DuoCast: Duo-Probabilistic Diffusion for Precipitation Nowcasting

Penghui Wen[1], Mengwei He[1], Patrick Filippi[2], Na Zhao[3], Feng Zhang[4], Thomas Francis Bishop[2], Zhiyong Wang[1], Kun Hu[5,*]

[1] School of Computer Science, The University of Sydney, Sydney, Australia
[2] School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia
[3] School of Mathematics, Shanghai University of Finance and Economics, Shanghai, China
[4]Department of Atmospheric and Oceanic Sciences, Institutes of Atmospheric Sciences, Fudan University, Shanghai, China
[5]School of Science, Edith Cowan University, Perth, Australia
{penghui.wen, mengwei.he, patrick.filippi}@sydney.edu.au, zhaona@shufe.edu.cn, fengzhang@fudan.edu.cn,
{thomas.bishop, zhiyong.wang}@sydney.edu.au, k.hu@ecu.edu.au

Abstract

Accurate short-term precipitation forecasting is critical for weather-sensitive decision-making in agriculture, transportation, and disaster response. Existing deep learning approaches often struggle to balance global structural consistency with local detail preservation, especially under complex meteorological conditions. We propose *DuoCast*, a dual-diffusion framework that decomposes precipitation forecasting into low- and high-frequency components modeled in orthogonal latent subspaces. We theoretically prove that this frequency decomposition reduces prediction error compared to conventional single branch U-Net diffusion models. In *DuoCast*, the low-frequency model captures large-scale trends via convolutional encoders conditioned on weather front dynamics, while the high-frequency model refines fine-scale variability using a self-attention-based architecture. Experiments on four benchmark radar datasets show that *DuoCast* outperforms state-of-the-art baselines, achieving superior accuracy in both spatial detail and temporal evolution.

Code — https://github.com/ph-w2000/DuoCast
Appendix — https://arxiv.org/abs/2412.01091

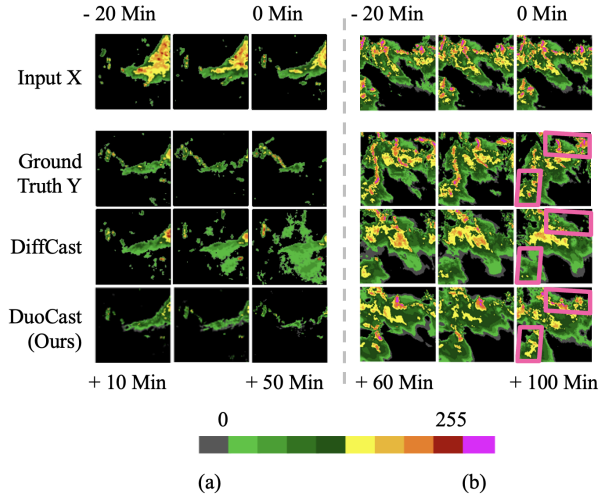

Figure 1: Challenges in precipitation nowcasting: a) precipitation with weather (warm) front patterns, and b) high-frequency features (micro-scale variability) within edge regions shown in pink parallelograms.

Introduction

Precipitation plays a vital role in weather systems, influencing temperature, humidity, and overall atmospheric conditions. Precipitation nowcasting, which predicts short-term rainfall using recent radar observations, is increasingly important for applications such as agriculture, transportation, and disaster response (Zhang et al. 2023). Its importance was underscored during the July 2025 Texas flash floods: National Weather Service (NWS) observed localized totals over 5 inches and rainfall rates of 2–3 inches per hour at many locations, yet forecasts slightly mispredicted the heaviest cells, resulting in flash flooding in central Texas and trapping hundreds of people (Moore et al. 2025). Conventional approaches to nowcasting rely on numerical weather prediction (NWP), which simulates atmospheric dynamics by solving partial differential equations (PDEs) derived from physical laws (Bauer, Thorpe, and Brunet 2015; Lorenc 1986; Alley, Emanuel, and Zhang 2019). However, the complexity of NWP renders them computationally expensive and inefficient for real-time forecasting (Bi et al. 2023).

Recent advances in data-driven deep learning techniques have shown greater potential than traditional NWP methods in precipitation prediction by leveraging large datasets and avoiding the need to solve complex physical equations. Deterministic models (Shi et al. 2017; Gao et al. 2022; Yan et al. 2024; Lin et al. 2025) aim to capture the overall motion of precipitation systems by generating mean-value forecasts. However, they often produce blurry predictions and lack fine-grained spatial detail (Ravuri et al. 2021). In contrast, probabilistic models (Ravuri et al. 2021; Zhang et al. 2023; Gao et al. 2024) introduce stochastic latent variables to better capture local-scale variability. While effective at modeling uncertainty, treating the entire system as fully stochastic can introduce excessive randomness and degrade forecasting

*Corresponding author.

precision (Yu et al. 2024). To address this, hybrid approaches such as DiffCast (Yu et al. 2024) combine global deterministic forecasting with local probabilistic refinement. Nevertheless, its CNN-based diffusion model struggles to preserve high-frequency information due to limited capacity (Yang et al. 2023), resulting in less spatially accurate precipitation intensity, as illustrated in Fig. 1(b).

Therefore, we introduce *DuoCast*, a dual-diffusion framework for precipitation nowcasting. *DuoCast* decomposes the forecasting task into two frequency-specific subspaces, where separate stochastic diffusion models capture low- and high-frequency components. We theoretically show that frequency decomposition reduces prediction error compared to standard U-Net-based diffusion with formal proof. Specifically, the low-frequency diffusion model captures large-scale precipitation trends via convolutions, conditioned on weather front representations extracted from historical radar data. These features integrate spatial modeling of air masses and temporal modeling of front evolution, with intensity-aware enhancement for structural regions. The high-frequency model operates in latent space using a self-attention backbone to model fine-scale variability, conditioned on both historical inputs and the low-frequency forecast. Evaluated on four benchmark radar datasets, *DuoCast* achieves state-of-the-art performance and yields substantial improvements in precipitation nowcasting accuracy.

Our work's key contributions are summarized as follows:

- We propose *DuoCast*, a dual-diffusion framework that decomposes precipitation forecasting into low- and high-frequency subspaces for frequency-aware modeling.
- We provide a theoretical analysis showing that frequency decomposition reduces prediction error compared to standard single-branch diffusion models.
- Extensive experiments on four benchmark radar datasets demonstrate that *DuoCast* achieves state-of-the-art performance in short-term precipitation nowcasting.

