Response to Reviewer 2 Comments

We sincerely thank the reviewer for your time, constructive feedback, and valuable suggestions, which greatly improved the quality of the manuscript. We carefully considered each comment and made corresponding revisions to the manuscript. The following is a point by point response to the reviewer's comments.

Comments and Suggestions for Authors:

1. The manuscript focuses on the incorporation of CBAM into a traditional U-Net architecture and compares with other off-the-shelf architectures. However, there is not much discussion of hyperparameter details of their chosen architecture and there is no discussion of the hyperparameters chosen for the competing models. If the architecture of these models is to remain the focus of this manuscript, these details (e.g. number of convolutional filters, block design, batch size, learning rate, optimizer, etc...) need to be listed and discussed. Ideally there would be some common point of comparison between these models controlling for sample throughput, number of trainable parameters, or compute usage. Currently, readers cannot conclude much from the given results since hyperparameters could vary greatly between each of these models.

Response:We sincerely thank the reviewer for this valuable comment. We agree that detailed descriptions of model configurations and training settings are essential for improving reproducibility and for enabling a fair interpretation of the comparative experiments.

In the revised manuscript, we have therefore expanded the description of the proposed CBAM-U-Net architecture. The encoder–decoder configuration, the number of convolutional filters at different feature levels, and the integration strategy of the CBAM modules are now explicitly described. We have also provided the main training hyperparameters, including the optimizer, learning rate, weight decay, learning-rate scheduler, batch size, maximum number of training epochs, and early-stopping criteria. These additions are intended to make the implementation of the proposed model more transparent and reproducible.

For the comparative models, including U-Net, Attention U-Net, MCDA-U-Net, SegFormer, and SRViT, we have also clarified their main architectural configurations and the corresponding experimental settings. The models were implemented according to their respective architectural designs and adapted to the same seven-channel satellite-to-radar reflectivity reconstruction task. To make the comparison more transparent, the revised manuscript now reports the relevant model-specific settings while maintaining the same dataset partition and evaluation protocol across the experiments.

In addition to the architectural and training hyperparameters, we agree with the reviewer that a common quantitative measure of model complexity is useful for assessing whether the observed performance differences are simply associated with differences in model capacity or computational cost. We therefore further calculated the number of trainable parameters and the computational complexity of all evaluated models under the same input configuration of 1×7×301×214. The comparison is summarized below.

Model	Params	GFLOPs
U-Net	31.040	93.061
Attention U-Net	31.391	95.159
MCDA-U-Net	3.457	13.692
SegFormer	2.588	10.819
SRViT	5.032	5.376
CBAM-U-Net	31.256	93.077

This comparison is particularly informative for assessing the additional cost introduced by CBAM relative to its U-Net backbone. Under the same input configuration of 1×7×301×214, the standard U-Net contains 31.040 M trainable parameters and requires approximately 93.061 GFLOPs for a single forward pass, whereas CBAM-U-Net contains 31.256 M parameters and requires approximately 93.077 GFLOPs. Thus, incorporating CBAM increases the number of trainable parameters by only approximately 0.70% and the computational complexity by approximately 0.02%, indicating that CBAM introduces only marginal additional overhead relative to the original U-Net backbone.

We acknowledge that the evaluated architectures have inherently different computational characteristics and that the primary objective of this study is not to establish a comprehensive computational-efficiency benchmark among different network families. Nevertheless, by adding both the implementation details and the quantitative comparison of Params and GFLOPs, the revised manuscript now provides a clearer common basis for interpreting the relationship between model complexity and retrieval performance. These revisions have been incorporated into the revised manuscript (Page 13–14, Lines 360–406;).

3. Model

3.1. Model Architecture and Experimental Setup

U-Net-style encoder-decoder architectures have been widely used for meteorological field reconstruction due to their ability to extract multi-scale spatial features and preserve spatial structures.(Ronneberger et al., 2015; Yu et al., 2023b; Veillette et al., 2020) However, severe convective echoes exhibit highly complex structural characteristics, typically featuring pronounced central intensity gradients and intricate edge textures. For the present satellite-to-radar reconstruction task, localized high-intensity echo structures remain challenging to reproduce accurately. This motivates the investigation of additional feature-recalibration mechanisms to improve the representation of spectrally and spatially informative features. Various attention-based architectures have been introduced to enhance feature representation in meteorological applications. For example, Attention U-Net has been applied to multichannel Doppler-radar precipitation recognition to emphasize informative features (Chen et al., 2023), while dual-attention recurrent networks have

been developed to integrate complementary satellite infrared and radar information (Wang et al., 2025). Transformer-based architectures have also been explored for estimating radar reflectivity from satellite observations by modeling broader spatial dependencies (Stock et al., 2024). These approaches demonstrate the potential of attention mechanisms from different perspectives. For the present multispectral satellite-to-radar reconstruction task, both discrimination among different infrared channels and spatial localization of convective echoes are important. CBAM sequentially applies channel and spatial attention within a lightweight convolutional module (Woo et al., 2018), and has also been introduced into weather-radar echo correction tasks (Wang et al., 2026). Therefore, CBAM was selected as a task-oriented feature-recalibration module because it jointly considers channel-wise and spatial feature importance while remaining convenient to integrate into the U-Net backbone.

Based on the above considerations, this study adopts a lightweight CBAM-enhanced U-Net architecture for the satellite-to-radar reflectivity reconstruction task. By seamlessly integrating the Convolutional Block Attention Module (CBAM) into the standard U-Net backbone, this framework is specifically engineered to process the 7-channel input tensor derived from FY-4A/AGRI observations (comprising five independent infrared bands and two composite brightness temperature difference features). The primary objective is to establish an end-to-end nonlinear mapping from multidimensional satellite cloud-top characteristics to the target Radar Composite Reflectivity (CR).

For the overall May–November 2023 experiment, the FY-4A samples were randomly partitioned at the individual-sample level into training, validation, and test subsets, accounting for approximately 70%, 10%, and 20% of the samples, respectively. For the three period-specific experiments, the samples within each corresponding period were separately partitioned following the same strategy, and the models were independently trained and evaluated for each period.

The proposed CBAM-U-Net follows the standard encoder-decoder structure of U-Net. The encoder consists of four convolutional stages with 64, 128, 256, and 512 filters, respectively, and symmetric decoder blocks are used for feature reconstruction. The CBAM module is inserted after the convolutional feature extraction blocks to perform sequential channel and spatial feature recalibration.The model was optimized using AdamW with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. A ReduceLROnPlateau scheduler reduced the learning rate by a factor of 0.5 when the validation loss plateaued for 10 epochs. Training used a batch size of 4 for a maximum of 150 epochs, with early stopping (patience 10) based on validation loss to prevent overfitting. All experiments were conducted on a single NVIDIA GeForce RTX 4090 GPU with PyTorch 2.0 and CUDA 11.8.The comparative models were implemented following their original architectural designs and were trained using the same dataset partition, loss function, optimizer, and training schedule for consistency.

2. The manuscript is not well-referenced and largely ignores previous literature in the field that explores the use of geostationary imager observations to emulate composite reflectivity from ground-based measurements. I suggest that this article includes a detailing of previous methods used, a discussion of their deficiencies, and how the proposed CBAM-U-Net differs from them.

Response:Thank you for this important comment. We agree that the use of CNN or U-Net-based architectures for retrieving precipitation intensity or radar-reflectivity-related quantities from geostationary multispectral observations is not new, and we have revised the manuscript to avoid implying otherwise. The novelty of the present study is therefore not claimed to lie in the general use of CNNs for satellite-based retrieval, nor in the CBAM module itself.

Instead, we have clarified that the contribution of this work lies in addressing a more specific problem that remains insufficiently examined: the reconstruction of localized high-intensity radar echoes over complex terrain using geostationary multispectral observations. In this context, we investigate whether a lightweight cascaded channel--spatial attention strategy within a U-Net backbone can improve the representation of localized severe echoes while retaining a relatively simple convolutional architecture.

To substantiate this distinction quantitatively, CBAM-U-Net is evaluated under the same training protocol against U-Net, Attention U-Net, MCDA-U-Net, SegFormer, and SRViT.The newly revised location is page 19 line 517 to 533. Although the domain-averaged RMSE improvement over U-Net is moderate (7.0043 to 6.8290 dBZ, approximately 2.5%), the improvement is considerably more pronounced for severe echoes. In the 45-70 dBZ interval, the POD increases from 0.2617 to 0.5296, the CSI from 0.3141 to 0.4384, and the FAR decreases from 0.5954 to 0.4357. In addition, we have added a strong-echo-specific intensity-error analysis for observed reflectivity≥45 dBZ. Compared with U-Net, CBAM-U-Net reduces MAE from 10.3802 to 9.3923 dBZ, RMSE from 14.9184 to 13.8787 dBZ, and the magnitude of the negative bias from 9.8769 to 8.6382 dBZ. Among all six evaluated models, CBAM-U-Net achieves the lowest strong-echo MAE and RMSE and the smallest magnitude of negative bias.The newly revised location is page 20 line 563 - 579.

4.1.3. *Strong-echo-specific intensity error analysis*

To further examine model performance for severe echoes beyond the categorical verification in Table 6, continuous intensity-error statistics were calculated specifically for pixels with observed composite reflectivity $Z_{obs} \geq 45$ dBZ. Unlike the 45–70 dBZ interval-based verification, this analysis retains observed severe-echo pixels regardless of the corresponding predicted intensity, thereby directly evaluating the ability of each model to reproduce high-reflectivity values and ensuring that substantial underestimation is explicitly reflected in the error statistics.

Table 7: Intensity-error statistics for pixels with observed composite reflectivity $Z_{obs} \geq 45$ dBZ.

