QFSNET: A QUATERNION FOURIER SPECTRAL NETWORK FOR DETERMINISTIC PRECIPITATION NOWCASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Precipitation nowcasting requires capturing complex, multi-scale spatiotemporal dynamics under strict latency constraints. However, existing deep learning paradigms face two fundamental bottlenecks: standard spectral representations fail to explicitly disentangle scale and orientation, leading to trajectory underfitting, and a critical mismatch exists between continuous pixel-wise optimization and discrete structural boundaries, causing models to blur sharp intensity transitions to minimize aggregate errors. To address these challenges, we propose QFSNet, a highly efficient deterministic nowcasting architecture anchored by a Quaternion Fourier Neural Operator. To resolve the directional misalignment, we introduce a lossless wavelet-quaternion representation that maps multi-scale details into a quaternion field, leveraging the non-commutative Hamilton product and a two-sided Quaternion Fourier Transform (QFT) to jointly model directional kinematics, and decouple the forecasting process into three specialized branches for fine-scale dynamics, macroscopic trends, and local phase transport. Furthermore, standard continuous optimization inherently causes structural blurring and the loss of high-contrast edges in predictions. To resolve this, we introduce a mathematically rigorous normal–tangent decomposition mechanism. This strictly decouples the boundary-reshaping dynamics required to restore sharp structural discontinuities from the intra-region intensity calibration , explicitly preventing the over-smoothing of abrupt precipitation transitions without introducing artificial artifacts. Extensive experiments across four benchmark datasets demonstrate that QFSNet achieves state-of-the-art performance, particularly in extreme precipitation detection, while requiring the lowest parameter count and computational cost, and maintaining the fastest inference speed.

1 INTRODUCTION

Precipitation nowcasting is critical for severe weather warning systems, requiring the accurate capture of complex spatiotemporal evolution under stringent latency constraints. The fundamental challenge lies in the multi-scale and multi-physics nature of precipitation dynamics: large-scale translational advection coexists with local non-linear deformation and rapid convective generation or dissipation. Consequently, errors in motion estimation, structural representation, and intensity prediction compound non-linearly over the forecast horizon.

Existing paradigms face distinct limitations. Spatial and optical flow methods Shi et al. (2015); Wang et al. (2017; 2018); Gao et al. (2022a); Tan et al. (2023); Gao et al. (2022b); Zhang et al. (2023); Ritvanen et al. (2023) struggle with long-range dependencies or blur high-frequency boundaries via pixel-space warping. Generative models Gao et al. (2023); Gong et al. (2024); Yu et al. (2024); Perrone Ribeiro & Faganeli Pucer (2026) incur prohibitive latency and hallucinate physically inconsistent textures. While spectral methods Pathak et al. (2022); Wu et al. (2024a;b); Lin et al. (2025) improve global efficiency, standard complex representations are constrained by a single imaginary unit. This restricts their capacity to disentangle heterogeneous directional phase shifts, inevitably causing severe phase interference. Furthermore, a deeper flaw persists across deterministic paradigms: the mismatch between continuous pixel-wise optimization and discrete meteorological transitions forces models to

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

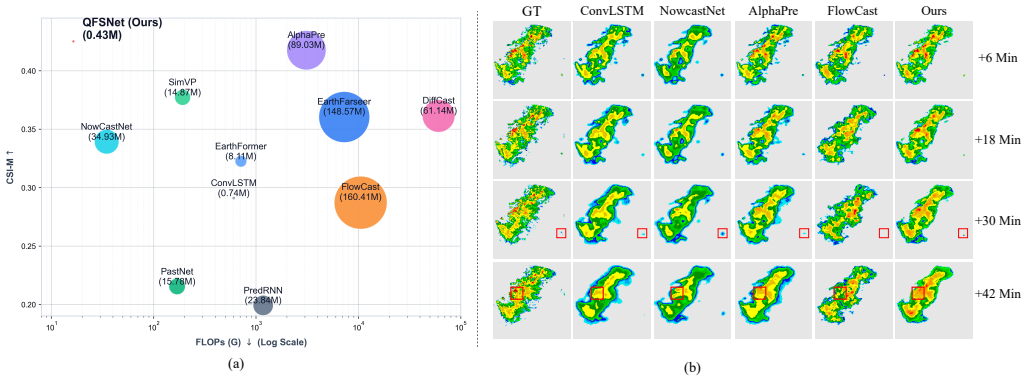

Figure 1: (a) Comparison of QFSNet with baselines. QFSNet achieves the best CSI with the fewest parameters and lowest FLOPs. (b) Visual examples on the Shanghai dataset. Recurrent/convolutional (ConvLSTM) and optical flow (NowcastNet) methods struggle with long-range dependencies and blur high-frequency boundaries. Frequency (AlphaPre) and generative (FlowCast) models fall short in geometric kinematics and physical essence. QFSNet efficiently captures both.

smooth critical intensity boundaries to minimize aggregate errors, severely degrading extreme event detection, as shown in Fig. 1b.

Addressing these bottlenecks requires preserving multi-scale details while circumventing directional phase conflicts. Wavelet transforms intrinsically offer lossless reconstruction and high-fidelity detail preservation Strang (1993); Sifuzzaman et al. (2009). Overcoming the directional constraints of standard complex implementations necessitates endowing this representation with a richer algebraic structure. Quaternion algebra, characterized by three imaginary units and a non-commutative Hamilton product, intrinsically encodes multi-directional rotations and disentangles orthogonal phase shifts Bulow & Sommer (2001). Mapping wavelet sub-bands into a quaternion field enables explicit modeling of directional kinematics while precluding phase interference. Furthermore, precipitation dynamics exhibit physical separability: macroscopic advection, local non-linear deformation, and abrupt intensity transitions that require distinct structural decoupling.

Guided by these insights, we propose **QFSNet**, a highly efficient deterministic nowcasting architecture. We anchor QFSNet with a Quaternion Fourier Neural Operator and a lossless wavelet-quaternion representation to fundamentally resolve directional phase conflicts. To enforce physical consistency and linear complexity, we explicitly decouple the forecasting process into three specialized branches for macroscopic advection, thermodynamic trends, and local deformations. Furthermore, to directly correct this structural blurring in the prediction space, we introduce a mathematically rigorous normal–tangent decomposition. This strictly isolates the normal structural displacement needed to restore sharp boundaries from the tangential monotone calibration required for precise intra-region intensity refinement. Collectively, these synergistic designs bridge the gap between efficient global spectral modeling and precise operational forecasting. Our main contributions are summarized as follows:

1. We introduce a novel, lossless wavelet-quaternion spectral representation that explicitly disentangles scale and orientation, facilitates structured Hamiltonian cross-band coupling, and enables axis-aware global phase evolution via a two-sided QFT.
2. We develop QFSNet, a highly efficient forecasting architecture that decouples fine-scale spectral evolution, macroscopic LL dynamics, and local non-linear phase transport, utilizing complementary global, local, and boundary-sensitive spectral operators.
3. We propose a mathematically rigorous normal–tangent decomposition that isolates boundary-reshaping dynamics from intra-region intensity calibration, explicitly preventing the over-smoothing of sharp structural discontinuities.
4. Experiments across four datasets demonstrate that QFSNet achieves state-of-the-art (SOTA) performance with the lowest parameter count and computational cost, while maintaining the fastest inference speed (Fig. 1a).

