MULTI-SCALE TRUNCATED DIFFUSION WITH HORIZON-TARGETED REFORECASTING FOR PRECIPITATION NOWCASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Precipitation nowcasting aims to forecast the distribution of rainfall over the upcoming 0–2 hours based on historical spatio-temporal sequences. However, current mainstream methods still struggle to accurately capture the evolutionary characteristics of precipitation, mitigate the progressive degradation of forecast accuracy with increasing lead time, and adapt to complex mountainous terrains. To address these challenges, we propose a multi-scale truncated diffusion with horizon-targeted reforecasting for precipitation nowcasting. Specifically, a multi-scale truncated diffusion module is designed to jointly model macroscopic precipitation evolution and fine-grained local structures through multi-scale spatio-temporal guidance, while shortening the reverse diffusion trajectory. Subsequently, a horizon-targeted reforecasting module is introduced to selectively reforecast the terminal frames, which are more susceptible to accumulated errors at longer lead times, thereby improving late-horizon forecast accuracy. Furthermore, to enhance precipitation forecasting capabilities in complex mountainous terrains, we construct the Southern Anhui Mountain Radar Dataset. Extensive experiments conducted on our self-constructed dataset, alongside two public datasets (SEVIR and CIKM), demonstrate that our method achieves state-of-the-art performance across all evaluation metrics, enabling robust precipitation nowcasting under complex mountainous conditions.

1 INTRODUCTION

Precipitation nowcasting typically refers to the high-resolution prediction of rainfall over the next 0–2 hours based on historical observational sequences Ravuri et al. (2021b). It enables rapid emergency decision-making for meteorological warnings, urban management, and disaster mitigation Shi et al. (2017); Pulkkinen et al. (2019). The key issue is to model the nonlinear and multiscale evolution of precipitation systems.

Existing precipitation nowcasting works can be broadly divided into physics equation-driven numerical weather prediction (NWP) and data-driven intelligent forecasting. NWP advances atmospheric states by numerically solving discretized dynamical and thermodynamic equations Lorenz (1963); Bauer et al. (2015). They suffer from prohibitive computational costs and extreme sensitivity to initial field perturbations. Empowered by their robust spatio-temporal feature modeling capabilities, intelligent forecasting methods are emerging as the mainstream research direction in this field Bi et al. (2023).

Intelligent forecasting can be broadly categorized into deterministic and probabilistic models. Deterministic models produce a single future sequence and efficiently capture large scale precipitation evolution. However, these methods still struggle to effectively represent fine-grained atmospheric features Gao et al. (2023). In contrast, probabilistic models learn the conditional distribution of future precipitation and generate multiple plausible forecasts to represent predictive uncertainty. Recently, as the most representative probabilistic models, diffusion models have demonstrated the ability to model complex precipitation distributions through an iterative denoising process, generating diverse and realistic predictions Ho et al. (2020). Consequently, our proposed method is also based on diffusion model.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Several diffusion methods are demonstrating significant advantages. For instance, PreDiff Gao et al. (2023) introduces a conditional latent diffusion model with a knowledge-alignment mechanism that adjusts the denoising transitions according to domain-specific priors. PostCast Gong et al. (2025) employs an unconditional diffusion prior with zero-shot blur-kernel estimation to refine blurry forecasts without paired retraining data. However, these methods still face challenges in accurately capturing the complex evolutionary characteristics of precipitation, especially the coordinated modeling of large-scale motion trends and fine-grained local variations. Additionally, CasCast Gong et al. (2024) proposed a cascaded modeling strategy that decouples the precipitation process into patterns of various scales, thereby mitigating the issue of forecast smoothing at high resolutions. DiffCast Yu et al. (2024) decomposes precipitation evolution into a deterministic global motion component and a stochastic local residual, and models the latter using residual diffusion. Nevertheless, existing methods generally adopt a one-pass forecasting paradigm and lack explicit mechanisms to correct errors accumulated during the forecasting process, resulting in progressive degradation of prediction accuracy with increasing lead time. Moreover, most existing precipitation forecasting datasets are collected from relatively flat regions and provide limited representation of terrain-induced precipitation variations. This restricts the ability of current models to generalize effectively to complex mountainous environments, where precipitation evolution is strongly influenced by heterogeneous topography.

To address the above limitations, we propose a multi-scale truncated diffusion with horizon-targeted reforecasting framework for precipitation nowcasting. First, we design a multi-scale truncated diffusion module that introduces multi-scale spatio-temporal guidance into different stages of the denoising process. This module shortens the reverse diffusion trajectory while effectively capturing both large-scale precipitation evolution and fine-grained local structures. Second, we introduce a horizon-targeted reforecasting module that selectively regenerates terminal forecast frames based on historical observations and preliminary predictions, aiming to alleviate the progressive degradation caused by accumulated errors at longer lead times. Finally, we construct the Southern Anhui Mountain Radar Dataset to facilitate precipitation nowcasting research under complex mountainous conditions and evaluate our method on this dataset together with two public benchmarks, CIKM and SEVIR.

The main contributions of this work are summarized as follows:

- We propose a multi-scale truncated diffusion framework for precipitation nowcasting, which introduces multi-scale spatio-temporal guidance into the denoising process. The proposed method shortens the reverse diffusion trajectory while effectively preserving both large-scale precipitation evolution and fine-grained local structures.
- We introduce a horizon-targeted reforecasting module that selectively regenerates terminal forecast frames based on historical observations and preliminary predictions. This design alleviates the progressive degradation caused by error accumulation at longer forecast horizons.
- We construct the Southern Anhui Mountain Radar Dataset to support precipitation nowcasting research under complex mountainous conditions. Extensive experiments on this dataset and two public benchmarks, CIKM and SEVIR, demonstrate the effectiveness of our method across multiple evaluation metrics.

