AMU-Net: Physics-Aware Adaptive Mamba U-Net for Efficient Multi-Lead Time Precipitation Nowcasting

Baitian Liu[1], Haiping Zhang[1], Dongjing Wang[1], Luan Xu[2], Ying Li[2], Feng Chen[2], and Xujian Fang[1,*]
[1]*Department of Computer Science and Technology, Hangzhou Dianzi University*, Hangzhou, China
{sonderliu, zhanghp, dongjing.wang, fangxj}@hdu.edu.cn
[2]*Zhejiang Institute of Meteorological Sciences*, Hangzhou, China
xuluan_new@126.com, liyingsdzb@163.com, fchen_zj@163.com

***Abstract*—Precipitation nowcasting is critical for disaster mitigation but struggles to capture the rapid, nonlinear changes in convective systems. Conventional deep learning methods face two main issues: error propagation from autoregressive prediction, which reduces long-term accuracy, and the tendency to blur and underestimate intensity due to deterministic loss functions that fail to capture extremes. To address this, we propose the Adaptive Mamba U-Net (AMU-Net), a physics-aware framework for multi-lead-time direct forecasting of extreme precipitation. AMU-Net adopts a Unified Direct Forecasting Paradigm with a Physics-Aware Time Embedding and Adaptive Layer Normalization to inject temporal information into the network, enabling a single model to dynamically modulate its feature statistics and adapt to physical laws across lead times. It integrates the Mamba module—a linear-complexity State Space Model—into a U-Net to capture long-range convective advection with global receptive fields without the prohibitive quadratic costs of Transformers. To reduce blurring, a Hybrid Generative Training Strategy combining spectral consistency and adversarial learning recovers high-fidelity textures. Extensive experiments on the SEVIR dataset show that AMU-Net significantly outperforms baselines across both meteorological skill scores and perceptual quality. Ablation studies further validate the necessity of each proposed component, confirming that the unified direct forecasting strategy and hybrid training are essential for reconciling computational efficiency with accurate detection of extreme weather events.**

***Index Terms*—Precipitation Nowcasting, Deep Learning, Mamba, Physics-aware Neural Networks**

I. Introduction

Short-term heavy precipitation forecasting, or nowcasting (0–2 hours), is crucial for mitigating natural disasters like flash floods and urban waterlogging, which threaten safety and infrastructure. While Numerical Weather Prediction (NWP) systems are standard for global and regional forecasting, they depend on complex physical models that cause inherent "spin-up" latency [1]. Consequently, NWP models fail to meet the real-time demands of minute-scale updates. To reduce this delay, radar echo extrapolation methods based on optical flow (e.g., PySTEPS [2]) have emerged as a faster alternative [3]. However, these methods assume motion conservation, which fails to capture rapid, nonlinear convective evolution and often results in insufficient accuracy for extreme weather predictions and warnings [4].

Deep Learning Weather Prediction (DLWP) has recently outperformed traditional methods by learning complex spatiotemporal dynamics from large datasets [5]. Despite these advancements, a critical challenge remains: the gap between short- and long-term forecast skills. Models often struggle across lead times as precipitation changes. Many DLWP models use an autoregressive strategy to extend forecast lead times [6]. However, this recursive process causes error accumulation, with minor early deviations in early predictions amplifying rapidly, leading to significant performance degradation at longer lead times [7].

Deep learning models excel at statistical fitting but are data-driven and lack physical constraints. Standard regression often causes models to approximate the mean, causing the "blurring effect" and underestimating extremes [8]. In nowcasting, these issues lead to loss of detail and failed alarms, making forecasts inadequate for operational standards [9]. From a neural architecture perspective, purely data-driven models often lack physical biases, like conservation laws. Embedding these priors into network topology, not just as loss terms, is an open problem [9], [10]. Our work introduces a modulation module to adapt weights based on physical context.

In addressing these issues, two major challenges remain in deep learning-based nowcasting:

1) **Error Accumulation in Autoregression.** Most existing models rely on a recursive (rolling) strategy to extend forecast lead times [10]–[13]. This process inevitably introduces error accumulation, where minor structural distortions in early frames amplify nonlinearly, causing significant performance degradation at longer horizons [7]. While training separate models for specific lead times (e.g., Pangu [11]) can mitigate this issue, such multi-model strategies are computationally expensive and difficult to maintain within a unified framework.

Xujian Fang* is the corresponding author. This research is supported by the National Natural Science Foundation of China through grants 42475038 and 62202131, the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China through grant LZJMY24D050007, the China Meteorological Administration through grant FPZJ2025-053, and the China Scholarship Council program with project ID 202305330010.

2) **The “Blurring Effect” and Intensity Underestimation.** Standard regression objectives (e.g., MSE) drive models to approximate the statistical mean of future states, resulting in oversmoothed predictions and systematic underestimation of extreme values [8], [14]. Although Generative Adversarial Networks (GANs) [9], [10] have been employed to reconstruct textures, pure GAN approaches can be unstable and prone to generating spatially inaccurate hallucinations [15].

We introduce **AMU-Net** (Adaptive Mamba U-Net), a physics-aware framework that balances efficiency and high-fidelity modeling of extreme precipitation. To address error accumulation in recursive methods, we propose a **Multi-Lead Time Direct Forecasting** mechanism with a *Physics-Aware Time Embedding* module that encodes forecast horizons into continuous instructions. By injecting these instructions into the network via *Adaptive Layer Normalization (AdaLN)*, the network dynamically adjusts feature statistics to match different lead times, avoiding iterative propagation and ensuring long-term forecast accuracy. We utilize Mamba as the core mixer to capture global spatiotemporal dependencies efficiently, overcoming Transformer limitations. Additionally, our extbfHybrid Generative Training Strategy enforces constraints in spectral (via FFT) and perceptual (via Adversarial Learning) domains to recover high-frequency features and extreme cores without losing spatial precision.