Related Work

Precipitation Nowcasting

Precipitation nowcasting has traditionally relied on numerical weather prediction (NWP), which solves atmospheric PDEs (Skamarock et al. 2008). Although effective, NWP is computationally intensive, often requiring hours on supercomputers (Bi et al. 2023). Recent deep learning models (Andrychowicz et al. 2023) offer competitive performance with significantly lower latency. They are generally categorized as either deterministic or probabilistic.

Deterministic Precipitation Models aim to predict the mean-based evolution of precipitation by modeling spatial and temporal dynamics. ConvLSTM (Shi et al. 2015) captures these dynamics via convolutional and recurrent structures, while PhyDNet (Guen and Thome 2020) introduces a physics-guided decomposition into stochastic and deterministic motion. FACL (Yan et al. 2024) replaces standard L2 loss with a loss tailored for signal-based spatiotemporal modeling. AlphaPre (Lin et al. 2025) disentangles phase and amplitude to separately model motion and intensity.

Probabilistic Precipitation Models use latent variables to capture the inherent uncertainty of future weather, enabling finer modeling of micro-scale phenomena. Representative methods include DGMR (Ravuri et al. 2021), which employs a generative adversarial network (GAN) with spatial and temporal discriminators, and NowcastNet (Zhang et al. 2023), which integrates physical priors into GAN-based prediction. A diffusion model generates images by progressively denoising random noise into coherent visuals (Wen et al. 2024; Lu et al. 2024). Building on this, Prediff (Gao et al. 2024) incorporates knowledge control to guide the sampling process. Although such approaches enhance forecast diversity, their fully stochastic nature can also introduce excessive variability, undermining predictive reliability.

Hybrid Precipitation Models combine deterministic backbones for global precipitation trends with probabilistic components to capture local variability, aiming to balance accuracy and diversity. For example, CasCast (Gong et al. 2024) and DiffCast (Yu et al. 2024) integrate trajectory forecasting with stochastic refinement. However, their limited treatment of frequency-specific structures and micro- and macro-patterns may hinder long-range performance.

Frequency Analysis and Modeling

Frequency-domain analysis has shown strong effectiveness in image analysis and generation (Tan et al. 2024; Woo et al. 2022; Wen et al. 2023; Chen et al. 2025), serving as a valuable complement to spatial-based approaches, which often struggle with artifacts. For instance, (Doloriel and Cheung 2024) apply frequency-based masking to extract shared features, while (Zhou et al. 2024) categorize frequency components into semantic, structural, and noise levels to identify regions with unique spectral characteristics.

Methodology

Problem Formulation

Precipitation nowcasting is formulated as a spatio-temporal forecasting task of single-channel radar echoes (Yu et al. 2024; Gao et al. 2024; Gong et al. 2024). Given past echo intensity observations, denoted as $\mathbf{X} \in \mathbb{R}^{S\times H\times W}$ as the condition, short-term precipitation nowcasting models' objective is to model the conditional probabilistic $p(\mathbf{Y}|\mathbf{X})$ of the future following frames $\mathbf{Y} \in \mathbb{R}^{S\times H\times W}$, where H and W define the spatial resolution of each frame and S is the number of temporal frames. For notation simplicity, we denote $d = S \times H \times W$. To extend predictions over longer horizons, an autoregressive strategy is employed.

Spectral Subspace Diffusion in DuoCast

To enhance precipitation estimation, *DuoCast* employs two distinct stochastic diffusion processes to model low- and high-frequency components in dedicated subspaces. The architecture overview is presented in Fig. 2. Specifically, we design two complementary diffusion models: low-frequency, which integrates weather front information to model low-frequency, large-scale precipitation patterns; and high-frequency, which captures high-frequency, localized variability to refine fine-grained precipitation details.

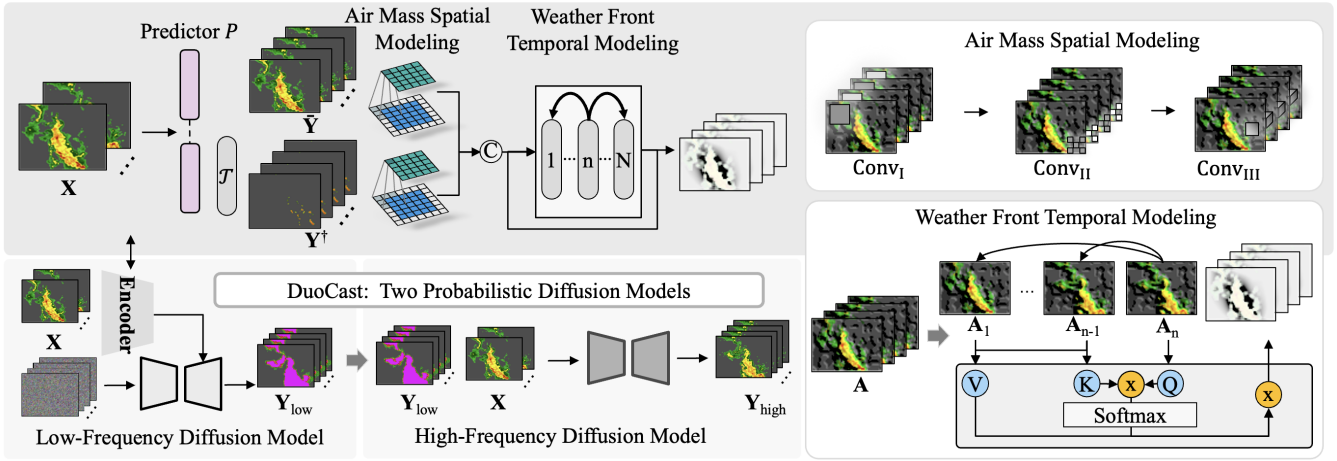

Figure 2: Overview of *DuoCast*. A two-stage diffusion model leveraging historical weather fronts signals: (i) a low-frequency convolutional stage capturing front-guided precipitation trends, and (ii) a high-frequency self-attention stage refining micro-scale variability for pixel-level forecasts.

