FoumCast: Adaptive Fourier-Meteorological Disentanglement and Diffusion for High-Fidelity Precipitation Nowcasting

1st Yangyang Xu
Hefei Institutes of Physical Science
University of Science and Technology of China
Hefei, China
https://orcid.org/0009-0007-3679-4390

Xianwei Meng*
Hefei Institutes of Physical Science
Hefei, China
https://orcid.org/0000-0001-8253-1785

2nd Lin Jia
Hefei Institutes of Physical Science
Hefei, China
https://orcid.org/0009-0009-2055-9235

***Abstract*—Accurate precipitation nowcasting is critical for disaster mitigation and urban resource management. However, existing deep learning models often struggle to balance spatial precision with temporal consistency, typically yielding over-smoothed predictions that lack fine-grained meteorological details. To address these challenges, we propose FoumCast, a novel frequency-domain guided framework that synergizes adaptive spectral-spatial disentanglement with conditional diffusion modeling. In the first stage, a deterministic component is optimized via the Fourier Amplitude and Correlation Loss. By explicitly supervising both spectral magnitude and phase, FoumCast accurately preserves rainfall intensity and global spatial positioning. A learnable spectral mask is introduced to adaptively partition the preliminary forecasts into low-frequency advective envelopes (representing macro-scale evolution) and high-frequency convective textures (capturing micro-scale dynamics), which are then integrated through a Cross-Attention Fusion Module for multi-scale temporal alignment. In the second stage, a conditional diffusion network leverages these spectral-aware features as priors to model residual uncertainties. Extensive evaluations on four benchmark radar datasets, demonstrate that FoumCast outperforms state-of-the-art baselines in terms of predictive accuracy, structural realism, and physical consistency.**

***Index Terms*—Precipitation Nowcasting, Frequency-Domain Learning, Spectral-Spatial Disentanglement, Conditional Diffusion Model**

I. Introduction

Precipitation, appearing in diverse forms such as rainfall, snowfall, and hail, plays a fundamental role in regulating atmospheric conditions including temperature and humidity. Precipitation nowcasting, which predicts short-term rainfall evolution (0–6 hours) from historical radar observations, has become indispensable for applications in agriculture, transportation, and disaster risk management [1].

Traditional numerical weather prediction (NWP) methods formulate atmospheric dynamics as partial differential equations (PDEs) and integrate them through large-scale numerical simulations [2]–[5]. However, due to the chaotic and nonlinear nature of precipitation systems, NWP-based approaches often suffer from high computational costs and limited accuracy. They struggle to capture fine-scale variability, tend to produce overly smoothed forecasts at longer lead times, and frequently underestimate extreme precipitation events—hindering their reliability in practical applications.

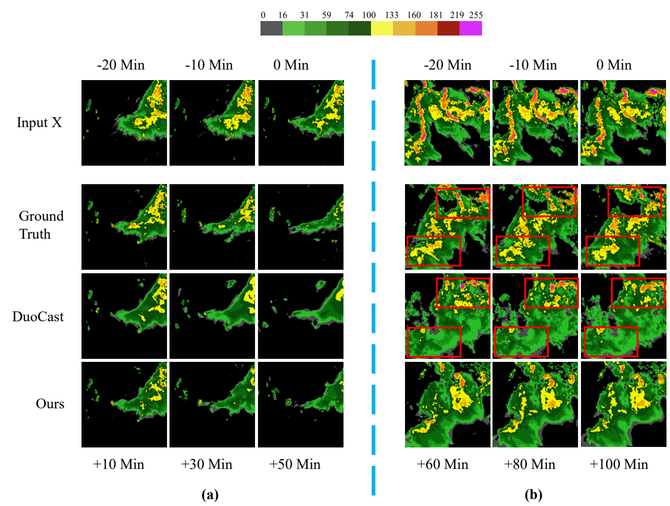

Fig. 1. We evaluate the performance of two models on precipitation forecasting. Figure (a) illustrates that our model provides more detailed rainfall structures in regions of small-scale precipitation, while Figure (b) demonstrates that, for large-scale rainfall events, our model achieves superior performance in both rainband localization and rainfall intensity prediction.

In recent years, data-driven deep learning methods have shown great promise for precipitation prediction, overcoming several limitations of traditional NWP models. By leveraging large-scale radar datasets, these approaches bypass explicit physical modeling and directly learn spatiotemporal rainfall dynamics. Deterministic models [6]–[9] capture the overall motion of precipitation systems but often produce blurry forecasts and lose fine-scale details as the prediction horizon extends [7], [10]. In contrast, probabilistic models [11], [12] introduce stochastic sampling to better represent local meteorological variability, yet treating the entire system as random often leads to uncontrollable uncertainty and reduced accuracy [13].

Notably, the most recent approaches combine deterministic and probabilistic models, achieving notable success in precipitation nowcasting. For instance, CasCast [13] employs a cascaded framework to simultaneously predict the evolution of mesoscale and small-scale precipitation systems, while DiffCast [14] decomposes the weather dynamics into global deterministic motion and local probabilistic variations. In essence, these methods refine the coarse forecasts generated by deterministic models with probabilistic components to enhance prediction accuracy. However, they generally overlook precipitation-specific characteristics. Certain meteorological factors play a decisive role in shaping rainfall intensity and duration. For example, convective lifting often produces localized heavy rainfall that is intense but short-lived, whereas frontal lifting—caused by the interaction of warm and cold air masses—gives rise to distinct types of precipitation [15]. Although DuoCast [16] accounts for these meteorological pattern transitions, it fails to disentangle high- and low-frequency components, resulting in blurred predictions along rainband boundaries and at the interfaces between different rainfall intensities, as shown in Fig. 1(a). Moreover, DuoCast struggles to precisely model the intrinsic correspondence between amplitude and rainfall intensity, as well as between phase and rainfall location, in the frequency domain. This leads to biased rainband positioning and the loss of areas with strong precipitation, as shown in Fig. 1(b).