Model	MAE_{45+} (dBZ)	$RMSE_{45+}$ (dBZ)	$Bias_{45+}$ (dBZ)
U-Net	10.3802	14.9184	-9.8769
Attention U-Net	10.0938	15.0011	-9.1911
MCDA-U-Net	12.5021	16.6286	-12.1219
SegFormer	13.4136	16.6481	-13.1976
SRViT	19.4452	21.5531	-19.3940
CBAM-U-Net	**9.3923**	**13.8787**	**-8.6382**

As shown in Table 7, all six models exhibit negative biases for observed severe echoes, indicating a general tendency to underestimate high-reflectivity intensities. Among all compared models, CBAM-U-Net achieves the lowest MAE and RMSE and the smallest magnitude of negative bias.

Compared with U-Net, the MAE decreases from 10.3802 to 9.3923 dBZ, corresponding to a reduction of approximately 9.5%, while the RMSE decreases from 14.9184 to 13.8787 dBZ, corresponding to a reduction of approximately 7.0%. In addition, the absolute bias decreases from 9.8769 to 8.6382 dBZ, corresponding to a reduction of approximately 12.5%. These results further demonstrate that CBAM-U-Net improves the reconstruction of high-intensity echoes and alleviates the systematic underestimation of severe-echo intensity.

The introduction (Page4 line119 - 133), experimental setup and data consistency (Page30 line784 - 785), and Conclusion (Page31-32 line811- 870) have been rewritten to more clearly compare this study with previous CNN based satellite retrieval work. The revised manuscript no longer presents CNN-based retrieval itself as the novelty; instead, it emphasizes the complex-terrain application, the targeted lightweight spectral--spatial enhancement of localized severe echoes, and the multi-perspective evaluation of severe-echo robustness and cross-sensor transferability.

(1) **Problem-oriented reconstruction under complex-terrain and severe-convection conditions**: Rather than revisiting the general feasibility of CNN-based satellite precipitation retrieval, this study focuses on radar composite reflectivity reconstruction over the complex terrain of Sichuan, with particular emphasis on localized high-intensity convective echoes.

(2) **Lightweight spectral–spatial enhancement of localized severe echoes**: A cascaded channel–spatial attention strategy is integrated into the U-Net backbone to sequentially recalibrate multispectral feature importance and spatially emphasize localized high-reflectivity structures. The objective is not to claim the CBAM mechanism itself as novel, but to investigate its task-specific value for mitigating severe-echo smoothing and intensity underestimation in satellite-to-radar reconstruction.

(3) **Multi-perspective evaluation of severe-echo reconstruction capability and cross-sensor transferability**: Beyond conventional domain-averaged metrics, the proposed framework is evaluated using intensity-stratified categorical verification, strong-echo-specific intensity errors for observed reflectivity ≥ 45 dBZ, different flood-season precipitation regimes, and an independent zero-shot FY-4B test.

The FY-4B experiment was conducted as a strict zero-shot cross-sensor evaluation. Neither model parameters nor normalization statistics were adapted using FY-4B data; all model weights and preprocessing statistics were inherited directly from the FY-4A training stage.

5. Conclusion

This study investigated radar composite reflectivity reconstruction from FY-4A/AGRI multispectral observations over the complex terrain of Sichuan Province, with particular emphasis on the reconstruction of localized high-intensity convective echoes. Rather than treating CNN-based satellite-to-radar retrieval as a new paradigm, this study examined the task-specific value of incorporating lightweight sequential channel–spatial feature recalibration into a U-Net backbone. The resulting CBAM-U-Net sequentially applies channel and spatial attention to enhance multispectral feature selection and spatial representation while retaining the basic encoder–decoder structure of U-Net.

Evaluated on an extensive dataset from the 2023 extended flood season in the Sichuan region, CBAM-U-Net demonstrated improvements in both retrieval accuracy and structural fidelity.It achieved the lowest pixel-wise errors, the highest PSNR (25.2701 dB), and the highest SSIM (0.7894) among the evaluated models.

Although the reduction in domain-averaged RMSE relative to U-Net was moderate, the advantage became more pronounced for severe echoes.

Beyond overall statistical metrics, the proposed model showed improved performance in identifying and reconstructing severe echo regions. Threshold-based verification revealed that CBAM-U-Net consistently outperformed baselines in the strong echo region (45–70 dBZ), achieving the highest Probability of Detection (POD = 0.5296) and Critical Success Index (CSI = 0.4384) alongside the lowest False Alarm Ratio (FAR = 0.4357). Compared with U-Net, these correspond to an approximately 102.4% increase in POD, a 39.6% increase in CSI, and a 26.8% reduction in FAR. These results indicate that the performance advantage of CBAM-U-Net is particularly pronounced for high-intensity echoes rather than being limited to improvements in domain-averaged error metrics.

The additional strong-echo-specific analysis further showed that the advantage of CBAM-U-Net extends beyond categorical detection to the reconstruction of echo intensity. For pixels with observed composite reflectivity $Z_{obs} \geq 45$ dBZ, CBAM-U-Net achieved the lowest MAE (9.3923 dBZ) and RMSE (13.8787 dBZ), together with the smallest magnitude of negative bias (−8.6382 dBZ), among the six evaluated models. Relative to U-Net, the MAE and RMSE were reduced by approximately 9.5% and 7.0%, respectively, while the magnitude of the negative bias was reduced by approximately 12.5%. These results indicate that systematic underestimation of severe echoes remains a common limitation, but that the sequential spectral–spatial recalibration used in CBAM-U-Net can partially alleviate this problem and improve the reconstruction of localized high-reflectivity structures.

The consistency of model performance was further examined under different flood-season precipitation regimes, while cross-sensor transferability was evaluated independently using FY-4B observations from July 2024.In the FY-4B experiment, the model parameters and normalization statistics derived from FY-4A were kept fixed, and no FY-4B data were used for training, calibration, or fine-tuning. Therefore, this experiment represents a strict zero-shot cross-sensor evaluation. The results indicate that the FY-4A-trained CBAM-U-Net retains comparatively strong reconstruction performance under changes in satellite platform, supporting its cross-sensor transferability.

Despite these improvements, several limitations remain. In particular, all evaluated models still exhibit negative bias for high-reflectivity echoes, indicating that the reconstruction of intense convective cores remains challenging. Future research will focus on multimodal feature integration by incorporating topographical data (e.g., Digital Elevation Models), cloud microphysical parameters, and Numerical Weather Prediction (NWP) background fields to construct a physically constrained fusion framework. Additionally, acknowledging the highly dynamic evolution of convective systems, incorporating recurrent or self-attention temporal modules (such as ConvLSTM or Spatiotemporal Transformers) will enforce temporal consistency and enhance the kinematic tracking of fast-moving storms. Finally, advancing the architectural paradigm by exploring CNN–Transformer hybrid networks with multi-scale cross-attention mechanisms could organically unify global structural context modeling with local fine-grained detail fidelity. A further limitation concerns the partitioning strategy of the FY-4A dataset. The current FY-4A experiments employed sample-level random partitioning and therefore did not explicitly enforce a temporal buffer or storm-event-level separation among the training, validation, and test subsets. Accordingly, the FY-4A results are primarily interpreted as comparative evaluations of the retrieval methods under consistent experimental settings, while the three period-specific experiments examine whether their relative performance remains consistent under different precipitation regimes. Future work will adopt temporally separated or storm-event-based partitioning strategies, in which samples associated with the same precipitation system are assigned exclusively to one subset, to provide a more stringent assessment of generalization to independent weather events.

3. Overall, the language in this paper is overly positive regarding the performance of the presented model. I recommend heavy revision of the text throughout and tempering much of the language used. For example... Line 577: “[...] the proposed model exhibited a superior capability in delineating sever convective cores [...]” seems too strong to me given the very similar performance to other U-Nets. Much of the text in this manuscript does not reflect the level of evidence provided in the experiments shown. There are several factual mistakes in the introduction and background.

Response:We thank the reviewer for this valuable comment. We agree that several statements in the original manuscript were stronger than directly supported by the

experimental results. We have therefore revised the Conclusion and related sections throughout the manuscript to adopt more objective descriptions.

For example, the wording in Section 3.4 was revised from expressions implying "optimal" or "superior" performance to more bounded descriptions such as "an effective balance" and "competitive performance and retrieval capability" (Page 17, Lines 487–493).

Through a rigorous comparative analysis of these distinct architectural paradigms, this research aims to ascertain whether CBAM-U-Net can achieve an effective balance between computational efficiency and the high-fidelity reconstruction of intricate structural details—particularly in real-world scenarios constrained by limited sample availability. Ultimately, this investigation aspires to demonstrate that the proposed CBAM-U-Net offers competitive performance and retrieval capability in retrieving severe convective echoes compared to increasingly complex, computationally intensive global modeling methodologies.

17

In the overall comparison, we further clarified that the observed CNN–Transformer differences may arise from multiple factors and should not be interpreted as evidence of an inherent architectural limitation (Page 20, Lines 535–544).

It is noteworthy that, in the present experiments, convolutional neural network (CNN)-based models generally outperform the Transformer-based models. The observed performance differences may be associated with multiple factors, including architectural inductive biases, model capacity, optimization characteristics, and the available training-sample size. The present experiments do not isolate the individual contribution of these factors.

Conversely, while SegFormer and SRViT exhibit robust overall modeling capabilities, their performance is lower than that of the evaluated CNN-based models under the current experimental configuration. Therefore, these results should be interpreted as an empirical comparison under the present dataset and training settings rather than as evidence of an inherent limitation of Transformer-based architectures.