The main contributions of this study are summarized as follows:

1) **Unified Direct Forecasting Paradigm.** We propose a single framework that conditions on embedded time instructions to synthesize future frames directly. This approach inherently eliminates recursive error accumulation and avoids the prohibitive cost of training separate models for different horizons.
2) **Physics-Aware Mamba Architecture.** We design a novel backbone integrating the linear-complexity Mamba mixer with AdaLN. This architecture captures global context while enabling dynamic feature adaptation to the distinct physical dynamics of requested lead times.
3) **Hybrid Generative Strategy.** We introduce a composite objective combining spectral consistency with adversarial training. This strategy effectively mitigates the blurring effect, ensuring faithful reconstruction of extreme precipitation events and sharp structural boundaries.

The rest of this paper is organized as follows: Section II reviews related work. Section III outlines the proposed methodology. Experimental settings and results are presented in Sections IV and V, respectively. Finally, Section VI concludes the work.

Our code is available at https://github.com/sonderlau/AMU-Net.

II. Related Works

A. Spatiotemporal Predictive Learning

Early deep-learning nowcasting relied on RNN-CNN hybrids such as ConvLSTM [16] and PredRNN [6] to capture spatiotemporal dynamics. Recently, simplified encoder-decoder architectures, such as SimVP [7], and radar-specific models, such as SSA-UNet [17] and multi-source fusion networks [18], have shown that pure CNN-based paradigms can achieve competitive performance while maintaining higher training efficiency.

A fundamental limitation is dependence on autoregressive forecasting, which accumulates errors, causing blurring over long horizons [9]. Multi-model methods like Pangu-Weather [11] help but are costly. Developing a unified, error-free forecasting framework remains a challenge.

B. Long-Range Modeling

Convective systems are governed by fluid dynamics, which require global context, a property that traditional CNNs with local receptive fields struggle to capture [19]. To model these long-range dependencies, Vision Transformers [20], meteorological variants such as EarthFormer [21], and hybrid transformer models [22] have been introduced. Despite their effectiveness, the quadratic complexity of self-attention imposes prohibitive memory and latency costs for high-resolution real-time tasks.

Recently, State Space Models (SSMs), particularly Mamba [23], have emerged as a powerful alternative. Mamba combines the global receptive field of Transformers with the linear complexity of RNNs. Its “Selective Scan” mechanism is remarkably analogous to atmospheric flow, enabling efficient propagation of relevant features (e.g., precipitation cores) while filtering noise.

C. Generative Nowcasting for Extreme Weather

Deterministic models optimized with pixel-wise regression losses (e.g., MSE) tend to produce spatially smoothed predictions that systematically underestimate extreme events [8], [14]. To reconstruct realistic textures, generative approaches such as DGMR [9] and NowcastNet [10] use adversarial training [24] to enforce perceptual realism.

Pure GANs can be unstable, producing plausible but spatially inaccurate hallucinations [15]. A hybrid approach is needed to balance pixel fidelity and structural sharpness [19]. This work combines spectral constraints with adversarial training to maintain frequency and physical intensity preservation.

III. Methodology

A. Problem Formulation

Given a historical radar sequence $X_{t-T+1:t} \in \mathbb{R}^{T\times H\times W}$, traditional deep learning approaches typically formulate long-term prediction as a recursive process. To obtain the prediction for a specific lead time τ, the model f_θ is applied iteratively:

$$\hat{X}_{t+\tau} = \underbrace{f_\theta(f_\theta(\cdots f_\theta(X_{t-T+1:t})))}_{\tau \text{ times}} \tag{1}$$

This recursive unrolling inherently introduces error accumulation, as predictions depend on estimated outputs from previous steps, leading to significant blurring and degradation at longer horizons [25].

To address this issue, we propose a multi-lead-time direct-forecasting paradigm. Instead of stepwise iteration, we learn a unified function F_θ that takes both the historical sequence and the target lead time τ as inputs:

$$\hat{X}_{t+\tau} = F_\theta(X_{t-T+1:t}\ , \tau) \tag{2}$$

Here, τ serves as an explicit conditioning variable embedded into the network. This mechanism allows the model to directly synthesize the target frame $\hat{X}_{t+\tau}$ without intermediate dependencies, thereby effectively eliminating the cumulative error inherent in recursive methods.

B. Overall Architecture

To achieve unified, multi-lead time direct forecasting, we propose the **Adaptive Mamba U-Net (AMU-Net)**. As illustrated in Figure 1, the framework is built on a hierarchical encoder-decoder backbone that captures the multiscale spatial characteristics of convective systems. The architecture operates through two synergistic streams:

1) *Spatiotemporal Feature Stream*: This stream is the model's backbone, employing a U-Net hierarchy where the encoder downsamples radar inputs to capture storm motion, and the decoder upsamples features to predict future precipitation. Skip connections merge multi-scale contexts, maintaining details critical for small convective cores.
2) *Global Condition Stream*: To address static weights' limitation with different forecast horizons, we add an active control mechanism (denoted as Time Embedding in Figure 1). Utilizing the *Physics-Aware Time Embedding* (Section III-C), this encodes temporal context into a high-dimensional broadcast, allowing the model to adapt its reasoning, applying stronger diffusion for long-term forecasts and a structural focus for short-term ones.

C. Physics-Aware Time Embedding

A unified model for direct forecasting across continuous lead times requires a mechanism to interpret the physical context of each prediction. We propose a Physics-Aware Time Embedding module that maps discrete temporal context into a constant, high-dimensional control vector, combining forecast lead time and periodic temporal context.

1) Lead Time Embedding: The forecast lead time τ serves as the explicit instruction for the prediction distance. Direct use of scalar values is suboptimal for deep neural networks due to limited representational capacity. Instead, we map these discrete time steps into a learnable manifold using an embedding layer to retrieve a semantic vector $C_\tau = \text{Embedding}(\tau)$. This allows the model to semantically cluster similar horizons (e.g., 10 min and 20 min) while distinguishing distinct physical regimes (e.g., 10 min vs. 120 min) in the latent space.