2 Related Works

2.1 Spatial Domain-Based Methods

Early deep learning approaches for precipitation nowcasting predominantly relied on recurrent architectures Shi et al. (2015); Wang et al. (2017; 2018; 2019); Guen & Thome (2020); Tang et al. (2023), which extended LSTMs with convolutional gates to capture local spatial dependencies. Subsequent studies proposed direct multi-frame mapping using pure convolutional networks Gao et al. (2022a); Tan et al. (2023; 2025) or Transformer-based networks Gao et al. (2022b); Bai et al. (2022). While this approach allows for more flexible processing of spatiotemporal features, its spatiotemporal operations treat directional frequency bands as independent channels. This lack of algebraic coupling among directional sub-bands fundamentally limits their ability to preserve sharp structural details, leading to severe over-smoothing. To further enhance the representation of local future details, generative models Ravuri et al. (2021); Gao et al. (2023); Gong et al. (2024); Yu et al. (2024); Perrone Ribeiro & Faganeli Pucer (2026) have been recently introduced. Despite their high fidelity in generating realistic textures, they incur massive computational overhead during inference and struggle to strictly preserve the deterministic mass conservation and sharp structural boundaries required for accurate physical modeling. Alternatively, to explicitly model physical motion, advection-based methods Zhang et al. (2023); Ritvanen et al. (2023) warp historical radar frames by predicting pixel-space optical flow. Although physically intuitive, pixel-space warping inevitably introduces interpolation artifacts that blur sharp echo boundaries. Furthermore, their reliance on linear or weakly non-linear flow fields makes them highly susceptible to errors when dealing with complex, non-linear trajectories and rapid convective intensity evolution.

2.2 Spectral Domain-Based Methods

To overcome the computational bottlenecks and local receptive field limitations of spatial convolutions, spectral domain methods have emerged as a powerful paradigm for spatiotemporal sequence prediction. By transforming data into the frequency domain, these methods efficiently capture global contextual patterns and high-frequency details via global multiplications Pathak et al. (2022); Wu et al. (2024a;b), or decouple frequency-domain information to maintain global coherence Yan et al. (2024); Lin et al. (2025). Despite their success in global pattern recognition, existing spectral methods predominantly operate in standard complex or real-valued spectral spaces. In such unified systems, the kinematics of the sequence are inferred holistically without explicit structural separation. Crucially, standard complex spectra fail to disentangle distinct motion directions. When multiple convective cells with heterogeneous moving directions interact, these models attempt to reconcile conflicting phase shifts within a single algebraic system. This structural limitation leads to phase misalignment and trajectory underfitting, severely degrading the prediction accuracy for complex, multi-directional weather systems.

In contrast, our QFSNet introduces a unified Quaternion Fourier framework. By packing directional wavelet sub-bands into a quaternion field, the non-commutative Hamilton product natively resolves these directional conflicts, explicitly modeling heterogeneous phase transport without trajectory underfitting.

3 Preliminary

3.1 Problem Formulation and Notation

Let $\boldsymbol{X} \in \mathbb{R}^{T_{\text{in}} \times C \times H \times W}$ be the input radar sequence of T_{in} frames, C physical channels, and spatial resolution $H \times W$. We aim to predict the future absolute intensity field $\widehat{\boldsymbol{Y}} \in \mathbb{R}^{T_{\text{out}} \times C \times H \times W}$. We denote the quaternion algebra by $\mathbb{H}$ with basis $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$. The ordered Hamilton product is $\otimes_H$, and ordered circular Hamilton convolution is $\star$. Spatial dimensions after one-level wavelet analysis are $H_w = H/2$ and $W_w = W/2$.

3.2 WAVELET-QUATERNION REPRESENTATION

Orthonormal wavelet quaternion packing. For a local spatial block, we compute the standard orthonormal wavelet coefficients, i.e., the low-frequency scalar L and the horizontal, vertical, and diagonal vector details $D^{\mathrm{h}}, D^{\mathrm{v}}, D^{\mathrm{d}}$, and pack them into a single quaternion field:

$$Q = L + \mathbf{i}D^{\mathrm{h}} + \mathbf{j}D^{\mathrm{v}} + \mathbf{k}D^{\mathrm{d}} \in \mathbb{H}^{H_w \times W_w}. \tag{1}$$

This packing leverages the non-commutative Hamilton product to induce cross-orientation interactions among directional details without auxiliary kernels, while the scalar-vector decomposition naturally separates the large-scale state from high-frequency content.

Two-sided QFT. The orthonormal two-sided QFT of Q is defined as

$$\mathcal{F}_Q[Q](\omega_y, \omega_x) = \frac{1}{\sqrt{N_w}} \sum_{y=0}^{H_w-1} \sum_{x=0}^{W_w-1} e^{-\mathbf{i}2\pi\omega_y y/H_w} Q(y,x) e^{-\mathbf{j}2\pi\omega_x x/W_w}, \tag{2}$$

where $N_w = H_w W_w$. The left $\mathbf{i}$ and right $\mathbf{j}$ exponentials preserve the non-commutative geometry of the packed bands.

4 METHOD

4.1 OVERALL DESIGN

QFSNet first maps the observed input radar sequence to a wavelet-quaternion trajectory $Q^{\mathrm{in}}_{1:T_{\mathrm{in}}}$ via orthonormal packing, and subsequently computes its two-sided spectrum $Z^{\mathrm{in}}_{1:T_{\mathrm{in}}} = \mathcal{F}_Q(Q^{\mathrm{in}}_{1:T_{\mathrm{in}}})$. This shared spectral representation serves as the unified coordinate system for three parallel forecasting branches, which predict fine-scale directional details, macroscopic thermodynamic trends, and local motion residuals, respectively. Their outputs are additively fused and synthesized back to the spatial domain via $\mathcal{W}^{-1}$ to form a base forecast P^0_t. This base forecast is subsequently refined by the normal-tangent module to restore sharp structural boundaries, yielding the final prediction $\widehat{\boldsymbol{Y}}$.

4.2 QUATERNION FOURIER SPECTRAL OPERATORS

4.2.1 RADIAL SPECTRAL MIXER

For an input quaternion spectrum $Z \in \mathbb{H}^{C_{\mathrm{in}} \times H_w \times W_w}$, let $\{\varphi_m\}_{m=1}^{M}$ be triangular radial basis maps partitioning the frequency plane into M concentric annuli. With grouped quaternion weights $W_m \in \mathbb{H}^{C_{\mathrm{out}} \times C_{\mathrm{in}}}$, the output channel is computed as

$$[\mathcal{M}(Z)]_o(\omega_y, \omega_x) = \sum_{m=1}^{M} \varphi_m(\omega_y, \omega_x) \sum_{c \in \mathcal{G}(o)} W_{m,o,c} \otimes_H Z_c(\omega_y, \omega_x). \tag{3}$$

This provides a global receptive field with scale-adaptive modulation.

4.2.2 EXACT LOCAL CONVOLUTION VIA PAIR-COMPLEX FACTORIZATION

We deploy an exact bi-sided Hamilton convolution $\mathcal{C}(X) = w_{\mathrm{left}}(K_{\mathrm{left}} \star X) + w_{\mathrm{right}}(X \star K_{\mathrm{right}})$. Via pair-complex factorization, mapping the quaternion spectrum to a complex pair (A_q, B_q) reduces the spectral computation to pointwise complex multiplications:

$$A_{F \star G} = s_w(A_F A_G - B_F \widetilde{B}_G), \quad B_{F \star G} = s_w(A_F B_G + B_F \widetilde{A}_G), \tag{4}$$

where $\widetilde{Z}(\omega) = \overline{Z(-\omega)}$ is the reflected conjugate and s_w is the orthonormal scale. Both proposed operators achieve structured receptive fields with linear spatial complexity, avoiding the quadratic cost of self-attention.

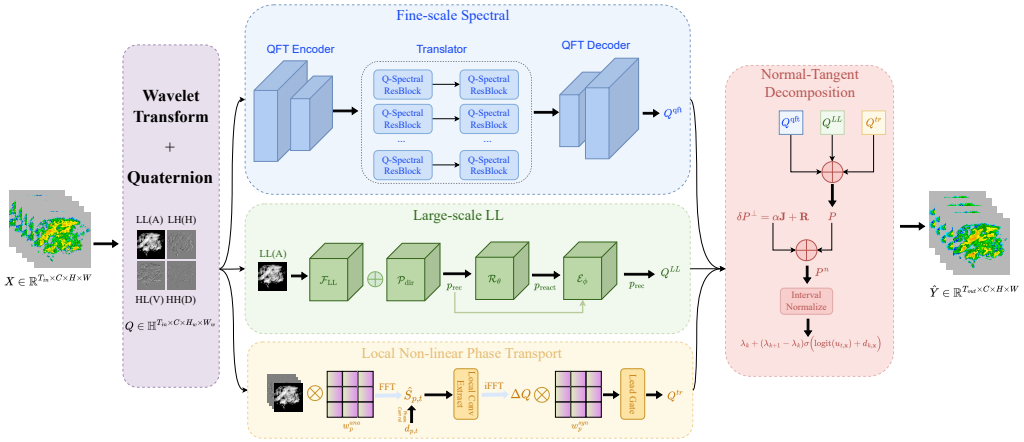

Figure 2: Overall architecture of QFSNet. Input radar frames undergo wavelet decomposition and quaternion packing, then feed into three parallel branches: Fine-scale Spectral, Large-scale LL, and Local Non-linear Phase Transport. The fused outputs are refined via a Normal-Tangent Decomposition module, where a normal update drives boundary reshaping and a tangent map calibrates intra-region intensity, yielding the final prediction $\hat{Y}$.