Similar wording was adopted in the period-specific analysis, Page 22–23, Lines 612–620, 668-672:

Table 9: Threshold-based categorical verification of comparative models during the pre-flood onset phase

Model	POD			FAR			CSI		
	0–25 dBZ	25–45 dBZ	45–70 dBZ	0–25 dBZ	25–45 dBZ	45–70 dBZ	0–25 dBZ	25–45 dBZ	45–70 dBZ
U-Net	0.9168	0.4581	0.1613	0.0970	0.3054	0.4755	0.8758	0.3813	0.1470
Attention-U-Net	0.9280	0.4622	0.1979	0.0950	0.3535	0.4495	0.8836	0.3995	0.1798
MCDA-U-Net	0.9462	0.4878	0.1896	0.0927	0.3452	0.4335	0.8711	0.3880	0.1656
SegFormer	0.9422	0.4663	0.0629	0.0918	0.4190	0.5558	0.8546	0.3490	0.0583
SRViT	0.9274	0.3321	0.0090	0.1222	0.5544	0.7061	0.8214	0.2350	0.0089
CBAM-U-Net	**0.9690**	**0.5032**	**0.3477**	**0.0864**	**0.3023**	**0.4229**	**0.8877**	**0.4131**	**0.2495**

Table 9 further reveals the performance differences of models in handling local severe convection. For the severe echo interval (45–70 dBZ), which is the most critical for early warning, the CSI of CBAM-U-Net reaches 0.2495, substantially higher than that of U-Net (0.1470) and Transformer-based models (<0.06). These results indicate that, under the rapidly evolving convective conditions of the pre-flood period, CBAM-U-Net provides comparatively better

22

reconstruction of localized high-reflectivity echoes. The observed differences among the CNN- and Transformer-based models may be associated with multiple architectural and optimization factors, which are not separately isolated in the present experiments. Overall, CBAM-U-Net shows improved sensitivity to strong echo structures during this period.

In summary, the incorporation of the CBAM channel-spatial attention mechanism significantly mitigates inversion errors and improves feature fitting and structural preservation capabilities, indicating improved performance under the complex meteorological conditions prevalent during the primary flood season.

Table 11: Threshold-based evaluation of comparative models during the primary flood season

Model	POD			FAR			CSI		
	0–25 dBZ	25–45 dBZ	45–70 dBZ	0–25 dBZ	25–45 dBZ	45–70 dBZ	0–25 dBZ	25–45 dBZ	45–70 dBZ
U-Net	0.9193	0.6858	0.4085	0.1100	0.2539	0.3172	0.8336	0.5560	0.3394
Attention-U-Net	0.9231	0.6668	0.4147	0.1116	0.2497	0.4002	0.8364	0.5510	0.3376
MCDA-U-Net	0.8899	0.6604	0.3915	0.1240	0.3446	0.3814	0.7904	0.4902	0.2471
SegFormer	0.8261	0.6885	0.1928	0.1209	0.4333	0.5358	0.7418	0.4511	0.1577
SRViT	0.7448	0.5703	0.0614	0.1820	0.5651	0.7115	0.6390	0.3276	0.0533
CBAM-U-Net	**0.9342**	**0.6914**	**0.4585**	**0.0973**	**0.2434**	**0.3347**	**0.8405**	**0.5646**	**0.3845**

As can be seen from the POD, FAR, and CSI metrics across different precipitation intensities in Table 11, CBAM-U-Net shows strong overall performance across low, moderate, and high-reflectivity intervals, with the most pronounced advantage observed in the 45–70 dBZ range.

Page 27, Lines 720–726:

(1) **Quantitative indicator analysis**

To evaluate the model performance at this stage, Table 12 presents the regression errors of each model, while Table 13 illustrates the results of threshold-based categorical verification results.

Table 12: Evaluation metrics of comparative models during the post-flood transition phase

Model	RMSE	MAE	R^2	PSNR	SSIM
U-Net	6.2935	4.6386	0.6291	20.9241	0.7618
Attention-U-Net	6.1694	4.5344	0.6482	21.1536	0.7769
MCDA-U-Net	6.7651	5.0422	0.5714	20.2964	0.7403
SegFormer	7.0974	5.3753	0.5283	19.8800	0.7098
SRViT	9.1217	7.1482	0.2208	17.7004	0.6309
CBAM-U-Net	**6.1294**	**4.4336**	**0.6431**	**26.0172**	**0.7935**

As can be seen from the quantitative indicators in Table 12, the radar echo inversion performance of various models exhibits significant variability during the post-flood transition phase. CBAM-U-Net achieves the lowest RMSE and MAE and the highest SSIM among the evaluated models, with values of 6.1294, 4.4336, and 0.7935, respectively. These results indicate that the model provides comparatively better reconstruction of the actual echo distribution during the post-flood transition phase while preserving a higher degree of structural similarity.

In summary, the integration of the joint channel-spatial attention mechanism within CBAM-U-Net supports its applicability to radar echo inversion tasks during the post-flood transition phase, showing improved performance in prediction accuracy, error mitigation, and the restoration of image structure.

Page 28, 739-740:

The differences among the models are most pronounced in the 45–70 dBZ strong echo region, where CBAM-U-Net achieves the highest POD (0.3807) and CSI (0.2992) in the 45–70 dBZ interval, although its FAR (0.3825) is not the lowest among the evaluated models. In contrast,

Page 29, 757-758:

These results indicate that the CBAM-U-Net model can effectively improve the representation of severe convection features and the reconstruction accuracy of radar echo structures.

The FY-4B experiment is now described specifically as a cross-sensor transferability evaluation rather than evidence of operational readiness (Page 29–30, Lines 759–785).

4.3. Model generalization test across satellite platforms

To objectively assess the model's robustness outside the distribution of training samples and its transfer capability in cross-sensor evaluation scenarios, this study performed a cross-domain test utilizing FY-4B satellite data.

4. One claimed contribution of this paper is "addressing the operational challenge of radar coverage blind spots over complex topography. However, none of the shown experiments specifically address this. Generalization is only determined by running the model on a different satellite sensor. No experiments evaluate the estimation of precipitation structures in radar blind spots.

Response:We sincerely thank the reviewer for pointing out this inconsistency. We agree that the experiments presented in this study do not directly evaluate radar-blind-spot reconstruction, and therefore the original wording overstated the demonstrated capability of the proposed framework.

Accordingly, we removed statements claiming that the method "addresses" or "compensates for" radar coverage blind spots. For example, in the Abstract, the potential application is now described more conservatively as providing "complementary reflectivity information" in regions with limited radar observations, rather than directly filling radar coverage gaps (Page 1, Lines 14–17).

Abstract

To investigate the potential of satellite-based radar composite reflectivity reconstruction under complex terrain conditions, this study proposes an end-to-end deep learning framework to reconstruct radar composite reflectivity (CR) from FY-4A satellite multi-channel observations. We introduce CBAM-U-Net by embedding a lightweight Convolutional Block Attention Module into a U-Net backbone. This dual-dimensional mechanism adaptively recalibrates multispectral features and enhances the representation of localized intense convective structures. Evaluated on a spatiotemporally matched satellite-radar dataset containing 14,023 samples from Sichuan Province (May–November 2023), CBAM-U-Net achieves the best overall performance among the evaluated CNN- and Transformer-based models in retrieval accuracy (RMSE = 6.8290 dBZ, R^2 = 0.6277) and structural fidelity (SSIM = 0.7894). Particularly, within the challenging severe echo regime (45–70 dBZ), the model achieves the highest Probability of Detection (POD = 0.5296) and Critical Success Index (CSI = 0.4384) among the evaluated models. Furthermore, cross-sensor evaluation using FY-4B observations demonstrates the transferability of the proposed framework under observational differences between satellite platforms. This research highlights the potential of integrating satellite multispectral features with attention-enhanced networks to provide complementary reflectivity information for severe convective weather monitoring, especially in regions with limited radar observations.

The objective in the Introduction is now described as reconstructing radar composite reflectivity over complex terrain using FY-4A observations, rather than compensating for radar blind spots (Page 3, Lines 109–112).

To address the aforementioned challenges, this study proposes an attention-enhanced end-to-end reconstruction framework, termed CBAM-U-Net, designed to reconstruct radar composite reflectivity over complex terrain using multi-channel observations from the FY-4A geostationary satellite. Focusing on the topographically complex Sichuan Province during the 2023 extended flood season, we seamlessly integrate the Convolutional Block Attention Module (CBAM) into a U-Net backbone(Woo et al., 2018). By serially combining channel and spatial attention mechanisms within the same feature layer, the framework adaptively recalibrates the responses of critical spectral bands (e.g., water vapor channels, split-window brightness temperature differences) while

The corresponding contribution statement has also been reformulated to focus on radar composite reflectivity reconstruction under complex-terrain and severe-convection conditions (Page 4, Lines 119–133).

(1) **Problem-oriented reconstruction under complex-terrain and severe-convection conditions**: Rather than revisiting the general feasibility of CNN-based satellite precipitation retrieval, this study focuses on radar composite reflectivity reconstruction over the complex terrain of Sichuan, with particular emphasis on localized high-intensity convective echoes.

(2) **Lightweight spectral–spatial enhancement of localized severe echoes**: A cascaded channel–spatial attention strategy is integrated into the U-Net backbone to sequentially recalibrate multispectral feature importance and spatially emphasize localized high-reflectivity structures. The objective is not to claim the CBAM mechanism itself as novel, but to investigate its task-specific value for mitigating severe-echo smoothing and intensity underestimation in satellite-to-radar reconstruction.

(3) **Multi-perspective evaluation of severe-echo reconstruction capability and cross-sensor transferability**: Beyond conventional domain-averaged metrics, the proposed framework is evaluated using intensity-stratified categorical verification, strong-echo-specific intensity errors for observed reflectivity $\geq$ 45 dBZ, different flood-season precipitation regimes, and an independent zero-shot FY-4B test.

In addition, the FY-4B experiment is now interpreted strictly as a zero-shot cross-sensor transferability evaluation and is not presented as evidence of radar-blind-spot reconstruction or operational deployment capability.

Specific Comments:

Comments 1:Line 8: What is meant by “optimal” here? Is this with respect to the other models shown?