2 Related Work

2.1 Precipitation Nowcasting Model

Precipitation nowcasting methods can be broadly categorized into physics-based numerical weather prediction (NWP) and data-driven forecasting approaches. NWP methods forecast atmospheric evolution by solving discretized physical equations Bauer et al. (2015); Lorenz (1963), but they usually suffer from high computational costs and sensitivity to initial conditions. In contrast, data-driven methods have recently become dominant due to their strong capability in learning complex spatio-temporal precipitation patterns Ravuri et al. (2021b). These methods can be further divided into deterministic, probabilistic, and hybrid approaches.

Deterministic models aim to capture the overall motion and intensity evolution of precipitation systems. ConvLSTM SHI et al. (2015) and PhyDNet Guen & Thome (2020) explore spatio-temporal feature learning and physics-guided representation disentanglement, respectively. More recently, AlphaPre Lin et al. (2025) further improves deterministic forecasting by modeling radar echo evolution from different perspectives.

Probabilistic models focus on generating diverse future predictions by learning the uncertainty distribution of precipitation evolution. DGMR Ravuri et al. (2021a) introduces adversarial generation with spatio-temporal discrimination, while NowcastNet Zhang et al. (2023) incorporates physical priors into a generative framework. PreDiff Gao et al. (2023) further employs conditional diffusion with knowledge alignment to improve probabilistic radar forecasting.

Hybrid architectures combine deterministic prediction with probabilistic refinement to balance large-scale motion consistency and fine-grained details. DiffCast Yu et al. (2024) decomposes precipitation evolution into deterministic motion and stochastic residual components, while CasCast Gong et al. (2024) adopts cross-scale latent modeling to generate multi-scale precipitation patterns.

2.2 Diffusion-based Spatio-Temporal Forecasting

Diffusion models have recently emerged as powerful generative frameworks for modeling complex data distributions through iterative denoising processes Ho et al. (2020). Beyond image synthesis, diffusion-based approaches have been extended to spatio-temporal prediction tasks by modeling future uncertainty and high-frequency variations. For example, STDiff applies spatio-temporal diffusion for stochastic video prediction Ye & Bilodeau (2024), while LDT performs probabilistic time-series forecasting in a compressed latent space Feng et al. (2024). These advances demonstrate the capability of diffusion models in capturing complex temporal dynamics.

Recently, diffusion models have been increasingly explored for precipitation nowcasting. PreDiff introduces conditional latent diffusion with knowledge alignment for probabilistic radar forecasting Gao et al. (2023). DuoCast further separates low- and high-frequency precipitation components using dual diffusion branches to balance large-scale evolution and local variations Wen et al. (2026). However, existing diffusion-based forecasting methods still face challenges in simultaneously modeling multi-scale precipitation evolution and mitigating error accumulation over long forecast horizons.

3 METHODS

3.1 Problem Definition

Precipitation nowcasting is formulated as a spatio-temporal forecasting task based on radar observations SHI et al. (2015). Given a sequence of historical radar frames $\mathbf{X} \in \mathbb{R}^{T_{\text{in}} \times C \times H \times W}$, the objective is to model the conditional distribution $p_\theta(\mathbf{Y} \mid \mathbf{X})$ of the future precipitation sequence $\mathbf{Y} \in \mathbb{R}^{T_{\text{out}} \times C \times H \times W}$, where T_{in} and T_{out} denote the numbers of input and forecast frames, respectively, and C, H, and W represent the channel number and spatial resolution of each frame. Accordingly, the model generates $\hat{\mathbf{Y}} \sim p_\theta(\mathbf{Y} \mid \mathbf{X})$ to characterize the future spatio-temporal evolution of precipitation.

3.2 Overall Framework

As illustrated in Fig. 1, the proposed framework is a diffusion-based precipitation nowcasting approach consisting of a multi-scale truncated diffusion module and a horizon-targeted reforecasting module. Given historical radar observations $\mathbf{X}$, the framework first employs a backbone predictor to generate a coarse forecast $\mathbf{Y}^d$. The historical frames, coarse forecast, and optical-flow fields are then integrated to construct a composite spatio-temporal representation. A spatio-temporal Mamba encoder extracts multi-scale contextual features from this representation, which guide the fragment-wise truncated diffusion process to generate residual predictions. These residuals are combined with the corresponding coarse forecast fragments to obtain the preliminary prediction. To further improve

