Response to Reviewer 1 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:

1.①This is far from the first article presenting the use of CNN for retrieving precipitation intensity (or equivalent radar reflectivity) from geostationary multispectral infrared images. The authors need to justify the novelty of their work and to highlight what distinguishes it from previous similar works. ②The evaluation part is rudimentary, it consists in a listing of standard performance metrics, repeated three times for three different periods. This contribution is stereotyped and incremental rather than novel.

Response①: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 line 119 - 133), experimental setup and data consistency (Page30 line 784 - 785), and Conclusion (Page31-32 line 811- 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.

32

Response②:Thank you for this important comment. We agree that the presentation in the original manuscript may have given the impression that the three temporal subsets merely repeated the same standard evaluation metrics. We have therefore revised Section 4.2 to clarify the purpose of this analysis.

The three periods were not introduced as independent repetitions of the overall model evaluation, but were used to represent different precipitation regimes associated with the Sichuan flood season. 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 can alter cloud development, precipitation intensity, and the relationship between satellite infrared observations and radar-derived reflectivity. Therefore, independently training and evaluating the models within each meteorologically distinct period provides a cross-regime robustness assessment of whether the relative performance advantage remains consistent under different precipitation regimes.

To make this purpose clearer, we have reorganized the discussion in Section~4.2 and now emphasize the consistency of the severe-echo retrieval advantage across the three precipitation regimes rather than simply repeating the aggregate metrics. In particular, for the 45-70 dBZ severe-echo interval, CBAM-U-Net achieves CSI values of 0.2495, 0.3845, and 0.2992 during the pre-flood, primary flood, and post-flood periods, respectively. Relative to the best-performing competing baseline in each period, these values correspond to improvements of approximately 38.8%, 13.3%, and 15.6%, respectively. These results indicate that the improvement in severe-echo reconstruction is consistently maintained under different precipitation regimes rather than being confined to a specific temporal subset.

Accordingly, the revised manuscript now presents the period-based analysis explicitly as a cross-regime robustness evaluation, and the corresponding explanation has been added to Section 4.2.

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.

2.The abstract and introduction “promise” things that are not really delivered. It is claimed that the objective is to retrieve the "3D intra-cloud precipitation structure" but only the composite reflectivity is retrieved, the vertical variability is not resolved. The algorithm is supposedly intended for filling gaps in radar coverage but the manuscript does not go all the way to demonstrate its utilization as such, no case of "filled" radar images is showed.
Response:We appreciate the reviewer’s valuable comment. We agree that the original manuscript used some descriptions that could imply a three-dimensional reconstruction capability, while the proposed method actually retrieves two-dimensional radar composite reflectivity fields. We have revised the related statements throughout the manuscript to avoid ambiguity. Specifically, terms related to “3D intra-cloud precipitation structure” have been replaced by “radar composite

reflectivity" or "radar-derived precipitation echo structures."

We also agree that the original description of "filling radar gaps" was too strong because this study does not explicitly perform radar gap-filling experiments. Therefore, we have revised the manuscript to describe the method as providing complementary radar-like reflectivity information in regions with limited radar observations rather than directly filling radar coverage gaps. Corresponding modifications have been made in the Abstract(Page 1 line1- 17), Introduction(Page1-4 line 31 - 133), and Conclusions (Page 30-31 line 803 - 863).

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.

Keywords: FY-4A, CBAM, Radar composite reflectivity, complex terrain, deep learning reconstruction, severe convection monitoring

landslides, and urban waterlogging, posing significant threats to socioeconomic stability(Yan et al., 2022). Consequently, establishing a meteorological monitoring system characterized by high spatiotemporal resolution and robust reliability is critical for disaster prevention. Currently, while radiosondes provide precise atmospheric vertical profiling, their sparse spatial distribution and low temporal frequency (typically twice daily) severely hinder the real-time capture of rapidly evolving mesoscale convective systems(Wang et al., 2020b). Therefore, operational nowcasting predominantly relies on weather radars. By providing high-resolution backscattering information of precipitation particles, radar networks serve as the fundamental data source for severe convection monitoring(Yongguang et al., 2015). However, radar networks suffer from intrinsic spatiotemporal limitations. In topographically complex regions—such as the western Sichuan Plateau, where elevation fluctuations exceed 3000 m—radar beams can be blocked on any side depending on the radar position. In our study area, blockage tends to be particularly severe on the leeward sides of mountains(Yu et al., 2023b). This topographic blockage, coupled with the sparse distribution of stations and the recurrent "attenuation shadow" effect caused by heavy precipitation, inevitably creates extensive observation blind zones(Wang et al., 2013, 2020b; Yu et al., 2023b; Wang et al., 2026). Thus, achieving continuous monitoring of severe convective systems relying solely on ground-based radar networks in complex terrain remains exceedingly difficult.

In contrast, geostationary meteorological satellites effectively overcome these geographical barriers. Benefiting from geosynchronous orbits, these satellites deliver continuous large-scale all-weather observations with a typical temporal resolution of 5–15 minutes across most global regions. Multi-channel satellite radiative data can effectively characterize the macroscopic spatial 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 10–12 μm thermal infrared window bands serve as indicators of deep convection and vigorous updrafts, and such brightness temperature reductions match elevated radar composite reflectivity in space and time. Nevertheless, high-altitude cirrus clouds also exhibit very low infrared brightness temperatures but yield negligible radar reflectivity below typical detection thresholds. This inherent physical linkage provides a solid theoretical basis for mapping multi-channel satellite radiative features to radar reflectivity fields.

However, constructing such a precise mapping relationship poses substantial challenges. Traditional physical retrieval techniques and simplistic statistical models (e.g., linear regression, polynomial fitting) may be limited by the highly complex, nonlinear, and ill-posed relationship 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 techniques.Acting as intelligent deep learning models with robust automatic feature extraction capabilities, Convolutional Neural Networks (CNNs)—particularly encoder-decoder architectures like U-Net—have demonstrated strong capabilities in meteorological image reconstruction due to their multi-scale feature extraction ability(Veillette et al., 2018; Duan et al., 2021; Hilburn et al., 2021; Wang et al., 2025).