Response:Thank you for pointing this out. The term “optimal” was intended to mean the best performance among the models evaluated in this study, rather than a theoretically optimal or universally best result. To avoid ambiguity, we have replaced “optimal” with “the highest among the evaluated models” in the Abstract. (Page 1 line 11-12).

Abstract

To investigate the potential of satellite-based radar composite reflectivity reconstruction under complex terrain conditions, this study proposes an end-to-end deep learning framework to reconstruct radar composite reflectivity (CR) from FY-4A satellite multi-channel observations. We introduce CBAM-U-Net by embedding a lightweight Convolutional Block Attention Module into a U-Net backbone. This dual-dimensional mechanism adaptively recalibrates multispectral features and enhances the representation of localized intense convective structures. Evaluated on a spatiotemporally matched satellite-radar dataset containing 14,023 samples from Sichuan Province (May–November 2023), CBAM-U-Net achieves the best overall performance among the evaluated CNN- and Transformer-based models in retrieval accuracy (RMSE = 6.8290 dBZ, R^2 = 0.6277) and structural fidelity (SSIM = 0.7894). Particularly, within the challenging severe echo regime (45–70 dBZ), the model achieves the highest Probability of Detection (POD = 0.5296) and Critical Success Index (CSI = 0.4384) among the evaluated models. Furthermore, cross-sensor evaluation using FY-4B observations demonstrates the transferability of the proposed framework under observational differences between satellite platforms. This research highlights the potential of integrating satellite multispectral features with attention-enhanced networks to provide complementary reflectivity information for severe convective weather monitoring, especially in regions with limited radar observations.

Comments 2:Line 30: “Although satellites primarily detect cloud-top thermal radiation and microphysical properties [...]” Please be more specific about which kinds of sensors are described by this statement. I would agree this is correct for passive IR imaging instruments. However, this is not true for satellite cloud lidars and radars. You might also clarify that you mean ground-based radars in the latter part of this sentence.

Response:We have revised this single sentence to address both comments without splitting the paragraph.

To resolve the overgeneralized term "satellites", we explicitly specify that our satellite observations come from the AGRI infrared imager carried by FY-4A. A parenthetical clarification is added to acknowledge that other satellite platforms carry space-borne radars capable of retrieving full vertical cloud microphysical information, which distinguishes our passive cloud-top measurements from active satellite radar products.

We fully agree with the reviewer that scattering traits of hydrometeors belong to intrinsic microphysical properties. The text is rephrased to clarify that ground radars measure backscattered echo signals generated from such microphysical scattering behaviour, rather than measuring the scattering properties of particles directly, reconciling the physical definition with radar observational quantities. All adjustments are embedded within the original sentence to maintain textual integrity (Page 2 line 44-54).

distribution and evolution of cloud systems. The AGRI sensor carried by FY-4A mainly captures thermal radiance emitted from cloud tops; other satellite payloads such as space-borne radars are capable of retrieving complete vertical cloud microphysical structures. Ground-based weather radars, by contrast, record backscattered echo signals generated from the intrinsic microphysical scattering properties of hydrometeors inside and below cloud layers. Previous comparisons between space-borne precipitation radar and ground-based radar have shown that differences in observing geometry, radar frequency, and hydrometeor scattering characteristics can affect the consistency of radar reflectivity measurements(Kou et al., 2023). Despite these fundamental observational differences, robust physical and statistical relationships can be identified between satellite observations and ground-based radar measurements during severe convective events (Roberts and Rutledge, 2003; Wang et al., 2022). Specifically, prominent cooling signals in the

Comments 3:Line 34: “aligns spatiotemporally [...]” This is not necessarily true for all IR channels (e.g., water vapor in the 5-7 μm range, CO2 at ~13-14 μm, and ozone at ~9.6 μm.

Response:We sincerely appreciate the reviewer for these highly accurate and constructive physical comments. We fully agree that the relationship between depressed infrared brightness temperature and enhanced radar reflectivity is not universally valid, and we have revised the relevant description accordingly.

First, we have clarified that the association between low brightness temperature and deep convection only applies to the 10–12 μm thermal infrared window channels, as other infrared channels are primarily sensitive to absorptions from water vapor, ozone, and carbon dioxide rather than convective cloud-top emission.

Second, we have added an important caveat that high-altitude cirrus clouds, although characterized by very low infrared brightness temperatures, are not accompanied by convective updrafts and typically exhibit radar reflectivity below typical detection thresholds. Thus, the spatiotemporal correspondence only holds for precipitating deep convective systems, not for all cold cloud features.

The relevant sentence in the manuscript has been carefully revised to include these constraints and avoid overgeneralization (Page 2 line 54-58).

54 (Roberts and Rutledge, 2003; Wang et al., 2022). Specifically, prominent cooling signals in the
55 10–12 μm thermal infrared window bands serve as indicators of deep convection and vigorous
56 updrafts, and such brightness temperature reductions match elevated radar composite reflectivity in
57 space and time. Nevertheless, high-altitude cirrus clouds also exhibit very low infrared brightness
58 temperatures but yield negligible radar reflectivity below typical detection thresholds. This

Comments 4:Line 38: Please describe these conventional models that are being contrasted here. What “physical retrieval techniques” exist to solve this task?

Response:We thank the reviewer for pointing out that the descriptions of conventional and physical retrieval approaches were insufficiently specific. We agree that the original statement was too general.

In the revised manuscript, we have clarified that the conventional approaches refer mainly to empirical and statistical retrieval methods, which commonly use manually designed spectral, textural, or meteorological predictors. We have also added a brief description of physically based retrieval approaches, including methods relying on radiative transfer simulations, cloud microphysical assumptions, and parameterized relationships.

Meanwhile, we have revised the wording to avoid implying that all traditional approaches are inadequate. The revised text now emphasizes that these approaches may face challenges in representing complex nonlinear relationships within multispectral satellite observations (Page 2-3 line 64-73).

between cloud-top radiative characteristics and radar-derived composite reflectivity fields. Because traditional empirical and statistical retrieval approaches typically rely on manually designed spectral, textural, or meteorological predictors to establish relationships between satellite observations and precipitation-related variables, they may have limited capability in representing the complex nonlinear relationships embedded in multispectral satellite datasets. In addition, physically based retrieval approaches, which rely on radiative transfer simulations, cloud microphysical

assumptions, or parameterized relationships, may be affected by uncertainties in cloud processes and the availability of ancillary information. Recent deep learning approaches provide an alternative framework by automatically learning hierarchical representations from multispectral observations. Consequently, recent studies have increasingly turned to advanced deep learning

Comments 5:Line 50: I do not agree that this is a characteristic of U-Nets. A counterfactual to this would be any diffusion model that uses a U-Net architecture. My expectation would be that the loss function used would have a much stronger impact on the "smoothness" of the predictions.

Response:We fully agree with the reviewer's valuable comment regarding the source of the smoothing artifact. We acknowledge that the smoothing tendency is not an inherent flaw of the U-Net architecture itself, but largely originates from the choice of objective function during model training. As pointed out by the reviewer, many existing retrieval algorithms adopt U-Net backbones together with optimized loss designs to avoid over-smoothing.

Accordingly, we have revised the relevant sentence in the manuscript to correct the misleading causal description. We clarify that standard U-Net tends to produce smoothed outputs when trained with conventional pixel-wise loss functions, which leads to blurred convective boundaries and underestimated peak reflectivity in intense echo cores. This issue motivates our adoption of the CBAM attention module and an intensity-weighted MSE loss to mitigate the underestimation of strong echoes in this work. We have also incorporated the three references suggested by the reviewer to provide appropriate support for this discussion (Page 3 line 91-95).

convective cores of merely a few kilometers.Standard U-Net architectures, while capable of recovering macroscopic morphology, may exhibit a persistent "smoothing effect" when trained with conventional pixel-wise regression objectives, particularly for strong echo regions, leading to blurred convective boundaries and systematic underestimation of peak intensities within strong echo cores (Guilloteau et al., 2023, 2025; Glawion et al., 2025). Compounding this retrieval

Comments 6:Line 64: Please provide the citation for CBAM.

Response:Thank you for pointing out this omission. The original CBAM paper by Woo et al. (2018) has now been cited at the first introduction of CBAM in the manuscript, in addition to the citations provided in the detailed descriptions of the channel- and spatial-attention modules (Page 3 line 114).

satellite. Focusing on the topographically complex Sichuan Province during the 2023 extended flood season, we seamlessly integrate the Convolutional Block Attention Module (CBAM) into a U-Net backbone(Woo et al., 2018). By serially combining channel and spatial attention mechanisms within the same feature layer, the framework adaptively recalibrates the responses of critical spectral bands (e.g., water vapor channels, split-window brightness temperature differences) while

Comments 7:Line 70 and throughout: I don't agree that what has been developed here qualifies as an "expert system." Most definitions I see involve the emulating human experts.

Response:We agree with the reviewer that the term "expert system" is inappropriate for the proposed method because the model does not emulate a human expert through an explicit knowledge- or rule-based reasoning system. We have therefore removed the term "expert system" and replaced it with "retrieval framework" (Page 4 line 136).

The remainder of this paper is organized as follows: Section 2 introduces the multi-source datasets, physical feature selection, and rigorous preprocessing procedures. Section 3 details the architecture of the proposed CBAM-U-Net retrieval framework and its training strategies. Section 4 presents comprehensive comparative experiments, staged evaluations during the flood season, and typical case studies. Finally, Section 5 summarizes the findings, discusses limitations, and outlines future research directions.

Comments 8:Line 80: It is not obvious to me how "immense value for real-time disaster early warning operations" follows from "robust cross-sensor zero-shot generalization." Please add some explanation to this. This is also not an especially large dataset. If it were extremely large, we would expect to see much better performance from the transformers than we do here.