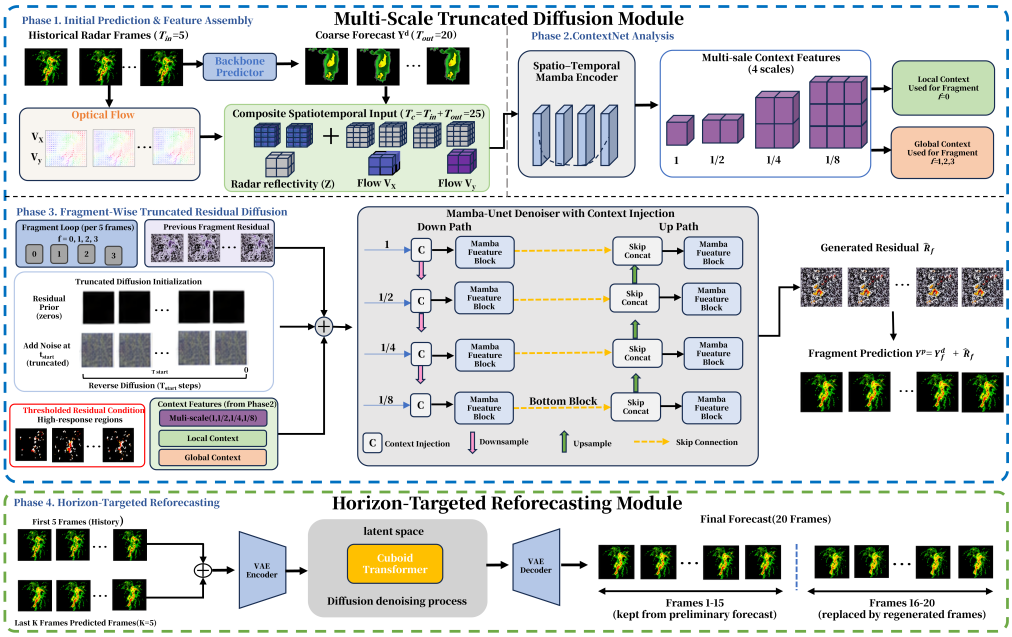


Figure 1: Overview of the proposed framework for precipitation nowcasting, consisting of multi-scale truncated diffusion and horizon-targeted reforecasting.

long-horizon forecasting accuracy, the horizon-targeted reforecasting module takes the historical observations and terminal frames of the preliminary prediction as input. It selectively regenerates these terminal frames in the latent space through a diffusion process with a Cuboid Transformer denoiser, and replaces the corresponding preliminary frames to produce the final forecast.

3.3 Multi-Scale Truncated Diffusion

3.3.1 Initial Prediction and Multi-Scale Context Extraction

Given the historical radar sequence $\mathbf{X}$, a backbone predictor first generates a coarse forecast $\mathbf{Y}^d$:

$$\mathbf{Y}^d = \mathcal{B}_\phi(\mathbf{X}). \tag{1}$$

The coarse forecast provides the overall motion trend and large-scale intensity evolution of precipitation. To incorporate explicit motion information, dense optical-flow fields $\mathbf{V}_x$ and $\mathbf{V}_y$ are computed between consecutive radar frames. The historical observations and optical-flow fields are concatenated along the channel dimension, while two zero-valued channels are appended to each coarse forecast frame to match their dimensions. These components are then assembled along the temporal dimension to form a composite spatio-temporal input:

$$\mathbf{S} = \mathrm{Concat}_t\left([\mathbf{X}, \mathbf{V}_x, \mathbf{V}_y], [\mathbf{Y}^d, \mathbf{0}, \mathbf{0}]\right). \tag{2}$$

The composite input is processed by a spatio-temporal Mamba encoder to extract multi-scale context features. Features are generated at four spatial resolutions, including 1, $1/2$, $1/4$, and $1/8$ of the original resolution. The feature corresponding to the first coarse forecast step is retained as local context to guide the initial fragment prediction, while features from the complete sequence are aggregated to provide global context for subsequent fragments. This design enables the diffusion model to exploit both short-term local evolution and long-range temporal dependencies.

3.3.2 FRAGMENT-WISE TRUNCATED RESIDUAL DIFFUSION

The future sequence is divided into F non-overlapping temporal fragments, each containing L frames. In our implementation, $L = T_{\text{in}} = 5$ and $F = 4$. Instead of directly generating radar frames, the proposed method models the residual between the ground-truth fragment and the coarse forecast:

$$\mathbf{R}_f = \mathbf{Y}_f - \mathbf{Y}_f^d. \tag{3}$$

For each fragment, the previous residual prediction is used as a temporal condition. During training, this condition is obtained from the previous ground-truth residual, while the generated residual is used during inference. In addition, a thresholded residual condition is constructed by retaining only high-response regions, which emphasizes areas where the coarse forecast fails to capture strong precipitation signals.

The noisy residual, residual condition, and thresholded residual condition are concatenated and fed into a Mamba-UNet denoiser. Multi-scale context features extracted from the spatio-temporal encoder are injected into corresponding resolution levels. Diffusion-step and fragment-index embeddings are introduced to distinguish different noise levels and forecast stages. Mamba feature blocks are employed in both the downsampling and upsampling paths to capture spatial dependencies within each fragment.

Since the residual distribution is concentrated around zero, the reverse diffusion process can start from an intermediate noise level rather than the maximum diffusion step. Specifically, a zero residual prior is perturbed at t_{start}:

$$\mathbf{R}_{f,t_{\text{start}}}^{(0)} = \sqrt{1 - \bar{\alpha}_{t_{\text{start}}}}\boldsymbol{\epsilon}. \tag{4}$$

The denoiser then performs reverse diffusion from t_{start} to 0, reducing the sampling trajectory while maintaining the ability to recover fine-grained precipitation details. The generated residual is added to the corresponding coarse forecast fragment:

$$\mathbf{Y}_f^p = \mathbf{Y}_f^d + \hat{\mathbf{R}}_f. \tag{5}$$

All fragment predictions are concatenated along the temporal dimension to obtain the preliminary forecast $\mathbf{Y}^p$.

3.4 HORIZON-TARGETED REFORECASTING

Forecast errors generally increase with longer prediction horizons, causing the terminal frames of preliminary forecasts to suffer from accumulated inaccuracies. To address this issue, we introduce a horizon-targeted reforecasting module that selectively regenerates the final K frames instead of forecasting the entire sequence again. In our implementation, $K = T_{\text{in}} = 5$.

The historical sequence and the terminal preliminary forecast frames are combined to construct the reforecasting condition:

$$\mathbf{C}^r = \mathbf{X} + \mathbf{Y}^p_{T_{\text{out}}-K+1:T_{\text{out}}}. \tag{6}$$

A pretrained VAE encoder maps the condition and target frames into a compact latent space. A latent diffusion model with a Cuboid Transformer denoiser is then employed to regenerate the latent representations of the terminal forecast frames.