Accurate precipitation nowcasting remains a challenging task due to highly nonlinear dynamics and multiscale spatial dependencies. Existing deep learning approaches exhibit a clear trade-off: deterministic models often produce over-smoothed forecasts with blurred convective details, while diffusion-based models may suffer from stochastic instability and spatiotemporal inconsistency. To address this issue, we propose FoumCast, a two-stage framework that combines frequency-domain structural modeling with conditional diffusion refinement. FoumCast is motivated by the observation that precipitation can be decomposed into stable large-scale advective patterns and transient small-scale convective structures. In the first stage, a deterministic backbone is trained with a Fourier Amplitude and Correlation Loss (FACL), which constrains both spectral magnitude and phase to better preserve rainfall intensity and global spatial alignment. The preliminary forecasts are then decomposed into low- and high-frequency components, processed by specialized convolutional blocks for large-scale motion and fine-grained texture, and fused through a Cross-Attention Fusion Mechanism (CAFM). In the second stage, a Conditional Diffusion Module further refines the predictions by modeling residual uncertainty under the guidance of spectral priors. This design improves visual realism while maintaining temporal coherence. Experiments on four benchmark datasets—SEVIR, MeteoNet, Shanghai Radar, and CIKM—show that FoumCast consistently outperforms state-of-the-art methods.

The main contributions of this work are summarized as follows:

- We design FoumCast, a novel dual-stage model that integrates frequency-domain learning and conditional diffusion. This framework effectively balances global rainfall motion modeling with fine-grained meteorological detail refinement.
- We improve forecasting accuracy by incorporating the Fourier Amplitude and Correlation Loss together with frequency disentanglement between low- and high-frequency components.
- Extensive evaluations across four benchmark radar datasets verify that FoumCast outperforms existing methods, yielding superior CSI, HSS, and SSIM scores and generating precipitation maps that are both physically coherent and visually detailed.

II. Related Work

A. Deep Learning Models for Precipitation Nowcasting

Precipitation nowcasting methods are broadly categorized into deterministic and probabilistic approaches. Deterministic models aim to capture the overall motion of precipitation systems. Early recurrent-based models like ConvLSTM [8] and ConvGRU [9] integrated convolutions into recurrent units, followed by extensions like PredRNN [17], MAU [18], and PhyDNet [10] to enhance temporal modeling. Recurrent-free designs such as SimVP [7] and Earthformer [6] further improved efficiency. However, these models often produce blurry forecasts due to the regression-to-the-mean effect of pixel-wise losses. Probabilistic models address this by capturing stochasticity. DGMR [11] and NowcastNet [1] utilize GANs for realism, while PreDiff [12] and DiffCast [14] employ diffusion processes to refine forecasts or model local variations. Despite their realism, these methods face challenges like training instability, high sampling costs, and excessive randomness that may degrade accuracy.

B. Fourier Transform Applications in Image Processing

Spectral analysis reveals that DNNs exhibit a spectral bias toward low-frequency functions [19], [20], motivating Fourier-based losses to preserve fine details. Examples include Frequency Domain Perceptual Loss [21], Focal Frequency Loss [22], and Fourier Space Loss [23]. Furthermore, amplitude–phase disentanglement has gained traction, as amplitude represents intensity while phase encodes structure. This paradigm has been applied to image enhancement [24], robustness [25], [26], and domain generalization [27]–[30]. Notably, FACL [31] introduced amplitude and correlation losses to improve spatial clarity in nowcasting, providing a foundation for our frequency-aware approach.

III. Method

A. Task Formulation

In line with previous studies [6], [8], [12], [14], we cast precipitation nowcasting as a spatio-temporal forecasting task of radar echoes. Specifically, given a sequence of past radar observations $\mathbf{X} = \{\mathbf{X}_t\}_{t=1}^{T_{\text{in}}} \in \mathbb{R}^{T_{\text{in}} \times C \times H \times W}$, where T_{in} denote the number of historical frames, H and W represent the spatial resolution, and C is the number of channels ($C = 1$ in radar echo data, corresponding to echo intensity), the objective is to predict the future sequence $\mathbf{Y} = \{\mathbf{Y}_t\}_{t=1}^{T_{\text{out}}} \in \mathbb{R}^{T_{\text{out}} \times C \times H \times W}$. Formally, the goal of precipitation nowcasting is to model the conditional distribution $p(\mathbf{Y}|\mathbf{X})$, thereby generating the most likely evolution of future radar echoes over the next T_{out} times steps.

B. Preliminary: Fourier Spectral Decomposition

Precipitation patterns are characterized by multi-scale properties: large-scale advection (low-frequency) governs the general movement of rainbands, while small-scale convection (high-frequency) dictates local intensity fluctuations. The 2D Discrete Fourier Transform (DFT) provides a theoretical framework to decouple these components in the frequency domain. For a precipitation field $\mathbf{Z} \in \mathbb{R}^{H \times W}$, its frequency representation $\mathcal{F}(\mathbf{Z})$ is defined as:

$$\mathcal{F}(\mathbf{Z})(u,v) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \mathbf{Z}(h,w) e^{-i2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}, \tag{1}$$

where (u, v) coordinates the spectral space. In this domain, the amplitude $A(u, v) = |\mathcal{F}(\mathbf{Z})|$ reflects the energy distribution corresponding to precipitation intensity, while the phase $\phi(u, v) = \angle\mathcal{F}(\mathbf{Z})$ encodes the spatial structure and arrangement. In spectral analysis, an optical shift is typically employed to centralize the zero-frequency component at $(u_0, v_0) = (H/2, W/2)$. Within this centralized spectrum, a low-pass filter M partitions the signal based on a cutoff radius D_c:

$$\begin{aligned} &\mathbf{Z}_{\text{low}} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{Z}) \odot M)\,, \\ &M(u,v) = \begin{cases} 1, & \sqrt{(u-u_0)^2 + (v-v_0)^2} \le D_c \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

where $\odot$ denotes the Hadamard product. Consequently, the high-frequency component is represented by the residual $\mathbf{Z}_{\text{high}} = \mathbf{Z} - \mathbf{Z}_{\text{low}}$. This decomposition facilitates the separate modeling of stable macro-scale atmospheric structures and stochastic micro-scale variations.

C. The Proposed Model

The proposed FoumCast is a two-stage framework integrating frequency-domain optimization with conditional diffusion modeling. An overview of the framework is illustrated in Fig. 2. In the first stage, a Preliminary Forecaster utilizes a deterministic backbone optimized by FACL to generate initial predictions. These forecasts undergo spectral disentanglement into low- and high-frequency components, which are then fused via cross-attention to produce robust spectral representations. In the second stage, a Conditional Diffusion Network takes the preliminary prediction as a condition to model residual uncertainties and refine fine-grained rainfall structures. By decoupling frequency-scale dynamics, FoumCast enhances control over both spatial arrangement and intensity precision, yielding high-fidelity precipitation forecasts.