Let $\mathcal{F}$ denote the Fourier transform on $L^2(\mathbb{R}^d)$, and write $\widehat{f} := \mathcal{F}f$ for the transform of f. Given a cutoff $\omega_c > 0$, define the low-frequency subspace $\mathcal{B}_{\omega_c} \subset L^2(\mathbb{R}^d)$ as

$$\mathcal{B}_{\omega_c} := \left\{ f \in L^2(\mathbb{R}^d) : \operatorname{supp}(\widehat{f}) \subseteq \left\{ \xi \in \mathbb{R}^d : |\xi| \leq \omega_c \right\} \right\},$$

and denote its orthogonal complement as $\mathcal{B}^{\perp}_{\omega_c}$. Define the corresponding orthogonal projections by

$$P_{\mathcal{B}_{\omega_c}} f := \mathcal{F}^{-1}(\chi_{\omega_c} \widehat{f}), P_{\mathcal{B}^{\perp}_{\omega_c}} f := f - P_{\mathcal{B}_{\omega_c}} f, \quad (1)$$

where $\chi_{\omega_c}(\xi) := \mathbf{1}_{\{|\xi| \leq \omega_c\}}(\xi)$ is the frequency-domain indicator function. These projections satisfy the standard orthogonal projection properties:

$$P^2_{\mathcal{B}_{\omega_c}} = P_{\mathcal{B}_{\omega_c}}, P^2_{\mathcal{B}^{\perp}_{\omega_c}} = P_{\mathcal{B}^{\perp}_{\omega_c}}, P_{\mathcal{B}_{\omega_c}} P_{\mathcal{B}^{\perp}_{\omega_c}} = 0. \quad (2)$$

To this end, $\mathbf{Y}$ can be decomposed into $\mathbf{Y}_{\text{low}} = P_{\mathcal{B}_{\omega_c}}(\mathbf{Y})$ and $\mathbf{Y}_{\text{high}} = P_{\mathcal{B}^{\perp}_{\omega_c}}(\mathbf{Y})$. The two diffusion models estimate $\mathbf{Y}$ in two stages, for $\mathbf{Y}_{\text{low}}$ and $\mathbf{Y}_{\text{high}}$ to capture the low- and high-frequency components, respectively.

Low-Frequency Subspace Diffusion. Low-frequency components capture the overall scale and structure of precipitation systems, as macro trends. To estimate $\mathbf{Y}_{\text{low}}$ in the low-frequency subspace, a low-frequency diffusion network $\epsilon_{\theta_{\text{low}}}$ is employed, where θ_{low} indicates the parameters of the diffusion model. For forward diffusion, we sample from $q(\mathbf{Y}^{\text{t}}_{\text{low}}|\mathbf{Y}^0_{\text{low}})$ in a closed form at an arbitrary timestep $t \in T$ as $q(\mathbf{Y}^{\text{t}}_{\text{low}}|\mathbf{Y}^{\text{t-1}}_{\text{low}}) = \mathcal{N}(\mathbf{Y}^{\text{t}}_{\text{low}}; \sqrt{1-\beta_t}\mathbf{Y}^{\text{t-1}}_{\text{low}}, \beta_t \mathbf{I})$. The reverse diffusion process is formulated as:

$$p_{\theta_{\text{low}}}(\mathbf{Y}_{\text{low}}|\mathbf{X}). \quad (3)$$

To optimize $\epsilon_{\theta_{\text{low}}}$, an objective function is defined as follows:

$$\mathcal{L}_{\text{low}} = \mathbb{E}_{\mathbf{Y}^0_{\text{low}},\epsilon,t} \left\| \epsilon - \epsilon_{\theta_{\text{low}}} \left(\mathbf{Y}^t_{\text{low}}, t, \mathbf{X} \right) \right\|^2. \quad (4)$$

High-Frequency Subspace Diffusion. High-frequency components correspond to localized precipitation intensities, shown as micro-scale variations. To estimate $\mathbf{Y}_{\text{high}}$, we introduce a high-frequency diffusion model $\epsilon_{\theta_{\text{high}}}$, which follows a similar formulation as the low-frequency counterpart, as defined in Equations 3 and 4. To optimize $\epsilon_{\theta_{\text{high}}}$, an objective function is defined as follows:

$$\mathcal{L}_{\text{high}} = \mathbb{E}_{\mathbf{Y}^0_{\text{high}},\epsilon,t} \left\| \epsilon - \epsilon_{\theta_{\text{high}}} \left(\mathbf{Y}^t_{\text{high}}, t, \mathbf{X} \right) \right\|^2. \quad (5)$$

Foundations of Spectral Subspace Diffusion

Spectral Decay of Finite Support Convolutions. While convolutional backbones like U-Net are commonly used in diffusion models, they are limited in generating high-frequency components—particularly crucial in precipitation nowcasting. We first illustrate the limitations of existing architectures in modeling high-frequency patterns, and then explain how a two-stage design operating in distinct spectral subspaces addresses this issue.

Lemma 1 (Polynomial Fourier decay of *bounded variation* (BV) kernels)**.** *Let $k \in L^1(\mathbb{R}^d)$ be a convolution kernel with compact support and finite total variation, i.e. $k \in BV(\mathbb{R}^d)$. Then there exists a constant $C_{\text{BV}} > 0$ (depending only on $\text{TV}(k)$) such that*

$$\left|\widehat{k}(\xi)\right| \leq C_{\text{BV}}/(1+|\xi|), \qquad \forall\, \xi \in \mathbb{R}^d.$$

The standard architectural constraints of CNNs (finite kernel size and finite-precision weights) guarantee that the theoretic hypotheses of the lemma are met. Therefore the polynomial Fourier-decay result is rigorously applicable to virtually all convolution kernels used in modern deep-learning. Suppose a depth-L CNN (e.g., UNet or repeated UNet ϵ_θ used in Diffusion) is formed by L consecutive convolutional kernels $k_1, \ldots, k_L$ with each $k_\ell \in BV(\mathbb{R}^d)$, and define the composite frequency response:

$$H_L(\xi) := \prod_{\ell=1}^{L} \widehat{k_\ell}(\xi). \quad (6)$$

Theorem 1 (Spectral envelope under bounded variation)**.** *Under the above assumptions, there exists $C > 0$ such that*

$$|H_L(\xi)| \le C^L (1 + |\xi|)^{-L}, \qquad \forall\, \xi \in \mathbb{R}^d. \tag{7}$$

Sharp Tail Bound. We now derive an *input–dependent* high–frequency estimate that retains the true spectrum.