Response:We agree with the reviewer that robust cross-sensor generalization alone does not establish operational value for real-time disaster early warning. The original statement therefore extended beyond the evidence provided by the experiment. We have removed this claim and now interpret the FY-4B experiment specifically as an evaluation of cross-sensor transferability (Page 4 line 129-133).

(3) **Multi-perspective evaluation of severe-echo reconstruction capability and cross-sensor transferability**: Beyond conventional domain-averaged metrics, the proposed framework is evaluated using intensity-stratified categorical verification, strong-echo-specific intensity errors for observed reflectivity ≥ 45 dBZ, different flood-season precipitation regimes, and an independent zero-shot FY-4B test.

Comments 9:Line 133: "To further augment the model's feature representation, two BTD formulations (BTD1 and BTD2) are engineered." A difference between two channels is something that a neural network can learn extremely easily in just a single fully-connected layer. Therefore, it's not obvious to me that these differences would add any value provided that the training data set is large. A following sentence "This physically informed feature engineering approach provides a solid foundation for [...]" overstates the importance of adding these differences to such a large neural network, in my opinion.

Response:We thank the reviewer for this insightful comment. We agree that channel-difference relationships such as BTD1 and BTD2 are mathematically simple and, in principle, can be learned directly by a sufficiently expressive neural network

from the original multispectral channels. Therefore, we agree that the original wording overstated the importance of these engineered features.

In the revised manuscript, we have accordingly clarified the role of the BTD features. Rather than treating them as indispensable predictors or as a major source of model improvement, we now describe BTD1 and BTD2 as auxiliary physically motivated inputs that explicitly provide spectral contrasts associated with cloud-top thermal characteristics, upper-level moisture conditions, and cloud-particle phase differences. Their purpose is to introduce known radiative and microphysical relationships into the input representation, rather than to provide information that the neural network would otherwise be unable to learn.

From this perspective, explicitly supplying these spectral contrasts may reduce the burden on the network to reconstruct such simple physically meaningful relationships from the raw channels during representation learning. However, we do not claim that the BTD features are essential or that they constitute the primary source of the performance improvement observed in this study. We have therefore revised the manuscript to use more cautious wording and removed the previous statement that this feature-engineering strategy provides a "solid foundation" for the retrieval framework.

The corresponding text has been revised on Page 6, Lines 205–209.

204 To further augment the model's feature representation, two BTD formulations (BTD1 and
205 BTD2) are engineered. These physically motivated auxiliary variables are introduced to provide
206 additional spectral contrast information related to cloud-top thermal and spectral characteristics.
207 Although such channel differences can also be potentially learned by deep neural networks from the
208 original multispectral inputs, explicitly providing these features may facilitate the representation
209 learning process (Yang et al., 2023). A comprehensive overview of the selected channel parameters
210 and their corresponding physical significance is presented in Table 1.

Comments 10:Lines 140 – 163: The ground truth for the model appears to be a 2-D array of column-wise maximum reflectivity. This seems to be at odds with the stated goal of estimated "3D 'intra-cloud precipitation structure'" in Line 123.

Response:We agree with the reviewer. The original wording incorrectly implied that the model retrieves a three-dimensional intra-cloud precipitation structure, whereas the target variable is a two-dimensional radar composite reflectivity field representing the column-wise maximum reflectivity. We have corrected this statement throughout the manuscript. The revised manuscript now consistently describes the task as reconstruction of two-dimensional radar composite reflectivity from FY-4A multispectral observations (Page 6 line 186-188).

186 the integration of atmospheric radiative transfer theory and cloud microphysics. The objective is
187 to establish a robust mapping relationship to infer the column-wise maximum radar composite
188 reflectivity from the 2D cloud-top thermal radiation field. Recognizing that strong radar reflectivity

Comments 11:Lines 164 – 174: How are the training and testing set separated? There

is not enough information here to confirm that there is no leakage between these sets.
Response:Thank you very much for this important and constructive comment. We agree that the dataset partitioning procedure and the independence of the training, validation, and test subsets should be clearly described.

In the original FY-4A experiments, the samples were randomly partitioned at the individual-sample level into training, validation, and test subsets, accounting for approximately 70%, 10%, and 20% of the samples, respectively. For the three period-specific experiments, the samples within each corresponding period were separately partitioned using the same strategy, and the models were independently trained and evaluated for each period. We have now clarified this procedure in the revised manuscript.

We acknowledge that the sample-level random partitioning used in the FY-4A experiments did not explicitly enforce storm-event-level separation. Therefore, temporally adjacent samples potentially associated with the same or closely related precipitation systems may occur in different subsets. Following the reviewer's comment, we have revised the manuscript to avoid interpreting the FY-4A random-split results as a strict storm-event-independent generalization assessment.

The three period-specific experiments have also been more clearly defined in the revised manuscript. These experiments involve separate training and evaluation for each precipitation period and are intended to examine whether the relative performance advantage of the proposed architecture remains consistent under different precipitation regimes, rather than to demonstrate generalization to independent storm events.

In addition, the manuscript includes an independent cross-sensor evaluation using FY-4B observations from July 2024. The FY-4B data were acquired from a different satellite platform and time period and were not involved in model training, validation, calibration, or fine-tuning. We have strengthened the corresponding description to clarify that this experiment provides an independent zero-shot evaluation complementary to the FY-4A experiments.

Finally, we have explicitly acknowledged the lack of storm-event-level separation in the FY-4A experiments as a limitation of the present study. In future work, temporally separated or storm-event-based partitioning will be adopted, such that samples associated with the same precipitation system are assigned exclusively to a single subset. This will enable a more stringent assessment of model generalization to independent weather events.

We sincerely thank the reviewer for pointing out this important issue, which has helped us clarify the scope and interpretation of the current experiments and improve the rigor of the manuscript.

The modifications in this section can be found on page 14, lines 390-394:

For the overall May–November 2023 experiment, the FY-4A samples were randomly partitioned at the individual-sample level into training, validation, and test subsets, accounting for approximately 70%, 10%, and 20% of the samples, respectively. For the three period-specific experiments, the samples within each corresponding period were separately partitioned following the same strategy, and the models were independently trained and evaluated for each period.

The modifications in this section can be found on page 21, lines 580-592:

4.2. Analysis of flood season classification

The period-specific experiments conducted in this study were designed as a cross-regime robustness analysis rather than as repeated evaluations of a single trained model across different time periods. The three periods represent distinct precipitation regimes associated with the Sichuan flood season, and separate training and evaluation experiments were conducted for each period to examine whether the relative performance advantage of the proposed method remains consistent under varying meteorological conditions. Specifically, the pre-flood period is characterized by rapidly developing convective systems, the primary flood season by more frequent severe convection and heavy precipitation events, and the post-flood period by transitional precipitation processes. These differences are associated with variations in cloud development, precipitation intensity, and the relationship between satellite infrared observations and radar-derived reflectivity. Therefore, the period-specific experiments provide complementary evidence that the relative effectiveness of the proposed architecture remains consistent across diverse precipitation regimes.

The modifications in this section can be found on page 30, lines 783-785:

The FY-4B experiment was conducted as a strict zero-shot cross-sensor evaluation. Neither model parameters nor normalization statistics were adapted using FY-4B data; all model weights and preprocessing statistics were inherited directly from the FY-4A training stage.

The modifications in this section can be found on page 32, lines 862-870:

concerns the partitioning strategy of the FY-4A dataset. The current FY-4A experiments employed sample-level random partitioning and therefore did not explicitly enforce a temporal buffer or storm-event-level separation among the training, validation, and test subsets. Accordingly, the FY-4A results are primarily interpreted as comparative evaluations of the retrieval methods under consistent experimental settings, while the three period-specific experiments examine whether their relative performance remains consistent under different precipitation regimes. Future work will adopt temporally separated or storm-event-based partitioning strategies, in which samples associated with the same precipitation system are assigned exclusively to one subset, to provide a more stringent assessment of generalization to independent weather events.

Comments 12:Lines 164 – 174: How is parallax handled in the matching of the radar and satellite data? Elevated cloud-tops and precipitation structures could be mismatched if this area is at a high enough viewing angle from FY-4A.

Response:We thank the reviewer for raising this important issue regarding the potential parallax effect in geostationary satellite observations.

The FY-4A data used in this study were processed following the standard preprocessing procedures, including radiometric calibration and geographic registration. However, we acknowledge that these procedures do not constitute an explicit cloud-top parallax correction.

In the current study, the target variable is radar composite reflectivity, which represents the maximum reflectivity within the atmospheric column rather than the exact three-dimensional position of cloud-top structures. Therefore, the FY-4A observations and radar products were matched on a common geographic grid using the FY-4A grid as the spatial reference. No additional cloud-height-dependent parallax correction was applied.

We have clarified this point in the revised manuscript and added the potential parallax uncertainty as a limitation. Future work will investigate parallax-aware satellite-radar matching approaches, particularly for high cloud-top and deep convective systems (Page 11, lines 324-332).

The potential parallax effect associated with geostationary satellite observations was considered as a source of spatial uncertainty in the satellite–radar matching. In this study, the prediction target is radar composite reflectivity, which represents the maximum radar reflectivity within the atmospheric column, whereas the infrared observations from FY-4A are primarily associated with cloud-top radiative characteristics. Satellite and radar observations were therefore co-registered on a common geographic grid using the FY-4A grid as the spatial reference. No explicit cloud-top parallax correction was applied in the current dataset construction. Consequently, residual spatial displacement may remain, particularly for high-altitude cloud systems and deep convective clouds, and this is recognized as a limitation of the present study.

Comments 13:Lines 175 – 200: I do not feel that this section is necessary to be included in the article unless it is not present in existing publicly available FY-4A/AGRI literature. While the calculation of latitude and longitude can be a somewhat involved process, as this text shows, this does not contribute much to the stated objective of the manuscript. I feel that a citation to a FY-4A user's manual (or similar) would be sufficient.