2) Periodic Temporal Context: Convective weather systems rely on strong thermodynamic cycles. Precipitation patterns vary with daily and seasonal changes. For example, thermal convection peaks in the afternoon with maximum solar heating, and rainfall patterns differ across seasons.

To incorporate these physical priors, we introduce auxiliary conditions: day progress (t_{day}) and year progress (t_{year}). Since time is a cyclic variable, simple scalar normalization would introduce numerical discontinuities at the boundary (e.g., 23:59 vs. 00:00). To preserve cyclical continuity, we employ sine-cosine encoding:

$$C_{env} = \text{MLP}\left([\sin(2\pi t_{day}), \cos(2\pi t_{day}), \sin(2\pi t_{year}), \cos(2\pi t_{year})]\right) \tag{3}$$

where $[\cdot]$ denotes concatenation. Injecting C_{env} informs the network of the environmental state (e.g., a summer afternoon or a winter night), enabling the synthesis of precipitation fields aligned with current thermodynamic conditions.

3) Global Condition Fusion via AdaLN: To synthesize a unified control signal, we integrate the forecast target (C_τ) with the environmental context (C_{env}) via a Multi-Layer Perceptron (MLP) to generate the global instruction $C_{global} = \text{MLP}([C_\tau, C_{env}])$.

To inject this instruction into the backbone, we employ Adaptive Layer Normalization (AdaLN). Unlike standard normalization with fixed parameters, AdaLN dynamically regresses the affine coefficients from C_{global}. Formally, for an input feature map $\mathbf{Z}$, the modulation parameters $\boldsymbol{\gamma}, \boldsymbol{\beta}$ are computed via linear projections:

$$\boldsymbol{\gamma}, \boldsymbol{\beta} = \mathbf{W} \cdot C_{global} + \mathbf{b} \tag{4}$$

The modulated feature is then obtained by:

$$\text{AdaLN}(\mathbf{Z}, C_{global}) = (1 + \boldsymbol{\gamma}) \odot \frac{\mathbf{Z} - \mu}{\sigma} + \boldsymbol{\beta} \tag{5}$$

where μ and σ denote channel-wise statistics. Treating $(1+\boldsymbol{\gamma})$ as a scaling gate, this mechanism directly injects the physical forecast instruction into the feature space before mixing.

D. The Mamba Mixer Block

To achieve large-scale receptive fields for convective advection tracking within real-time constraints, we use the Mamba [23] module as the main feature extractor (see Figure 1, bottom left).

1) *Mamba Choice*: Traditional CNNs are limited by local receptive fields, making them inefficient at modeling long-range dependencies. Conversely, Transformers offer global context but suffer from quadratic complexity with respect to sequence length, creating a bottleneck for high-resolution radar data. Mamba functions as a discretized State Space Model (SSM) that achieves linear complexity.
2) *Machanism*: The system dynamics are governed by the continuous ordinary differential equation $\dot{\mathbf{h}}(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}x(t)$. In Mamba, this is discretized into a recursive form $\mathbf{h}_t = \overline{\mathbf{A}}\mathbf{h}_{t-1} + \overline{\mathbf{B}}x_t$ via a learned timescale parameter $\boldsymbol{\Delta}$.