Theorem 2 (Sharp capacity bottleneck)**.** *Let $\mathbf{Y} \in L^2(\mathbb{R}^d)$ be any target, $\omega_c = \Omega(\epsilon)$, $\mathbf{Y}_{low} = P_{\mathcal{B}_{\omega_c}}(\mathbf{Y})$ and $\mathbf{Y}_{high} = P_{\mathcal{B}^{\perp}_{\omega_c}}(\mathbf{Y}) = \mathbf{Y} - \mathbf{Y}_{low}$. For every generation $g = \mathbf{Y}^t$ based on a CNN with input $f = \mathbf{Y}^0$,*

$$\|\mathbf{Y} - \mathbf{Y}^t\|_2 \;\ge\; \left[\|\mathbf{Y}_{high}\|_2 - \sqrt{\varepsilon}\,\|\mathbf{Y}^0\|_2\right]_+.$$

Corollary 1 (Irreducible high-frequency error)**.** *If $\|\mathbf{Y}_{high}\|_2 > \sqrt{\varepsilon}\,\|\mathbf{Y}^0\|_2$, then every CNN-based estimation $\mathbf{Y}^t$ obeys*

$$\|\mathbf{Y} - \mathbf{Y}^t\|_2^2 \;\ge\; \left(\|\mathbf{Y}_{high}\|_2 - \sqrt{\varepsilon}\,\|\mathbf{Y}^0\|_2\right)^2 \;>\; 0,$$

so the approximation error cannot be eliminated by any depth-L CNN.

Two–Stage Diffusion Modeling. For any target $\mathbf{Y} = \mathbf{Y}_{\text{low}} + \mathbf{Y}_{\text{high}}$, where $\mathbf{Y}_{\text{low}} \in \mathcal{B}_{\omega_c}$ and $\mathbf{Y}_{\text{high}} \in \mathcal{B}^{\perp}_{\omega_c}$, assume the first and second diffusion model generation $g^{(1)}_\theta \colon \mathcal{B}_{\omega_c} \to \mathcal{B}_{\omega_c}$ and $g^{(2)}_\phi \colon \mathcal{B}^{\perp}_{\omega_c} \to \mathcal{B}^{\perp}_{\omega_c}$, respectively. Orthogonality yields:

$$\|\mathbf{Y} - g^{(1)}_\theta - g^{(2)}_\phi\|_2^2 = \|\mathbf{Y}_{\text{low}} - g^{(1)}_\theta\|_2^2 + \|\mathbf{Y}_{\text{high}} - g^{(2)}_\phi\|_2^2.$$

Corollary 2 (Two–stage universal approximation)**.** *Suppose the family $\mathcal{G}_\theta \subset \mathcal{B}_{\omega_c}$ is dense in $\mathcal{B}_{\omega_c}$, and $\mathcal{G}_\phi \subset \mathcal{B}^{\perp}_{\omega_c}$ is dense in $\mathcal{B}^{\perp}_{\omega_c}$. Then, for every $\mathbf{Y} = \mathbf{Y}_{low} + \mathbf{Y}_{high} \in L^2$,*

$$\inf_{\theta \in \mathcal{G}_\theta,\, \phi \in \mathcal{G}_\phi} \left\|\mathbf{Y} - g^{(1)}_\theta - g^{(2)}_\phi\right\|_2 = 0.$$

We next investigate two (approximated) architectural designs for the low- and high-frequency diffusion models.

Low-Frequency Diffusion Model

Weather fronts, which evolve gradually and span broad spatial regions, are key to shaping precipitation systems (Catto et al. 2012). For instance, warm fronts bring steady light rain, cold fronts induce abrupt heavy rainfall, and occluded fronts result in widespread precipitation. Inadequate modeling of them can lead to significant forecasting errors. As shown in Fig. 1(a), the context indicates a weakening warm front, yet DiffCast erroneously predicts intensifying rainfall. Therefore, the low-frequency diffusion captures these trends with convolutions to guide prediction, conditioned on front dynamics derived from underlying air mass interactions.

We first introduce a base encoder that extracts precipitation trends from the input sequence $\mathbf{X}$. It focuses on two key aspects: the spatial distribution of general precipitation and the identification of high-intensity precipitation regions.

Specifically, a convolutional predictor P produces an initial forecast $\bar{\mathbf{Y}} = P(\mathbf{X})$, capturing the general spatial distribution of precipitation. To further isolate high-intensity regions that are indicative of strong air mass interactions, a thresholding operator $\mathcal{T}$ with cutoff θ_{int} is applied: $\mathbf{Y}^\dagger = \mathcal{T}(\bar{\mathbf{Y}}, \theta_{\text{int}})$. These two outputs serve as proxies for identifying air mass locations and inferring evolving weather fronts. The predictor is trained via the following objective:

$$\mathcal{L}_P = \mathbb{E}\left\|\mathbf{Y} - \bar{\mathbf{Y}}\right\|^2. \tag{8}$$

Air Mass Spatial Modeling. To model the spatial organization of weather fronts, we introduce a convolutional module that learns the distribution and interactions of underlying air masses. As air mass boundaries often coincide with sharp gradients in precipitation (Lagerquist, Mcgovern, and Gagne II 2019), this module extracts spatial cues relevant to front structure. Specifically, we apply three tailored convolutional layers to the preliminary forecasts $\bar{\mathbf{Y}}$ and $\mathbf{Y}^\dagger$ to capture local context, multi-scale structure, and temporal anchoring, forming a spatial representation $\mathbf{A}$.

1) Surrounding Mass Context - $\text{Conv}_{\text{I}}(\cdot)$: We first apply large-kernel depth-wise convolutions to extract broad spatial contexts of air mass influence, capturing smooth but widespread structures such as pressure boundaries.

2) Occluded Front Dynamics - $\text{Conv}_{\text{II}}(\cdot)$: To detect multi-scale interactions like occluded fronts, we apply dilated convolutions, which enhance sensitivity to overlapping cold and warm air regions at varying spatial scales.

3) Temporal Anchoring - $\text{Conv}_{\text{III}}(\cdot)$: Lastly, we incorporate temporal locality via pixel-wise convolution across time, which encodes the evolution of precipitation intensity at each spatial location.

The operations are applied separately to the general and high-intensity forecasts:

$$\begin{aligned} \bar{\mathbf{A}} &= \text{Conv}_{\text{III}}(\text{Conv}_{\text{II}}(\text{Conv}_{\text{I}}(\bar{\mathbf{Y}}))), \\ \mathbf{A}^\dagger &= \text{Conv}_{\text{III}}(\text{Conv}_{\text{II}}(\text{Conv}_{\text{I}}(\mathbf{Y}^\dagger))). \end{aligned} \tag{9}$$

We concatenate the outputs to obtain the spatial air mass representation as front-aware spatial features:

$$\mathbf{A} = \text{Concat}(\bar{\mathbf{A}}, \mathbf{A}^\dagger).$$

Weather Front Temporal Modeling. To capture the temporal evolution of weather fronts, we devise a lightweight convolution-based cross-attention mechanism that relates the current spatial features to both historical structure and recent trends. For each frame $\mathbf{A}_n$, we compute attention over two reference frames: (1) the leading frame $\mathbf{A}_1$, which serves as a static structural prior, and (2) the previous frame $\mathbf{A}_{n-1}$, which reflects recent changes in air mass interactions. The attention query is derived from the current frame, while the key and value come from the concatenation of the leading and previous frames. Formally: Cross-Attention($\mathbf{Q} = \mathbf{W_Q} \circledast \mathbf{A}_n$,$\mathbf{K} = \mathbf{W_K} \circledast [\mathbf{A}_1 \oplus \mathbf{A}_{n-1}]$,$\mathbf{V} = \mathbf{W_V} \circledast [\mathbf{A}_1 \oplus \mathbf{A}_{n-1}]$), where $\circledast$ is the depth-wise convolution, $\oplus$ is concatenation along the channel dimension, and $\mathbf{W_Q}$, $\mathbf{W_K}$, and $\mathbf{W_V}$ are learnable projection weights.