4.3 Spectral Evolution Branches

4.3.1 Fine-scale QFT Branch

As illustrated in Fig. 2, the fine-scale branch takes the shared spectrum Z^{in} as input. Its QFT Encoder–Translator–Decoder produces a predicted fine-scale spectrum Z_t^{fine}, which is returned to the wavelet-quaternion coordinate as $\widehat{Q}_t^{\text{fine}} = \mathcal{F}_Q^{-1}(Z_t^{\text{fine}})$. Each layer of the encoder and decoder consists of a Q-Spectral ResBlock and an upsampling or downsampling operation. Each Q-Spectral ResBlock within the Translator comprises layer normalization, the quaternion spectral operators from Sec. 4.2, a quaternion ModReLU activation, and a residual connection with grouped channel shuffle. Given an intermediate spectrum Z, one block computes:

$$Z_1 = \mathcal{N}_1(Z), \quad Z_{\text{spec}} = \mathcal{M}(Z_1) + \eta\,\mathcal{C}(Z_1), \quad Z_2 = \sigma_{\mathbb{H}}(Z_{\text{spec}}),$$
$$Z_{\text{out}} = Z + \rho \cdot \text{Shuffle}\Big(\mathcal{P}\big(\mathcal{N}_2(Z_2)\big)\Big), \tag{5}$$

where $\mathcal{N}_{1,2}$ are normalizations, $\sigma_{\mathbb{H}}$ is quaternion ModReLU, $\mathcal{P}$ is a 1×1 grouped projection, and $\eta \in \{0, 1\}$ activates $\mathcal{C}$ only at the translator bottleneck.

4.3.2 Independent Large-scale LL Branch

The LL branch uses the LL coefficients of Q^{in} together with three descriptors: temporal tendency, local curvature, and the mismatch with the coarse kinematic baseline P_t^{base} derived from the fine branch. We pack them into

$$U_t^{\text{LL}} = X_t^{\text{LL}} + D_t\,\mathbf{i} + C_t\,\mathbf{j} + E_t\,\mathbf{k}. \tag{6}$$

A spectral reaction $\mathcal{R}_\theta$ and recurrent evolution $\mathcal{E}_\phi$ operate on U_t^{LL} and produce the scalar LL forecast:

$$X_t^{\text{LL}} = P_t^{\text{base}} + \text{Re}\big(\mathcal{R}_\theta(U_t^{\text{LL}})\big) + \text{Re}\big(\mathcal{E}_\phi(U_t^{\text{LL}})\big). \tag{7}$$

We embed this scalar field as $\widehat{Q}_t^{\text{LL}} = X_t^{\text{LL}} + 0\,\mathbf{i} + 0\,\mathbf{j} + 0\,\mathbf{k}$ and use a bounded gate to prevent it from overwriting the fine-scale directional detail.

4.3.3 Local Non-linear Phase Transport

To capture spatially heterogeneous motion, the phase branch takes the last observed wavelet-quaternion field $Q_{T_{\text{in}}}^{\text{in}}$. We partition it into P overlapping patches with Gaussian analysis windows

w_p^{ana}, giving $Q_p^{\mathrm{local}} = w_p^{\mathrm{ana}} \odot Q_{T_{\mathrm{in}}}^{\mathrm{in}}$. Normalized cross-power spectra and soft-argmax phase correlation estimate the mean velocity $\mathbf{v}_p$ and curvature $\mathbf{c}_p$. The resulting displacement $\mathbf{d}_{p,t}$ defines an exact two-sided QFT shift of the local spectrum $\widehat{Q}_p = \mathcal{F}_Q[Q_p^{\mathrm{local}}]$:

$$\hat{Q}_{p,t}^{\mathrm{tr}} = E_{\mathbf{i}}(-\theta_{y,t}) \otimes_H \hat{Q}_p \otimes_H E_{\mathbf{j}}(-\theta_{x,t}), \quad \theta_{y,t} = 2\pi d_{p,t,y}\omega_y / H_w. \tag{8}$$

The rigorous derivation of this shift theorem is provided in Appendix A.2. The patch residuals $\Delta Q_{p,t} = \mathcal{F}_Q^{-1}[\widehat{Q}_{p,t}^{\mathrm{tr}}] - Q_p^{\mathrm{local}}$ are combined by confidence-weighted overlap-add, yielding the branch output ΔQ_t^{tr}. The three branch outputs are fused explicitly as

$$Q_t^{\mathrm{fuse}} = \widehat{Q}_t^{\mathrm{fine}} + \widehat{Q}_t^{\mathrm{LL}} + \Delta Q_t^{\mathrm{tr}}, \qquad P_t^0 = \mathcal{W}^{-1}\big(Q_t^{\mathrm{fuse}}\big). \tag{9}$$

The complementary image-space contributions of the three branches are visualized in the top row of Fig. 3. The fine-scale and phase-transport outputs provide signed directional corrections, whereas the LL branch supplies the dominant large-scale intensity envelope.

4.4 Normal–Tangent Decomposition for Structure-Aware Refinement

Since continuous loss functions cause over-smoothing in the base forecast P_t^0, we decouple the total correction $\delta P = \widehat{\boldsymbol{Y}} - P_t^0$ into two geometrically distinct operations: $\delta P = \delta P^{\perp} + \delta P^{\parallel}$.

Normal update: conservative boundary reshaping ($\delta P^{\perp}$). The normal correction drives the spatial redistribution of mass to reshape sharp structural contours. It is constructed to localize corrections near abrupt intensity transitions $\{\lambda_k\}$ via a mass-conserving spatial redistribution:

$$\delta P^{\perp} = \alpha_{\mathrm{shape}} \mathbf{J}_t + \mathbf{R}_t, \tag{10}$$

where the spatial shape flux $\mathbf{J}_t$ is defined as:

$$\mathbf{J}_t = \mathrm{ZeroMean}\left(\mathbf{J}_t^{\mathrm{learn}} + \frac{1}{K} \sum_{k=1}^{K} s_k \cdot 4E_{t,k}(1 - E_{t,k}) \right). \tag{11}$$

Here, $E_{t,k} = \sigma(\kappa(P_t^0 - \lambda_k))$ is a soft boundary mask. The term $4E_{t,k}(1 - E_{t,k})$ strictly localizes the correction to pixels near the critical transitions λ_k. Concurrently, $\mathbf{R}_t = b_t M_t$ represents the mass reaction, where $\langle M_t \rangle_{xy} = 1$ ensures $\langle \mathbf{R}_t \rangle_{xy} = b_t$. The rigorous proof of this strict localization and mass conservation is deferred to Appendix A.1.

Tangent calibration: intra-region refinement ($\delta P^{\parallel}$). Taking the intermediate state $P_t^{\mathrm{n}} = P_t^0 + \delta P^{\perp}$, the tangent update refines continuous intensity within the specific homogeneous region bounded by the restored contours $[\lambda_k, \lambda_{k+1})$. The pixel value is normalized to a relative position $u \in (0, 1)$:

$$u_{t,\mathbf{x}} = \frac{P_{t,\mathbf{x}}^{\mathrm{n}} - \lambda_k}{\lambda_{k+1} - \lambda_k}. \tag{12}$$

A compact network predicts a logit shift $d_{k,\mathbf{x}}$, and the final prediction is obtained via a monotone map:

$$\widehat{Y}_{t,\mathbf{x}} = \lambda_k + (\lambda_{k+1} - \lambda_k)\sigma\big(\mathrm{logit}(u_{t,\mathbf{x}}) + d_{k,\mathbf{x}}\big). \tag{13}$$

Since $\sigma(\cdot) \in (0, 1)$, this map strictly confines the output to $[\lambda_k, \lambda_{k+1})$, preventing spurious edge blurring or structural artifacts.