During training, the ground-truth terminal frames are used to construct latent training pairs. During inference, they are replaced with the corresponding preliminary predictions, ensuring that the reforecasting process only relies on historical observations and model-generated forecasts.

The decoded frames replace the terminal part of the preliminary prediction, while the preceding frames remain unchanged:

$$\hat{\mathbf{Y}} = \left[\mathbf{Y}^{p}_{1:T_{\text{out}}-K}, \hat{\mathbf{Y}}^{r}_{T_{\text{out}}-K+1:T_{\text{out}}} \right]. \tag{7}$$

By focusing computational resources on the most error-prone terminal frames, the horizon-targeted reforecasting module improves long-horizon prediction accuracy without modifying earlier forecasts.

3.5 Training Objective

The proposed framework is optimized using three objectives. The backbone loss supervises the coarse deterministic forecast. The residual diffusion loss trains the multi-scale truncated diffusion module to recover precipitation residuals. The horizon-targeted reforecasting loss optimizes the latent diffusion process for refining the terminal forecast frames. The overall training objective is formulated as:

$$\mathcal{L} = \lambda_b \mathcal{L}_b + \lambda_r \mathcal{L}_r + \lambda_h \mathcal{L}_h. \tag{8}$$

The coefficients λ_b, λ_r, and λ_h balance the contributions of the backbone prediction loss, residual diffusion loss, and horizon-targeted reforecasting loss.

4 Experiments

4.1 Datasets

SEVIR. SEVIR is a large-scale multimodal meteorological dataset collected over the contiguous United States Veillette et al. (2020). It contains more than 10,000 severe-weather events and provides spatially and temporally aligned satellite imagery, NEXRAD-based Vertically Integrated Liquid (VIL) mosaics, and lightning observations. Each event covers a $384\,\mathrm{km} \times 384\,\mathrm{km}$ region over four hours at five-minute intervals. Following standard precipitation-nowcasting protocols, we use the VIL modality and resize the radar fields to 128×128.

CIKM. The CIKM AnalytiCup 2017 dataset contains 14,000 radar sequences collected over a $101\,\mathrm{km} \times 101\,\mathrm{km}$ region. Each sequence comprises 15 consecutive observations at six-minute intervals and includes radar reflectivity at four altitudes (0.5, 1.5, 2.5, and 3.5 km). We use the official training and test partitions and form the validation set from the training data.

Southern Anhui Mountain Radar Dataset. To evaluate nowcasting in complex terrain, we construct a radar dataset from observations provided by the Anhui Meteorological Observatory. It contains heavy-rainfall events occurring during the flood seasons from 2022 to 2025. The study domain spans $114.0°$–$120.35°\mathrm{E}$ and $28.65°$–$35.05°\mathrm{N}$ and includes terrain elevations ranging from approximately 50 m to more than 1,800 m. After quality control, temporal alignment, and spatial resampling, the radar fields are represented on a 128×128 grid at 5-km spatial and 10-minute temporal resolution. Reflectivity values are normalized to $[-1, 1]$. The processed samples are divided into training, validation, and test sets in an 8:1:1 ratio. This dataset complements the two public benchmarks by emphasizing precipitation systems affected by heterogeneous mountainous terrain.

4.2 Implementation Details

All models are implemented in PyTorch and evaluated on SEVIR, CIKM, and the Southern Anhui Mountain Radar Dataset. Unless otherwise stated, SimVP is used as the deterministic backbone. Our model is trained for 40 epochs on an NVIDIA A800 GPU using AdamW with a batch size of 8, an initial learning rate of 1.0×10^{-4}, and $(\beta_1, \beta_2) = (0.9, 0.95)$. The residual diffusion process uses a sigmoid noise schedule with 1,000 training steps and the v-prediction objective. At inference time, the multi-scale truncated diffusion module uses five-step DDIM sampling from the selected truncation level. The horizon-targeted reforecasting module employs a pretrained, frozen VAE and 50-step DDIM sampling to regenerate the final $K = 5$ frames. The truncation level and the loss weights λ_b, λ_r, and λ_h are selected on the validation set and then fixed for test evaluation.