Response:Thank you for this suggestion. We agree that the detailed navigation equations are not central to the scientific objective of the manuscript. To improve the focus and readability of the main text, the detailed FY-4A AGRI latitude – longitude conversion procedure has been moved to Appendix A(Page 36), while the main methodology section now retains only a concise description of the geolocation procedure and its role in satellite – radar matching (Page 9 256-258).

(1) **FY-4A AGRI L1 geometric positioning and latitude-longitude conversion**

The detailed geometric positioning and latitude-longitude conversion procedure for FY-4A AGRI L1 data is documented in Appendix A.

Appendix A. FY-4A AGRI L1 geometric positioning and latitude-longitude conversion

The FY-4A/AGRI Level 1 data are typically structured according to line and column indices within the framework of a geostationary orbit projection. This organization necessitates the conversion of pixel line and column indices into corresponding latitude and longitude coordinates through the application of navigation parameters. In alignment with the data format and navigation specifications outlined for FY-4A, this paper delineates the subsequent procedure for the inverse calculation of latitude and longitude coordinates.

Initially, an intermediate angular variable (expressed in radians) is defined, derived from the row and column indices:

$$x = \frac{\pi \cdot (c - \mathrm{COFF})}{180 \cdot 2^{-16} \cdot \mathrm{CFAC}} \tag{A.1}$$

$$y = \frac{\pi (l - \mathrm{LOFF})}{180 \cdot 2^{-16} \cdot \mathrm{LFAC}} \tag{A.2}$$

Where c and l represent the column and row numbers of the pixel, respectively; COFF/CFAC denote the column-direction offset and scale factor, and LOFF/LFAC represent the row-direction offset and scale factor (In practical applications, it is advisable to first extract the navigation parameters stored within each raw FY-4A HDF file to avoid manually hardcoding geographic transformation coefficients.) For the FY-4A AGRI 4 km resolution product, typical values commonly reported in open research are COFF = LOFF = 1373.5 and CFAC = LFAC = 10233137, after which other intermediate variables can be calculated as follows:

$$S_d = \sqrt{(h\cos(x)\cos(y))^2 - \left(\cos^2(y) + \frac{ea^2}{eb^2}\sin^2(y)\right)(h^2 - ea^2)} \tag{A.3}$$

$$S_n = \frac{h\cos(x)\cos(y) - S_d}{\cos^2(y) + \frac{ea^2}{eb^2}\sin^2(y)} \tag{A.4}$$

$$S_1 = h - S_n\cos(x)\cos(y) \tag{A.5}$$

$$S_2 = S_n\sin(x)\cos(y) \tag{A.6}$$

$$S_3 = -S_n\sin(y) \tag{A.7}$$

$$S_{xy} = \sqrt{S_1^2 + S_2^2} \tag{A.8}$$

where h denotes the distance from the geocenter to the center of mass of the satellite, and e_a and e_b represent the Earth's semi-major and semi-minor axes, respectively. The specific values of these parameters are set as follows: h = 42164 km, e_a = 6378.137 km, and e_b = 6356.7523 km.

Comments 14:Line 218: Wouldn't the resolution of the sensor vary as a function of viewing angle? What is the approximate viewing angle at this location?

Response: We thank the reviewer for pointing out the influence of geostationary viewing geometry on the effective spatial resolution.

We agree that the spatial resolution of FY-4A/AGRI is not strictly constant across the full disk because of the variation in viewing angle. In the original manuscript, the statement that one FY-4A pixel corresponds to 4×4 radar pixels was intended to describe the grid relationship between the two datasets, but it may have been interpreted as an exact physical resolution equivalence. During the 2023 study period, FY-4A was located at approximately 104.7°E. Based on the geostationary viewing geometry, the satellite viewing zenith angle over the study domain (26°–34.5°N,

97°–109°E) is approximately 31°–41°, with a value of about 35° near the center of the study region.Therefore, the effective ground sampling distance varies moderately across the study area.

We have revised the manuscript to clarify that the 0.04° and 0.01° values represent the geographic grids of the FY-4A and SWAN products, respectively. The radar composite reflectivity data were aggregated onto the FY-4A grid before model training to ensure spatial consistency (Page 10–11, lines 283–332).

Regarding spatial matching, the FY-4A/AGRI observations used in this study are provided on an approximately 0.04° geographic grid, whereas the SWAN radar products are available on a 0.01° grid. Since FY-4A is a geostationary satellite, the effective ground sampling distance varies with viewing geometry. Therefore, the 0.04° and 0.01° values represent the respective data grids rather than strictly constant physical resolutions. To achieve spatial consistency between satellite observations and radar targets, the radar composite reflectivity data were aggregated onto the FY-4A grid before model training.

To preserve localized high-intensity echo structures during spatial aggregation from the finer radar grid to the coarser satellite grid, this study adopts a max-pooling strategy for satellite–radar spatial co-registration. Specifically, each 0.04° FY-4A grid cell contains approximately a 4×4 patch of 0.01° SWAN radar pixels, and the maximum radar composite reflectivity value within this patch is assigned as the corresponding target value. It should be emphasized that this horizontal aggregation is distinct from the vertical maximum operation used to construct the original radar composite reflectivity product. Therefore, the resulting target represents a maximum-preserving coarse-grid representation of the original radar composite reflectivity rather than a redefinition of composite reflectivity itself. Compared with spatial averaging, which may smooth localized extreme echoes and reduce peak reflectivity values, max pooling retains the strongest sub-grid echo within each matched satellite footprint. Nevertheless, this operation may also shift the coarse-grid reflectivity distribution toward higher values and increase the apparent coverage of strong echoes; therefore, its influence is further evaluated through a sensitivity experiment using an alternative aggregation strategy.

$$\mathrm{CR}_{0.04^\circ}^{\max}(i,j) = \max_{(m,n)\in\Omega_{i,j}} \mathrm{CR}_{0.01^\circ}(m,n) \tag{3}$$

where $\Omega_{i,j}$ denotes the set of radar sub-pixels located within the (i,j)-th FY-4A grid cell. The

max-aggregation strategy is intended to preserve localized peak reflectivity information that may otherwise be weakened during spatial averaging. Its influence on the statistical distribution of the radar targets and on the retrieval performance, particularly for strong echoes (e.g., ≥ 45 dBZ), is further examined through aggregation-sensitivity experiments.

To assess the sensitivity of the retrieval results to the horizontal aggregation strategy, an additional comparison experiment was conducted using linear-reflectivity mean aggregation (Mean-Z) as an alternative to Max aggregation. For a fair comparison, the same FY-4A input data, training/validation/test split, model architectures, training hyperparameters, and evaluation metrics were retained, with only the radar-label aggregation strategy being changed. Both U-Net and CBAM-U-Net were trained separately using Max and Mean-Z targets.

Table 2: Sensitivity of retrieval performance to different radar spatial aggregation strategies.

Aggregation	Model	RMSE	MAE	R^2	SSIM	$POD_{\geq 45}$	$FAR_{\geq 45}$	$CSI_{\geq 45}$
Max	U-Net	6.8109	4.9437	0.6223	0.7812	0.3242	0.3260	0.2803
Max	**CBAM-U-Net**	**6.6989**	**4.8605**	**0.6346**	**0.7832**	**0.3686**	**0.3460**	**0.3085**
Mean-Z	U-Net	5.8742	4.1834	0.6582	0.8044	0.2234	0.2854	0.2051
Mean-Z	**CBAM-U-Net**	**5.7764**	**4.1028**	**0.6694**	**0.8067**	**0.3027**	**0.3435**	**0.2613**

As shown in Table 2, CBAM-U-Net shows consistent improvements in RMSE, R^2, POD, and CSI relative to the corresponding U-Net baseline under both aggregation strategies, although FAR increases. Under Max aggregation, the RMSE decreases from 6.8109 to 6.6989 dBZ and the CSI for echoes ≥ 45 dBZ increases from 0.2803 to 0.3085. Under Mean-Z aggregation, the RMSE similarly decreases from 5.8742 to 5.7764 dBZ, while the CSI increases from 0.2051 to 0.2613. Although the absolute metric values differ between the two aggregation schemes because they produce different coarse-grid target distributions, the relative improvement of CBAM-U-Net over U-Net is retained. This indicates that the performance improvement introduced by CBAM is not solely dependent on the use of Max aggregation.

The potential parallax effect associated with geostationary satellite observations was considered as a source of spatial uncertainty in the satellite–radar matching. In this study, the prediction target is radar composite reflectivity, which represents the maximum radar reflectivity within the atmospheric column, whereas the infrared observations from FY-4A are primarily associated with cloud-top radiative characteristics. Satellite and radar observations were therefore co-registered on a common geographic grid using the FY-4A grid as the spatial reference. No explicit cloud-top parallax correction was applied in the current dataset construction. Consequently, residual spatial displacement may remain, particularly for high-altitude cloud systems and deep convective clouds, and this is recognized as a limitation of the present study.

Comments 15:Line 220: Why wouldn't an average or interpolation be sufficient? This choice of max-pooling needs to be better motivated and needs to be made sufficiently clear throughout the manuscript that this filter is applied to the ground-truth data and changes the interpretation of the resulting model.

Response:Thank you for this comment. We have revised the manuscript to clarify both the motivation for Max aggregation and its effect on the interpretation of the model target.

Spatial averaging is suitable when the objective is to represent the mean reflectivity within a coarse satellite grid cell. However, in this study, we aim to retain localized high-intensity convective echoes, which may occupy only a small fraction of the approximately 0.04° FY-4A grid cell. Averaging can substantially reduce these localized peak values, while Max aggregation retains the strongest sub-grid radar echo. Similarly, interpolation primarily provides a spatially smoothed estimate on the target

grid and does not explicitly preserve sub-grid extreme reflectivity values.