Gradient orthogonality via identity-tangent parameterization. By the chain rule, the Jacobian of the tangent map with respect to the normal state is:

$$\frac{\partial \widehat{Y}_{t,\mathbf{x}}}{\partial P_{t,\mathbf{x}}^{\mathrm{n}}} = \sigma'\big(\mathrm{logit}(u) + d_k\big) \cdot \frac{1}{u(1-u)}. \tag{14}$$

Crucially, the network readout is initialized to zero ($d_k = 0$). At initialization, the Jacobian evaluates exactly to 1. This identity-tangent property ensures that during early training, the gradient flows through the tangent stage as an exact identity mapping, remaining orthogonal to the normal stage's learning. The detailed derivation of this Jacobian via the chain rule is provided in Appendix A.1. The bottom row of Fig. 3 illustrates this decomposition on a representative forecast lead. The normal update restores the geometry of sharp structural boundaries, while the tangent update refines continuous intensity within the local region without introducing spurious structural artifacts.

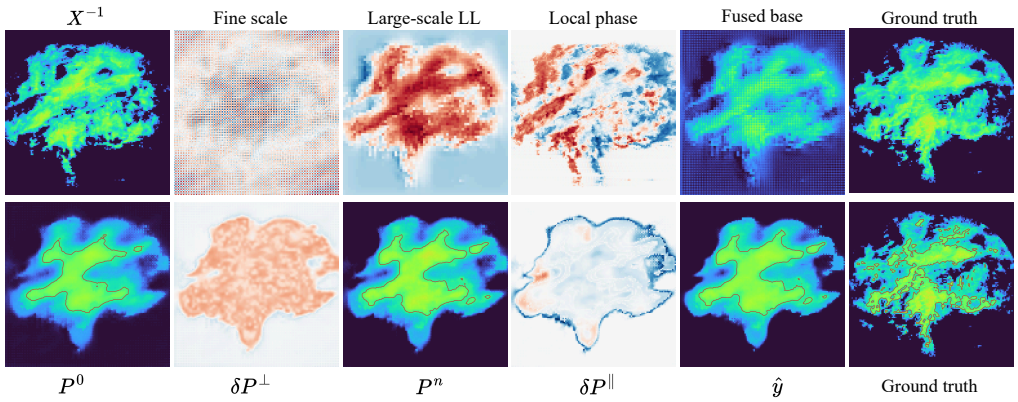

Figure 3: Visualization of branch contributions and normal–tangent refinement. **Top:** Last input frame, signed image-space contributions of the fine-scale and LL branches, phase-transport increment, fused synthesis, and ground truth. **Bottom:** Refinement from base forecast P^0 through normal update $\delta P^{\perp}$ and intermediate state P^n, to the final prediction $\widehat{Y}$. Red contours indicate the operational warning threshold.

4.5 Boundary-Aware Asymmetric Event Loss

$$\mathcal{L} = \mathcal{L}_{\text{MAE}} + \lambda_e \mathcal{L}_{\text{event}}, \qquad \mathcal{L}_{\text{event}} = \frac{1}{|\mathcal{J}|} \sum_{l \in \mathcal{J}} \frac{1}{|T_{out}|} \sum_{t \in T_{out}} \left(1 - \frac{h_{t,l}}{h_{t,l} + m_{t,l} + \alpha_{\text{FA}} f_{t,l} + \epsilon_e} \right). \tag{15}$$

The first term preserves continuous intensity accuracy. The second aggregates soft event overlap across critical intensity transitions $\mathcal{J}$ and T_{out}, using differentiable hit, miss, and false-alarm masses $h_{t,l}$, $m_{t,l}$, and $f_{t,l}$. The coefficient α_{FA} controls the relative cost of false alarms, while λ_e balances event localization against pixel-wise reconstruction. This asymmetric form directly emphasizes boundary exceedances without discarding amplitude information.

5 Experiments

5.1 Experimental Setup

Datasets and protocol. We evaluate QFSNet on Shanghai (Chen et al., 2020), SEVIR (Veillette et al., 2020), Meteo (Larvor & Berthomier, 2021), and CIKM AnalytiCup 2017 (Yao & Li, 2017). All inputs are single-channel 128×128 radar sequences scaled to $[0, 1]$. Five observed frames are used to predict 20 future frames, except on CIKM, where 10 frames are predicted. For further details, please refer to the Appendix D.

Baselines. We compare with ConvLSTM (Shi et al., 2015), PredRNN (Wang et al., 2017), SimVP (Gao et al., 2022a), EarthFormer (Gao et al., 2022b), PastNet (Wu et al., 2024b), EarthFarseer (Wu et al., 2024a), NowcastNet (Zhang et al., 2023), DiffCast (Yu et al., 2024), AlphaPre (Lin et al., 2025), and FlowCast (Perrone Ribeiro & Faganeli Pucer, 2026), covering recurrent, convolutional, Transformer, spectral, physics-informed, generative paradigms.

Metrics. We report mean critical success index (CSI-M), CSI at the two highest dataset-specific thresholds, probability of detection (POD), structural similarity (SSIM), continuous ranked probability score (CRPS), and mean absolute error (MAE). Higher CSI, POD, and SSIM are better; lower CRPS and MAE are better. Definitions and threshold values are provided in Appendix D.2. The complete layer-wise architecture, tensor shapes, channel mappings, and parameter allocation are given in Appendix C.

Table 1: Predictive quality and computational efficiency. Best and second-best available results are bold and underlined, respectively.

Method	Shanghai							SEVIR						
	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓	CSI-M↑	CSI-181↑	CSI-219↑	POD↑	SSIM↑	CRPS↓	MAE↓
ConvLSTM	0.2910	0.2376	0.0768	0.3274	0.6639	0.0249	2.2415	0.2782	0.0803	0.0019	0.3408	0.6302	0.0393	10.0450
PredRNN	0.1992	0.1610	0.0575	0.2441	0.6532	0.0317	2.8550	0.1621	0.0188	0.0004	0.2092	0.5541	0.0511	13.0274
SimVP	0.3772	0.3443	0.2110	0.4449	0.7572	0.0209	1.8788	0.3042	0.1150	0.0500	0.3703	0.6274	0.0400	10.1979
EarthFormer	0.3225	0.2808	0.1491	0.3786	0.7290	0.0237	2.1368	0.2435	0.0519	0.0136	0.3080	0.5952	0.0457	11.6523
NowcastNet	0.3392	0.2897	0.1606	0.3896	0.7531	0.0205	1.8471	0.2410	0.0519	0.0149	0.2833	0.6054	0.0431	10.9878
EarthFarseer	0.3604	0.3172	0.1913	0.4178	0.7564	0.0210	1.8874	0.2764	0.0835	0.0365	0.3310	0.5997	0.0416	10.6085
DiffCast	0.3614	**0.3989**	0.2346	0.4860	0.7542	0.0233	2.0980	0.3046	0.1301	0.0582	0.4024	0.6604	0.0393	10.0171
PastNet	0.2154	0.1371	0.0314	0.2313	0.7454	0.0226	2.0370	0.2679	0.0689	0.0193	0.3207	0.6087	0.0414	10.5555
AlphaPre	0.4176	0.3853	0.2614	0.5008	0.6450	0.0235	2.1153	**0.3355**	0.1402	0.0557	0.4012	**0.6746**	**0.0350**	**8.9321**
FlowCast	0.2872	0.2463	0.1570	0.3920	0.7585	0.0254	2.2896	0.2965	0.1168	0.0467	0.3884	0.6721	0.0382	9.7329
QFSNet	**0.4252**	0.3882	**0.2869**	**0.6020**	**0.8040**	**0.0192**	**1.7280**	0.3270	**0.1595**	**0.0876**	**0.4389**	0.6721	0.0381	9.6239