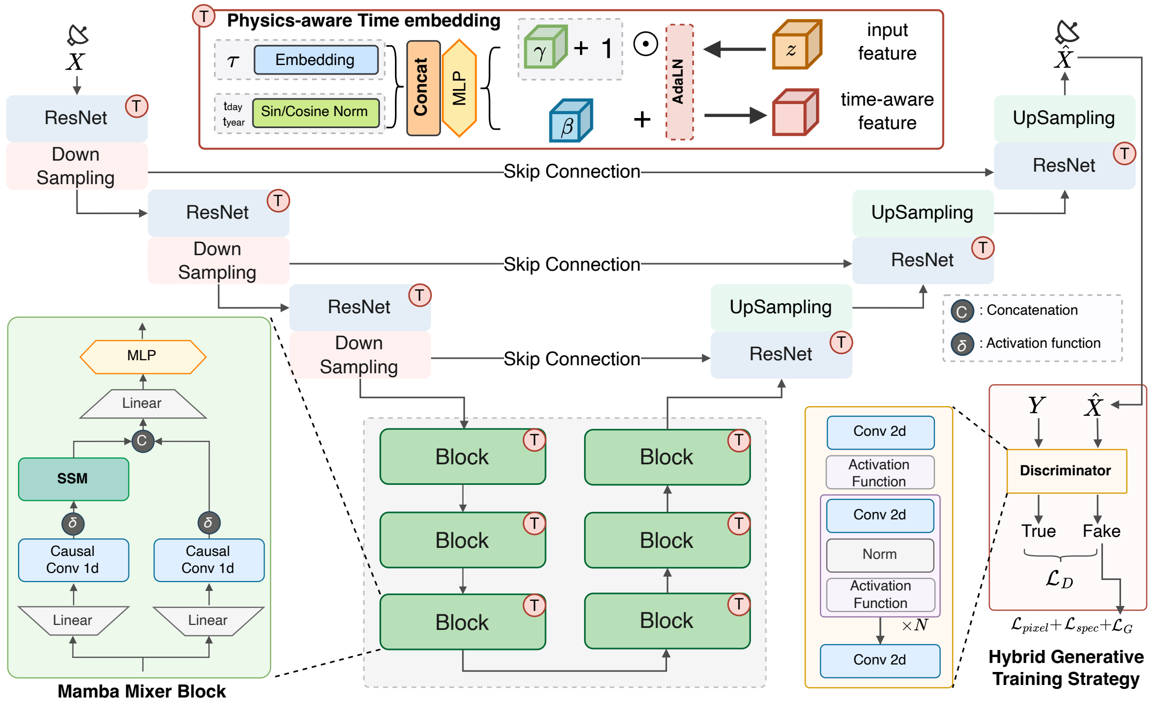

Fig. 1. The overall architecture of the proposed AMU-Net, designed as a physics-aware neural architecture. **(Top)** The framework operates via two synergistic streams: the *Spatiotemporal Feature Stream* (U-Net backbone) for feature extraction, and the *Global Condition Stream* (**Physics-Aware Time Embedding**, red box). The latter functions as a dynamic hypernetwork, encoding temporal inputs (τ, t_{day}, t_{year}) into affine parameters (γ, β) to modulate the backbone via AdaLN. **(Bottom Left)** The detailed internal structure of the **Mamba Mixer Block**, which utilizes linear-complexity State Space Models (SSM) and Causal Convolutions for efficient long-sequence modeling. **(Bottom Right)** The **Hybrid Generative Training Strategy**, illustrating the Discriminator architecture and the composite objective function ($\mathcal{L}_{total}$) employed to enforce spectral consistency and perceptual realism.

The key innovation is the selective scan mechanism, in which parameters ($\mathbf{\Delta}$, $\overline{\mathbf{B}}$, $\mathbf{C}$) are learned as input-dependent functions. This allows the model to propagate relevant signals, like intense precipitation cores, while reducing noise. We flatten radar features into 1D sequences and use bidirectional scanning to capture spatiotemporal dependencies.

E. Hybrid Generative Training Strategy

Regression nowcasting faces the "blurring effect" [8], where models focused on pixel-wise accuracy (e.g., MSE) tend to converge on the statistical mean, suppressing high-frequency details and underestimating extreme events. We tackle this with a hybrid objective that imposes constraints in both frequency and perceptual domains.

1) *Spectral Consistency Loss*: Precipitation fields exhibit rich multi-scale variability. To explicitly constrain the model's frequency-domain fidelity, we introduce a spectral consistency loss using the 2D Fast Fourier Transform (FFT):

$$\mathcal{L}_{spec} = \mathbb{E}\left[\left\| |\mathcal{F}(\hat{X})| - |\mathcal{F}(Y)| \right\|^2\right] \tag{6}$$

where $|\cdot|$ represents the magnitude spectrum. Minimizing $\mathcal{L}_{spec}$ helps the model match the energy of atmospheric turbulence across all wavenumbers, preventing high-frequency loss.

2) *Adversarial Learning Objective*: To further enhance textural fidelity, we adopt an adversarial training scheme with a PatchGAN discriminator (D), as shown in Figure 1 (bottom right). The discriminator learns to distinguish real radar sequences (Y) from generated ones ($\hat{X}$), and the generator optimizes a non-saturating adversarial loss to synthesize realistic patterns:

$$\mathcal{L}_D = -\frac{1}{2}\mathbb{E}_Y[\log D(Y)] - \frac{1}{2}\mathbb{E}_{\hat{X}}[\log(1 - D(\hat{X}))] \tag{7}$$

$$\mathcal{L}_G = -\mathbb{E}_{\hat{X}}[\log D(\hat{X})] \tag{8}$$

This adversarial game prompts the model to generate plausible details in regions of uncertainty.

The final objective balances deterministic accuracy, spectral integrity, and perceptual realism:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{pixel} + \lambda_2\mathcal{L}_{spec} + \lambda_3\mathcal{L}_G \tag{9}$$

where $\mathcal{L}_{pixel}$ is the MSE loss, and $\lambda_{1,2,3}$ are hyperparameters balancing the trade-off between metrics.

IV. Experimental Setup

A. Dataset

We use the SEVIR [26] Vertically Integrated Liquid (VIL) product, which provides high-resolution observations (384×384, 1 km). Raw 8-bit values are normalized to $[0, 1]$ for training. The dataset is chronologically partitioned into 36,271 training, 4,030 validation, and 4,477 test sequences.

B. Baselines

To comprehensively validate AMU-Net, we compare it against four representative baselines spanning different technical paradigms: (1) U-Net [27], a standard CNN backbone used to isolate the performance gains attributable to our proposed modules; (2) ConvLSTM [16], a classic autoregressive baseline chosen to demonstrate the advantages of our direct forecasting strategy; (3) SmaAt-UNet [28], an attention-equipped variant used to benchmark the efficiency trade-offs between our Mamba mixer and attention mechanisms; and (4) SimVP [7], a state-of-the-art video prediction model serving as a firm reference for temporal modeling quality.

C. Evaluation Metrics

We employ three complementary metrics to assess performance. *Root Mean Squared Error* quantifies pixel-level numerical deviation: $RMSE = \sqrt{\frac{1}{n}\sum(\hat{y}-y)^2}$. *Critical Success Index* (CSI) and *Heidke Skill Score* (HSS) evaluate classification skill across six thresholds $(16, 74, 133, 160, 181, 219)$, with a focus on high intensities $(181, 219)$ for extreme weather. CSI measures the hit rate, while HSS accounts for correct negatives and random chance. *Structural Similarity Index Measure* (SSIM) [29] assesses perceptual fidelity by modeling luminance, contrast, and structure, explicitly measuring the spatial integrity of convective systems.

We use two metrics for physical consistency. *Radially Averaged Power Spectral Density* (RAPSD) measures kinetic energy across scales via 2D FFT on the precipitation field, averaging into 1D radial bins. The *Peak Intensity Ratio* evaluates the preservation of extreme events, defined as the ratio of maximum predicted to maximum truth intensity $(\max(\hat{X})/\max(Y))$. A ratio close to 1 indicates the model avoids intensity collapse.