Method	SEVIR					CIKM					Southern Anhui Mountain Radar Dataset				
	CSI-M↑	CSI-181↑	CSI-219↑	HSS↑	SSIM↑	CSI-M↑	CSI-35↑	CSI-40↑	HSS↑	SSIM↑	CSI-M↑	CSI-30↑	CSI-40↑	HSS↑	SSIM↑
MAU (2021)	0.3076	0.1071	0.0516	0.3863	0.6505	0.3039	0.2054	0.1241	0.3928	0.6325	0.2317	0.2834	0.0947	0.3412	0.5310
SimVP (2022)	0.3108	0.1106	0.0517	0.3924	0.6508	0.3052	0.2044	0.1321	0.3955	0.6538	0.2403	0.2913	0.1053	0.3541	0.5524
FourCastNet (2022)	0.2686	0.0717	0.0339	0.3355	0.5976	0.2980	0.1849	0.1015	0.3801	0.4359	0.2211	0.2652	0.0803	0.3185	0.4421
Earthformer (2022)	0.2892	0.0844	0.0245	0.3665	0.6633	0.3077	0.2039	0.1369	0.4001	0.6267	0.2444	0.3000	0.1084	0.3512	0.5413
PhyDNet (2020)	0.3017	0.1040	0.0278	0.3812	0.6532	0.3038	0.2052	0.1287	0.3931	0.6541	0.2353	0.2780	0.0909	0.3458	0.5467
EarthFarseer (2024)	0.3004	0.0992	0.0413	0.3829	0.6327	0.3000	0.2046	0.1259	0.3911	0.6373	0.2387	0.3013	0.0994	0.3475	0.5352
NowcastNet (2023)	0.2791	0.0770	0.0351	0.3512	0.6839	0.2991	0.1940	0.1188	0.3865	0.6713	0.2303	0.2717	0.0864	0.3342	0.5715
PreDiff (2024)	0.2744	0.0627	0.0235	0.3592	0.5881	0.2842	0.1848	0.1002	0.3481	0.5582	0.2258	0.2694	0.0815	0.3264	0.4908
DiffCast (2024)	0.3050	0.1300	0.0582	0.3996	0.6482	0.3159	0.2009	0.1457	0.4085	0.6499	0.2517	0.3144	0.1167	0.3662	0.5471
CasCast (2024)	0.2878	0.0954	0.0312	0.3563	0.5680	0.3021	0.1974	0.1142	0.3806	0.5241	0.2366	0.2822	0.0883	0.3391	0.4825
FACL (2024)	0.3161	0.1349	0.0655	0.4033	0.5686	0.3105	0.1993	0.1354	0.3996	0.6442	0.2485	0.3090	0.1113	0.3598	0.5389
AlphaPre (2025)	0.3259	0.1332	0.0545	0.4110	0.6884	0.3194	0.2068	0.1416	0.4137	0.6568	0.2567	0.3279	0.1318	0.3785	0.5832
DuoCast (2026)	0.3375	0.1818	0.1074	0.4318	0.6827	0.3179	0.2275	0.1562	0.4217	0.6659	0.2611	0.3353	0.1447	0.3876	0.5694
Ours	**0.3382**	**0.1900**	**0.1130**	**0.4327**	**0.6938**	**0.3511**	**0.2387**	**0.1872**	**0.4442**	**0.6804**	**0.2738**	**0.3494**	**0.1643**	**0.3914**	**0.5896**

Table 1: Quantitative comparison on SEVIR, CIKM, and the Southern Anhui Mountain Radar Dataset. The best results are highlighted in bold, and the second-best results are underlined.

4.3 Evaluation Metrics

Following previous studies Yu et al. (2024); Lin et al. (2025); Gong et al. (2024), we evaluate nowcasting performance using the Critical Success Index (CSI), Heidke Skill Score (HSS), and Structural Similarity Index (SSIM). For each dataset, the predictions and ground truth are converted into binary event maps using dataset-specific thresholds. CSI measures the overlap between predicted and observed precipitation regions, while HSS evaluates forecast skill relative to random chance. We report CSI-M by averaging CSI over all forecast frames and adopted thresholds, and additionally report CSI at the two highest thresholds to evaluate high-intensity precipitation forecasting. HSS is also averaged over all forecast frames and thresholds. SSIM measures the structural similarity between the predicted and observed radar fields. Higher values indicate better performance for all metrics.

4.4 Comparison with State-of-the-Art

Table 1 compares our method with representative deterministic, probabilistic, and hybrid nowcasting approaches. Our method obtains the best value for every reported metric on all three datasets, indicating that the improvements are consistent across categorical skill scores and structural similarity.

On SEVIR, our method achieves a CSI-M of 0.3382, an HSS of 0.4327, and an SSIM of 0.6938. Although the gains in the aggregate metrics are modest relative to the strongest baseline, the improvements are clearer at severe-precipitation thresholds: CSI-181 increases from 0.1818 to 0.1900 and CSI-219 from 0.1074 to 0.1130, corresponding to relative gains of 4.5% and 5.2%, respectively. This result shows that the proposed framework improves the detection of high-intensity echoes while preserving the overall spatial structure.

The gains are larger on CIKM. Our method improves CSI-M from 0.3194 to 0.3511 and CSI-40 from 0.1562 to 0.1872 over the respective second-best results, yielding relative improvements of 9.9% and 19.8%. It also reaches 0.2387 in CSI-35, 0.4442 in HSS, and 0.6804 in SSIM. The simultaneous improvement in threshold-based scores and SSIM indicates better localization and intensity estimation without sacrificing structural fidelity.

On the Southern Anhui Mountain Radar Dataset, our method improves CSI-M from 0.2611 to 0.2738, CSI-30 from 0.3353 to 0.3494, and CSI-40 from 0.1447 to 0.1643. The relative gain at the most demanding threshold is 13.5%. It also achieves the highest HSS (0.3914) and SSIM (0.5896). These results are particularly relevant because mountainous precipitation is characterized by irregular displacement, rapid intensity changes, and fragmented echo structures. The consistent ranking across three datasets with different regions and sampling intervals supports the robustness of the proposed design.

Variant	SEVIR					CIKM					Southern Anhui Mountain Radar Dataset				
	CSI-M↑	CSI-181↑	CSI-219↑	HSS↑	SSIM↑	CSI-M↑	CSI-35↑	CSI-40↑	HSS↑	SSIM↑	CSI-M↑	CSI-30↑	CSI-40↑	HSS↑	SSIM↑
w/o Optical Flow	0.3294	0.1812	0.1045	0.4256	0.6861	0.3420	0.2275	0.1738	0.4367	0.6726	0.2665	0.3340	0.1519	0.3847	0.5826
w/o Local–Global Context	0.3175	0.1678	0.0948	0.4105	0.6685	0.3304	0.2112	0.1571	0.4216	0.6556	0.2572	0.3081	0.1372	0.3715	0.5680
w/o Horizon-Targeted Reforecasting	0.3338	0.1766	0.0978	0.4320	0.6869	0.3465	0.2226	0.1629	0.4433	0.6738	0.2702	0.3264	0.1421	0.3905	0.5838
Full Model	**0.3382**	**0.1900**	**0.1130**	**0.4327**	**0.6938**	**0.3511**	**0.2387**	**0.1872**	**0.4442**	**0.6804**	**0.2738**	**0.3494**	**0.1643**	**0.3914**	**0.5896**

Table 2: Ablation results on SEVIR, CIKM, and the Southern Anhui Mountain Radar Dataset. The best results are highlighted in bold.