D. Preliminary Forecaster

The Preliminary Forecaster focuses on learning a deterministic, physically interpretable mapping from past observations $\mathbf{X}$ to future precipitation patterns. To maintain meteorological coherence and spectral fidelity, this stage operates across three modules: a Fourier-aware deterministic component, a spectral decomposition unit, and a cross-attention fusion mechanism.

Fourier-Aware Deterministic Component. To preserve both spatial structural coherence and spectral sharpness, we integrate the Fourier Amplitude and Correlation Loss (FACL) into the training of the deterministic component $\mathcal{P}(\cdot)$ (e.g. SimVP). FACL consists of two terms: Fourier Amplitude Loss (FAL) for spectral magnitude alignment and Fourier Correlation Loss (FCL) for phase and global structural consistency:

$$\begin{aligned} \mathcal{L}_{\text{FAL}}(\mathbf{Y}, \hat{\mathbf{Y}}) &= \Big\| \, |\mathcal{F}(\hat{\mathbf{Y}})| - |\mathcal{F}(\mathbf{Y})| \, \Big\|_2^2, \\ \mathcal{L}_{\text{FCL}}(\mathbf{Y}, \hat{\mathbf{Y}}) &= 1 - \frac{\sum \text{Re}\Big[\mathcal{F}(\hat{\mathbf{Y}}) \cdot \overline{\mathcal{F}(\mathbf{Y})}\Big]}{\sqrt{\sum |\mathcal{F}(\hat{\mathbf{Y}})|^2 \, \sum |\mathcal{F}(\mathbf{Y})|^2}}. \end{aligned} \tag{3}$$

where $\mathcal{F}(\cdot)$ denotes the 2D FFT. To balance structural stability and detail enhancement, we adopt a scheduling strategy:

$$\mathcal{L}_{\text{FACL}}(\mathbf{Y}, \hat{\mathbf{Y}}, t) = \begin{cases} \mathcal{L}_{\text{FAL}}(\mathbf{Y}, \hat{\mathbf{Y}}), & \text{if } p > P(t) \\ \mathcal{L}_{\text{FCL}}(\mathbf{Y}, \hat{\mathbf{Y}}), & \text{otherwise} \end{cases} \tag{4}$$

where $P(t)$ gradually decreases from 1 to 0 during training, shifting the focus from structural consistency to high-frequency refinement. The preliminary output is given by $\mathbf{Y}^{\text{pre}} = \mathcal{P}(\mathbf{X})$.

Adaptive Spectral Disentanglement and Dual-Path Spectral Encoder. To address the limitations of empirical thresholding, we propose an adaptive spectral disentanglement mechanism that dynamically partitions frequency components based on real-time atmospheric scale characteristics. Given the frequency representation $\mathbf{Y}^{\text{pre}}$, we implement a Learnable Spectral Mask ($\mathcal{M}_{\boldsymbol{\theta}}$), optimized during training to identify the transition between stable advection and stochastic convection:

$$\begin{aligned} \mathbf{Y}^{\text{pre}}_{\text{low}} &= \mathcal{F}^{-1}(\mathcal{F}(\mathbf{Y}^{\text{pre}}) \odot \mathcal{M}_{\boldsymbol{\theta}})\,, \\ \mathbf{Y}^{\text{pre}}_{\text{high}} &= \mathbf{Y}^{\text{pre}} - \mathbf{Y}^{\text{pre}}_{\text{low}}. \end{aligned} \tag{5}$$

Unlike fixed truncation, $\mathcal{M}_{\boldsymbol{\theta}}$ is parameterized by a lightweight global context encoder that adjusts the soft-cutoff radius D_c according to the input energy distribution. This ensures that the disentanglement remains robust across different weather regimes (e.g. widespread stratiform rain vs. localized convective cells). Following spectral decomposition, the components are processed via a Dual-Path Spectral Encoder. To ensure robust feature learning under varying batch sizes and nonlinear dynamics, each path employs a three-layer structure equipped with Group Normalization (GN) and SiLU activation. We define a fundamental convolutional block $\Psi_k^d(\cdot)$ as the sequence:

$$\Psi_k^d(\cdot) = \text{SiLU}(\text{GN}(\text{Conv}^d_{k \times k}(\cdot))), \tag{6}$$