These issues are particularly relevant over Sichuan, where heterogeneous terrain and rapidly evolving convective systems produce radar echo fields with strong spatial gradients and highly localized intense cores. Conventional encoder–decoder architectures can recover the broad spatial morphology of radar echoes but may smooth localized high-reflectivity structures and underestimate peak intensities. Therefore, the question addressed in this study is not whether CNNs can retrieve radar reflectivity from geostationary satellite observations, but whether a lightweight spectral–spatial feature recalibration strategy can improve the representation of localized severe echoes under complex-terrain conditions while maintaining robustness across different observation regimes. The limited understanding of these issues arises from several interconnected challenges.Foremost among these is the detailed characterization of multi-scale echo structures. Severe convective echoes exhibit profound spatial heterogeneity, ranging from expansive mesoscale systems spanning hundreds of kilometers to highly localized, intense 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 challenge are the inherent difficulties associated with spatiotemporal alignment and data quality control. Geostationary satellites and radars inherently differ in spatial resolution (approximately 4 km versus 1 km) and observation latency.

Inadequate spatiotemporal registration and the inclusion of low-quality samples (e.g., radar mosaic gaps) introduce additional noise and uncertainties into model learning, in addition to random noise. Furthermore, existing methodologies frequently struggle to efficiently utilize multi-channelradiometric information(Wang et al., 2020a; Sun et al., 2021; Wang et al., 2025; Yang et al., 2023; Liu et al., 2026; Yu et al., 2023a; He et al., 2022). While satellite data provide rich cloud-top thermodynamic signatures—for instance, water vapor channels are highly sensitive to deep convection, whereas infrared window channels correlate with cloud thickness—how to effectively incorporate lightweight attention mechanisms into satellite-to-radar reflectivity reconstruction to simultaneously enhance spectral feature representation and localized convective echo reconstruction remains an important research problem.

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 improving the representation of strong echo regions and their spatial gradients.

The primary contributions of this paper are summarized as follows:

(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 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.

3.Several statements regarding the instruments capabilities or the cloud physics are inaccurate or are oversimplifications, or elude important aspects of the retrieval physics (see line-specific comments attached).

Response:We sincerely appreciate the reviewer's careful examination of the physical interpretation and instrument descriptions in our manuscript. We agree that several statements regarding satellite observation capabilities, cloud physics, and the relationship between satellite-derived radiometric information and radar reflectivity were overly simplified and required more precise descriptions.

In the revised manuscript, we have carefully reviewed and modified the relevant statements throughout the Abstract, Introduction, and Methodology sections. Specifically, we have clarified

(1) the observational characteristics and limitations of FY-4A AGRI infrared measurements,

(2) the indirect physical relationship between cloud-top radiative signatures and radar-observed precipitation echoes,

(3) that infrared observations primarily provide information from cloud-top regions, rather than directly resolving internal cloud microphysical structures.

We have also revised several expressions to avoid overgeneralization when describing infrared brightness temperature, cloud properties, and their connections to precipitation processes.

In addition, the abbreviation RCRF (Radar Composite Reflectivity) has been uniformly replaced with CR throughout the entire manuscript for conciseness and consistency with standard radar meteorology terminology.

4.Some methodological choices need to be further justified and their implication must be discussed. For example the choice of using a max-pooling operator when aggregating the radar composite reflectivity from 0.01 deg to 0.04 deg.

Response:Thank you for this important comment. We have substantially revised the description of the radar--satellite spatial matching procedure and further discussed the implications of using Max aggregation. The 0.01°SWAN radar composite reflectivity is aggregated onto the approximately 0.04° FY-4A grid using a maximum-preserving operator. This choice was made to retain localized high-reflectivity convective cores that may be weakened by spatial averaging during the transition from the finer radar grid to the coarser satellite grid.

We also acknowledge that Max aggregation is not a neutral resampling operation, as it may shift the coarse-grid reflectivity distribution toward higher values and increase the apparent coverage of strong echoes. Therefore, we added a sensitivity experiment using linear-reflectivity mean aggregation (Mean-Z) as an alternative. Under Max aggregation, CBAM-U-Net reduced RMSE from 6.8109 to 6.6989 dBZ and increased the CSI for echoes ＞45 dBZ from 0.2803 to 0.3085 relative to U-Net. Under Mean-Z aggregation, CBAM-U-Net similarly reduced RMSE from 5.8742 to 5.7764 dBZ and increased CSI from 0.2051 to 0.2613. These results indicate that the relative improvement introduced by CBAM is retained under both aggregation strategies, although the absolute metric values differ because the two aggregation methods produce different coarse-grid target distributions.

The corresponding methodological explanation, sensitivity experiment, and discussion have been added to the revised manuscript (Page 10-11 line 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.

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

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., $\geq$ 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 $\geq$ 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.

5.It is not clear how the dataset is partitioned into training, validation, and testing subsets. The only information provided is that “the dataset was partitioned into training, validation, and testing subsets following a standard 7:2:1 ratio.” This raises concerns about the independence of the three subsets. If samples in the different subsets are in close spatial and/or temporal proximity, or potentially originate from the same storm systems, the subsets cannot be considered statistically independent. The authors should clarify how the partitioning was performed and, in particular, whether measures were taken to prevent samples associated with the same or closely related storm systems from being distributed across different subsets. Otherwise, the reported model performance may be affected by information leakage and may not accurately reflect the model’s ability to generalize to independent cases.

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.
To further examine whether the relative performance of the proposed model is maintained under different meteorological backgrounds, we additionally conducted three 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 distinct and represent different precipitation and convective regimes in Sichuan. For each period, the samples belonging to that period were separately partitioned, and the models were independently trained and evaluated. The pre-flood period is dominated by rapidly developing and relatively localized convective systems, the primary flood season contains more frequent severe convection and heavy precipitation, whereas the post-flood period is characterized by transitional precipitation processes.

The consistent relative advantage of CBAM-U-Net across these three independently conducted period-specific experiments provides complementary evidence that the performance improvement of the proposed architecture is not confined to a single precipitation regime or meteorological background. In this sense, the three-stage experiments provide a cross-regime robustness assessment across temporally distinct weather periods and different convective environments. However, we emphasize that these experiments do not completely resolve the reviewer's concern regarding storm-event-level independence, because sample-level random partitioning was still applied within each individual period. Therefore, we do not regard them as a substitute for strict storm-event-based validation, but rather as complementary evidence of model robustness under different precipitation regimes.

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.

32

The specific questions and responses in the article are as follows:

1.Line 23“radar beams are severely blocked on the leeward sides of mountains.”--Why specifically on the leeward side? Beam blockage can happen on any side, it only depends on the radar location.

Response:We fully agree with the reviewer that radar beam blockage caused by terrain obstacles can occur in any azimuth and is fundamentally controlled by the relative positions between the radar site and surrounding topography; there is no universal rule that blockage preferentially appears on the leeward side in all regions.