4.5 Ablation Studies

To isolate the contribution of each design choice, we evaluate three variants under the same settings: removing optical-flow inputs, removing the local–global context used by multi-scale truncated diffusion, and removing horizon-targeted reforecasting so that the preliminary sequence is returned directly.

As shown in Table 2, the full model performs best in every setting. Removing optical flow decreases CSI-M from 0.3382 to 0.3294 on SEVIR, from 0.3511 to 0.3420 on CIKM, and from 0.2738 to 0.2665 on the Southern Anhui dataset. The corresponding high-threshold CSI values also decline, showing that explicit motion cues improve echo displacement modeling and temporal continuity.

Removing local–global context causes the largest overall degradation. On SEVIR, CSI-M falls to 0.3175, while CSI-181 and CSI-219 decrease to 0.1678 and 0.0948. Similar drops occur on CIKM and the Southern Anhui dataset, and SSIM decreases on all three benchmarks. This result confirms that the two context types are complementary: local context anchors the transition into the first forecast fragment, whereas global context supplies sequence-level guidance to later fragments.

Removing horizon-targeted reforecasting has a smaller effect on CSI-M and HSS but produces a clearer reduction in high-threshold CSI and SSIM. Specifically, CSI-219 decreases from 0.1130 to 0.0978 on SEVIR, CSI-40 from 0.1872 to 0.1629 on CIKM, and CSI-40 from 0.1643 to 0.1421 on the Southern Anhui dataset. This pattern matches the module's intended role: it focuses on the terminal frames, where accumulated errors most strongly affect intense echoes and fine-scale structure.

Overall, optical flow improves motion representation, local–global context provides the largest gain in general forecast quality, and horizon-targeted reforecasting primarily strengthens high-intensity and late-horizon predictions. Their consistent effects across all three datasets demonstrate that the components address complementary sources of forecast error.

4.6 Qualitative Analysis

Figure 2 presents qualitative comparisons on two representative cases from SEVIR and the Southern Anhui Mountain Radar Dataset. As shown in Fig. 2(a), the precipitation system in SEVIR undergoes continuous displacement and deformation while maintaining several localized high-intensity echo regions. SimVP captures the overall motion trend but produces increasingly smooth forecasts, causing sharp boundaries and high-intensity structures to gradually disappear. DiffCast preserves more local variations, but some predicted echoes become fragmented and deviate from the ground-truth spatial distribution at longer lead times. CasCast and AlphaPre retain the major precipitation regions, although the intensity and morphology of the central echoes are progressively weakened or distorted. DuoCast generates relatively sharper structures, but fine-scale echo patterns still deteriorate in the later forecast frames.

In contrast, our method better preserves both the large-scale displacement and the local precipitation morphology. The predicted echo regions remain closer to the ground truth in terms of spatial extent, boundary shape, and high-intensity cores, while small-scale structures are retained more consistently over time. This indicates that the proposed multi-scale truncated diffusion effectively complements the coarse forecast with fine-grained residual information, reducing the excessive smoothing commonly observed in deterministic predictions.

Figure 2(b) further illustrates the forecasting performance under the more complex precipitation patterns of the Southern Anhui Mountain Radar Dataset. The observed system contains an elongated

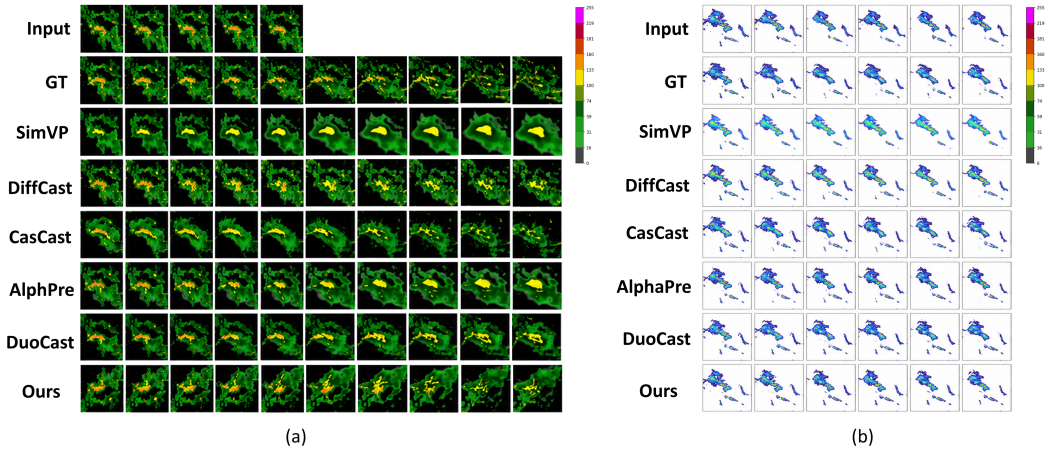


Figure 2: Qualitative comparison of precipitation nowcasting methods on (a) SEVIR and (b) the Southern Anhui Mountain Radar Dataset. In each panel, the first row shows the historical input sequence, the second row shows the ground-truth future sequence, and the remaining rows show the predictions generated by SimVP, DiffCast, CasCast, AlphaPre, DuoCast, and our method, respectively.

main precipitation structure together with several scattered echo regions, whose positions and intensities vary noticeably over time. SimVP tends to smooth these localized structures, while DiffCast and CasCast preserve more details but introduce varying degrees of structural distortion or spatial inconsistency. AlphaPre and DuoCast maintain the dominant precipitation regions more effectively, yet some small-scale echoes and local intensity variations are still weakened in later frames.