D. Implementation Details

Experiments used PyTorch on an NVIDIA H100 GPU. The input has 6 frames sampled every 10 minutes (60 minutes total), with target prediction lead times $\tau \in \{10, 20, 30, 60, 120\}$ minutes. For baselines, SimVP-D trains five models for each lead time, while our AMU-Net enables unified multi-lead forecasting by sampling τ during training. Optimization uses AdamW (batch size 8, initial lr 1×10^{-4}, cosine annealing) with loss weights $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.1$. Default configurations are used for all baselines.

V. Results and Analysis

A. Performance Comparison

As shown in Table I and Figure 2, two trends emerge: first, direct forecasting (Type D) outperforms autoregressive (Type A), confirming that bypassing iterative rollout reduces long-term error. Second, AMU-Net achieves state-of-the-art results across most skill metrics (CSI, HSS, SSIM).

We observe two notable trade-offs inherent to our design. (1) *Generalization vs. Specialization:* AMU-Net exhibits a marginal lag at the +10 min horizon compared with single-step baselines. This is a natural cost of our unified strategy, which generalizes across diverse lead times within a single model, whereas baselines are optimized exclusively for immediate prediction. However, AMU-Net rapidly establishes dominance as lead times extend. (2) *Perception vs. Distortion:* Compared to the direct forecasting baseline, SimVP-D, AMU-Net exhibits an expected drop in RMSE performance. This stems from the classic "double penalty" effect [30] in forecasting. While MSE-optimized models (e.g., SimVP) favor blurry, conservative outputs to strictly minimize pixel displacement errors, our approach prioritizes sharp, physically realistic structures. Consequently, AMU-Net trades slight increases in pixel-wise RMSE for significantly superior performance in high-intensity detection tasks (CSI-181 and CSI-219).

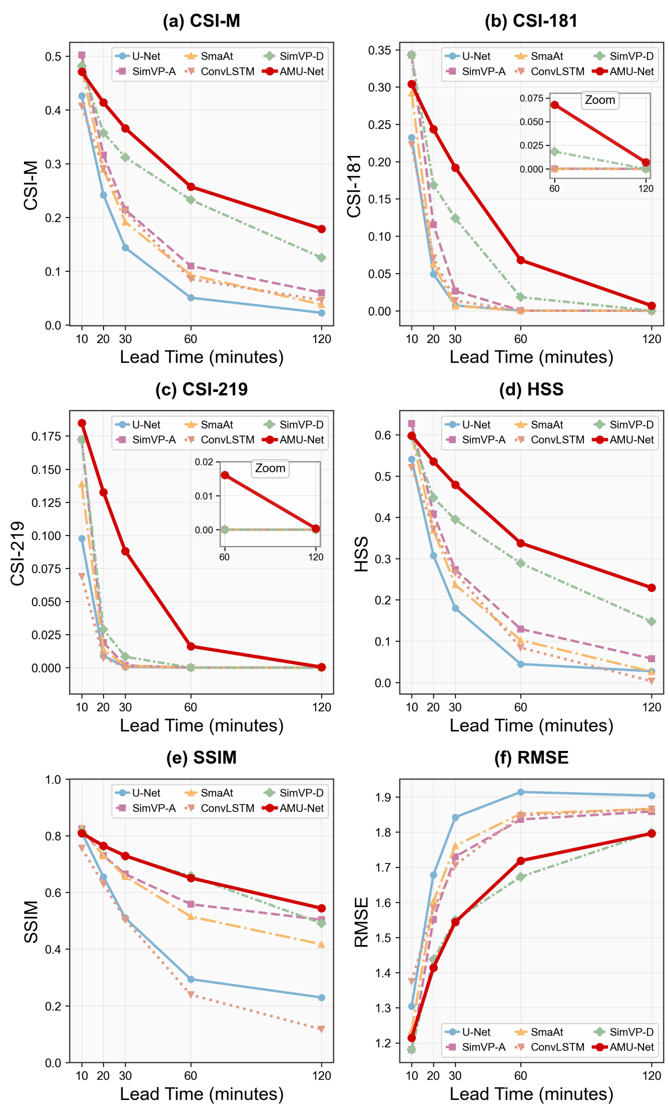

Fig. 2. Quantitative performance comparison of AMU-Net against baseline models across varying forecast lead times (10 to 120 minutes). The evaluation encompasses six key metrics: (a) CSI-M, (b) CSI-181, (c) CSI-219, (d) HSS, (e) SSIM, and (f) RMSE. For RMSE, lower values indicate better performance, whereas for all other metrics, higher values are desirable.

TABLE I

Comparison of forecasting performance averaged across all lead times. **Bold** and underlined values indicate the best and second-best results, respectively. A and D denote the autoregressive and direct prediction paradigms, respectively.

Method	Type	CSI-M ↑	CSI-181 ↑	CSI-219 ↑	HSS ↑	SSIM ↑	RMSE ↓
U-Net	A	0.177012	0.057763	0.021412	0.219943	0.499882	1.728615
ConvLSTM	A	0.208181	0.061383	0.015365	0.249260	0.449136	1.675426
SmaAt-UNet	A	0.218061	0.072126	0.030743	0.266653	0.629083	1.663072
SimVP	A	0.240389	0.097006	0.038562	0.299160	0.656183	1.631478
SimVP	D	0.301888	0.130788	0.041860	0.374920	0.690751	**1.527996**
AMU-Net	D	**0.337438**	**0.162821**	**0.084406**	**0.435862**	**0.699657**	1.537527

B. Qualitative Visualization

Figure 3 shows a high-intensity flash flood. At +10 to +20 min, AMU-Net uniquely captures extreme cores (pink), unlike baselines' blurring. After 60 min, autoregressive models like ConvLSTM experience structural collapse due to error buildup, whereas direct forecasting reduces this issue. AMU-Net outperforms the multi-model SimVP-D with a single architecture. Even at +120 min, it maintains sharp boundaries and peaks, demonstrating that our hybrid approach overcomes over-smoothing in standard regression.