High-Frequency Diffusion Model

The high-frequency diffusion model ϵ_{θ_2} captures high-frequency precipitation patterns using a fully self-attention-based architecture in latent space. By dynamically reweighting spatial features, it preserves fine-scale details more effectively than fixed convolutional filters (Rombach et al.

2022). Complementing the coarse predictions from low-frequency diffusion, it enhances local variability and short-term dynamics, particularly at extended lead times. The model is conditioned on two inputs: the structural prior $\hat{\mathbf{Y}}_{\text{low}}$ from low-frequency and the historical observations $\mathbf{X}$. These are encoded into a latent representation $\mathbf{Z}$ via a pre-trained autoencoder (Kingma and Welling 2013). The model finally generates a high-frequency correction $\hat{\mathbf{Y}}_{\text{high}}$.

Optimization

We follow a standardized diffusion training process for both the low-frequency and high-frequency diffusion models:

$$\mathcal{L} = \lambda_1 \mathcal{L}_P + \lambda_2 \mathcal{L}_{\text{low}} + \lambda_3 \mathcal{L}_{\text{high}}, \tag{10}$$

where λ_1, λ_2 and λ_3 are hyperparameters.

Experiments & Discussions

Experimental Settings

Datasets To evaluate *DuoCast*'s effectiveness in generating precise precipitation maps, we performed experiments using four radar echo datasets: SEVIR (Veillette, Samsi, and Mattioli 2020), MeteoNet (Larvor and Berthomier 2021), Shanghai_Radar (Chen et al. 2020), and CIKM[1].

Method	SEVIR				
	↑CSI-M	↑CSI-181	↑CSI-219	↑HSS	↑SSIM
DuoCast	**0.338**	**0.182**	**0.107**	**0.432**	**0.683**
High-Freq	0.310	0.158	0.050	0.401	0.611
Low-Freq	0.327	0.164	0.066	0.410	0.642
Low-Frequency Backbone Exploration					
Attention	0.310	0.143	0.043	0.393	0.608
Conv	0.327	0.164	0.066	0.410	0.642
- Conv_{I}	0.316	0.155	0.051	0.405	0.631
- Conv_{II}	0.317	0.159	0.048	0.400	0.632
- Conv_{III}	0.320	0.156	0.050	0.404	0.622
- $\mathbf{Y}^{\dagger}$	0.316	0.146	0.048	0.400	0.619
- Airmass	0.309	0.143	0.051	0.388	0.625
- Front	0.311	0.138	0.050	0.383	0.628

Table 1: Ablation study on key components. *Attention* and *Conv* denote low-frequency models with self-attention and convolution backbones, respectively. *Airmass* and *Front* are air mass spatial and weather front temporal component. A dash (–) indicates component removal.

Evaluation Metrics Following SOTA (Yu et al. 2024; Lin et al. 2025; Gong et al. 2024), we evaluate the nowcasting with the average Critical Success Index (CSI), Heidke Skill Score (HSS), Structural Similarity Index Measure (SSIM). The CSI, similar to the Intersection over Union (IoU), measures the degree of pixel-wise alignment between predictions and ground truth after thresholding them into binary (0/1) matrices, while HSS reflects improvement over random guessing. Following (Lin et al. 2025), we measure and average CSI (CSI-M) and CSI with two highest thresholds to measure the performance of high-intensity precipitation forecasting. SSIM assesses prediction quality.

[1]https://tianchi.aliyun.com/dataset/1085

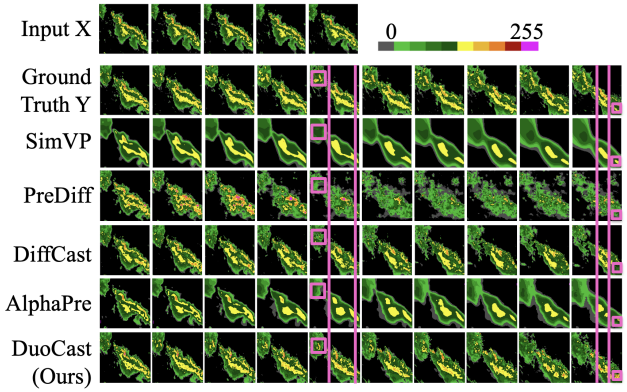

Figure 3: Qualitative comparison with SoTA on a SEVIR event. *DuoCast* captures finer micro-scale details (pink box) and preserves more consistent evolution trends (pink line).

Implementation Details *DuoCast* was trained with Adam optimizer (learning rate 1×10^{-4}) and 1,000 diffusion steps per low-/high-frequency branch. We used a two-stage schedule: first train the low-frequency model, then jointly train both. The learning rate was fixed across stages; to stabilize training we applied loss weights. In the first stage, the loss weight coefficients were $\lambda_1 = \lambda_2 = 0.5$. In the second stage, the loss weight were $\lambda_1 = \lambda_2 = 0.1$ and $\lambda_3 = 0.5$. The SimVP model is used as initial predictor P. The cutoff θ_{init} is set as the high-intensity precipitation threshold defined in the datasets.

Compared with the State-of-the-Art

To assess the performance of our *DuoCast* framework in precipitation nowcasting, we compare it against a range of state-of-the-art methods, including probabilistic, deterministic, and hybrid approaches.

Quantitative Analysis Based on the results in Table 2, we observe the following: i) Our *DuoCast* framework achieves 2–4% improvements in CSI and 2–5% in HSS across four datasets. It also matches state-of-the-art performance in SSIM and MSE, showing its effectiveness in enhancing prediction accuracy. ii) The CSI at higher thresholds reflects *DuoCast*'s capability to forecast high-intensity precipitation accurately. Furthermore, Fig. 6 (a) illustrates the variation of CSI across different prediction time steps for recent methods, highlighting that our model consistently outperforms others across nearly all time steps.