Method	Meteo							CIKM						
	CSI-M↑	CSI-24↑	CSI-32↑	POD↑	SSIM↑	CRPS↓	MAE↓	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓
ConvLSTM	0.2830	0.2330	0.0717	0.3297	0.7508	0.0167	1.5021	0.3050	0.1765	0.1015	0.3721	0.3916	0.0516	4.6497
PredRNN	0.1684	0.1312	0.0323	0.2167	0.7242	0.0205	1.8466	0.1317	0.0249	0.0044	0.1690	0.5427	0.0607	4.8525
SimVP	0.3422	0.3138	0.1347	0.4046	0.7922	0.0146	1.3098	0.2976	0.1929	0.1214	0.3581	0.6344	0.0398	**3.1807**
EarthFormer	0.2644	0.2100	0.0597	0.3094	0.7557	0.0165	1.4860	0.3036	0.1682	0.1088	0.4126	0.5777	0.0458	4.1189
NowcastNet	0.3321	0.2874	0.1021	0.3761	**0.8189**	**0.0128**	1.1514	0.3386	0.2171	0.1627	0.4130	0.6232	0.0413	3.7141
EarthFarseer	0.3300	0.2900	0.1284	0.3859	0.7829	0.0149	1.3387	0.3047	0.2048	0.1330	0.3829	0.6255	0.0406	3.2482
DiffCast	0.3352	0.3203	0.1829	0.4624	0.7808	0.0157	1.3923	0.3389	0.2231	0.1782	0.4494	0.6194	0.0432	3.8892
PastNet	0.3262	0.2982	0.1175	0.3855	0.7967	0.0145	1.3011	0.2871	0.1807	0.1095	0.3412	0.6364	0.0399	3.1897
AlphaPre	0.3821	0.3630	0.2001	0.4593	0.7965	0.0144	1.2936	0.2920	0.1920	0.1194	0.3569	0.6353	0.0407	3.2528
FlowCast	0.3177	**0.3812**	0.1540	0.4400	0.7922	0.0155	1.3923	0.3545	0.2431	0.1899	0.4842	0.6486	0.6486	3.8240
QFSNet	**0.3876**	0.3797	**0.2233**	**0.5643**	0.8067	0.0136	**1.1258**	**0.3814**	**0.2728**	**0.2165**	**0.5246**	**0.6667**	**0.0393**	3.5400

Metric	ConvLSTM	PredRNN	SimVP	EarthFormer	NowcastNet	EarthFarseer	DiffCast	PastNet	AlphaPre	FlowCast	**QFSNet**
FLOPs (G)↓	605.91	1171.72	191.88	711.05	34.65	7257.89	60852.22	169.00	3101.81	10520.12	**8.04**
Latency (ms)↓	32.72	53.89	60.81	347.02	22.36	516.91	14052.34	66.76	121.91	1298.60	**20.21**
Params (M)↓	0.74	23.84	14.87	8.11	34.93	148.57	61.14	15.78	89.03	160.41	**0.18**

5.2 Quantitative Results

Table 1 demonstrates that QFSNet achieves state-of-the-art extreme precipitation detection across four benchmarks, alongside competitive continuous fidelity and exceptional computational efficiency. QFSNet achieves CSI-40 of 0.2869 on Shanghai and CSI-219 of 0.0876 on SEVIR, outperforming AlphaPre by 9.8% and 57.3%, respectively. Consistent improvements extend to Meteo and CIKM, where CSI-32 reaches 0.2233 and CSI-40 reaches 0.2165, exceeding the strongest corresponding baselines by 11.6% and 14.0%, respectively. Moreover, QFSNet obtains the highest POD on all four datasets, demonstrating consistently stronger precipitation event recall across the evaluated benchmarks. These results support the effectiveness of directional spectral modeling in extreme event forecasting across diverse benchmarks. Crucially, these detection gains coexist with competitive continuous intensity fidelity. QFSNet achieves the lowest MAE on Shanghai and Meteo, and the highest SSIM on Shanghai and CIKM, although AlphaPre leads SEVIR in both metrics. These findings are consistent with the intended balance between boundary preservation and intra-region intensity calibration. Furthermore, QFSNet demonstrates exceptional computational efficiency compared to generative baselines that incur prohibitive inference latency. Our architecture requires merely 0.18 million parameters and 8.04 GFLOPs, representing the lowest computational footprint among all compared methods. This enables a deterministic inference latency of 20.21 milliseconds, establishing QFSNet as a highly practical solution that bridges the gap between high-fidelity extreme-event forecasting and the stringent real-time constraints of operational meteorology.

5.3 Ablation Studies

We conduct comprehensive ablations on the Shanghai dataset to validate our core designs. Table 2 summarizes the key results, while details of additional control experiments on components, capacity, hyperparameters, and random initialization are provided in Appendix E. We analyze the physical and geometric implications of each structural perturbation as follows.

Wavelet–Quaternion Representation and Directional Decoupling. Removing wavelet packing or replacing the two-sided QFT with a standard complex FFT causes a disproportionate drop in POD relative to the MAE increase. This indicates that standard complex spectra fail to disentangle heterogeneous directional phase shifts, leading to severe phase interference and blurred structural

Table 2: Key ablations on Shanghai. CSI-M and POD are averaged over thresholds $(20, 30, 35, 40)$.

Variant	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓
Full QFSNet	**0.4252**	**0.3882**	**0.2869**	**0.6020**	**0.8040**	**0.0192**	**1.7280**
W/o wavelet–quaternion packing	0.3993	0.3656	0.2695	0.5214	0.7820	0.0214	1.9291
Real/complex Fourier representation	0.4036	0.3689	0.2736	0.5236	0.7845	0.0210	1.8927
W/o fine spectral branch	0.3832	0.3428	0.2478	0.5553	0.7867	0.0218	1.9607
W/o LL branch	0.2507	0.2367	0.1817	0.3217	0.7659	0.0259	2.3355
W/o local phase branch	0.4173	0.3810	0.2811	0.5891	0.7994	0.0199	1.7866
W/o normal transport	0.4127	0.3775	0.2773	0.5954	0.7982	0.0200	1.8015
W/o post-normal tangent correction	0.4116	0.3743	0.2750	0.5499	0.7992	0.0197	1.7762
W/o both final corrections	0.4105	0.3756	0.2774	0.5569	0.7998	0.0196	1.7631

boundaries. By contrast, organizing directional sub-bands into a non-commutative quaternion field via the Hamilton product natively resolves these directional conflicts, preserving sharp structural discontinuities without phase cancellation.

Physical Separability and Complementary Branches. The tri-branch architecture respects the distinct physical scales of precipitation dynamics. Removing the Large-scale LL branch causes a catastrophic collapse across all metrics, confirming its irreplaceable role in macroscopic thermodynamic intensity evolution. Without this large-scale envelope, high-frequency branches lack the physical anchor to recover the baseline. Conversely, ablating the Fine-scale Spectral branch primarily degrades global directional coherence, validating its necessity for continuous high-frequency details. Removing the Local Phase Transport branch yields a minor drop in global metrics but consistent degradation in boundary-sensitive scores. This validates its specialized role as a localized corrective mechanism for spatially heterogeneous motion residuals and non-linear deformations that global spectral operators struggle to resolve.

Geometric Decoupling and the Over-smoothing Dilemma. The ablations of the Normal–Tangent Decomposition expose the inherent over-smoothing of continuous pixel-wise optimization. Notably, removing both corrections suffers from severe boundary blurring. Specifically, removing the Normal transport prevents the mass-conserving spatial redistribution required to reshape sharp structural contours. More critically, removing the Tangent correction causes the most severe POD drop. This reveals that while the normal update restores the geometry of the boundaries, the tangent update is indispensable for ensuring physical consistency within those restored regions. Without intra-region monotone calibration, the model produces spurious intensity fluctuations, causing borderline predictions to fail. Together, this orthogonality reconciles continuous neural optimization with the preservation of sharp physical structures.