We now explicitly state throughout the revised manuscript that the Max operator is applied to the radar ground-truth labels during spatial co-registration, rather than to the model output. We also clarify that this operation changes the interpretation of the prediction target: the model trained with Max aggregation retrieves a maximum-preserving coarse-grid representation of radar composite reflectivity rather than the spatially averaged reflectivity within each FY-4A grid cell.

To evaluate whether this choice materially affects the model conclusion, we additionally performed experiments using Mean-Z aggregation. The relative improvement of CBAM-U-Net over U-Net was retained under both Max and Mean-Z aggregation, supporting the robustness of the proposed model improvement with respect to the aggregation strategy (Page 10–11, lines 283–332).

Regarding spatial matching, the FY-4A/AGRI observations used in this study are provided on an approximately 0.04° geographic grid, whereas the SWAN radar products are available on a 0.01° grid. Since FY-4A is a geostationary satellite, the effective ground sampling distance varies with viewing geometry. Therefore, the 0.04° and 0.01° values represent the respective data grids rather than strictly constant physical resolutions. To achieve spatial consistency between satellite observations and radar targets, the radar composite reflectivity data were aggregated onto the FY-4A grid before model training.

To preserve localized high-intensity echo structures during spatial aggregation from the finer radar grid to the coarser satellite grid, this study adopts a max-pooling strategy for satellite–radar spatial co-registration. Specifically, each 0.04° FY-4A grid cell contains approximately a 4×4 patch of 0.01° SWAN radar pixels, and the maximum radar composite reflectivity value within this patch is assigned as the corresponding target value. It should be emphasized that this horizontal aggregation is distinct from the vertical maximum operation used to construct the original radar composite reflectivity product. Therefore, the resulting target represents a maximum-preserving coarse-grid representation of the original radar composite reflectivity rather than a redefinition of composite reflectivity itself. Compared with spatial averaging, which may smooth localized extreme echoes and reduce peak reflectivity values, max pooling retains the strongest sub-grid echo within each matched satellite footprint. Nevertheless, this operation may also shift the coarse-grid reflectivity distribution toward higher values and increase the apparent coverage of strong echoes; therefore, its influence is further evaluated through a sensitivity experiment using an alternative aggregation strategy.

$$\mathrm{CR}_{0.04^\circ}^{\max}(i,j) = \max_{(m,n)\in\Omega_{i,j}} \mathrm{CR}_{0.01^\circ}(m,n) \tag{3}$$

where $\Omega_{i,j}$ denotes the set of radar sub-pixels located within the (i,j)-th FY-4A grid cell. The

max-aggregation strategy is intended to preserve localized peak reflectivity information that may otherwise be weakened during spatial averaging. Its influence on the statistical distribution of the radar targets and on the retrieval performance, particularly for strong echoes (e.g., ≥ 45 dBZ), is further examined through aggregation-sensitivity experiments.

To assess the sensitivity of the retrieval results to the horizontal aggregation strategy, an additional comparison experiment was conducted using linear-reflectivity mean aggregation (Mean-Z) as an alternative to Max aggregation. For a fair comparison, the same FY-4A input data, training/validation/test split, model architectures, training hyperparameters, and evaluation metrics were retained, with only the radar-label aggregation strategy being changed. Both U-Net and CBAM-U-Net were trained separately using Max and Mean-Z targets.

Table 2: Sensitivity of retrieval performance to different radar spatial aggregation strategies.

Aggregation	Model	RMSE	MAE	R^2	SSIM	$POD_{\geq 45}$	$FAR_{\geq 45}$	$CSI_{\geq 45}$
Max	U-Net	6.8109	4.9437	0.6223	0.7812	0.3242	0.3260	0.2803
Max	**CBAM-U-Net**	**6.6989**	**4.8605**	**0.6346**	**0.7832**	**0.3686**	**0.3460**	**0.3085**
Mean-Z	U-Net	5.8742	4.1834	0.6582	0.8044	0.2234	0.2854	0.2051
Mean-Z	**CBAM-U-Net**	**5.7764**	**4.1028**	**0.6694**	**0.8067**	**0.3027**	**0.3435**	**0.2613**

As shown in Table 2, CBAM-U-Net shows consistent improvements in RMSE, R^2, POD, and CSI relative to the corresponding U-Net baseline under both aggregation strategies, although FAR increases. Under Max aggregation, the RMSE decreases from 6.8109 to 6.6989 dBZ and the CSI for echoes ≥ 45 dBZ increases from 0.2803 to 0.3085. Under Mean-Z aggregation, the RMSE similarly decreases from 5.8742 to 5.7764 dBZ, while the CSI increases from 0.2051 to 0.2613. Although the absolute metric values differ between the two aggregation schemes because they produce different coarse-grid target distributions, the relative improvement of CBAM-U-Net over U-Net is retained. This indicates that the performance improvement introduced by CBAM is not solely dependent on the use of Max aggregation.

The potential parallax effect associated with geostationary satellite observations was considered as a source of spatial uncertainty in the satellite–radar matching. In this study, the prediction target is radar composite reflectivity, which represents the maximum radar reflectivity within the atmospheric column, whereas the infrared observations from FY-4A are primarily associated with cloud-top radiative characteristics. Satellite and radar observations were therefore co-registered on a common geographic grid using the FY-4A grid as the spatial reference. No explicit cloud-top parallax correction was applied in the current dataset construction. Consequently, residual spatial displacement may remain, particularly for high-altitude cloud systems and deep convective clouds, and this is recognized as a limitation of the present study.

Comments 16:Line 244: “It is important to [...]” Since Batch Normalization does not appear to be used in this work, the following two sentences can be removed, in my opinion.

Response 16:Response: We thank the reviewer for this comment. Although Batch Normalization layers are used within the convolutional blocks of the proposed CBAM-U-Net, the normalization discussed in this section refers specifically to input-feature standardization rather than network-level normalization. To avoid confusion, we have removed the general discussion of Batch Normalization and retained only the description of input feature standardization (Page 12, lines 350–352).

This step mitigates the discrepancies in magnitude and numerical scale across different channels (e.g., the distinct value ranges associated with water vapor channels versus window region channels), thereby enhancing the optimization process. The normalization performed in this study is a data preprocessing step applied to input variables, aiming to reduce scale discrepancies among different channels and improve the numerical stability of model optimization.

Comments 17:Line 259: “[...] U-Net models frequently struggle to reconcile

macroscopic global structures with localized peak details" This is not necessarily a characteristic of U-Nets in general. Diffusion models are often U-Nets and do not have this issue.

Response:We agree with the reviewer that the original statement incorrectly generalized this behavior as an inherent limitation of U-Net architectures. We have removed this generalization. The revised manuscript now describes the difficulty specifically in the context of the present satellite-to-radar reconstruction task, where localized high-intensity echoes remain challenging to reproduce accurately, without attributing this limitation intrinsically to the U-Net architecture (Page 12, lines 357-361).

364 intensity gradients and intricate edge textures. For the present satellite-to-radar reconstruction
365 task, localized high-intensity echo structures remain challenging to reproduce accurately. This
366 motivates the investigation of additional feature-recalibration mechanisms to improve the represen-
367 tation of spectrally and spatially informative features. Various attention-based architectures have
368 been introduced to enhance feature representation in meteorological applications. For example,
369 Attention U-Net has been applied to multichannel Doppler-radar precipitation recognition to
370 emphasize informative features (Chen et al., 2023), while dual-attention recurrent networks have

Comments 18:Line 255-275: This section needs to be better referenced. Which other attention mechanisms have been introduced? Why are the suboptimal compared to CBAM? What is the citation for CBAM?

Response:Thank you for this helpful comment. We have revised this section to provide additional references and to clarify the motivation for selecting CBAM. Specifically, representative attention mechanisms, including attention-gating mechanisms, and Transformer-based self-attention, are now introduced and properly referenced. The original citation for CBAM (Woo et al., 2018) has also been added.

We have also revised the wording to avoid implying that other attention mechanisms are universally inferior to CBAM. Instead, CBAM was selected because the present multispectral satellite-to-radar reconstruction task requires both channel-wise discrimination among infrared observations and spatial localization of convective echoes. CBAM sequentially performs channel and spatial feature recalibration within a lightweight module that can be readily integrated into the U-Net backbone. The corresponding discussion has been added to the revised manuscript (Page 13 - 14, lines 368-383).

368 been introduced to enhance feature representation in meteorological applications. For example,
369 Attention U-Net has been applied to multichannel Doppler-radar precipitation recognition to
370 emphasize informative features (Chen et al., 2023), while dual-attention recurrent networks have

been developed to integrate complementary satellite infrared and radar information (Wang et al., 2025). Transformer-based architectures have also been explored for estimating radar reflectivity from satellite observations by modeling broader spatial dependencies (Stock et al., 2024). These approaches demonstrate the potential of attention mechanisms from different perspectives. For the present multispectral satellite-to-radar reconstruction task, both discrimination among different infrared channels and spatial localization of convective echoes are important. CBAM sequentially applies channel and spatial attention within a lightweight convolutional module (Woo et al., 2018), and has also been introduced into weather-radar echo correction tasks (Wang et al., 2026). Therefore, CBAM was selected as a task-oriented feature-recalibration module because it jointly considers channel-wise and spatial feature importance while remaining convenient to integrate into the U-Net backbone.

Based on the above considerations, this study adopts a lightweight CBAM-enhanced U-Net architecture for the satellite-to-radar reflectivity reconstruction task. By seamlessly integrating

Comments 19:Line 273: How are these datasets separately? Ideally neighboring times are not included in the training and testing sets. There should be a time buffer in between training and testing to ensure no leakage.

Response:Thank you very much for this important and constructive comment. We agree that the dataset partitioning procedure and the temporal independence of the training, validation, and test subsets should be clearly described. In the current FY-4A experiments, no explicit temporal buffer was imposed between the training, validation, and test subsets.