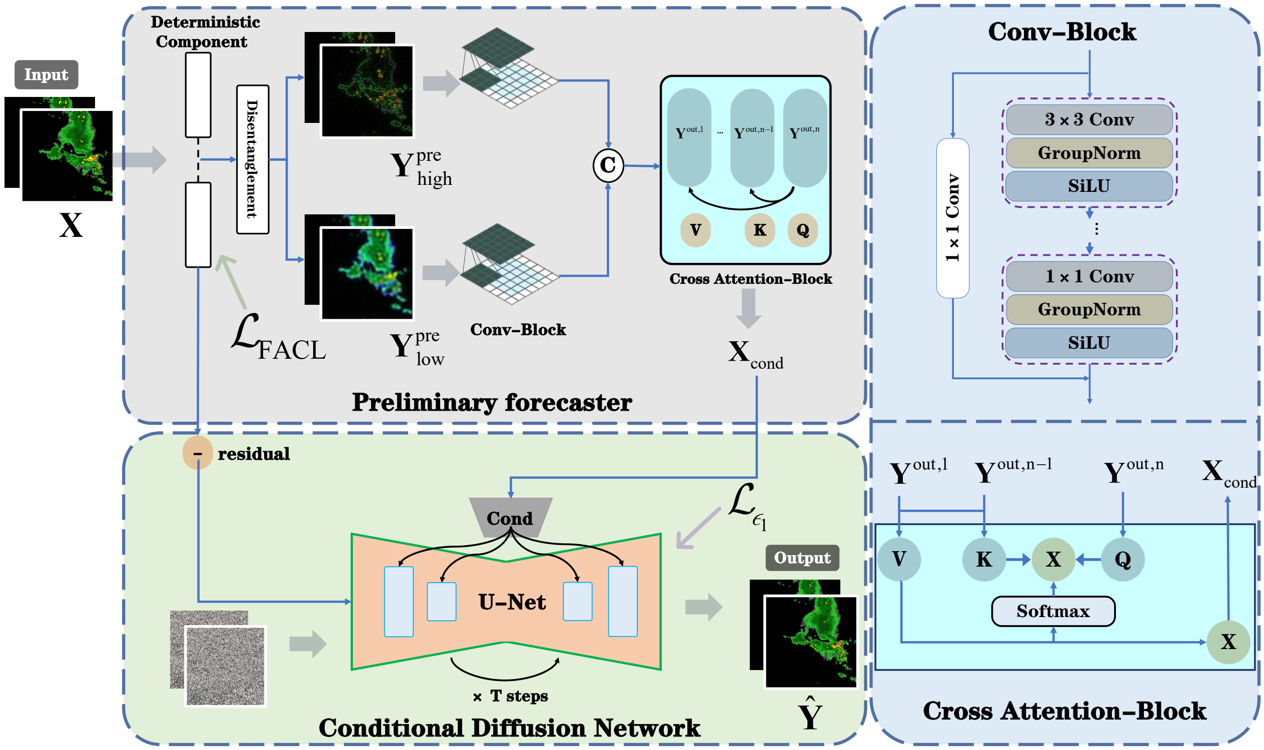

Fig. 2. Overview of the proposed FoumCast framework. FoumCast consists of a Fourier-aware deterministic forecaster and a conditional diffusion network. The first stage employs FACL loss–guided training and spectral disentanglement to separate high- and low-frequency rainfall components, which are refined by Conv-Blocks and fused through a cross-attention mechanism for temporal–spectral alignment. The second stage uses a conditional diffusion model to reconstruct fine-scale rainfall details and improve prediction fidelity.

where k denotes the kernel size and d is the dilation rate. The dual-path processing is formulated as a Context-Extraction-Projection cascade:

$$\begin{aligned}\mathbf{F}_{\text{low}} &= \text{Conv}_{1\times1}(\mathbf{Y}^{\text{pre}}_{\text{low}}) + \Psi_{1\times1}(\Psi^{d=2}_{5\times5}(\Psi_{3\times3}(\mathbf{Y}^{\text{pre}}_{\text{low}}))),\\ \mathbf{F}_{\text{high}} &= \text{Conv}_{1\times1}(\mathbf{Y}^{\text{pre}}_{\text{high}}) + \Psi_{1\times1}(\Psi_{3\times3}(\Psi_{3\times3}(\mathbf{Y}^{\text{pre}}_{\text{high}}))).\end{aligned} \tag{7}$$

To further enhance the feature representation, a residual connection is added, where the block's input is passed through a 1×1 convolution and then combined with the block's representation, allowing the network to capture and incorporate residual information. Then, the output is given by $\mathbf{Y}^{\text{out}} = \text{Concat}(\mathbf{F}_{\text{low}}, \mathbf{F}_{\text{high}})$, which provides a spectrally enriched conditioning representation for the subsequent diffusion process.

Cross-Attention Fusion for Spatio-Temporal and Spectral Alignment. To capture the complex temporal dependencies and multiscale spectral correlations inherent in precipitation evolution, we propose a Cross-Attention Fusion Module (CAFM). The CAFM is designed to align and synchronize spatial-spectral representations across varying temporal contexts. By performing non-local feature interactions, the module mitigates phase misalignment and preserves physical consistency between successive frames, thereby enhancing the temporal coherence of the forecast. CAFM constructs a cross-attention mechanism by leveraging multi-temporal cues to bridge historical trajectories with future states. The Query ($\mathbf{Q}$) is projected from the hidden state of the current prediction, while the Key ($\mathbf{K}$) and Value ($\mathbf{V}$) are generated from a combination of the most recent observation and initial historical frames. This configuration allows the model to adaptively weight historical motion priors and spectral contexts, ensuring that the synthesized rainfall fields remain spatially aligned and temporally synchronized with the underlying meteorological flow. Specifically:

$$\begin{aligned}\mathbf{Q} &= \mathbf{W}_Q \circledast \mathbf{Y}^{\text{out,n}},\\ \mathbf{K} &= \mathbf{W}_K \circledast \text{Concat}(\mathbf{Y}^{\text{out,n}-1}, \mathbf{Y}^{\text{out,1}}),\\ \mathbf{V} &= \mathbf{W}_V \circledast \text{Concat}(\mathbf{Y}^{\text{out,n}-1}, \mathbf{Y}^{\text{out,1}}),\\ \mathbf{X}_{\text{cond}} &= \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}.\end{aligned} \tag{8}$$

where $\circledast$ represents convolution operator, and d represents the dimension of the embeddings. This design enables the model to align temporal dynamics across different stages of rainfall evolution while maintaining coherence between spectral components.

E. Conditional Diffusion Network

While providing a robust baseline for rainfall evolution, the deterministic stage often misses the stochastic, multi-scale nuances of precipitation. FoumCast addresses this by employing a conditional diffusion network to model the residual error between preliminary forecasts and ground-truth fields.

This generative refinement constrains the stochastic sampling process with spectral-aware priors, effectively restoring local spatial fidelity and resolving the pervasive over-smoothing issues of traditional regression-based methods.

We follow the stochastic residual prediction mechanism [14] to compute the residual $\mathbf{R} = \mathbf{Y} - \mathbf{Y}^{\text{pre}}$. During the forward diffusion process, Gaussian noise is progressively added to $\mathbf{R}$ following a Markov chain:

$$\begin{aligned} q(\mathbf{R}^t \mid \mathbf{R}^{t-1}) &= \mathcal{N}\Big(\sqrt{1-\beta_t}\,\mathbf{R}^{t-1}, \beta_t \mathbf{I}\Big), \\ q(\mathbf{R}^t \mid \mathbf{R}^{0}) &= \mathcal{N}\Big(\sqrt{\hat{\alpha}_t}\,\mathbf{R}^{0}, (1-\hat{\alpha}_t)\,\mathbf{I}\Big). \end{aligned} \tag{9}$$

where β_t controls the variance schedule of noise injection, and the timestep $t \in T$. Using the notation $\alpha_t = 1 - \beta_t$ and $\hat{\alpha}_t = \prod_{s=0}^{t} \alpha_s$. During the reverse process, we adopt an autoregressive forecasting manner across both the diffusion step t and the forecast frame index l. At each temporal index l, the target frame $\mathbf{R}_l$ is predicted conditionally based on the preceding frame representation $\mathbf{R}_{l-1}$ and its conditional input $\mathbf{X}_{\text{cond},l}$. Since the $\mathbf{R}_{l-1}$ is unavailable during inference, we instead use its estimation $\hat{\mathbf{R}}_{l-1}$ obtained from the previous generation step to guide the current frame prediction. Formally, this process can be expressed as:

$$p_\theta(\mathbf{R}_l^{t-1} \mid \mathbf{R}_l^{t}, \hat{\mathbf{R}}_{l-1}, \mathbf{X}_{\text{cond},l-1}), \tag{10}$$