6 Conclusion

In this paper, we presented QFSNet, a highly efficient deterministic precipitation nowcasting framework anchored by a Quaternion Fourier Neural Operator. By introducing a lossless wavelet-quaternion representation and a two-sided QFT, QFSNet explicitly disentangles scale and orientation, resolving the phase misalignment inherent in standard complex spectral methods. The tri-branch architecture strategically decouples fine-scale directional dynamics, macroscopic thermodynamic trends, and local non-linear phase transport, achieving comprehensive spatiotemporal modeling without the quadratic cost of global attention. Crucially, the normal–tangent decomposition directly mitigates the over-smoothing effect of continuous optimization, explicitly restoring sharp precipitation boundaries and ensuring structure-aware intensity predictions, thereby bridging the gap between pixel-wise regression and the preservation of high-contrast physical discontinuities. Extensive evaluations across four datasets demonstrate that QFSNet achieves state-of-the-art performance with substantial gains in extreme precipitation detection, while requiring the lowest parameter count and computational cost among all baselines.

AI USE STATEMENT

In this work, we used generative AI tools for assist in the writing of proofs, implement methods, assist with translation. We have not used generative AI tools for assist in the formulate mathematical claims, design or provide feedback on research methodology or experiments, provide critical ingredients for proving mathematical claims, design or provide feedback on research methodology or experiments, interpret results. and generate synthetic data sets, help develop theoretical models or conceptual frameworks, propose or refine hypotheses, clean and reformat dataset, support qualitative and thematic data analysis are not applicable to this work. Additionally, we used generative AI tools for create or edit software code, propose a title or keywords for a research paper. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by all authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Cong Bai, Feng Sun, Jinglin Zhang, Yi Song, and Shengyong Chen. Rainformer: Features extraction balanced network for radar-based precipitation nowcasting. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2022.3162882.

T. Bulow and G. Sommer. Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case. *IEEE Transactions on Signal Processing*, 49(11):2844–2852, 2001. doi: 10.1109/78.960432.

Lei Chen, Yuan Cao, Leiming Ma, and Junping Zhang. A deep learning-based methodology for precipitation nowcasting with radar. *Earth and Space Science*, 7(2):e2019EA000812, 2020.

Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z Li. Simvp: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3170–3180, 2022a.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Bernie Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35:25390–25403, 2022b.

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

Junchao Gong, Lei Bai, Peng Ye, Wanghan Xu, Na Liu, Jianhua Dai, Xiaokang Yang, and Wanli Ouyang. Cascast: Skillful high-resolution precipitation nowcasting via cascaded modelling. *arXiv preprint arXiv:2402.04290*, 2024.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11474–11484, 2020.

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

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

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Bernardo Perrone Ribeiro and Jana Faganeli Pucer. Flowcast: Advancing precipitation nowcasting with conditional flow matching. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and

A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 156863–156883, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/fe112a9042ed19ec515ed700d394a8bb-Paper-Conference.pdf.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson Ellen Clancy, Alberto Arribas†, and Shakir Mohamed. Skillful precipitation nowcasting using deep generative models of radar. *Nature*, 597:672–677, 2021.

Jenna Ritvanen, Bent Harnist, Miguel Aldana, Terhi Mäkinen, and Seppo Pulkkinen. Advection-free convolutional neural network for convective rainfall nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:1654–1667, 2023.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

M. Sifuzzaman, M. R. Islam, and M. Z. Ali. Application of wavelet transform and its advantages compared to fourier transform. *Journal of Physical Sciences*, 2009.

Gilbert Strang. Wavelet transforms versus fourier transforms. *Bull. Amer. Math. Soc.*, 1993.

Cheng Tan, Zhangyang Gao, Lirong Wu, Yongjie Xu, Jun Xia, Siyuan Li, and Stan Z Li. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18770–18782, 2023.

Cheng Tan, Zhangyang Gao, Siyuan Li, and Stan Z. Li. Simvpv2: Towards simple yet powerful spatiotemporal predictive learning. *IEEE Transactions on Multimedia*, 27:5170–5184, 2025. doi: 10.1109/TMM.2025.3543051.

Song Tang, Chuang Li, Pu Zhang, and RongNian Tang. Swinlstm: Improving spatiotemporal prediction accuracy using swin transformer and lstm. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13470–13479, 2023.

Mark Veillette, Siddharth Samsi, and Chris Mattioli. Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, 33:22009–22019, 2020.

Yunbo Wang, Mingsheng Long, Jianmin Wang, Zhifeng Gao, and Philip S Yu. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Advances in neural information processing systems*, 30, 2017.

Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, and S Yu Philip. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In *International conference on machine learning*, pp. 5123–5132. PMLR, 2018.

Yunbo Wang, Jianjin Zhang, Hongyu Zhu, Mingsheng Long, Jianmin Wang, and Philip S Yu. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9154–9162, 2019.

Hao Wu, Yuxuan Liang, Wei Xiong, Zhengyang Zhou, Wei Huang, Shilong Wang, and Kun Wang. Earthfarseer: Versatile spatio-temporal dynamical systems modeling in one model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15906–15914, 2024a.

Hao Wu, Fan Xu, Chong Chen, Xian-Sheng Hua, Xiao Luo, and Haixin Wang. Pastnet: Introducing physical inductive biases for spatio-temporal video prediction. In *Proceedings of the 32nd ACM international conference on multimedia*, pp. 2917–2926, 2024b.

Chiu-Wai Yan, Shi Quan Foo, Van Hoan Trinh, Dit-Yan Yeung, Ka-Hing Wong, and Wai-Kin Wong. Fourier amplitude and correlation loss: Beyond using l2 loss for skillful precipitation nowcasting. *Advances in Neural Information Processing Systems*, 37:100007–100041, 2024.

Yichen Yao and Zhongjie Li. Cikm analyticup 2017: short-term precipitation forecasting based on radar reflectivity images. In *Proceedings of the Conference on Information and Knowledge Management, Short-Term Quantitative Precipitation Forecasting Challenge, Singapore*, pp. 6–10, 2017.

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

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

A Mathematical Proofs and Theorems

A.1 Proofs for Normal-Tangent Decomposition

To substantiate the mathematical rigor of the normal-tangent decomposition mechanism introduced in Sec. 4.4, we provide rigorous derivations for its strict interval boundedness, exact identity Jacobian initialization, and mass-conserving boundary localization.

Strict interval boundedness of the tangent map.

By definition, the relative position is $u = \frac{P^{\mathrm{n}} - \lambda_k}{\lambda_{k+1} - \lambda_k}$. Since $P^{\mathrm{n}} \in [\lambda_k, \lambda_{k+1}]$, it follows that $u \in [0, 1]$. The tangent map is given by:

$$\widehat{Y}_{t,\mathbf{x}} = \lambda_k + (\lambda_{k+1} - \lambda_k)\sigma\big(\mathrm{logit}(u) + d_{k,\mathbf{x}}\big) \tag{16}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ satisfies $0 < \sigma\big(\mathrm{logit}(u) + d_{k,\mathbf{x}}\big) < 1$. Therefore, $\lambda_k < \widehat{Y}_{t,\mathbf{x}} < \lambda_{k+1}$. This strictly confines the output within the target warning interval, mathematically guaranteeing that the tangent calibration cannot inadvertently cause spurious boundary crossings.

Exact identity Jacobian at initialization.

Let $z = \mathrm{logit}(u) + d_k$. By the chain rule, the derivative of $\widehat{Y}$ with respect to P^{n} is:

$$\frac{\partial \widehat{Y}}{\partial P^{\mathrm{n}}} = \frac{\partial \widehat{Y}}{\partial z} \cdot \frac{\partial z}{\partial u} \cdot \frac{\partial u}{\partial P^{\mathrm{n}}} \tag{17}$$

where $\frac{\partial \widehat{Y}}{\partial z} = (\lambda_{k+1} - \lambda_k)\sigma'(z)$, $\frac{\partial z}{\partial u} = \frac{\partial}{\partial u} \ln\left(\frac{u}{1-u}\right) = \frac{1}{u(1-u)}$, and $\frac{\partial u}{\partial P^{\mathrm{n}}} = \frac{1}{\lambda_{k+1} - \lambda_k}$. Simplifying these terms yields $\frac{\partial \widehat{Y}}{\partial P^{\mathrm{n}}} = \sigma'\big(\mathrm{logit}(u) + d_k\big) \cdot \frac{1}{u(1-u)}$. For $d_k = 0$, let $v = \mathrm{logit}(u)$, then $\sigma(v) = \frac{1}{1+e^{-v}} = u$. Differentiating yields $\sigma'(v) = \sigma(v)(1 - \sigma(v)) = u(1-u)$. Substituting this back gives $\left.\frac{\partial \widehat{Y}}{\partial P^{\mathrm{n}}}\right|_{d_k=0} = u(1-u) \cdot \frac{1}{u(1-u)} = 1$. Thus, the tangent module acts as an exact identity mapping at initialization, ensuring strict gradient orthogonality to the normal stage.