Our method produces forecasts that more closely follow the spatial evolution of the ground truth, preserving the continuity of the main precipitation structure as well as surrounding scattered echoes. In particular, the terminal forecasts exhibit less structural degradation and better retention of localized precipitation features. These results are consistent with the intended roles of the proposed components: optical-flow cues provide explicit motion guidance, multi-scale spatio-temporal context improves residual reconstruction at different spatial resolutions, and horizon-targeted reforecasting further alleviates the accumulated errors in the later forecast frames. Overall, the qualitative observations on both datasets agree with the improvements in high-threshold CSI and SSIM reported in the quantitative experiments.

5 Conclusion

This work addresses two key challenges in precipitation nowcasting: capturing multi-scale precipitation evolution and maintaining forecast accuracy at longer lead times. We propose a unified framework that combines multi-scale truncated residual diffusion with horizon-targeted reforecasting. The former refines coarse predictions through optical-flow-guided, fragment-wise residual modeling with multi-scale spatio-temporal context, while the latter selectively revises terminal frames to mitigate accumulated errors. We further introduce the Southern Anhui Mountain Radar Dataset for evaluating precipitation nowcasting under complex mountainous conditions. Extensive experiments on SEVIR, CIKM, and the proposed dataset demonstrate consistent improvements over strong deterministic, probabilistic, and hybrid baselines across CSI, HSS, and SSIM, with particularly clear gains for high-intensity precipitation. Ablation studies confirm the complementary contributions of motion cues, multi-scale contextual guidance, and targeted terminal-frame refinement, demonstrating the effectiveness of the proposed framework across diverse precipitation regimes.

AI USE STATEMENT

Generative AI tools were used for language editing, structural revision, and LaTeX formatting assistance. All technical content, experimental design, results, and conclusions were checked by the authors.

REFERENCES

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525:47–55, 2015. doi: 10.1038/nature14956.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619:533–538, 2023. doi: 10.1038/s41586-023-06185-3.

Shibo Feng, Chunyan Miao, Zhong Zhang, and Peilin Zhao. Latent diffusion transformer for probabilistic time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 11979–11987, 2024. doi: 10.1609/aaai.v38i11.29085.

Zhihan Gao, Xingjian Shi, Boran Han, Hao Wang, Xiaoyong Jin, Danielle Maddix, Yi Zhu, Mu Li, and Yuyang Bernie Wang. Prediff: Precipitation nowcasting with latent diffusion models. *Advances in Neural Information Processing Systems*, 36:78621–78656, 2023.

Junchao Gong, Lei Bai, Peng Ye, Wanghan Xu, Na Liu, Jianhua Dai, Xiaokang Yang, and Wanli Ouyang. Cascast: Skillful high-resolution precipitation nowcasting via cascaded modelling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 15809–15822. PMLR, 2024.

Junchao Gong, Siwei Tu, Weidong Yang, Ben Fei, Kun Chen, Wenlong Zhang, Xiaokang Yang, Wanli Ouyang, and LEI BAI. Postcast: Generalizable postprocessing for precipitation nowcasting via unsupervised blurriness modeling. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 87005–87026, 2025.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11471–11481, 2020.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Kenghong Lin, Baoquan Zhang, Demin Yu, Wenzhi Feng, Shidong Chen, Feifan Gao, Xutao Li, and Yunming Ye. Alphapre: Amplitude-phase disentanglement model for precipitation nowcasting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17841–17850, 2025.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2): 130–141, 1963.

Seppo Pulkkinen, Daniele Nerini, Andrés A. Pérez Hortal, Carlos Velasco-Forero, Alan Seed, Urs Germann, and Loris Foresti. Pysteps: An open-source python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, 12:4185–4219, 2019. doi: 10.5194/gmd-12-4185-2019.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021a.

Suman Ravuri, Karel Lenc, Matthew Willson, Remi Lam, Piotr Mirowski, Sheleem Kashem, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Ellen Clancy, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597: 672–677, 2021b. doi: 10.1038/s41586-021-03854-z.

Xingjian SHI, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun WOO. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-Chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5617–5627, 2017.

Mark S. Veillette, Siddharth Samsi, and Christopher J. Mattioli. SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. In *Advances in Neural Information Processing Systems*, volume 33, pp. 22009–22019, 2020.

Penghui Wen, Mengwei He, Patrick Filippi, Na Zhao, Feng Zhang, Thomas Francis Bishop, Zhiyong Wang, and Kun Hu. DuoCast: Duo-probabilistic diffusion for precipitation nowcasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 39442–39450, 2026. doi: 10.1609/aaai.v40i46.41294.

Xi Ye and Guillaume-Alexandre Bilodeau. STDiff: Spatio-temporal diffusion for continuous stochastic video prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 6666–6674, 2024. doi: 10.1609/aaai.v38i7.28489.

Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, and Xunlai Chen. Diffcast: A unified framework via residual diffusion for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27758–27767, 2024.

Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I Jordan, and Jianmin Wang. Skilful nowcasting of extreme precipitation with nowcastnet. *Nature*, 619 (7970):526–532, 2023.

Supplementary Material

Multi-Scale Truncated Diffusion with Horizon-Targeted Reforecasting for Precipitation Nowcasting

A Implementation Details

A.1 Network Configuration

Table A.1 summarizes the default configuration of the proposed framework. A deterministic backbone first produces a coarse forecast. Multi-scale truncated diffusion then models the forecast residuals in temporal fragments under optical-flow and multi-scale spatio-temporal guidance. Finally, horizon-targeted reforecasting regenerates the last five frames in latent space while leaving the preceding frames unchanged.