In the revised manuscript, we have adjusted the relevant description to avoid overgeneralization. We clarify that prevalent beam blockage on the leeward side is a regional topographic feature within our research domain, instead of presenting it as a general conclusion applicable to all radar sites. We emphasize that this spatial distribution is tied to the specific mountain layout and radar deployment in this study (Page 2 line 31-35).

monitoring(Yongguang et al., 2015). However, radar networks suffer from intrinsic spatiotemporal limitations. In topographically complex regions—such as the western Sichuan Plateau, where elevation fluctuations exceed 3000 m—radar beams can be blocked on any side depending on the radar position. In our study area, blockage tends to be particularly severe on the leeward sides of mountains(Yu et al., 2023b). This topographic blockage, coupled with the sparse distribution of

2.Line 28.“Leveraging their geosynchronous orbit,they facilitate large-scale, all-weather continuous observations at a minute-level temporal resolution.”--Only 5 to 15 minutes in most areas across the globe.

Response:We thank the reviewer for correcting the temporal sampling description. We have revised the sentence to clarify that the standard revisit interval of the geostationary satellite system ranges from 5 to 15 minutes for most global regions, removing the ambiguous "minute-level" expression that implied sub-minute sampling frequency (Page2 line41-43).

barriers. Benefiting from geosynchronous orbits, these satellites deliver continuous large-scale all-weather observations with a typical temporal resolution of 5–15 minutes across most global regions. Multi-channel satellite radiative data can effectively characterize the macroscopic spatial

3.Line 29“Although satellites primar-ily detect cloud-top thermal radiation and microphysical properties, whereas radars measure the scattering characteristics of

precipitation particles within and beneath clouds, a strong physical and statistical correlation exists between the two during severe convective events (Roberts and Rutledge, 2003; Wang et al., 2022b).”--There are different type of instruments (including radars) onboard different satellites, capable of observing different aspects of cloud's macrophysics and microphysics.The scattering characteristics of precipitation particles is a microphysical property.
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

4.Line 32.“Specifically, the pronounced decline in satellite infrared brightness temperature—indicative of deep convection and intense updrafts—aligns spatiotemporally with corresponding increases in radar composite reflectivity.” --This is only true for the thermal infrared chanels (10-12 microns). Other infrared channels respond to water vapor, ozone and CO2 among others.Not always. Some high altitude clouds (e..g. cirrus) are not associated with convective updrafts and show very low IR brightness temperatures and also low radar reflectivity (below the detection level of most weather radars).
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).

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

5.Line 38. "Because conventional models rely heavily on manually crafted feature extrac- tion rules, they struggle to fully exploit the high-dimensional spatiotemporal correlations embedded in multi-channel satellite datasets."--What are "conventional models?" Today, the majority of operational models rely on convolutional deep neural networks. In 2026 CNNs are standard.

Response:We thank the reviewer for this important clarification. We agree that the term "conventional models" in the original manuscript was too broad and potentially misleading, particularly because convolutional neural networks are now widely used as standard approaches in operational and research applications.

In the original manuscript, "conventional models" was intended to refer specifically to traditional empirical and statistical retrieval approaches, such as linear or polynomial regression and heuristic retrieval methods based on manually designed spectral, textural, or meteorological predictors, as well as physically based retrieval approaches relying on radiative transfer calculations, cloud microphysical assumptions, threshold relationships, or parameterized empirical formulations. It was not intended to refer to modern CNN-based deep learning models.

To avoid this ambiguity, we have revised the corresponding paragraph and removed the overly general expression "conventional models." The revised text now explicitly distinguishes traditional empirical/statistical and physically based retrieval approaches from modern deep-learning methods, including convolutional neural networks. Specifically, we now state that traditional empirical and statistical retrieval methods often rely on manually designed predictors, whereas physically based retrieval approaches depend on radiative-transfer simulations and parameterized physical assumptions. Deep-learning approaches are then introduced separately as data-driven methods capable of automatically learning hierarchical representations from multispectral observations.

We appreciate the reviewer for pointing out this terminology issue, which has helped us make the methodological distinction more precise and consistent with the current state of the field.(Page 2–3, Lines 62–73).

polynomial fitting) frequently fail due to the highly complex, nonlinear, and ill-posed relationship 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

2

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

6.Line 49. “Standard U-Net architectures, while capable of recovering macroscopic morphology, frequently suffer from a persistent "smoothing effect," leading to the severe blurring of convective boundaries and the systematic underestimation of peak intensities within strong echo cores.”--The "smoothing" effect is not caused by the U-net architecture, it generally comes from the objective function. There exist several non-smooth retrievals algorithms that use U-net architectures. See references below:
https://doi.org/10.1109/LGRS.2023.3284278
https://doi.org/10.1109/TGRS.2025.3548518
https://doi.org/10.1038/s41612-025-01103-y

Response:We sincerely thank the reviewer for this important clarification and for providing the relevant references. We fully agree that the smoothing effect should not be attributed to the U-Net architecture itself. As the reviewer correctly pointed out, over-smoothed retrievals are often closely related to the choice of the training objective, particularly conventional pixel-wise regression losses, whereas U-Net-based frameworks combined with appropriately designed objectives can produce substantially sharper and more realistic retrievals.

Accordingly, we have revised the corresponding statement in the manuscript to remove the misleading causal attribution to the U-Net architecture. The revised text now states that standard U-Net architectures may exhibit a smoothing tendency when trained with conventional pixel-wise regression objectives, particularly for localized high-reflectivity structures, rather than implying that smoothing is an inherent property of U-Net itself. We have also incorporated the references suggested by the reviewer to support this clarification.

We would also like to clarify the role of CBAM in the present study. We agree that modifying the objective function—for example, through weighted, perceptual, adversarial, or other structure-aware losses—is an important route for reducing over-smoothing. However, our study investigates the problem from a complementary feature-representation perspective rather than from loss-function redesign. Under the same training objective used for the baseline models, CBAM sequentially applies channel and spatial attention to adaptively enhance informative multispectral responses and spatial features associated with localized high-reflectivity regions. This feature recalibration helps preserve strong-echo cores, spatial gradients, and echo boundaries during encoding and decoding, thereby alleviating, rather than fundamentally eliminating, the blurring and peak-intensity underestimation observed with the baseline U-Net.