This conditional input encodes large-scale precipitation evolution and provides a physical prior that guides the stochastic refinement process. Finally, the model is optimized via a reparameterized objective following the denoising score-matching principle:

$$\mathcal{L}_\epsilon = \mathbb{E}_{\mathbf{R}_l^t, \boldsymbol{\epsilon}, t}\left[\left\|\boldsymbol{\epsilon} - \epsilon_\theta(\mathbf{R}_l^t, t, \mathbf{X}_{\text{cond},l-1}, \hat{\mathbf{R}}_{l-1})\right\|_2^2\right], \tag{11}$$

where ϵ_θ denotes the network's estimate of the injected Gaussian noise, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The network learns to predict the added noise conditioned on both the noised input and the preliminary context. Finally, repeating this L_{out} times leads us to an estimated residual sequence $\hat{\mathbf{R}} = [\hat{\mathbf{R}}_l]_{l=1}^{L_{out}}$. Specifically, $\mathbf{R}_0$ is initialized to 0. Then, we compute the final prediction result $\hat{\mathbf{Y}}$ as:

$$\hat{\mathbf{Y}} = \hat{\mathbf{R}} + \mathbf{Y}^{\text{pre}}, \tag{12}$$

F. Training Strategy

FoumCast is trained in an end-to-end joint optimization manner. While the adaptive spectral mask $\mathcal{M}_{\boldsymbol{\theta}}$ and dual-path encoder lack explicit auxiliary supervision, they are implicitly optimized via backpropagation from the total loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{FACL}} + \lambda_2 \mathcal{L}_\epsilon \tag{13}$$

During training, gradients from the deterministic component loss ($\mathcal{L}_{FACL}$) and the diffusion refinement loss ($\mathcal{L}_\epsilon$) flow back through the entire architecture. This allows the model to dynamically refine the frequency partition boundary and path-specific features to minimize phase mismatch and energy loss.

IV. Experiment

A. Experimental Setting

Datasets. We evaluate FoumCast on four diverse datasets: (1) SEVIR [32], a large-scale radar dataset from the U.S., using the standard 13,000 training and 2,000 testing sequences; (2) MeteoNet [33], a meteorological dataset from France; (3) Shanghai Radar [34] contains continuous radar echo observations obtained through volume scans in Pudong, Shanghai; and (4) CIKM [1] AnalytiCup 2017 dataset records precipitation events in Guangdong, China, over an area of 101 km$\times$101 km. All sequences are cropped to 128×128 pixels, with a lead time of 60 to 120 minutes.

Evaluation Metrics. Following [6], [12]–[14], we assess performance using skill scores and perceptual metrics. We report the Critical Success Index (CSI) and Heidke Skill Score (HSS) to evaluate categorical accuracy. Specifically, CSI is computed at multiple reflectivity thresholds: $\{16, 74, 133, 160, 181, 219\}$ for SEVIR; $\{12, 18, 24, 32\}$ for MeteoNet; and $\{20, 30, 35, 40\}$ for Shanghai and CIKM datasets. We also report the mean CSI (CSI-M) and pooled CSI (4×4, 16×16) to evaluate regional extreme precipitation capturing. For perceptual and visual fidelity, SSIM and LPIPS are employed to measure structural and distribution similarity between predictions and ground truth.

Baseline. To assess the quality of high-resolution precipitation forecasts generated by our FoumCast framework, we conduct comparisons against five deterministic models (ConvGRU [9], SimVP [7], Earthformer [6], MAU [18], and PhyDNet [10]) four probabilistic models (STRPM [35], and PreDiff [12]), DiffCast [14], as well as DuoCast [16], a hybrid approach integrating both probabilistic and deterministic components.

Implementation Details. FoumCast is implemented in PyTorch and trained on NVIDIA 4090 GPUs. Our framework is optimized for 20 epochs using Adam with a learning rate of 10^{-4} and batch size of 4. As for diffusion setting, we set the diffusion steps as 1000 and denoising steps for inference as 250 with DDIM. The learnable spectral mask $\mathcal{M}_{\boldsymbol{\theta}}$ is initialized with a cutoff frequency $D_c = 10$, which approximates a 12 km physical scale in our spatial domain. Loss weights for $\mathcal{L}_{\text{FACL}}$ and $\mathcal{L}_\epsilon$ are empirically set to 0.005 and 0.5, respectively.

B. Experimental Results

Quantitative Comparison. Table I summarizes the performance of FoumCast against strong baselines. FoumCast consistently achieves state-of-the-art or competitive results across four metrics. As Table II illustrated, our model yields substantial performance gains at high-reflectivity thresholds. Specifically, FoumCast outperforms the second-best model (DuoCast) by approximately 4.3% in CSI-219 on the SEVIR dataset. The improvement is even more pronounced on the Shanghai Radar dataset, where FoumCast achieves a CSI-M of 0.5167 and improves the CSI-40 score by an absolute 4.4%. These results indicate that FoumCast effectively mitigates the

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

TABLE I

COMPARISON ON SEVIR, METEONET, SHANGHAI RADAR AND CIKM DATASETS. POOL4 AND POOL16 INDICATE 4-GRID AGGREGATIONS AND 16-GRID AGGREGATIONS, RESPECTIVELY. ↑ MEANS HIGHER IS BETTER, ↓ MEANS LOWER IS BETTER. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD AND THE SECOND-BEST RESULTS ARE UNDERLINED.