Item	Configuration
Backbone predictor	SimVP by default
Input–output protocol	$5 \rightarrow 20$ frames by default; $5 \rightarrow 10$ on CIKM
Fragment length	$L = 5$
Number of fragments	$F = 4$ by default; $F = 2$ on CIKM
Context scales	1, $1/2$, $1/4$, and $1/8$
Residual denoiser	Mamba-UNet
Diffusion objective	v-prediction
Noise schedule	Sigmoid schedule
Training diffusion steps	1,000
Truncation timestep	$t_{\text{start}} = 200$
Residual sampling	Five-step DDIM
Reforecasted frames	Final $K = 5$ frames
Reforecasting denoiser	Cuboid Transformer
Reforecasting sampling	50-step DDIM
Reforecasting latent shape	$5 \times 16 \times 16 \times 64$ for a 128×128 input
Optimizer	AdamW, $(\beta_1, \beta_2) = (0.9, 0.95)$
Learning rate	1.0×10^{-4}
Batch size	8
Training epochs	40

Table A.1: Default model and training configuration. Dataset-specific sequence lengths are used for CIKM.

B Model Complexity Analysis

Table B.1 compares the parameter count and computational cost of the proposed framework with representative nowcasting methods. Our model contains 94.40M trainable parameters and requires 0.32 TFLOPs under the adopted inference setting. Its complexity is close to that of DuoCast, while the main paper reports better forecasting performance across SEVIR, CIKM, and the Southern Anhui Mountain Radar Dataset.

Among diffusion-based approaches, the proposed method requires substantially fewer operations than PreDiff and DiffCast. This reduction results from two complementary designs. First, residual generation starts from an intermediate diffusion timestep and follows a shortened reverse trajectory. Second, horizon-targeted reforecasting operates in latent space and regenerates only the terminal frames rather than the entire sequence. Consequently, multi-scale refinement and long-horizon correction are introduced without requiring a full diffusion trajectory over all forecast frames.

Method	# Param.	TFLOPs	Year
MAU	20.13M	0.09	2021
SimVP	44.25M	0.05	2022
FourCastNet	82.95M	0.16	2022
Earthformer	34.61M	0.04	2022
PhyDNet	11.80M	0.08	2020
EarthFarseer	33.60M	0.12	2024
NowcastNet	97.80M	1.20	2023
PreDiff	135.24M	2.80	2024
DiffCast	58.33M	72.49	2024
CasCast	86.50M	0.45	2024
FACL	14.38M	0.02	2024
AlphaPre	89.03M	1.56	2025
DuoCast	94.18M	0.30	2026
Ours	**94.40M**	**0.32**	**2026**

Table B.1: Model-complexity comparison with representative nowcasting methods. FLOPs depend on the adopted sampling schedule and implementation and should be interpreted together with the corresponding inference settings.

The reported parameter count includes the deterministic backbone, the spatio-temporal context encoder, the residual diffusion denoiser, and the latent reforecasting network. The pretrained VAE is frozen and is therefore excluded from the trainable parameter count, although its encoder and decoder still contribute to inference memory and computation. Overall, the proposed framework provides a practical balance between forecast quality and computational complexity.

C Additional Quantitative Analysis

C.1 Effect of the Truncation Timestep on SEVIR

t_{start}	CSI-M	CSI-181	CSI-219	HSS	SSIM
100	0.3356	0.1861	0.1089	0.4301	0.6913
200	**0.3382**	**0.1900**	**0.1130**	**0.4327**	**0.6938**
400	0.3371	0.1877	0.1108	0.4316	0.6922
800	0.3317	0.1781	0.1017	0.4255	0.6853

Table C.1: Sensitivity to the truncation timestep on SEVIR. The best result for each metric is shown in bold.

Table C.1 evaluates the effect of the starting timestep used for truncated reverse diffusion. Increasing t_{start} from 100 to 200 improves all metrics, including CSI-181 from 0.1861 to 0.1900 and CSI-219 from 0.1089 to 0.1130. The additional perturbation at $t_{\text{start}} = 200$ gives the denoiser sufficient flexibility to recover residual structures that are missing from the deterministic forecast.

Further increasing the starting timestep gradually reduces performance. At $t_{\text{start}} = 800$, CSI-219 falls to 0.1017 and SSIM to 0.6853, indicating that excessive corruption makes high-intensity and fine-scale residuals more difficult to reconstruct. These results support the use of an intermediate starting point rather than either a nearly deterministic refinement or a full reverse trajectory. We therefore use $t_{\text{start}} = 200$ as the default setting.

C.2 Lead-Time Analysis of Horizon-Targeted Reforecasting

Forecast frames	Preliminary forecast	Full model	Gain
1–5	0.2021	0.2021	+0.0000
6–10	0.1648	0.1648	+0.0000
11–15	0.1286	0.1286	+0.0000
16–20	0.0730	0.1455	+0.0725

Table C.2: Lead-time analysis of CSI-40 on the Southern Anhui Mountain Radar Dataset. Horizon-targeted reforecasting modifies only the final five frames.

Table C.2 examines CSI-40 over four forecast fragments. The preliminary forecast degrades steadily with lead time, falling from 0.2021 in frames 1–5 to 0.0730 in frames 16–20. This decline highlights the difficulty of maintaining intense precipitation echoes over long horizons in mountainous regions.

By design, horizon-targeted reforecasting replaces only frames 16–20; therefore, the first 15 frames remain identical to the preliminary forecast. For the terminal fragment, CSI-40 increases from 0.0730 to 0.1455, an absolute gain of 0.0725. The result directly supports the motivation for selective reforecasting: computational effort is concentrated on the portion of the sequence most affected by accumulated displacement and intensity errors without disturbing the earlier predictions.