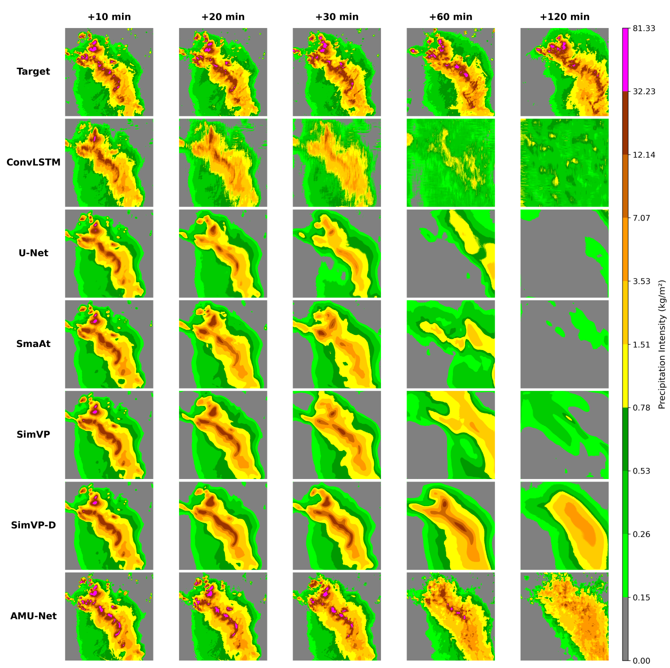

Fig. 3. Qualitative comparison of model predictions on a flash flood event.

C. Physical Consistency Comparision

The results visualized in Figure 4 (a-b) show AMU-Net's superior ability to preserve physical realism over baselines. At +10 min, all models match the ground truth, but by +120 min, baselines such as U-Net, SimVP, and SmaAt-UNet (lighter curves) exhibit blurring, whereas AMU-Net (red line) remains close to the ground truth, preserving high-frequency details even in long-term forecasts.

Figure 4 (c) reveals that autoregressive models like ConvLSTM and SimVP-D decay rapidly, dropping below 0.2 at 60 minutes. Still, AMU-Net retains over 50% of peak intensity at +120 min, confirming its effectiveness in reducing blurring and capturing extreme weather events.

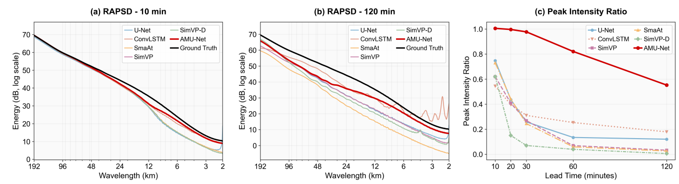

Fig. 4. Analysis of physical consistency and extreme value preservation on the SEVIR dataset. **(a)-(b)** Comparison of RAPSD at +10 min and +120 min lead times. The x-axis denotes wavelength (km) on a logarithmic scale, with the right side showing high-frequency details. **(c)** Evolution of Peak Intensity Ratio across forecast horizons.

D. Accuracy and Efficiency Trade-off

To evaluate the trade-off between predictive skill and operational efficiency, we analyze HSS alongside theoretical MACs. As detailed in Table II and Figure 5, AMU-Net demonstrates a superior efficiency-performance ratio. Unlike autoregressive baselines (e.g., ConvLSTM), which incur linearly escalating costs from iterative inference, AMU-Net maintains a constant computational overhead regardless of the lead time. Furthermore, it avoids the storage redundancy of SimVP-D, which requires quintupling the parameter count ($\times 5$) for separate models. Consequently, AMU-Net delivers state-of-the-art accuracy with fixed, low-latency costs, positioning it as an optimal solution for resource-constrained environments.

TABLE II

Comparison of model efficiency in terms of Parameters, MACs, and GPU memory cost. All statistics are reported for a single inference step (batch size = 1).

Method	Params (M)	MACs (G)	GPU memory (MB)
U-Net	39.402	181.25	1299
ConvLSTM	1.036	916.67	7254
SimVP	13.648	133.80	3399
SimVP-D	13.648×5	133.80	3399
SmaAt-UNet	7.973	43.24	3524
AMU-Net	7.542	200.88	5646

E. Interpreting the Behavior of Time Embedding

To validate the Physics-Aware Time Embedding, we visualized the learned affine parameters (γ and $boldsymbol\beta$) in the AdaLN layers. As shown in Figure 6, a clear range contraction occurs as lead time increases. Short horizons (10–30 min) show broad parameter ranges, indicating strong modulation for rapid, high-frequency changes. Longer lead times show concentrated distributions, suggesting the model lessens dependence on specific temporal instructions with increased atmospheric uncertainty, shifting toward intrinsic spatial representations.

Furthermore, a comparative analysis reveals a hierarchical control mechanism in which multiplicative scaling dominates. The scale parameters (γ) consistently span a significantly wider dynamic range than the additive shifts (β). This shows

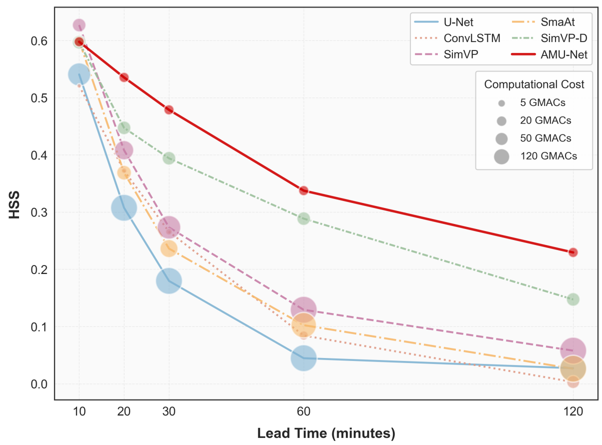

Fig. 5. Efficiency-Performance comparison across varying lead times. The y-axis denotes HSS performance, and bubble size indicates accumulated MACs.