Method	SEVIR						MeteoNet					
	↑CSI	↑CSI-pool4	↑CSI-pool16	↑HSS	↓LPIPS	↑SSIM	↑CSI	↑CSI-pool4	↑CSI-pool16	↑HSS	↓LPIPS	↑SSIM
ConvGRU	0.2416	0.2554	0.3050	0.2834	0.3766	0.6532	0.3400	0.3578	0.4473	0.4667	0.2950	0.7832
SimVP	0.2662	0.2844	0.3452	0.3369	0.3914	0.6304	0.3346	0.3383	0.4143	0.4568	0.3523	0.7557
Earthformer	0.2413	0.2601	0.2810	0.3073	0.4401	0.6773	0.3296	0.3428	0.4333	0.4604	0.3718	0.7899
MAU	0.2463	0.2566	0.2861	0.3004	0.3933	0.6361	0.3232	0.3304	0.4165	0.4451	0.3089	0.7897
PhyDnet	0.2560	0.2685	0.3005	0.3124	0.3785	0.6764	0.3384	0.3824	0.4986	0.4673	0.2941	**0.8022**
STRPM	0.2512	0.3243	0.4959	0.3277	0.2577	0.6513	0.2606	0.4138	0.6882	0.3688	0.2004	0.5996
Prediff	0.2504	0.3041	0.4028	0.2986	0.2851	0.5185	0.2657	0.3854	0.5692	0.3782	0.1543	0.7059
DiffCast	0.2757	**0.3797**	0.5296	0.3584	0.1845	0.6320	0.3472	**0.5066**	**0.7200**	0.4802	0.1234	0.7788
DuoCast	0.3122	0.3759	0.4969	0.3953	0.2011	0.6602	**0.3541**	0.5023	0.6159	0.5097	0.1965	0.7893
Ours	**0.3229**	0.3785	**0.5299**	**0.4061**	**0.1820**	**0.6774**	0.3421	0.4560	0.6384	**0.5160**	**0.1213**	0.7971

Method	Shanghai Radar						CIKM					
	↑CSI	↑CSI-pool4	↑CSI-pool16	↑HSS	↓LPIPS	↑SSIM	↑CSI	↑CSI-pool4	↑CSI-pool16	↑HSS	↓LPIPS	↑SSIM
ConvGRU	0.3612	0.4439	0.5596	0.4899	0.2564	0.7795	0.3092	0.3533	0.4668	0.4007	0.3135	0.6601
SimVP	0.3841	0.4467	0.5603	0.5183	0.2984	0.7764	0.3021	0.3530	0.4677	0.3948	0.3134	0.6324
Earthformer	0.3575	0.4008	0.4863	0.4843	0.2564	0.7750	0.3153	0.3547	0.4927	0.3828	0.3857	0.6510
MAU	0.3996	0.4695	0.5777	0.5356	0.2735	0.7303	0.2962	0.3152	0.4144	0.3660	0.3999	0.6277
PhyDnet	0.3653	0.4552	0.5920	0.4973	0.1989	0.7751	0.3073	0.3442	0.4655	0.3931	0.3631	0.6540
STRPM	0.3606	0.4944	0.6783	0.4931	0.1681	0.7724	0.2984	0.3590	0.5020	0.3870	0.2397	0.6443
Prediff	0.3583	0.4389	0.5448	0.4849	0.1696	0.7557	0.3043	0.3681	0.5117	0.3967	**0.2201**	0.6418
DiffCast	0.3671	0.4907	0.6493	0.4989	**0.1574**	0.7780	0.3111	0.3836	**0.5550**	0.4220	0.2227	0.6156
DuoCast	0.5071	0.5451	0.6334	0.6343	0.2271	0.7808	0.3159	0.3910	0.4801	0.4221	0.2439	0.6257
Ours	**0.5167**	**0.5530**	**0.6810**	**0.6431**	0.1848	**0.7850**	**0.3258**	**0.4068**	0.5145	**0.4472**	0.2213	**0.6646**

TABLE II

COMPARISON OF CSI PERFORMANCE AT DIFFERENT THRESHOLDS ACROSS VARIOUS MODELS AND DATASETS.

Method	SEVIR			MeteoNet		
	↑CSI-M	↑CSI-181	↑CSI-219	↑CSI-M	↑CSI-24	↑CSI-32
ConvGRU	0.2416	0.0879	0.0350	0.3400	0.2990	0.1431
SimVP	0.2662	0.1106	0.0517	0.3346	0.3002	0.1130
Earthformer	0.2413	0.0844	0.0245	0.3296	0.2884	0.1237
MAU	0.2463	0.1071	0.0516	0.3232	0.2839	0.0997
PhyDnet	0.2560	0.1040	0.0278	0.3384	0.3194	0.1366
STRPM	0.2512	0.1043	0.0359	0.2606	0.2438	0.1210
Prediff	0.2504	0.1170	0.0428	0.2657	0.2354	0.1192
DiffCast	0.2757	0.1300	0.0582	0.3472	**0.3340**	**0.1808**
DuoCast	0.3122	0.1334	0.0621	**0.3541**	0.3306	0.1798
Ours	**0.3229**	**0.1507**	**0.0648**	0.3421	0.3239	0.1790

Method	Shanghai Radar			CIKM		
	↑CSI-M	↑CSI-35	↑CSI-40	↑CSI-M	↑CSI-35	↑CSI-40
ConvGRU	0.3612	0.3163	0.2062	0.3092	0.2009	0.1259
SimVP	0.3841	0.3549	0.2382	0.3021	0.2044	0.1321
Earthformer	0.3575	0.3178	0.1872	0.3153	0.2039	0.1369
MAU	0.3996	0.3621	0.2417	0.2962	0.2054	0.1241
PhyDnet	0.3653	0.3236	0.2176	0.3073	0.2052	0.1287
STRPM	0.3606	0.3243	0.2159	0.2984	0.2038	0.1285
Prediff	0.3583	0.3241	0.2028	0.3043	0.2064	0.1292
DiffCast	0.3671	0.3440	0.2606	0.3111	0.2009	0.1457
DuoCast	0.5071	0.4691	0.4624	0.3159	**0.2073**	0.1526
Ours	**0.5167**	**0.5132**	**0.5069**	**0.3258**	0.2066	**0.1554**

intensity attenuation common in deterministic models. On SEVIR and Shanghai datasets, FoumCast surpasses both deterministic and diffusion-based baselines in pixel-wise accuracy and pooled scores, indicating superior recovery of precipitation spatial extent. While performance is slightly lower on MeteoNet due to sparse echoes limiting spectral discriminability, FoumCast remains robust across diverse climatic regions. Fig. 3 illustrates that although performance naturally declines with increasing lead time, FoumCast maintains a consistent margin over competitors, demonstrating its effectiveness in long-term forecasting.

Qualitative Analysis. Visual comparisons in Fig. 4 highlight two key advantages:

- Structural Preservation: Unlike deterministic models (e.g. SimVP) that suffer from over-smoothing, FoumCast retains sharp rainband edges and high-intensity cores, aligning with its superior SSIM and pooled CSI.
- Micro-scale Realism: Compared to diffusion baselines like DuoCast, FoumCast generates finer convective textures and demonstrates higher sensitivity to intensity variations, as evidenced by lower LPIPS.

