into Sections 3.1 Moving-MNIST dataset and 3.2 Radar echo dataset of the revised manuscript, and we hope these details will be helpful for reproducing the results.

3.Experimental Setup

The training/testing split (21,103 vs. 5,275 images) is provided, but the temporal continuity of the radar data is not discussed. Were sequences sampled randomly or chronologically? Random sampling could lead to data leakage if consecutive frames are split across sets.

The choice of batch size (4) and learning rate (0.0001) is stated, but no justification or ablation is provided.

The loss function used for training is not mentioned. Is it MSE, a combination of MSE and adversarial loss, or something else? This is essential for understanding the optimization process.

Response:

We have supplemented and clarified the relevant content in response to each of the points raised, as detailed below.

Regarding the temporal continuity of the radar data and the potential risk of data leakage, we divided the training set and test set according to chronological order rather than random sampling. In the experimental code, the training set and test set were determined by two predefined files, dataset_train.csv and dataset_testA.csv, respectively. The corresponding frame index ranges of these two files do not overlap: the training set consists of earlier radar image sequences, while the test set corresponds to later sequences. Therefore, consecutive frames were not assigned simultaneously to the training and test sets, which effectively avoids data leakage caused by overlap between temporally adjacent samples. This clarification regarding dataset partitioning has been added to Section 3.2 Radar echo dataset of the revised manuscript.

In our experiments, the batch size was set to 4 and the initial learning rate was set to 0.0001. These parameters were not chosen arbitrarily, but were determined based on multiple rounds of experimental testing, considering training stability, convergence performance, and computational resource constraints. Since radar echo prediction involves relatively high spatial resolution and long temporal input sequences, using a larger batch size would significantly increase GPU memory consumption and could negatively affect training stability. Therefore, adopting a small batch size is a practical and reasonable choice. The learning rate was also selected after repeated trials to ensure a stable training process. In addition, the learning rate was not kept constant throughout training. We employed an exponential decay strategy (ExponentialLR), under which the learning rate was multiplied by a decay factor after a certain number of iterations during the later training stage, enabling finer parameter updates and improving convergence stability.

The main focus of this study is on improving the model architecture and enhancing its performance in radar echo prediction. Therefore, the manuscript mainly reports the final experimental parameter configuration, rather than providing an extensive discussion of the hyperparameter selection process. Following the reviewer's suggestion, we have added a brief but necessary explanation at the beginning of Section 4. Experimental Procedure and Results in the revised manuscript to make this part clearer and more complete.

In addition, regarding the loss function, we agree that the original manuscript lacked a clear description, and we sincerely appreciate the reviewer for identifying this important issue. In this study, the training objective was the mean squared error (MSE, or L2 loss). In the implementation, L1 loss was also computed as an auxiliary monitoring metric during training; however, backpropagation and parameter updates were performed solely based on the MSE loss. No adversarial loss or other composite loss terms were introduced. This information has also been explicitly added at the beginning of Section 4. Experimental Procedure and Results in the revised manuscript so that readers can better understand the model optimization process.

4.Evaluation Metrics.

The use of CSI and HSS is appropriate and well explained. However, the manuscript does not discuss whether these metrics are computed on a pixel-wise basis or after some form of spatial smoothing.

Statistical significance testing is absent. Given the stochastic nature of deep learning training, it is important to report confidence intervals or perform significance tests (e.g., bootstrap or t-tests) to ensure that observed improvements are not due to random initialization.

Response:

In our implementation, the CSI and HSS metrics were computed directly on a pixel-by-pixel (grid-point) basis. Specifically, the observed and predicted precipitation fields were first binarized at each grid point according to a predefined threshold, after which TP, FP, FN, and TN were counted over all grid points. The CSI and HSS scores were then calculated using their standard formulas. No additional spatial smoothing or convolutional filtering was applied to the precipitation fields before computing these metrics. Therefore, the reported results reflect forecast performance at the original grid scale. This clarification regarding the computation of CSI and HSS has been added to Section 4.2.1 Evaluation indicators of the revised manuscript.

Regarding the lack of statistical significance analysis, we appreciate the reviewer's valuable suggestion. In response, we decided to supplement the manuscript with confidence intervals for the CSI and HSS results. Specifically, we repeated the experiments for MRDF-Net using five different random seeds, and the results have been reported in the form of mean ± standard deviation, so as to more clearly reflect the variability of the model under different random initializations. The corresponding revisions have been incorporated as follows:

(1)In Table 3, Table 4, and Table 5, the CSI and HSS results for the 1-hour forecast have been supplemented with confidence intervals in the form of mean ± standard deviation.