Qualitative Analysis Fig. 3 presents a qualitative comparison of our model with various SoTA methods for a precipitation event. It can be observed that SimVP and AlphaPre struggle to capture micro-scale variability, resulting in fewer fine-grained details over longer lead times, as deterministic models cannot effectively represent stochastic behaviors. While the probabilistic model Prediff capture finer micro-scale details, its predictions deviate significantly in trend, as probabilistic models often introduce excessive, uncontrollable randomness. In particular, Prediff forecasts a weakened precipitation band in the central region, resulting in an

Method	SEVIR					MeteoNet				
	↑CSI-M	↑CSI-181	↑CSI-219	↑HSS	↑SSIM	↑CSI-M	↑CSI-24	↑CSI-32	↑HSS	↑SSIM
MAU (2021)	0.3076	0.1071	0.0516	0.3863	0.6505	0.3233	0.2839	0.0997	0.4452	0.7897
SimVP (2022)	0.3108	0.1106	0.0517	0.3924	0.6508	0.3351	0.3002	0.1130	0.4573	0.7804
FourCastNet (2022)	0.2686	0.0717	0.0339	0.3355	0.5976	0.3027	0.2533	0.1085	0.4216	0.6450
Earthformer (2022)	0.2892	0.0844	0.0245	0.3665	0.6633	0.3205	0.2884	0.1237	0.4491	0.7772
PhyDNet (2020)	0.3017	0.1040	0.0278	0.3812	0.6532	0.3384	0.3194	0.1366	0.4673	0.7823
EarthFarseer (2024)	0.3004	0.0992	0.0413	0.3829	0.6327	0.3404	0.3170	0.1372	0.4726	0.7542
NowcastNet (2023)	0.2791	0.0770	0.0351	0.3512	0.6839	0.3427	0.3206	0.1598	0.4751	0.7879
PreDiff (2024)	0.2744	0.0627	0.0235	0.3592	0.5881	0.2944	0.2641	0.1202	0.4462	0.7085
DiffCast (2024)	0.3050	0.1300	0.0582	0.3996	0.6482	0.3512	0.3340	0.1808	0.4846	0.7887
CasCast (2024)	0.2878	0.0954	0.0312	0.3563	0.5680	0.3317	0.3021	0.1351	0.4682	0.7622
FACL (2024)	0.3161	0.1349	0.0655	0.4033	0.5686	0.3612	0.3341	0.1801	0.5006	0.7881
AlphaPre (2025)	0.3259	0.1332	0.0545	0.4110	**0.6884**	0.3824	0.3633	0.2002	0.5164	0.7968
Ours	**0.3375**	**0.1818**	**0.1074**	**0.4318**	0.6827	**0.3892**	**0.3841**	**0.2381**	**0.5297**	**0.7981**

Method	Shanghai Radar					CIKM				
	↑CSI-M	↑CSI-35	↑CSI-40	↑HSS	↑SSIM	↑CSI-M	↑CSI-35	↑CSI-40	↑HSS	↑SSIM
MAU (2021)	0.3983	0.3621	0.2417	0.5346	0.7195	0.3039	0.2054	0.1241	0.3928	0.6325
SimVP (2022)	0.3850	0.3549	0.2382	0.5194	0.7795	0.3052	0.2044	0.1321	0.3955	0.6538
FourCastNet (2022)	0.3571	0.3108	0.2073	0.4868	0.5598	0.2980	0.1849	0.1015	0.3801	0.4359
Earthformer (2022)	0.3503	0.3178	0.1872	0.4844	0.7298	0.3077	0.2039	0.1369	0.4001	0.6267
PhyDNet (2020)	0.3654	0.3236	0.2176	0.4957	0.7751	0.3038	0.2052	0.1287	0.3931	0.6541
EarthFarseer (2024)	0.3926	0.3608	0.2343	0.5330	0.5405	0.3000	0.2046	0.1259	0.3911	0.6373
NowcastNet (2023)	0.3953	0.3608	0.2450	0.5334	0.7902	0.2991	0.1940	0.1188	0.3865	0.6713
PreDiff (2024)	0.3504	0.3026	0.1855	0.4991	0.6981	0.2842	0.1848	0.1002	0.3481	0.5582
DiffCast (2024)	0.4089	0.3740	0.2606	0.5476	0.7879	0.3159	0.2009	0.1457	0.4085	0.6499
CasCast (2024)	0.3651	0.3291	0.2192	0.4971	0.7250	0.3021	0.1974	0.1142	0.3806	0.5241
FACL (2024)	0.3940	0.3511	0.2355	0.5308	0.7601	0.3105	0.1993	0.1354	0.3996	0.6442
AlphaPre (2025)	0.4178	0.3854	0.2615	0.5534	**0.7951**	**0.3194**	0.2068	0.1416	0.4137	0.6568
Ours	**0.4252**	**0.3933**	**0.2821**	**0.5643**	0.7788	0.3179	**0.2275**	**0.1562**	**0.4217**	**0.6659**

Table 2: Quantitative comparison across different methods, datasets and evaluation metrics.

incorrect reduction in predicted rainfall. Although DiffCast performs well in capturing both trend evolution and micro-scale variability in first a few frames, it cannot preserve high-frequency components, leading to overlooked regions on the map. Specifically, at both 50 and 100 minutes prediction, it misses the rainfall in the top-left and bottom-right regions. In contrast, *DuoCast* effectively captures both the overall trend of weather patterns and the finer micro-scale variability, with edge regions particularly catered.

Ablation Study

Low-Frequency Diffusion Model. To validate the effectiveness of the proposed mechanisms in *DuoCast*, we compare our low-frequency model with the baseline DiffCast. As shown in Table 1, our model achieves higher CSI and HSS scores, demonstrating strong forecasting performance even with low-frequency modeling alone. Qualitatively, Fig. 4(a) shows a cold front event where central high-precipitation regions expand over time. While the ground truth reflects this growth, DiffCast fails to capture the trend, predicting diminishing intensity. In contrast, our model accurately forecasts the expanding region, consistent with input observations. Fig. 4(b) presents a warm front event where the central yellow precipitation area gradually dissipates. Again, DiffCast incorrectly predicts intensified rainfall, whereas our model captures the decreasing trend more faithfully. These results show that the low-frequency model effectively captures large-scale precipitation by leveraging weather front dynamics. Additionally, we explore using a self-attention backbone, but it leads to a performance drop, because self-attention emphasizes inter-token fine-grained patterns making it less effective at preserving low-frequency structures.