These improvements stem from: (i) FACL-guided training which enforces spectral alignment; (ii) Spectral disentanglement via dual-path blocks to preserve multiscale features; and (iii) Conditional diffusion that refines residuals within a spectral-aware context.

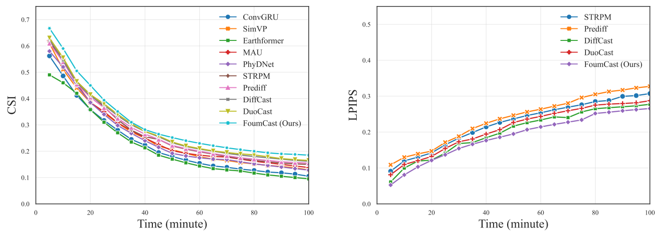

Fig. 3. Comparison of CSI and LPIPS metrics on the SEVIR dataset.

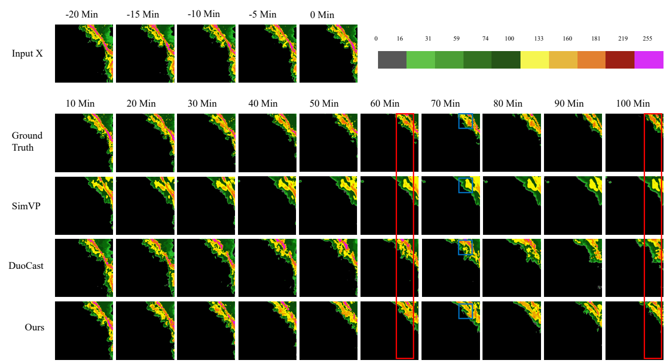

Fig. 4. A visual comparison example on a precipitation event from SEVIR.

C. Ablation Study

Effectiveness of FACL-Loss. As shown in Table III, removing FACL leads to a significant degradation in CSI-pool and SSIM, indicating weakened structural consistency. Without FACL, the backbone tends toward mean-value solutions, causing blurred edges and underestimated peak intensities, as shown in Fig. 5. Unlike pixel-level MSE, FACL ensures spectral fidelity: FAL preserves high-frequency energy (sharpness), while FCL enforces global phase alignment (structure). This global gradient coupling allows the model to capture the geometric integrity of rainbands rather than fitting local intensities, providing a superior starting point for the diffusion stage.

Effectiveness of Spectral Decomposition. Replacing the adaptive spectral disentanglement with a standard convolutional encoder (w/o Decomposition) results in a consistent drop in CSI and SSIM. Visually, the model becomes insensitive to localized intensity bursts. This confirms that explicitly separating low-frequency advection from high-frequency convection allows the network to learn scale-specific dynamics. By modeling these distinct meteorological flows separately, FoumCast stabilizes optimization and prevents the smoothing of convective textures.

Effectiveness of Cross-Attention Fusion. Replacing the CAFM with simple concatenation (w/o Cross-Attention) leads to temporal misalignment, where predicted rainbands lag behind the ground truth, as shown in Fig. 5. The cross-attention mechanism is crucial for synchronizing temporal cues with spectral representations, ensuring phase consistency across frames. Quantitatively, the increase in LPIPS without CAFM confirms its role in maintaining realistic rainfall evolution.

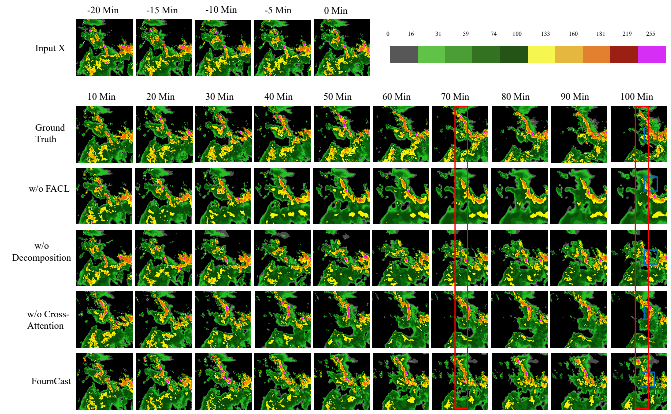

Fig. 5. Ablation study visualization on the SEVIR dataset.

TABLE III
THE ABLATION STUDY RESULTS FOR THE COMPONENTS OF FOUMCAST.

SEVIR	CSI-M↑	CSI-160↑	CSI-181↑	HSS↑	LPIPS↓	SSIM↑
w/o FACL	0.2732	0.1145	0.0589	0.3367	0.3933	0.6464
w/o Decomposition	0.3044	0.1365	0.0597	0.3814	0.2297	0.6498
w/o Cross-Attention	0.3114	0.1432	0.0609	0.3928	0.2169	0.6610
FoumCast	**0.3229**	**0.1507**	**0.0648**	**0.4061**	**0.1820**	**0.6774**
Shanghai	CSI-M↑	CSI-35↑	CSI-40↑	HSS↑	LPIPS↓	SSIM↑
w/o FACL	0.3626	0.3230	0.2415	0.4522	0.2231	0.7669
w/o Decomposition	0.4692	0.4607	0.4538	0.6309	0.1928	0.7768
w/o Cross-Attention	0.4851	0.4713	0.4510	0.6282	0.1983	0.7736
FoumCast	**0.5167**	**0.5132**	**0.5069**	**0.6431**	**0.1848**	**0.7850**

Effectiveness of Adaptive Masking. To assess the impact of frequency partitioning, we compare the fixed threshold ($D_c = 10$) with our adaptive spectral mask $\mathcal{M}_{\boldsymbol{\theta}}$. As visualized in Fig. 6, the fixed threshold occasionally leads to spectral leakage or over-filtering, resulting in either blurred convective cores or fragmented rainbands. In contrast, the adaptive mask dynamically aligns the frequency boundary with the input's energy distribution. This results in superior structural fidelity, where large-scale morphology is preserved without sacrificing the sharpness of localized high-intensity centers.