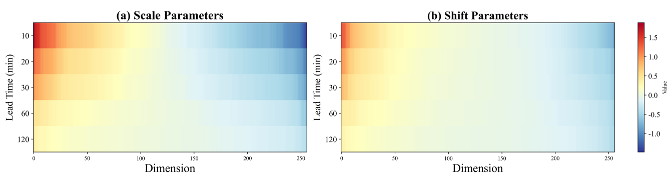

Fig. 6. Heatmap visualization of the affine modulation parameters (γ and β) across forecast horizons. Subplots (a) and (b) illustrate the scale and shift coefficients, respectively.

dot-product interactions mainly switch regimes, with minor additive shifts as bias. Multiplicative modulation better encodes complex multi-lead time dynamics than additive bias. The visualization confirms the architecture's strong explainability, as it learns to disentangle temporal physical instructions from spatial representations in the latent space.

F. Ablation Study

1) Effectiveness of Adaptive Temporal Modulation: To validate AdaLN, we compared it with a baseline that concatenates time embeddings with input features. Results shown in Figure 7 indicate that AdaLN's modulation strategy outperforms the baseline, especially at longer lead times (+60 and +120 min). While the baseline has slight advantages at short horizons, it lacks explicit statistical gating to distinguish physical regimes. AdaLN adapts internal feature statistics to forecast horizons, maintaining long-term accuracy.

2) Superiority of Mamba over Attention: We chose Mamba over standard Multi-Head Self-Attention based on a study with identical hyperparameters. As shown in Tables III and IV, Mamba offers a better cost-to-performance ratio. Its linear complexity results in lower GPU memory and MAC usage than quadratic Attention. Mamba also outperforms Attention in meteorological skill scores (CSI, HSS), indicating that while Attention excels at global similarity (higher SSIM), Mamba's linear scanning captures spatiotemporal dynamics more effectively, which is vital for high-resolution nowcasting. However, its overall performance remains unresolved [31].

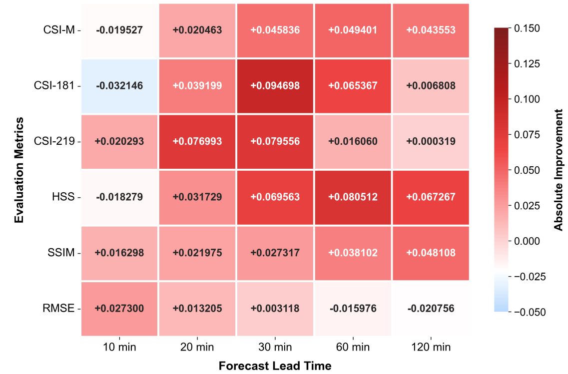

Fig. 7. Performance differential heatmap comparing the proposed modulation method with the concatenation baseline. Values represent the absolute improvement (or relative reduction for RMSE) achieved by AMU-Net. Red indicates superior performance.

TABLE III
COMPUTATIONAL COST COMPARISON OF MAMBA AND ATTENTION VARIATIONS.

Variation	Params (M)	MACs (G)	GPU memory (MB)
Attention	7.709	268.77	6284
Mamba	7.542	200.88	5646

3) Impact of Training Strategy: We measured the impact of adversarial training by comparing it with a deterministic baseline. As shown in Table V, the adversarial module enhances forecasting (CSI, HSS) and perceptual quality (SSIM), but slightly increases RMSE. This illustrates the "perception-distortion trade-off": deterministic models minimize RMSE at the expense of blurred predictions. At the same time, our generative approach prioritizes sharpness and accuracy, especially for high-intensity event detection (CSI-219).

VI. CONCLUSIONS

We propose AMU-Net, a physics-aware framework balancing efficiency and high-fidelity precipitation nowcasting. It integrates the Mamba mixer into a U-Net, achieving global receptive fields while lowering inference costs. The Physics-Aware Time Embedding encodes temporal instructions via Adaptive Layer Normalization, enabling multi-lead-time fore-

TABLE IV
MEAN PERFORMANCE METRICS OF THE MAMBA AND ATTENTION VARIATIONS. BETTER VALUES ARE IN **BOLD**.

Variation	CSI-M ↑	CSI-181 ↑	CSI-219 ↑	HSS ↑	SSIM ↑	RMSE ↓
Attention	0.302371	0.121487	0.029108	0.386839	**0.712677**	1.651351
Mamba	**0.337438**	**0.162821**	**0.084406**	**0.435862**	0.699657	**1.537527**

TABLE V
PERFORMANCE COMPARISON WITH AND WITHOUT THE PROPOSED HYBRID GENERATIVE TRAINING STRATEGY. **BOLD** VALUES INDICATE BETTER PERFORMANCE.

Method	CSI-M ↑	CSI-181 ↑	CSI-219 ↑	HSS ↑	SSIM ↑	RMSE ↓
w/o strategy	0.291856	0.109431	0.020949	0.370251	0.683465	**1.424136**
w strategy	**0.337438**	**0.162821**	**0.084406**	**0.435862**	**0.699657**	1.537527

casting and removing recursive error. Our Hybrid Generative Training combines spectral and adversarial constraints to reduce smoothing bias for high-intensity convective cores. Experiments on the SEVIR dataset show AMU-Net outperforms state-of-the-art baselines in skill scores and perceptual metrics.

Despite advances, limitations remain. The unified model has a slight performance drop at +10 min. AMU-Net improves extreme event detection, but long-term prediction of severe convection remains challenging, with only limited improvements. While inference is efficient, training complex multi-lead dynamics requires significant computation.

Future work will develop lightweight paradigms to enhance training and extend the framework to finer temporal resolutions. We also aim to capture extreme events better to improve early disaster warnings.