High-Frequency Diffusion Model. As shown in Table 1, removing the high-frequency diffusion model leads to a notable drop in CSI, HSS, and SSIM, highlighting its contribution to forecasting accuracy. Fig. 4(c) illustrates this effect: without the high-frequency component, the predicted precipitation band in the top-right region becomes increasingly blurred and blocky over time, especially between 80 and 100 minutes, missing fine-grained intensity variations. In contrast, incorporating the high-frequency model yields sharper predictions with more accurate spatial extent and intensity across multiple precipitation levels. These results demonstrate that the high-frequency diffusion process effectively refines coarse low-frequency outputs, enhancing micro-scale variability over extended lead times.

Air Mass Modeling. As shown in Table 1, removing air mass spatial modeling leads to a notable drop in CSI, particularly CSI-M, indicating its importance in capturing the spatial evolution of weather fronts. Additionally, we analyze the effectiveness of convolution in air mass spatial modeling. As shown in the Fig. 5, Conv_{I}, captures the spatial dynamics of surrounding air mass effects, highlighted by the heatmap. The frontal convolution, Conv_{II}, focuses on occluded fronts, which occur in critical transition areas where cold and warm

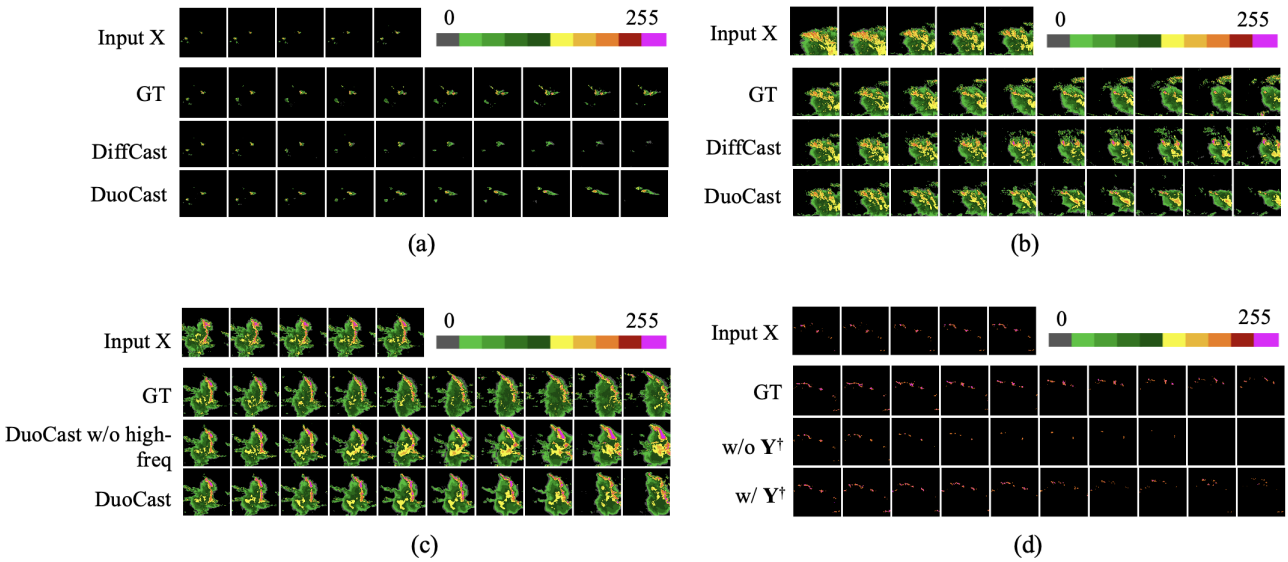

Figure 4: Ablation study with qualitative SEVIR examples. (a) and (b) demonstrate the effectiveness of *DuoCast* in leveraging weather fronts, specifically cold and warm fronts, respectively. (c) highlights the capability of the high-frequency model in refining micro-scale variability. (d) showcases the impact of $\mathbf{Y}^{\dagger}$ for high-intensity regions prediction.

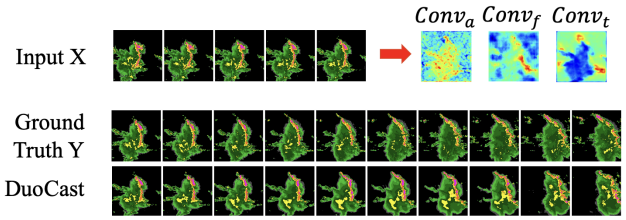

Figure 5: Visualization of air mass spatial modeling, highlighting captured weather fronts.

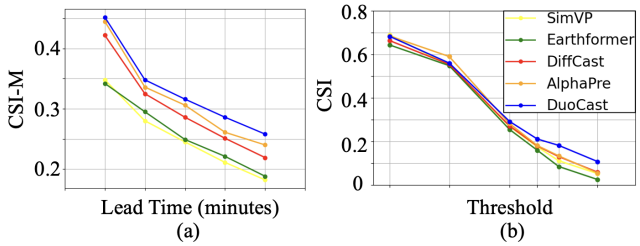

Figure 6: CSI across different (a) lead time steps and (b) intensity thresholds for different methods on SEVIR.

air masses interact, as depicted by the figure's red band. The temporal convection convolution, Conv_{III}, captures pixel-level temporal convection, with the red areas indicating potential temporal motion. Quantitatively, as in Table 1, every component in air mass spatial modeling contributes to superior performance.

Weather Front Modeling. Similarly, omitting weather front temporal modeling results in decreased CSI and HSS, reflecting its role in preserving temporal consistency. Additional ablation results for the temporal modeling component are provided in the supplementary material. These findings underscore the necessity of incorporating temporal front modeling for accurate precipitation forecasting.

$\mathbf{Y}^{\dagger}$ for High-Intensity Precipitation To assess the role of $\mathbf{Y}^{\dagger}$ in modeling high-intensity precipitation, we conduct targeted experiments on SEVIR. As shown in Table 1, removing $\mathbf{Y}^{\dagger}$ results in substantial performance drops for severe rainfall events, with CSI-181 and CSI-219 decreasing by 10.8% and 27.0%, respectively. Figure 4(d) visualizes high-intensity events (values ≥ 181), where predictions with $\mathbf{Y}^{\dagger}$ align closely with the ground truth, consistently maintaining the precipitation band over time. In contrast, models without $\mathbf{Y}^{\dagger}$ increasingly miss key regions, and the predicted band degrades notably after 30 minutes. These results highlight the importance of $\mathbf{Y}^{\dagger}$ in learning temporal consistency and structure in severe precipitation. Figure 6(b) shows consistently higher CSI across thresholds, confirming our model's strength in forecasting high-intensity rainfall.