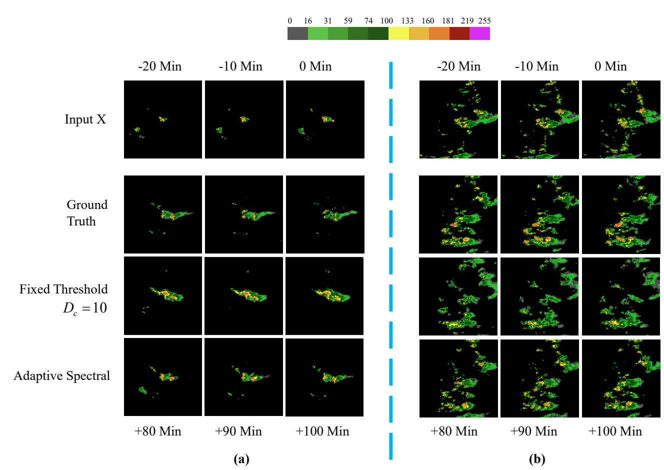

Fig. 6. Ablation experiment compares the effect of the Adaptive Masking.

V. Conclusion

In this paper, we presented FoumCast, a novel framework that synergizes Fourier-driven spectral disentanglement with conditional diffusion modeling for precipitation nowcasting. By explicitly partitioning the frequency domain into low-frequency advective envelopes and high-frequency convective textures, our model bridges the gap between macro-scale meteorological evolution and micro-scale stochastic dynamics. The introduction of a cross-attention fusion mechanism further enhances temporal and spectral alignment, while the conditional diffusion network refines residual uncertainties and local rainfall structures. Extensive experiments on four real-world datasets verifies the effectiveness of the proposed framework.

References

[1] 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, 2023.

[2] R. B. Alley, K. A. Emanuel, and F. Zhang, "Advances in weather prediction," *Science*, vol. 363, no. 6425, pp. 342–344, 2019.

[3] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.

[4] 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, 2023.

[5] A. C. Lorenc, "Analysis methods for numerical weather prediction," *Quarterly Journal of the Royal Meteorological Society*, vol. 112, no. 474, pp. 1177–1194, 1986.

[6] Z. Gao, X. Shi, H. Wang, Y. Zhu, Y. B. Wang, M. Li, and D.-Y. Yeung, "Earthformer: Exploring space-time transformers for earth system forecasting," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25390–25403, 2022.

[7] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "Simvp: Simpler yet better video prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3170–3180, 2022.

[8] 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," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

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

[10] V. L. Guen and N. 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. 11474–11484, 2020.

[11] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, *et al.*, "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, 2021.

[12] Z. Gao, X. Shi, B. Han, H. Wang, X. Jin, D. Maddix, Y. Zhu, M. Li, and Y. B. Wang, "Prediff: Precipitation nowcasting with latent diffusion models," *Advances in Neural Information Processing Systems*, vol. 36, pp. 78621–78656, 2023.

[13] J. Gong, L. Bai, P. Ye, W. Xu, N. Liu, J. Dai, X. Yang, and W. Ouyang, "Cascast: Skillful high-resolution precipitation nowcasting via cascaded modelling," *arXiv preprint arXiv:2402.04290*, 2024.

[14] D. Yu, X. Li, Y. Ye, B. Zhang, C. Luo, K. Dai, R. Wang, and X. 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.

[15] J. Catto, C. Jakob, G. Berry, and N. Nicholls, "Relating global precipitation to atmospheric fronts," *Geophysical Research Letters*, vol. 39, no. 10, 2012.

[16] P. Wen, L. Bai, M. He, P. Filippi, F. Zhang, T. F. Bishop, Z. Wang, and K. Hu, "Duocast: Duo-probabilistic meteorology-aware model for extended precipitation nowcasting," *arXiv preprint arXiv:2412.01091*, 2024.

[17] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[18] Z. Chang, X. Zhang, S. Wang, S. Ma, Y. Ye, X. Xinguang, and W. Gao, "Mau: A motion-aware unit for video prediction and beyond," *Advances in Neural Information Processing Systems*, vol. 34, pp. 26950–26962, 2021.

[19] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in *International Conference on Machine Learning*, pp. 5301–5310, PMLR, 2019.

[20] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[21] S. D. Sims, "Frequency domain-based perceptual loss for super resolution," in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2020.

[22] L. Jiang, B. Dai, W. Wu, and C. C. Loy, "Focal frequency loss for image reconstruction and synthesis," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 13919–13929, 2021.

[23] D. Fuoli, L. Van Gool, and R. Timofte, "Fourier space losses for efficient perceptual image super-resolution," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2360–2369, 2021.

[24] C. Li, C.-L. Guo, M. Zhou, Z. Liang, S. Zhou, R. Feng, and C. C. Loy, "Embedding fourier for ultra-high-definition low-light image enhancement," *arXiv preprint arXiv:2302.11831*, 2023.

[25] G. Chen, P. Peng, L. Ma, J. Li, L. Du, and Y. Tian, "Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 458–467, 2021.

[26] H. Liu, X. Li, W. Zhou, Y. Chen, Y. He, H. Xue, W. Zhang, and N. Yu, "Spatial-phase shallow learning: rethinking face forgery detection in frequency domain," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 772–781, 2021.

[27] H. Huang, H. Kang, S. Liu, O. Salvado, T. Rakotoarivelo, D. Wang, and T. Liu, "Paddles: Phase-amplitude spectrum disentangled early stopping for learning with noisy labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16719–16730, 2023.

[28] Q. Xu, R. Zhang, Y. Zhang, Y. Wang, and Q. Tian, "A fourier-based framework for domain generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14383–14392, 2021.

[29] D. Yin, R. Gontijo Lopes, J. Shlens, E. D. Cubuk, and J. Gilmer, "A fourier perspective on model robustness in computer vision," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[30] C. Zhao, W. Cai, C. Dong, and C. Hu, "Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8281–8291, 2024.

[31] C.-W. Yan, S. Q. Foo, V. H. Trinh, D.-Y. Yeung, K.-H. Wong, and W.-K. Wong, "Fourier amplitude and correlation loss: Beyond using l2 loss for skillful precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 37, pp. 100007–100041, 2024.

[32] M. Veillette, S. Samsi, and C. Mattioli, "Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology," *Advances in Neural Information Processing Systems*, vol. 33, pp. 22009–22019, 2020.

[33] G. Larvor, L. Berthomier, V. Chabot, B. Le Pape, B. Pradel, and L. Perez, "Meteonet, an open reference weather dataset," 2020.

[34] L. Chen, Y. Cao, L. Ma, and J. Zhang, "A deep learning-based methodology for precipitation nowcasting with radar," *Earth and Space Science*, vol. 7, no. 2, p. e2019EA000812, 2020.

[35] Z. Chang, X. Zhang, S. Wang, S. Ma, and W. Gao, "Strpm: A spatiotemporal residual predictive model for high-resolution video prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13946–13955, 2022.