Importantly, all comparative models in our experiments were trained using the same dataset partition, loss function, optimizer, and training schedule. Therefore, the relative improvement of CBAM-U-Net over the standard U-Net under the same objective function provides evidence that enhanced channel–spatial feature representation can complement loss-function-based approaches in reducing strong-echo smoothing and intensity underestimation.

We appreciate the reviewer’s comment, which helped us distinguish more clearly between the contribution of the training objective and that of the network feature-representation mechanism.

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

7.Line 54. “Inadequate spatiotemporal registration and the inclusion of low-quality samples (e.g., radar mosaic gaps) introduce se-vere systematic biases into the learning process.”--Why systematic biases? Aren't these types of error mostly random?

Response:We appreciate the reviewer’s careful observation. The errors caused by imperfect spatiotemporal registration and radar mosaic gaps are indeed predominantly random rather than systematic, and the original description was imprecise. We have revised the manuscript to replace “systematic biases” with “noise and uncertainties” to accurately characterize the nature of these errors (Page 3 line 99-101).

Inadequate spatiotemporal registration and the inclusion of low-quality samples (e.g., radar mosaic gaps) introduce additional noise and uncertainties into model learning, in addition to random noise. Furthermore, existing methodologies frequently struggle to efficiently utilize

8.Line 61. “To address the aforementioned challenges, this study proposes an intelligent, end-to-end reconstruction framework, termed CBAM-U-Net, designed specifically to compensate for radar blind spots over complex terrain utilizing multi-channel observations from the FY-4A geostationary satellite.”--The paper itself doesn't showcase the filling of the blind spots. Also for this particular application it would make sense to utilize an "inpainting" approach with the infrared brightness temperature serving as auxiliary covariate fields.

Response:We appreciate the reviewer’s valuable comment.In the revised manuscript, the description has been adjusted to focus on radar composite reflectivity reconstruction over complex terrain using FY-4A observations, rather than claiming direct radar blind spot compensation (Page 3 line 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

9.Line 77. “massive dataset (14,023 samples).”--What does constitute a sample? If a sample equals a single pixel or single field of view, then 14 000 is not at all a "massive" dataset by satellite remote sensing standards. For comparison, the radar onboard the GPM satellite produces 388 000 samples per orbit and has completed more than 70 000 orbits since 2015.

Response:We appreciate the reviewer’s insightful comment and clarify the definition of samples in this study to eliminate ambiguity. Different from pixel-level or single field-of-view samples adopted in satellite orbital observation studies, one sample in our work is defined as a complete spatiotemporally matched pair of FY-4A satellite multi-channel observations and ground-based radar composite reflectivity mosaics for the study domain. A total of 14,023 matched sample pairs were retained for regional-scale radar echo reconstruction and evaluation. (Page 9 line 278-281).

echoes as much as possible. After screening according to this criterion, a total of 14,023 valid samples were finally obtained. The distribution of the valid ratio is presented in Figure 3.In this study, one sample denotes a spatiotemporally matched pair consisting of one FY-4A multispectral observation and the corresponding SWAN radar composite reflectivity mosaic over the study domain.

In the revised manuscript, the relevant description in the summary of contributions has been completely rewritten and updated, and the potentially misleading term “massive” has been removed (Page 4 line 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 ≥ 45 dBZ, different flood-season precipitation regimes, and an independent zero-shot FY-4B test.

10.Line 122. “The objective is to establish a robust mapping relationship to infer the 3D "intra-cloud precipitation structure”--Can the retrieval of the composite reflectivity really be considered a retrieval of the 3D structure? The vertical variability is not resolved.

Response:We thank the reviewer for this precise and critical comment. We fully agree that the retrieval of radar composite reflectivity cannot resolve detailed vertical precipitation variability and thus does not represent a complete 3D intra-cloud precipitation structure. Composite reflectivity is a two-dimensional column-maximum product obtained by selecting the maximum radar reflectivity along the vertical dimension at each horizontal location. Accordingly, we have revised the manuscript description to accurately clarify that our model infers column-wise maximum radar composite reflectivity from 2D cloud-top radiation observations, rather than full three-dimensional precipitation structural information. This revision eliminates inappropriate overstatement and strictly aligns the manuscript description with the actual model outputs and research scope (Page 6 line 186-188).

In data-driven predictive modeling, constructing a physically meaningful input space is paramount. The feature selection methodology employed in this investigation is deeply rooted in the integration of atmospheric radiative transfer theory and cloud microphysics. The objective is to establish a robust mapping relationship to infer the column-wise maximum radar composite reflectivity from the 2D cloud-top thermal radiation field. Recognizing that strong radar reflectivity

11.Line 130. “Additionally, the 3.72 μm (Ch08) and 8.5 μm (Ch11) channels leverage the refractive index disparities of hydrometeors across the short-wave and mid-wave infrared spectra.” --Only of the hydrometeors in the uppermost cloud layers. None of the optical channels (visible or IR) can penetrate optically thick clouds.

Response:We agree with the reviewer that optical and infrared channels cannot penetrate optically thick clouds, and therefore only sense hydrometeors in the uppermost cloud layers rather than the full cloud depth. We have revised the description accordingly to accurately reflect this physical constraint (Page 6 line 198-201).

Additionally, the 3.72 µm (Ch08) and 8.5 µm (Ch11) channels capture refractive index differences of hydrometeors within the uppermost cloud layers across the short-wave and mid-wave infrared spectra. As optical and infrared channels cannot penetrate optically thick clouds, these observations primarily reflect information from cloud tops. These channels provide vital microphysical indicators regarding cloud-top particle sizes and phase states (liquid water versus ice).

12.Line 134. “while effectively suppressing background interference in-herent to single-channel observations.”--What does this mean? I do not understand this part of the sentence.

Response:Thank you for pointing this out. We agree that the phrase “suppressing background interference inherent to single-channel observations” was ambiguous and insufficiently supported. We have therefore removed this statement. The revised text now states that the derived channel-difference variables provide additional spectral contrast information associated with cloud-top thermal and spectral characteristics and may facilitate feature representation in the deep learning model (Page 6 line 205-209).

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

13.Line 136. “core regions of intense convection.”--I am not sure what the "core regions of intense convection" is supposed to mean. But given the fact that optical wavelengths cannot penetrate deeply into thick clouds (deep convective clouds in particular), claiming that they can reveal microphysical processes in the "core" of convective areas may be considered an overstatement.