We acknowledge that the sample-level random partitioning used in the FY-4A experiments did not explicitly enforce storm-event-level separation. Therefore, temporally adjacent samples potentially associated with the same or closely related precipitation systems may be assigned to different subsets. We would like to clarify, however, that the test samples were not directly used during model training or normalization-statistics estimation. Thus, the limitation primarily concerns the lack of strict temporal independence between the subsets rather than direct leakage of test-set information into the training procedure. Following the reviewer's comment, we have revised the manuscript to avoid interpreting the FY-4A random-split results as a strict storm-event-independent generalization assessment (Page 14, Lines 390–406).

For the overall May–November 2023 experiment, the FY-4A samples were randomly partitioned at the individual-sample level into training, validation, and test subsets, accounting for approximately 70%, 10%, and 20% of the samples, respectively. For the three period-specific experiments, the samples within each corresponding period were separately partitioned following the same strategy, and the models were independently trained and evaluated for each period.

The proposed CBAM-U-Net follows the standard encoder-decoder structure of U-Net. The encoder consists of four convolutional stages with 64, 128, 256, and 512 filters, respectively, and symmetric decoder blocks are used for feature reconstruction. The CBAM module is inserted after the convolutional feature extraction blocks to perform sequential channel and spatial feature recalibration.The model was optimized using AdamW with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. A ReduceLROnPlateau scheduler reduced the learning rate by a factor of 0.5 when the validation loss plateaued for 10 epochs. Training used a batch size of 4 for a maximum of 150 epochs, with early stopping (patience 10) based on validation loss to prevent overfitting. All experiments were conducted on a single NVIDIA GeForce RTX 4090 GPU with PyTorch 2.0 and CUDA 11.8.The comparative models were implemented following their original architectural designs and were trained using the same dataset partition, loss function, optimizer, and training schedule for consistency.

To provide an additional robustness assessment beyond the overall May–November experiment, we also conducted three separate period-specific experiments corresponding to the pre-flood onset phase (May–June), the primary flood season (July–August), and the post-flood transition phase (September–November). These three periods are temporally non-overlapping and represent different precipitation regimes and convective environments in Sichuan. For each period, the samples belonging to that corresponding period were separately partitioned, and the models were independently trained and evaluated (Page 21, Lines 580–592).

The relative performance advantage of CBAM-U-Net remains consistent across these three temporally distinct precipitation regimes. This provides complementary evidence that the observed improvement is not confined to a single meteorological period or a specific precipitation background, but persists under different weather regimes and convective conditions. However, because sample-level random partitioning was still applied within each individual period, these experiments do not fully guarantee temporal independence between neighboring samples. They should therefore be regarded as a cross-regime robustness assessment rather than as a strict storm-event-independent validation.

In addition, the manuscript includes an independent zero-shot cross-sensor evaluation using FY-4B observations from July 2024. These data originate from a different satellite platform and a different year and were not used for model training, validation, normalization-statistics estimation, calibration, or fine-tuning. This experiment therefore provides an additional independent assessment of model transferability that is complementary to the FY-4A random-split experiments.

Finally, we have explicitly identified the absence of a temporal buffer and storm-event-level separation as a limitation of the present study. In future work, temporally blocked or storm-event-based partitioning strategies will be adopted so that neighboring samples, or samples associated with the same precipitation system, are assigned exclusively to a single subset. This will provide a more stringent assessment of model generalization to independent weather events.

We sincerely thank the reviewer for pointing out this issue, which has helped us clarify the scope and interpretation of the current generalization assessment.

4.2. *Analysis of flood season classification*

The period-specific experiments conducted in this study were designed as a cross-regime robustness analysis rather than as repeated evaluations of a single trained model across different time periods. The three periods represent distinct precipitation regimes associated with the Sichuan flood season, and separate training and evaluation experiments were conducted for each period to examine whether the relative performance advantage of the proposed method remains consistent under varying meteorological conditions. Specifically, the pre-flood period is characterized by rapidly developing convective systems, the primary flood season by more frequent severe convection and heavy precipitation events, and the post-flood period by transitional precipitation processes. These differences are associated with variations in cloud development, precipitation intensity, and the relationship between satellite infrared observations and radar-derived reflectivity. Therefore, the period-specific experiments provide complementary evidence that the relative effectiveness of the proposed architecture remains consistent across diverse precipitation regimes.

Comments 20:Line 274: What is meant by “enhancement effects?”

Response:We thank the reviewer for pointing out that the phrase “enhancement effects” was unclear. We agree that this expression did not explicitly define the aspect of improvement being evaluated. In the revised manuscript, we have removed this wording and replaced it with a more precise description of the evaluation objectives, focusing on retrieval performance under different precipitation intensity thresholds and reconstruction capability for high-intensity echoes (Page 14 line 390-406).

For the overall May–November 2023 experiment, the FY-4A samples were randomly partitioned at the individual-sample level into training, validation, and test subsets, accounting for approximately 70%, 10%, and 20% of the samples, respectively. For the three period-specific experiments, the samples within each corresponding period were separately partitioned following the same strategy, and the models were independently trained and evaluated for each period.

The proposed CBAM-U-Net follows the standard encoder-decoder structure of U-Net. The encoder consists of four convolutional stages with 64, 128, 256, and 512 filters, respectively, and symmetric decoder blocks are used for feature reconstruction. The CBAM module is inserted after the convolutional feature extraction blocks to perform sequential channel and spatial feature recalibration.The model was optimized using AdamW with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. A ReduceLROnPlateau scheduler reduced the learning rate by a factor of 0.5 when the validation loss plateaued for 10 epochs. Training used a batch size of 4 for a maximum of 150 epochs, with early stopping (patience 10) based on validation loss to prevent overfitting. All experiments were conducted on a single NVIDIA GeForce RTX 4090 GPU with PyTorch 2.0 and CUDA 11.8.The comparative models were implemented following their original architectural designs and were trained using the same dataset partition, loss function, optimizer, and training schedule for consistency.

Comments 21:Figure 4: How are the odd-numbered dimensions handled by the 2x2 max-pooling operation? E.g., for a 301x214 image, how does the max-pool result in 150x107?

Response:Thank you for pointing out this implementation detail. The 2x2 max-pooling operation uses a stride of 2 with floor rounding. Therefore, for an input feature map of 301x214, the output size is 150x107; the incomplete boundary element along the odd dimension is not included in an additional pooling window. We have added this clarification to the description associated with Figure 4 (Page 15 line 407-409).

The 2×2 max-pooling layers use a stride of 2 with floor rounding. Therefore, when an input feature map has an odd spatial dimension, the incomplete boundary window is not retained; for example, a 301×214 feature map is reduced to 150×107 after one pooling operation.

Comments 22:Line 316: “dynamically emphasizes the structurally significant regions of precipitation systems” How was it determined that the regions emphasized were structurally significant? How do we know that clear-sky regions are not also emphasized in the attention maps?

Response:We thank the reviewer for this insightful comment. We agree that the original description “dynamically emphasizes the structurally significant regions of precipitation systems” could imply that the attention maps were explicitly verified to correspond to physically meaningful structures, which was not directly demonstrated in this study.

Therefore, we have revised this description to avoid over-interpreting the attention mechanism. The revised text now describes the spatial attention operation as generating spatially varying weights to recalibrate feature responses at different locations, without assigning explicit physical significance to the highlighted regions (Page 16 line 444-446 , 450-452).

3.3. Spatial Attention Mechanism

Complementary to the channel attention, the spatial attention mechanism employed in this research utilizes the Spatial Attention Module (SAM) (Woo et al., 2018), While CAM focuses on "what" features are meaningful, SAM is explicitly designed to determine "where" these critical features are located. SAM generates spatially varying weights to recalibrate feature responses at different spatial locations. These learned weights are interpreted as feature-importance coefficients within the network rather than as direct indicators of specific meteorological phenomena. The primary operational workflow of SAM is outlined as follows:

Initially, to aggregate rich spatial information while significantly reducing computational overhead, average pooling and max pooling operations are applied simultaneously along the channel axis of the channel-refined feature map (the output from CAM). This procedure generates two distinct 2D spatial feature descriptors representing the average and maximum feature responses across all channels.

Comments 23:Line 372: “This observation [...]” There are many reasons why a U-Net could outperform a transformer here. One is the inductive bias in convolutional layers not present in transformers allowing them to perform better on smaller datasets. I see no reason why one would conclude that it is specifically the encoder-decoder design that creates this difference observed here.

Response:We agree with the reviewer that the observed performance difference cannot be uniquely attributed to the encoder–decoder design. The original interpretation was therefore overly specific.

We have revised this discussion to acknowledge that several factors may contribute to the relatively stronger performance of the convolution-based models in our experiments, including the local inductive bias of convolutional layers, model capacity, optimization characteristics, and the available training-sample size. We no

longer interpret these results as evidence that encoder–decoder architectures are intrinsically superior to Transformer-based architectures(Page 20 line 536-539).

535 It is noteworthy that, in the present experiments, convolutional neural network (CNN)-based
536 models generally outperform the Transformer-based models. The observed performance differences
537 may be associated with multiple factors, including architectural inductive biases, model capacity,
538 optimization characteristics, and the available training-sample size. The present experiments do
539 not isolate the individual contribution of these factors.

Comments 24:References:

- Several of the listed references do not have DOIs listed, which would ease finding the cited literature.
- Duan et al. 2021 appears to be duplicated.
- Sun et al. 2021 appears to be duplicated
- The year may be incorrect for article by Kou, L et al.

Response:Thank you for pointing out these issues. We have carefully checked and revised the reference list. The duplicated entries for Duan et al. (2021) and Sun et al. (2021) have been removed, and the corresponding in-text citations have been corrected accordingly. We also verified the publication information for Kou et al. and corrected the publication year. In addition, DOI information has been added to the references wherever available to facilitate access to the cited literature. The entire reference list has also been rechecked for consistency and accuracy.