Mass conservation and boundary localization in normal update.

Mass Conservation: The spatial shape flux $\mathbf{J}_t$ is passed through a $\mathrm{ZeroMean}$ operator, which by definition enforces $\langle \mathbf{J}_t \rangle_{xy} = \frac{1}{HW} \sum_{x,y} (\mathbf{J}_t)_{x,y} = 0$. Consequently, the spatial sum of the shape flux is strictly zero. The mass reaction term $\mathbf{R}_t = b_t M_t$ has a spatial mean of $\langle \mathbf{R}_t \rangle_{xy} = b_t \langle M_t \rangle_{xy} = b_t$. Therefore, the total mass change over the spatial domain is exactly $HW \cdot b_t$, proving that the shape redistribution is perfectly conservative.

Threshold Localization: The correction is modulated by the function $f(E) = 4E(1-E)$, where $E = \sigma(\kappa(P^0 - \lambda_k)) \in (0, 1)$. Taking the derivative of $f'(E) = 4 - 8E$ with respect to E and setting $f'(E) = 0$ yields the unique critical point $E = 0.5$. The second derivative $f''(E) = -8 < 0$ confirms that this extremum is a maximum with $f(0.5) = 1$. Therefore, the normal correction is strictly localized to the immediate vicinity of the boundary transition, leaving pixels far from the threshold unaffected by the shape flux.

A.2 Proof of the Two-sided QFT Shift Theorem

Here, we provide the rigorous derivation for the exact two-sided QFT shift property utilized in the Local Non-linear Phase Transport module. Let $Q \in \mathbb{H}^{H_w \times W_w}$ be a discrete quaternion field undergoing a spatial translation by $(\Delta y, \Delta x)$. By the definition of the discrete two-sided QFT and applying the change of variables, the translation-dependent phase terms can be factored out from the summation:

$$\mathcal{F}_Q[Q^{\mathrm{tr}}](\omega_y, \omega_x) = e^{-\mathbf{i}2\pi\omega_y \Delta y / H_w} \mathcal{F}_Q[Q](\omega_y, \omega_x) e^{-\mathbf{j}2\pi\omega_x \Delta x / W_w}. \tag{18}$$

Defining the phase angles and utilizing the quaternion Euler formula, the shifted spectrum is exactly formulated as $\mathcal{F}_Q[Q^{\mathrm{tr}}] = E_{\mathbf{i}}(-\theta_y) \otimes_H \mathcal{F}_Q[Q] \otimes_H E_{\mathbf{j}}(-\theta_x)$.

B Operator Complexity Analysis

Let $N = H_w W_w$ be the number of wavelet-grid locations, C_{in} and C_{out} the input and output quaternion channel counts, K the spatial kernel width, M the number of radial bases, and G the number of channel groups. Table 3 compares the dominant terms of one feature transformation.

Table 3: Operator-level complexity comparison. Counts suppress fixed quaternion arithmetic constants. The entry/exit QFT cost $\mathcal{O}(N(C_{\text{in}}+C_{\text{out}})\log N)$ is shared by the spectral trajectory. "Global" denotes a full-image receptive field. Depthwise rows assume $C_{\text{out}} = C_{\text{in}}$.

Operator	Parameters	Arithmetic	Receptive field	Structural interaction
Dense $K\times K$ conv	$\mathcal{O}(K^2 C_{\text{in}} C_{\text{out}})$	$\mathcal{O}(NK^2 C_{\text{in}} C_{\text{out}})$	Local $K\times K$	Unconstrained local mixing
Depthwise $K\times K$ conv	$\mathcal{O}(K^2 C_{\text{in}})$	$\mathcal{O}(NK^2 C_{\text{in}})$	Local $K\times K$	No native cross-channel
Global self-attention	$\mathcal{O}(C_{\text{in}}C_{\text{out}} + C_{\text{out}}^2)$	$\mathcal{O}(NC_{\text{in}}C_{\text{out}} + N^2 C_{\text{out}})$	Global	Data-dependent all-pair
Radial mixer (Ours)	$\mathcal{O}(MC_{\text{in}}C_{\text{out}}/G)$	$\mathcal{O}(MNC_{\text{in}}C_{\text{out}}/G)$	Global	Scale-shared Hamilton
Exact local conv (Ours)	$\mathcal{O}(K^2 C_{\text{in}})$	$\mathcal{O}(NC_{\text{in}})$	Local $K\times K$	Ordered left/right Hamilton

The comparison isolates why the proposed suite is efficient. Local convolution remains appropriate for short-range interactions, and we retain it only at the bottleneck. The radial mixer supplies global context without the quadratic spatial term of attention, while the exact convolution adds locality at linear spatial cost. Their complementary roles allow QFSNet to use a narrow hidden state instead of compensating with a deep convolutional stack or a dense attention map.

C Complete Model Specification

This section specifies the active QFSNet graph for the $T_{\text{in}} \to T_{\text{out}}$ forecasting protocol. Let B denote the batch size. The default reported instance uses $T_{\text{in}} = 5$, $T_{\text{out}} = 20$, $C = 1$, $H = W = 128$, and $H_w = W_w = 64$. A quaternion tensor is written as $[B, C_{\mathbb{H}}, 4, H, W]$, where $C_{\mathbb{H}}$ is the number of quaternion channels and the dimension of size four stores the $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$ components; it is not an additional learned channel dimension. Let C_{hid} denote the spectral width (number of quaternion channels).

C.1 End-to-end forward propagation

This section formalizes the active QFSNet graph for the $T_{\text{in}} \to T_{\text{out}}$ forecasting protocol as a step-by-step forward propagation algorithm. Let B denote the batch size. The default reported instance uses $T_{\text{in}} = 5$, $T_{\text{out}} = 20$, $C = 1$, $H = W = 128$, and $H_w = W_w = 64$. A quaternion tensor is written as $[B, C_{\mathbb{H}}, 4, H, W]$, where $C_{\mathbb{H}}$ is the number of quaternion channels and the dimension of size four stores the $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$ components.

The three forecasting branches operate in the unified wavelet–quaternion coordinate system. Their outputs are aggregated via element-wise addition before undergoing phase-consistency projection and spatial wavelet synthesis. The complete active computational path is summarized in Algorithm 1.

C.2 Fine-scale spectral branch

The fine branch uses eight radial basis functions and eight channel groups in all C_{hid}-channel residual blocks. Its Encoder, Translator, and Decoder preserve the 64×64 grid; "down" and "up" refer to the two sides of the translator stack rather than spatial resampling. Encoder/decoder attention depth is zero, so the corresponding containers are identities. Only the deepest translator block activates the exact 3×3 bi-sided Hamilton convolution. Table 4 lists every active block.

Each spectral residual block applies quaternion frequency normalization, an eight-basis radial Hamilton mixer, quaternion ModReLU, a grouped 1×1 Hamilton projection, channel shuffle, and a residual connection. The fine branch contains 163,917 trainable parameters.