Response:Thank you for this important comment. We agree that the previous expressions “core regions of intense convection” and “microphysical signatures” could overstate the information that can be inferred from satellite infrared observations, particularly because infrared measurements primarily characterize cloud-top radiative properties and cannot directly probe the internal microphysical structure of optically thick convective clouds. We have therefore removed these expressions and revised the text to refer only to additional spectral contrast information associated with cloud-top thermal and spectral characteristics (Page 6 line 205-209).

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

14.Line 140. “The Radar Composite Reflectivity (RCRF) product utilized in this study is derived from the Severe Weather Automatic Now- casting (SWAN) system, operated by the China Meteorological Administration (CMA).”--More information is needed here. Non-expert may not known how the composite reflectivity is derived from vertical reflectivity profiles. The explanations given on lines 149-152 should come earlier.Which frequencies exactly?

Response:We sincerely appreciate the reviewer’s constructive and rigorous comments. We have comprehensively revised the relevant paragraph to address these points, with modifications highlighted in red. We have moved the definition of composite reflectivity earlier, explicitly stating that it is obtained by taking the maximum reflectivity over the entire vertical atmospheric column (Page 7 line 211-225).

2.2.2. Radar Composite Reflectivity Data

Mathematically and physically, CR (unit: dBZ) is constructed by extracting the maximum reflectivity factor over all vertical heights z within each atmospheric column at a given horizontal location (x, y). This operation projects the three-dimensional radar volume data onto a two-dimensional horizontal grid, representing the strongest convective echo at each horizontal position.

$$\mathrm{CR}(x, y) = \max_{z} Z(x, y, z) \tag{1}$$

Compared to reflectivity at a single elevation angle (e.g., 0.5° or 1.5°) or a Constant Altitude Plan Position Indicator (CAPPI, e.g., 3 km), CR consolidates the peak intensity information throughout the entire vertical structure of the cloud system. This attribute makes it significantly more effective for identifying severe convective systems characterized by profound vertical development and their associated heavy precipitation centers. Consequently, this study designates CR as the retrieval target (ground truth) for the proposed deep learning framework. This selection provides a rigorous benchmark to evaluate the capability of FY-4A multi-channel observations in accurately characterizing and reconstructing precipitation echo structures across complex topography.

15.Line 143. “S-band and C-band”--How are the discrepancies between S-band and C-band reflectivity accounted when mapping the composite reflectivity?

Response:We thank the reviewer for the careful comment.

We have specified the exact frequencies of the radar network: S-band (2.8–3.0 GHz) and C-band (5.3–5.7 GHz).

We also added a description that the SWAN system applies attenuation correction for C-band radars and inter-calibration against S-band radars to ensure consistency between different radar frequencies.

All revisions are marked in red (Page 7 line 230-234).

nowcasting system for severe convective weather, SWAN integrates observational data from a dense network of over 200 S-band (2.8–3.0 GHz, 10 cm wavelength) and C-band (5.3–5.7 GHz, 5 cm wavelength) Doppler weather radars across China. To ensure consistency between S-band and C-band reflectivity measurements, the operational SWAN system implements attenuation correction for C-band radars and inter-calibration against S-band reference radars to reduce frequency-related biases prior to mosaicking. Through rigorous quality control algorithms—encompassing ground clutter mitigation, outlier smoothing, and multi-radar 3D mosaicking—SWAN generates

16.Line 154. “The formula is expressed as: $RCRF(r,\phi) = \max_{\theta\in\{\theta_1,\theta_2,\dots,\theta_n\}} Z(\theta,r,\varphi)$ ”--This definition is only valid if the range of theta values is small. I think it would make more sense to define it as the max along z in a (x,y,z) Cartesian coordinate system.

Response:We thank the reviewer for this important and rigorous comment. We have revised the definition of composite reflectivity (CR) to follow the Cartesian coordinate system, defining CR as the maximum reflectivity value along the vertical height z at each horizontal grid point(x,y). The original polar coordinate and elevation-angle-based formulation has been removed accordingly, to ensure physical consistency with standard operational radar products. All revisions are highlighted in red in the manuscript (Page 7).

$$\mathrm{CR}(x,y) = \max_{z} Z(x,y,z) \tag{1}$$

17.Line 174. “Title: FY-4A AGRI L1 Geometric Positioning and Latitude Longitudinal Conversion Section”--This section could be moved in supplementary material or appendices.

Response:A standard geolocation conversion algorithm is applied to convert FY-4A AGRI L1 pixel coordinates into regular latitude–longitude grids for consistent spatial matching with radar mosaics; full derivation formulas and calculation procedures are provided in Appendix A (Page 36).

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.

18.Line 185. “data file”--The current location can be found in Appendix A, lines 919-921.What volume of date does a "file" correspond to? Is a "file" and a "sample" the same thing. This is all very confusing.

Response:We apologize for the ambiguous terminology that caused confusion between satellite raw files and matched dataset samples. We have revised the manuscript to clearly distinguish the two concepts:

1.The term "data file" refers to the original FY-4A AGRI L1 4 km HDF product, where each individual file stores one complete full-disk satellite observation acquired at a single UTC timestamp (approximately 15-minute scanning cycle per file, 200 MB file size). Navigation geolocation coefficients (COFF, CFAC, LOFF, LFAC) are embedded as internal attributes inside each HDF file.

2.A "sample" refers to a spatially and temporally co-registered paired dataset combining one satellite observation snapshot and one SWAN radar composite reflectivity mosaic. Our full dataset contains 14,023 such matched sample pairs, which means one sample does not directly equal one raw satellite HDF file, as multiple satellite files may be discarded due to missing radar data or low-quality cloud conditions during matching preprocessing. We have added this clarification in the text to eliminate ambiguity for readers (Page 9 line 279-281, Page 36 line 1002-1004).

samples were finally obtained. The distribution of the valid ratio is presented in Figure 3.In this study, one sample denotes a spatiotemporally matched pair consisting of one FY-4A multispectral observation and the corresponding SWAN radar composite reflectivity mosaic over the study domain.

$$180 \cdot 2^{-16} \cdot \mathrm{LFAC}$$

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:

19.Line 119. “To preserve the core char-acteristics of strong convective systems during the upscaling process and avoid the smoothing of extreme values caused by averaging, this study adopts a max-pooling strategy for spatial matching. Specifically, the maximum value among the 16 radar observations at 0.01◦ resolution within a 0.04◦ grid cell is assigned as the label value for that grid cell”--This methodological choice is critical and needs some further justification. This basically is a new definition of the composite reflectivity.