References

[1] J. Sun, M. Xue, J. W. Wilson, I. Zawadzki, S. P. Ballard, J. Onvlee-Hooimeyer, P. Joe, D. M. Barker, P.-W. Li, B. Golding, M. Xu, and J. Pinto, "Use of nwp for nowcasting convective precipitation: Recent progress and challenges," *Bull. Am. Meteorol. Soc.*, vol. 95, no. 3, pp. 409–426, 2014.

[2] S. Pulkkinen, D. Nerini, A. A. Pérez Hortal, C. Velasco-Forero, A. Seed, U. Germann, and L. Foresti, "Pysteps: an open-source python library for probabilistic precipitation nowcasting (v1.0)," *Geosci. Model Dev.*, vol. 12, no. 10, pp. 4185–4219, 2019. [Online]. Available: https://gmd.copernicus.org/articles/12/4185/2019/

[3] A. Tafferner and C. Forster, *Weather Nowcasting and Short Term Forecasting*. Berlin, Germany: Springer Berlin Heidelberg, 2012, pp. 363–380. [Online]. Available: https://doi.org/10.1007/978-3-642-30183-4_22

[4] G. Ayzel, M. Heistermann, and T. Winterrath, "Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1)," *Geosci. Model Dev.*, vol. 12, no. 4, pp. 1387–1402, 2019. [Online]. Available: https://gmd.copernicus.org/articles/12/1387/2019/

[5] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadtler, "Can deep learning beat numerical weather prediction?" *Philos. Trans. R. Soc. A*, vol. 379, no. 2194, p. 20200097, Feb. 2021. [Online]. Available: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0097

[6] Y. Wang, H. Wu, J. Zhang, Z. Gao, J. Wang, P. S. Yu, and M. Long, "Predrnn: A recurrent neural network for spatiotemporal predictive learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2208–2225, 2023.

[7] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "SimVP: Simpler yet better video prediction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 3170–3180.

[8] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, 2016. [Online]. Available: https://iclr.cc/archive/www/doku.php%3Fid=iclr2016:main.html

[9] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas, and S. Mohamed, "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, p. 672–677, Sep. 2021. [Online]. Available: http://dx.doi.org/10.1038/s41586-021-03854-z

[10] Y. Zhang, M. Long, K. Chen, L. Xing, R. Jin, M. I. Jordan, and J. Wang, "Skilful nowcasting of extreme precipitation with NowcastNet," *Nature*, vol. 619, no. 7970, pp. 526–532, Jul. 2023. [Online]. Available: https://www.nature.com/articles/s41586-023-06184-4

[11] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3d neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, Jul. 2023. [Online]. Available: https://www.nature.com/articles/s41586-023-06185-3

[12] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato *et al.*, "GraphCast: Learning skillful medium-range global weather forecasting," *Science*, vol. 382, no. 6677, pp. 1416–1421, 2023.

[13] L. Chen, X. Zhong, F. Zhang *et al.*, "FuXi: A cascade machine learning forecasting system for 15-day global weather forecast," *npj Clim. Atmos. Sci.*, vol. 6, no. 1, p. 190, 2023.

[14] P. A. G. Watson, "Machine learning applications for weather and climate need greater focus on extremes," *Environ. Res. Lett.*, vol. 17, no. 11, p. 111004, nov 2022.

[15] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, USA, 2018, pp. 6228–6237.

[16] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 28, Montreal, QC, Canada, 2015, pp. 802–810.

[17] M. Turzi and S. Mehrkanoon, "Ssa-unet: Advanced precipitation nowcasting via channel shuffling," in *2025 International Joint Conference on Neural Networks (IJCNN)*, 2025, pp. 1–8.

[18] D. Liu, P. Zheng, H. Deng, X. Yuan, X. Wu, and J. Hu, "Multi-source feature fusion and spatio-temporal unet for precipitation nowcasting," in *2025 International Joint Conference on Neural Networks (IJCNN)*, 2025, pp. 1–8.

[19] R. Durall, M. Keuper, and J. Keuper, "Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 7887–7896.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, Virtual, 2021.

[21] Z. Gao, X. Shi, H. Wang, Y. Zhu, Y. Wang, M. Li, and D.-Y. Yeung, "Earthformer: exploring space-time transformers for earth system forecasting," in *International Conference on Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, USA, 2022. [Online]. Available: https://dl.acm.org/doi/10.5555/3600270.3602111

[22] L. Ma, B. Duan, Y. Li, Y. Fu, and D. Ma, "A transformer-based hybrid model for multi-indicator numerical prediction of marine environment," in *2025 International Joint Conference on Neural Networks (IJCNN)*, 2025, pp. 1–8.

[23] T. Dao and A. Gu, "Transformers are SSMs: generalized models and efficient algorithms through structured state space duality," in *International Conference on Machine Learning (ICML)*, Vienna, Austria, 2024, pp. 9338–9373. [Online]. Available: https://dl.acm.org/doi/10.5555/3692070.3692469

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, p. 139–144, Oct. 2020. [Online]. Available: https://doi.org/10.1145/3422622

[25] C. Hamzaçebi, D. Akay, and F. Kutay, "Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting," 2009.

[26] M. S. Veillette, S. Samsi, and C. J. Mattioli, "SEVIR: a storm event imagery dataset for deep learning applications in radar and satellite meteorology," in *34th International Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, BC, Canada, 2020, pp. 22 009–22 019.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Munich, Germany: Springer, 2015, pp. 234–241.

[28] K. Trebing, T. Stanczyk, and S. Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," *Pattern Recognit. Lett.*, vol. 145, pp. 178–186, 2021.

[29] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[30] C. Subich, S. Z. Husain, L. Separovic, and J. Yang, "Fixing the double penalty in data-driven weather forecasting through a modified spherical harmonic loss function," in *International Conference on Machine Learning (ICML))*, Vancouver, BC, Canada, 2025. [Online]. Available: https://icml.cc/virtual/2025/poster/44912

[31] W. Yu and X. Wang, "Mambaout: Do we really need mamba for vision?" in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2025, pp. 4484–4496.