Conclusion

We introduce *DuoCast*—a dual-probabilistic meteorology-aware model designed for preserving different frequency information. Low-frequency aims to capture the broader trend of weather patterns and high-frequency is to obtain high frequency micro-scale variability relevant to precipitation. Comprehensive experiments on four real-world datasets validate the effectiveness of the proposed framework.

Acknowledgments

This work was supported by the Australian Research Council (ARC) Linkage Project #LP230100294 and ECU Science Early Career and New Staff Grant Scheme.

References

Alley, R. B.; Emanuel, K. A.; and Zhang, F. 2019. Advances in weather prediction. *Science*, 363(6425): 342–344.

Andrychowicz, M.; Espeholt, L.; Li, D.; Merchant, S.; Merose, A.; Zyda, F.; Agrawal, S.; and Kalchbrenner, N. 2023. Deep learning for day forecasts from sparse observations. *arXiv preprint arXiv:2306.06079*.

Bauer, P.; Thorpe, A.; and Brunet, G. 2015. The quiet revolution of numerical weather prediction. *Nature*, 525(7567): 47–55.

Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; and Tian, Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970): 533–538.

Catto, J.; Jakob, C.; Berry, G.; and Nicholls, N. 2012. Relating global precipitation to atmospheric fronts. *Geophysical Research Letters*, 39(10).

Chen, H.; Feng, L.; Wu, W.; Zhu, X.; Leo, S.; and Hu, K. 2025. F2Net: A Frequency-Fused Network for Ultra-High Resolution Remote Sensing Segmentation. *arXiv preprint arXiv:2506.07847*.

Chen, L.; Cao, Y.; Ma, L.; and Zhang, J. 2020. A deep learning-based methodology for precipitation nowcasting with radar. *Earth and Space Science*, 7(2): e2019EA000812.

Doloriel, C. T.; and Cheung, N.-M. 2024. Frequency masking for universal deepfake detection. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 13466–13470. IEEE.

Gao, Z.; Shi, X.; Han, B.; Wang, H.; Jin, X.; Maddix, D.; Zhu, Y.; Li, M.; and Wang, Y. B. 2024. Prediff: Precipitation nowcasting with latent diffusion models. *Advances in Neural Information Processing Systems*, 36.

Gao, Z.; Tan, C.; Wu, L.; and Li, S. Z. 2022. Simvp: Simpler yet better video prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3170–3180.

Gong, J.; Bai, L.; Ye, P.; Xu, W.; Liu, N.; Dai, J.; Yang, X.; and Ouyang, W. 2024. CasCast: Skillful High-resolution Precipitation Nowcasting via Cascaded Modelling. *International Conference on Machine Learning*.

Guen, V. L.; and Thome, N. 2020. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11474–11484.

Kingma, D. P.; and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Lagerquist, R.; Mcgovern, A.; and Gagne II, D. J. 2019. Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting*, 34(4): 1137–1160.

Larvor, G.; and Berthomier, L. 2021. Meteonet: An open reference weather dataset for ai by météo-france. In *American Meteorological Society Meeting Abstracts*, volume 101, 1–ii.

Lin, K.; Zhang, B.; Yu, D.; Feng, W.; Chen, S.; Gao, F.; Li, X.; and Ye, Y. 2025. AlphaPre: Amplitude-Phase Disentanglement Model for Precipitation Nowcasting. In *Computer Vision and Pattern Recognition Conference (CVPR)*, 17841–17850.

Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474): 1177–1194.

Lu, Z.; Hu, K.; Wang, C.; Bai, L.; and Wang, Z. 2024. Autoregressive omni-aware outpainting for open-vocabulary 360-degree image generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 14211–14219.

Moore, B.; Bartlett, S.; Cordeira, J.; and Kalansky, J. 2025. CW3E Event Summary: Early July 2025 Central Texas Floods. https://cw3e.ucsd.edu/cw3e-event-summary-early-july-2025-central-texas-floods/. Center for Western Weather and Water Extremes (CW3E), Scripps Institution of Oceanography, UC San Diego. Accessed 2025-11-09.

Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878): 672–677.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28.

Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-k.; and Woo, W.-c. 2017. Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in Neural Information Processing Systems*, 30.

Skamarock, W. C.; Klemp, J.; Dudhia, J.; Gill, D. O.; Barker, D. M.; Duda, M. G.; Huang, X.-Y.; Wang, W.; and Powers, J. G. 2008. A description of the advanced research WRF. *National Center for Atmospheric Research*, 3.

Tan, C.; Zhao, Y.; Wei, S.; Gu, G.; Liu, P.; and Wei, Y. 2024. Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning. In *the AAAI Conference on Artificial Intelligence*, volume 38, 5052–5060.

Veillette, M.; Samsi, S.; and Mattioli, C. 2020. SEVIR: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, 33: 22009–22019.

Wen, P.; Hu, K.; Yua, D.; Ning, Z.; Li, C.; and Wang, Z. 2024. Radio Frequency Signal based Human Silhouette Segmentation: A Sequential Diffusion Approach. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE.

Wen, P.; Hu, K.; Yue, W.; Zhang, S.; Zhou, W.; and Wang, Z. 2023. Robust audio anti-spoofing with fusion-reconstruction learning on multi-order spectrograms. *Interspeech 2023*.

Woo, S.; et al. 2022. Add: Frequency attention and multi-view based knowledge distillation to detect low-quality compressed deepfake images. In *the AAAI conference on artificial intelligence*, volume 36, 122–130.

Yan, C.-W.; Foo, S. Q.; Trinh, V. H.; Yeung, D.-Y.; Wong, K.-H.; and Wong, W.-K. 2024. Fourier Amplitude and Correlation Loss: Beyond Using L2 Loss for Skillful Precipitation Nowcasting. In *NeurIPS*.

Yang, X.; Zhou, D.; Feng, J.; and Wang, X. 2023. Diffusion probabilistic model made slim. In *IEEE/CVF Conference on computer vision and pattern recognition*, 22552–22562.

Yu, D.; Li, X.; Ye, Y.; Zhang, B.; Luo, C.; Dai, K.; Wang, R.; and Chen, X. 2024. Diffcast: A unified framework via residual diffusion for precipitation nowcasting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27758–27767.

Zhang, Y.; Long, M.; Chen, K.; Xing, L.; Jin, R.; Jordan, M. I.; and Wang, J. 2023. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619(7970): 526–532.

Zhou, J.; Li, Y.; Wu, B.; Li, B.; Dong, J.; et al. 2024. Freqblender: Enhancing deepfake detection by blending frequency knowledge. *Advances in Neural Information Processing Systems*, 37: 44965–44988.