Response:Thank you for pointing this out. We agree that our previous wording could be interpreted as introducing a new physical definition of radar composite reflectivity. We have therefore revised this part of the manuscript to clearly distinguish the two operations.

The original SWAN product is already radar composite reflectivity, which is obtained through the maximum reflectivity along the vertical radar column at each horizontal location. In contrast, the Max operation used in this study is applied only in the horizontal direction for spatial matching between the 0.01° radar grid and the approximately 0.04° FY-4A grid. Therefore, this horizontal aggregation does not redefine radar composite reflectivity. Instead, the resulting target is now explicitly described as a maximum-preserving coarse-grid representation of the original radar composite reflectivity.

To make this distinction clearer, we also revised the notation to $\mathrm{CR}^{\max}_{0.04^\circ}(\mathrm{i},\mathrm{j})$, and explicitly state that the model trained with Max aggregation should be interpreted as retrieving this maximum-preserving coarse-grid representation rather than the spatially averaged reflectivity within each FY-4A grid cell.

In addition, a Mean-Z sensitivity experiment was added to quantify the influence of this methodological choice and to demonstrate that the relative improvement of CBAM-U-Net over U-Net is not solely dependent on Max aggregation (Page 10-11 line 283-332).

20.Line 251. “hierarchical events.”--What does this mean?

Response:We thank the reviewer for pointing out this ambiguous expression. The "hierarchical events" refers to convective precipitation events defined at multiple hierarchical radar reflectivity thresholds. We calculate categorical metrics (POD, FAR, CSI) separately for different dBZ thresholds to evaluate the model’s ability to capture weak, moderate and intense convection. We have revised the sentence to explicitly clarify that the third dimension evaluates detection performance for convective events across multiple reflectivity thresholds, eliminating ambiguity (Page 12 line 354-357).

2.3. Evaluation Method

To thoroughly assess model performance, this study establishes a comprehensive evaluation framework encompassing three dimensions: numerical regression accuracy, spatial structural consistency of reconstructed echo fields, and the ability to detect convective events under hierarchical reflectivity thresholds. (detailed formulas are provided in Table 3).

21.Page 12 Regarding the Question of PSNR SSIM Formula.--What is the actual value used for the max in the formula? The figures 7–13 show reflectivities up to 70 dB.Why "brightness"? In the case of radar images, x and y represent the reflectivity not the brightness.

Response:We appreciate the reviewer's careful review of Table 3.

We have accordingly modified the description of PSNR parameters and value ranges, with all changes marked in red.

It has been clearly stated in the expression that the MAX constant in the PSNR formula is fixed at 70 dBZ, which is consistent with the upper limit of radar reflectivity used in Figure 7-13. Additional clarification was provided, stating that the 20-40dB range represents the typical PSNR values observed in our echo reconstruction experiments, rather than the entire theoretical range of this metric.

All relevant revisions are highlighted in red for easy inspection (Page 13).

Table 3: Evaluation Metrics and Corresponding Formulas

Metric Formula	Parameter Description	Value Range	Physical Meaning
$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-y_{\mathrm{pred}})^2}$	y_i: Ground truth; y_{pred}: Prediction; N: Total samples; $\bar{y}_i$: Mean of ground truth	$[0,+\infty)$	Penalizes large errors, sensitive to outliers, widely used for regression evaluation.
$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\lvert y_i-y_{\mathrm{pred}}\rvert$		$[0,+\infty)$	Calculates average absolute error, robust to outliers with stable performance.
$\mathrm{R}^2=1-\frac{\sum_{i=1}^{N}(y_i-y_{\mathrm{pred}})^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2}$		$(-\infty,1]$	Quantifies goodness of fit; closer to 1 indicates stronger regression ability.
$\mathrm{PSNR}=10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right)$	MAX: Upper limit of radar reflectivity (70 dBZ); MSE: Mean Square Error of reflectivity	20–40 dB (typical range in this study)	Higher value indicates smaller echo distortion and better reconstruction quality.
$\mathrm{SSIM}=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}$	μ_x,μ_y: Mean reflectivity; σ_x^2,σ_y^2: Variance; σ_{xy}: Covariance; c_1,c_2: Stability constants	$[0,1]$	Measures similarity in reflectivity magnitude, contrast and structure; 1 = perfect consistency, 0 = no similarity.
$\mathrm{POD}=\frac{\mathrm{TP}}{\mathrm{TP+FN}}$	TP: True Positive; FN: False Negative; FP: False Positive	$[0,1]$	Reflects event capture ability; closer to 1 indicates higher hit rate.
$\mathrm{FAR}=\frac{\mathrm{FP}}{\mathrm{TP+FP}}$		$[0,1]$	Quantifies false alarm level; lower value indicates more reliable predictions.
$\mathrm{CSI}=\frac{\mathrm{TP}}{\mathrm{TP+FP+FN}}$		$[0,1]$	Comprehensively balances hit rate and false alarm rate, core metric for event evaluation.

22.Line 356. “Peak Signal-to-Noise Ratio (PSNR)”--Given the fact that the PSNR is basically a constant minus the log of the MSE, I find it strange to put them in two different categories. The two measures are strictly equivalent in the end.

Response:We fully agree with the reviewer’s mathematical reasoning that PSNR is a logarithmic transformation of MSE and they share identical core error information, so separating them into two independent evaluation dimensions is logically inappropriate. We have restructured the three evaluation categories in the revised manuscript:

All pixel-wise quantitative error metrics (RMSE, MAE, R^2 , PSNR) are integrated into one unified group, and we briefly note the mathematical link between PSNR and MSE to avoid misleading classification.

SSIM is separated as an independent structural metric, since it evaluates spatial morphology and texture instead of simple pixel intensity errors, which provides complementary information to RMSE/PSNR.

Threshold-based categorical radar verification metrics (POD, FAR, CSI) remain the third dimension for operational meteorological assessment. This revised classification eliminates the contradiction pointed out by the reviewer and clarifies the distinct physical information captured by each group of metrics (Page 19 line 510-518).

> (1) **Pixel-wise Numerical Regression Metrics.** Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), the Coefficient of Determination (R^2), and Peak Signal-to-Noise Ratio (PSNR). PSNR is derived via logarithmic transformation of MSE and thus shares identical pixel error information with RMSE. These metrics are utilized to quantify the absolute, pixel-level intensity deviations between the deep learning retrieved echoes and the ground-truth radar observations.
>
> (2) **Spatial Structural Similarity Metric.** Only the Structural Similarity Index (SSIM). Different from pixel error-based indicators above, SSIM independently evaluates macroscopic morphological consistency and spatial texture of convective echoes without relying on pointwise intensity bias.