(2)In Figure 8 and Figure 9, the corresponding confidence intervals have also been added for the CSI and HSS results under the 2-hour forecast setting.

5.Comparative Analysis.

The comparison with ViViT-Prob, MF-UFNO, and RadarDiT in Section 4.2.4 is valuable, but the discussion is brief. Why does RadarDiT outperform MRDF-Net at lower thresholds? Is it due to the diffusion process better capturing large-scale patterns? This could be explored further.

The authors claim that MRDF-Net reduces echo weakening, but this is only shown qualitatively in Figures 7 and 10. Quantitative measures of echo intensity preservation (e.g., intensity histograms or peak intensity error) would strengthen this claim.

Response:

We agree that the performance differences between RadarDiT and MRDF-Net under different reflectivity thresholds deserve further discussion, and we have expanded the analysis in Section 4.2.4 accordingly.

Following the reviewer's suggestion, we have added a more detailed analysis in Section 4.2.4 Comparison with Recent State-of-the-Art Models. Based on the experimental results, different models may place different emphases on modeling echo regions of different intensities. RadarDiT performs better at lower thresholds, which suggests that it has certain advantages in predicting low-echo and weak-echo regions. In contrast, MRDF-Net achieves better results at higher thresholds, indicating that it is more effective in forecasting high-echo and heavy-precipitation regions. From the perspective of operational applications, heavy-precipitation areas are often associated with higher-risk weather processes. Therefore, accurately predicting high-echo regions is of more direct practical importance, and this is precisely where MRDF-Net shows its advantage.

As for why RadarDiT performs better at lower thresholds while MRDF-Net excels at higher thresholds, we believe this reflects a fundamental difference in the modeling priorities of the two architectures. On one hand, RadarDiT, as a diffusion-based generative model, is inherently capable of producing spatially diverse outputs, which may confer an advantage in capturing the broad spatial distributions characteristic of weak-echo regions at low reflectivity thresholds. On the other hand, MRDF-Net is specifically designed around a multidimensional reconstruction convolution mechanism with dynamic gating, which prioritizes the accurate reconstruction of localized high-intensity echo structures. This architectural focus naturally leads to stronger performance under high reflectivity thresholds, at the potential cost of some sensitivity to low-intensity regions.

We also agree with the reviewer's suggestion that the claim that MRDF-Net can alleviate echo attenuation would benefit from more direct quantitative support. We would like to further clarify that this conclusion in the original manuscript was not based solely on the qualitative results shown in Figure 7 and Figure 10, but rather on a combination of quantitative evaluations using CSI/HSS under multiple thresholds and visual analysis. Specifically, Figure 6 already presents the curves of CSI and HSS over forecast lead time under different reflectivity thresholds (10, 20, and 40 dBZ) for the 1-hour forecast setting. These results clearly show that MRDF-Net maintains superior performance over time compared with other models, especially under the 40 dBZ threshold, which already demonstrates that MRDF-Net is more capable of mitigating echo attenuation than the competing models. Figure 7 and Figure 10 then provide further intuitive support for these quantitative findings from the perspective of spatial structure and morphological preservation.

Minor Issues

1.Abstract: The phrase "an area typically addressed by radar nowcasting" is vague. Consider rephrasing to "a task commonly addressed through radar-based nowcasting."

Response:

Revised as suggested.

2.Line 48: "high effective" should be "highly effective.

Response:

Corrected as suggested.

3.Line 120: "while and weights" contains a typo; should be "while weights."

Response:

Corrected as suggested.

4.Line 329: "misses, and misses" is redundant; should be "hits, false alarms, and misses."

Response:

Corrected as suggested.

5.Line 432: The phrase "a 6.07% increase over PredRNN" is ambiguous—clarify whether this is absolute or relative improvement.

Response:

The phrase “a 6.07% increase over PredRNN” referred to a relative improvement. This has been clarified in the revised manuscript. In addition, there were several similar expressions in the same paragraph, and we have revised them all consistently to avoid ambiguity.

6.Line 465: "excels at better preserving" is redundant; "excels at preserving" suffices.

Response:

The redundant phrase “excels at better preserving” has been addressed. Since the paragraph containing this sentence was rewritten during other revisions, this wording issue no longer exists in the revised manuscript.

7.References: Some references are incomplete or inconsistently formatted (e.g., missing journal names, volume numbers). Please ensure all follow the journal’s style guide.

Response:

We have carefully checked and revised all references throughout the manuscript to ensure consistency with the AMT journal formatting requirements. The revisions include standardizing author name formats, adjusting the placement of publication years, adding missing DOIs, correcting journal title abbreviations, and unifying the formatting of conference papers and preprint entries according to the journal’s style guidelines.