23.Line 376. “global” overall "rather than "global". This is a regional study, not a global study.

Response:We appreciate the reviewer’s precise comment. To avoid misunderstanding with "global" (which might imply global coverage), we have replaced it with "overall" in the revised manuscript, consistent with the regional nature of this study. The corresponding sentence has been updated accordingly (Page 20 line 540).

> 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

24.Line 382. “event-based verification.”--How is this event-based verification? What does constitute an event?

Response:This study divides radar echo intensity (dBZ) into three intervals for threshold-based pixel-wise categorical verificationWe appreciate this critical clarification.

In this work, the so-called "event" refers to pixels where radar reflectivity exceeds a predefined intensity threshold (e.g., ≥25 dBZ, ≥45 dBZ). We perform pixel-wise categorical verification rather than object-based tracking of isolated connected convective storms.

We partition reflectivity into three tiers (0–25 dBZ, 25–45 dBZ, 45–70 dBZ). For each threshold interval, every pixel is classified as echo or non-echo, and we compute POD, FAR and CSI at the pixel scale.

To avoid confusion with object-based storm verification, we have revised the manuscript and replaced the ambiguous phrase "event-based verification" with threshold-based pixel-wise categorical verification in the relevant evaluation descriptions and section heading. We also clarify in the revised text that we evaluate the model’s ability to correctly reconstruct pixels corresponding to weak, moderate and intense convective echoes (Page 20 line 547-548).

4.1.2. *Threshold-based pixel-wise categorical verification*

To further evaluate the model's capability to capture precipitation events of different intensities, this study divides radar echo intensity (dBZ) into three intervals for threshold-based pixel-wise categorical verification with the results presented in Table 6.

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 15: Threshold-based categorical verification of CBAM-U-Net and comparative models in the cross-satellite FY-4B test

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	0.8655	0.6292	0.1833	0.1493	0.3772	0.5428	0.7514	0.4555	0.2758
Attention-U-Net	0.9139	0.5287	0.2474	0.1733	0.3233	0.6039	0.7670	0.4221	0.3203
MCDA-U-Net	0.8445	0.5114	0.2815	0.1659	0.4415	0.6478	0.7125	0.4112	0.2815
SegFormer	0.8588	0.6151	0.1978	0.1532	0.3919	0.6558	0.7433	0.4405	0.1824
SRViT	0.7625	0.5646	0.1251	0.1915	0.5279	0.8129	0.6458	0.3461	0.2226
CBAM-U-Net	**0.9258**	**0.6248**	**0.3316**	**0.1369**	**0.3482**	**0.5664**	**0.7690**	**0.468**	**0.3723**

25.Page 21 about Case Study Explanation "local convective event recorded at 20:19:17 (Beijing Time)" Did this "event" only last one second?

Response:Regarding the convective event: The timestamp 20:19:17 BT represents an instantaneous snapshot within the lifecycle of the convective system, rather than the full duration of the event itself. We have revised all the caption text to clarify that this time is one sampling moment during the convective process, to avoid the misleading interpretation that the storm only lasted one second.

Page 23 line 623-624：

To intuitively demonstrate the inversion performance of the models during the pre-flood active period, Figures 7 and 8 present a comparative analysis of snapshots captured at 20:19:17 (Beijing Time) on May 10, 2023 during a typical local convective event.

Page 25 line 675-676：

Meanwhile, to visually demonstrate the inversion effect of the model during the active period, Figures 9 and 10 present a visual comparison for a typical convective event centered at approximately 18:43 Beijing Time on 6 August 2023.

Page 28 line 747-748：

For the comprehensive evaluation of model performance during the post-flood recession period based on test set samples, Figures 11 and 12 present a visual case analysis for a typical convective event at 14:23 BT on 20 September 2023 (Beijing Time)

26.Annotations for all individual case diagrams:The color legend label says "reflectance". Is it supposed to be reflectivity? in dB units?

Response:We acknowledge the terminology error. The word “reflectance” belongs to optical remote sensing and is inappropriate for radar data. It has been corrected to reflectivity, and we explicitly mark the unit dBZ in all color bar legends and figure captions throughout the revised manuscript.

(2) **Visual analysis of typical cases**

To intuitively demonstrate the inversion performance of the models during the pre-flood active period, Figures 7 and 8 present a comparative analysis of snapshots captured at 20:19:17 (Beijing Time) on May 10, 2023 during a typical local convective event.

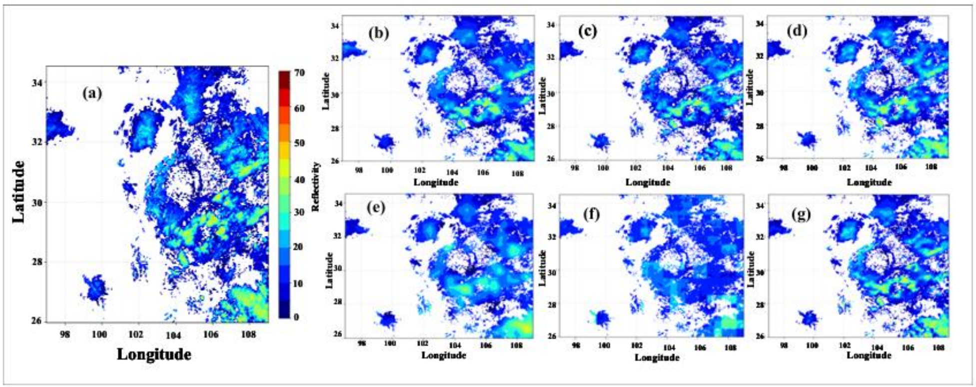

Figure 7: Panoramic overview of radar composite reflectivity(dBZ) at 20:19:17 on May 10, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net

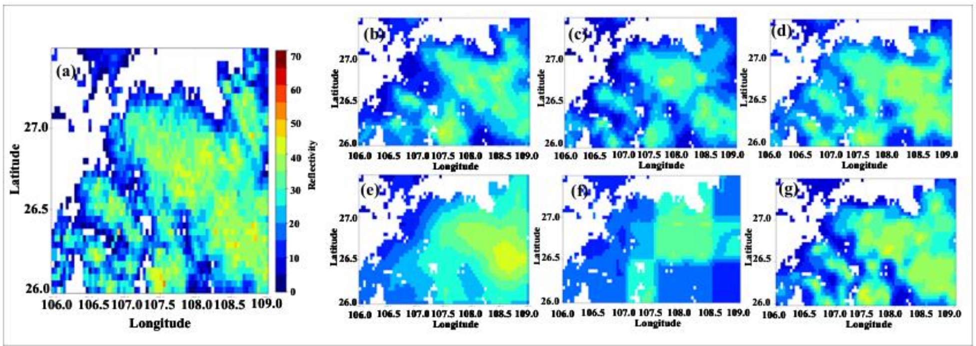

Figure 8: Local magnified view of radar composite reflectivity(dBZ) at 20:19:17 on May 10, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net

(2) **Visual analysis of typical cases**

Meanwhile, to visually demonstrate the inversion effect of the model during the active period, Figures 9 and 10 present a visual comparison for a typical convective event centered at approximately 18:43 Beijing Time on 6 August 2023.

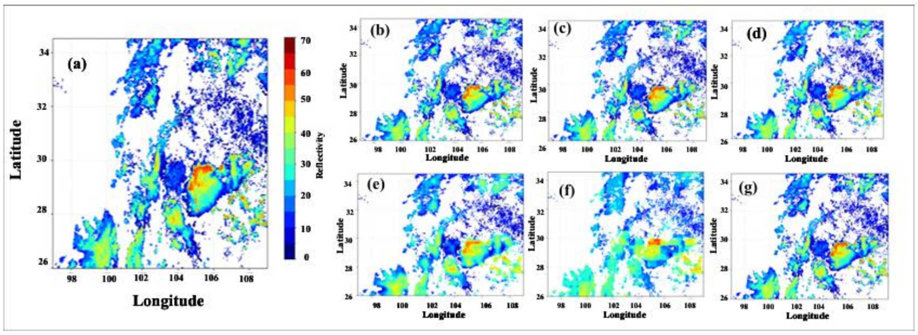

Figure 9: Radar combined reflectivity(dBZ) overview at 18:42:53 on August 6, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net

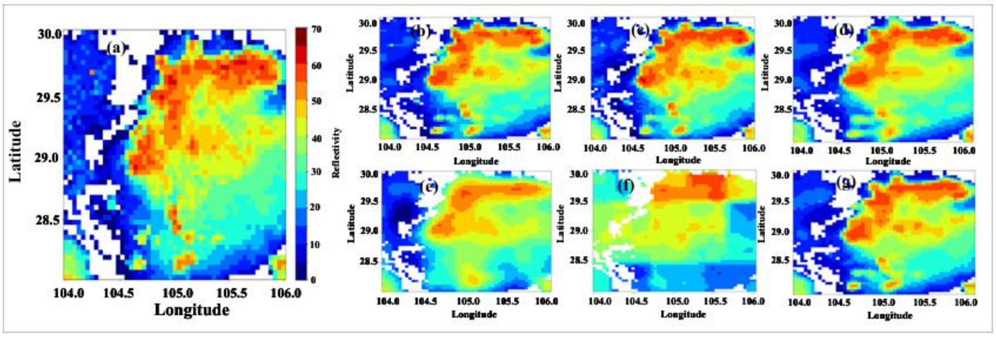

Figure 10: Radar combined reflectivity(dBZ) detailed view at 18:42:53 on August 6, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net

(2) **Visual analysis of typical cases**

For the comprehensive evaluation of model performance during the post-flood recession period based on test set samples, Figures 11 and 12 present a visual case analysis for a typical convective event at 14:23 BT on 20 September 2023 (Beijing Time)

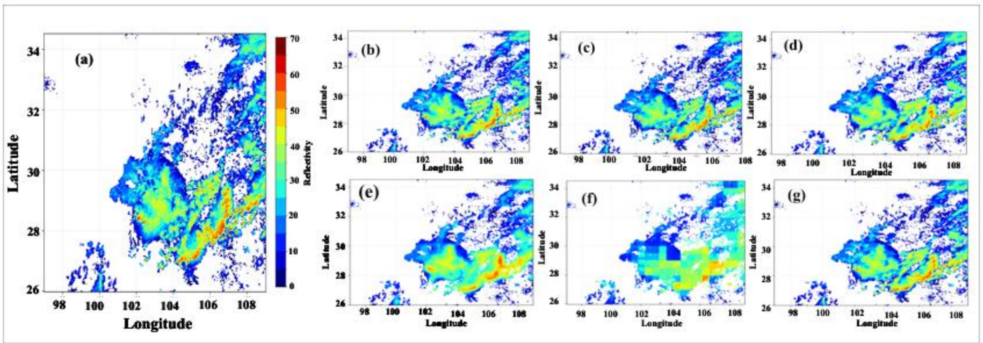

Figure 11: Panoramic overview of radar composite reflectivity(dBZ) at 14:23:35 (Beijing Time) on September 20, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net

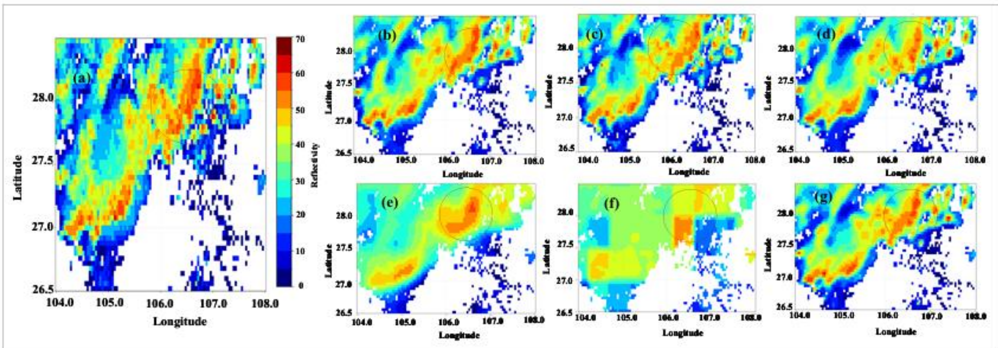

Figure 12: Radar composite reflectivity (dBZ) detailed view at 14:23:35 (Beijing Time) on September 20, 2023: (a) Ground truth (GT) radar composite reflectivity; (b) U-Net; (c) Attention-U-Net; (d) MCDA-U-Net; (e) SegFormer; (f) SRViT; (g) CBAM-U-Net(The area marked with a circle in the figure represents the core area of heavy precipitation, which demands high accuracy from the model)