Algorithm 1 End-to-End Forward Pass of QFSNet

Require: Input radar sequence $X \in \mathbb{R}^{B\times T_{\text{in}}\times C\times H\times W}$
Ensure: Forecast $\widehat{Y} \in \mathbb{R}^{B\times T_{\text{out}}\times C\times H\times W}$

1. Spectral Representation:
$X_{\text{norm}} \leftarrow \text{SpatiotemporalNorm}(X)$
$X_w \leftarrow \text{OneLevelHaarPack}(X_{\text{norm}})$ ▷ $X_w \in \mathbb{H}^{B\times T_{\text{in}}\times C\times H_w\times W_w}$
$X_q \leftarrow \text{TwoSidedQFT}(X_w)$ ▷ $X_q \in \mathbb{H}^{B\times T_{\text{in}}C\times H_w\times W_w}$

2. Parallel Branch Forecasting:
$Z_{\text{fine}} \leftarrow \text{FineBranch}(X_q)$ ▷ QFT encoder–translator–decoder
$Z_{\text{ll}} \leftarrow \text{LLBranch}(X_{\text{norm}})$ ▷ Scalar LL prediction and quaternion packing
$Z_{\text{phase}} \leftarrow \text{PhaseBranch}(X_q)$ ▷ Two-sided phase-shift increment

3. Fusion and Phase Consistency:
$Z_{\text{sum}} \leftarrow Z_{\text{fine}} + Z_{\text{ll}} + Z_{\text{phase}}$
$Z_{\text{cons}} \leftarrow \text{PhaseConsistency}(Z_{\text{sum}}, X_q)$ ▷ Global and 2×2 local projection

4. Spatial Synthesis and Calibration:
$P^0_{\text{synth}} \leftarrow \text{InverseHaarSynth}(Z_{\text{cons}})$
$P^0_{\text{read}} \leftarrow \text{ConvPixelShuffleReadout}(Z_{\text{cons}})$
$P^0 \leftarrow \text{BaseCalibration}(P^0_{\text{synth}}, P^0_{\text{read}})$ ▷ Intensity and moment heads

5. Physical Refinement:
$P^n \leftarrow \text{NormalStage}(P^0, X)$ ▷ Conservative shape/mass transport
$\widehat{Y} \leftarrow \text{TangentStage}(P^n, X)$ ▷ Shared conditional moment map
return $\widehat{Y}$

Table 4: Layer-wise specification of the fine-scale QFT branch. All tensors remain on the 64×64 wavelet grid.

Module	Submodule	Input	Output	Q-ch	Configuration
Encoder	Q-frequency Norm	$[B, T_{\text{in}}C, 4, H_w, W_w]$	$[B, T_{\text{in}}C, 4, H_w, W_w]$	$T_{\text{in}}C \to T_{\text{in}}C$	Quaternion normalization
	Spectral Linear	$[B, T_{\text{in}}C, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$T_{\text{in}}C \to C_{\text{hid}}$	Radial Hamilton Linear, ModReLU
	Spectral ResBlock	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Norm, radial Linear, pointwise Linear, shuffle
Translator	Down ResBlock $\times 2$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Spectral residual blocks
	Bottleneck ResBlock	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Radial Linear plus bi-sided 3×3 Conv
	Skip Linear $\times 2$	$2\times[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$2C_{\text{hid}} \to C_{\text{hid}}$	Concatenation, quaternion Linear
	Up ResBlock $\times 2$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Spectral residual blocks
Decoder	Spectral ResBlock	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Spectral residual block
	Q-frequency Norm	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$C_{\text{hid}} \to C_{\text{hid}}$	Quaternion normalization
	Output Spectral Linear	$[B, C_{\text{hid}}, 4, H_w, W_w]$	$[B, T_{\text{out}}C, 4, H_w, W_w]$	$C_{\text{hid}} \to T_{\text{out}}C$	Low-rank radial Hamilton Linear

C.3 LL, PHASE, FUSION, AND CALIBRATION MODULES

Table 5 details the remaining pre-normal modules. The phase branch returns a gated wavelet-coordinate motion residual relative to the last observation. The two phase-consistency projections operate after branch addition: the first uses one global region and the second uses a 2×2 regional partition. In Table 5, $\boldsymbol{v}$ and $\boldsymbol{c}$ denote local velocity and curvature, while r and r^c are their corresponding confidence scores. We denote the real-valued hidden widths of the fusion decoder and phase-consistency controller by C_{fus} and C_{pc}; their reported values are $C_{\mathrm{fus}} = 8$ and $C_{\mathrm{pc}} = 16$.

Table 5: Detailed specification of the LL, phase, fusion, and base-calibration modules. DW denotes depthwise convolution, $P = 16$ is the number of local phase regions, and BT_{out} folds batch and forecast lead.

Module	Submodule	Input shape	Output shape	Channels	Configuration
LL branch	Temporal Linear	$[B, T_{\mathrm{in}}, C, H_w, W_w]$	$[B, T_{\mathrm{out}}, C, H_w, W_w]$	$T_{\mathrm{in}} \to T_{\mathrm{out}}$	Per-pixel projection
	Spatial Conv	$[BT_{\mathrm{out}}, C, H_w, W_w]$	$[BT_{\mathrm{out}}, C, H_w, W_w]$	$C \to C$	5×5 smoothing
	Pointwise Conv	$[BT_{\mathrm{out}}, C, H_w, W_w]$	$[BT_{\mathrm{out}}, 4C, H_w, W_w]$	$C \to 4C$	1×1 Conv
	DW ResBlock $\times 4$	$[BT_{\mathrm{out}}, 4C, H_w, W_w]$	$[BT_{\mathrm{out}}, 4C, H_w, W_w]$	$4C \to 4C$	3×3, dilations $(1, 2, 3, 1)$
	Pointwise Conv	$[BT_{\mathrm{out}}, 4C, H_w, W_w]$	$[B, T_{\mathrm{out}}, C, H_w, W_w]$	$4C \to C$	Output projection and reshape
	Quaternion packing	$[B, T_{\mathrm{out}}, C, H_w, W_w]$	$[B, T_{\mathrm{out}}C, 4, H_w, W_w]$	$C \to C$ Q-ch	Scalar component only
Phase branch	Phase correlation	$Q_0^{\mathrm{in}} : [B, T_{\mathrm{in}}, C, H_w, W_w]$	$\boldsymbol{v}, \boldsymbol{c} : [B, P, C, 2]$; $r, r^c : [B, P, C]$	C in P regions	$T_{\mathrm{in}} - 1$ temporal pairs
	Bi-sided phase shift	$Z^{\mathrm{in}} : [B, T_{\mathrm{in}}C, 4, H_w, W_w]$; $Q^{\mathrm{in}} : [B, T_{\mathrm{in}}, C, 4, H_w, W_w]$	$[B, T_{\mathrm{out}}C, 4, H_w, W_w]$	$T_{\mathrm{in}}C \to T_{\mathrm{out}}C$	Gated overlap-added residual
Consistency	Global projection	$Q^f : [B, T_{\mathrm{out}}C, 4, H_w, W_w]$; $Q^{\mathrm{in}} : [B, T_{\mathrm{in}}, C, 4, H_w, W_w]$	$[B, T_{\mathrm{out}}C, 4, H_w, W_w]$	$T_{\mathrm{out}}C$	MLP $9 \to C_{\mathrm{pc}} \to 2$
	Local projection	$Q^f : [B, T_{\mathrm{out}}C, 4, H_w, W_w]$; $Q^{\mathrm{in}} : [B, T_{\mathrm{in}}, C, 4, H_w, W_w]$	$[B, T_{\mathrm{out}}C, 4, H_w, W_w]$	$T_{\mathrm{out}}C$	2×2 regions; same MLP width
Fusion	Inverse Haar	$[B, T_{\mathrm{out}}C, 4, H_w, W_w]$	$[B, T_{\mathrm{out}}, C, H, W]$	Q-ch→real	Fixed synthesis
	3×3 Conv	$[BT_{\mathrm{out}}, 4C, H_w, W_w]$	$[BT_{\mathrm{out}}, C_{\mathrm{fus}}, H_w, W_w]$	$4C \to C_{\mathrm{fus}}$	Group norm, SiLU
	Conv–PixelShuffle	$[BT_{\mathrm{out}}, C_{\mathrm{fus}}, H_w, W_w]$	$[BT_{\mathrm{out}}, C_{\mathrm{fus}}, H, W]$	$C_{\mathrm{fus}} \to C_{\mathrm{fus}}$	Conv to $4C_{\mathrm{fus}}$, upsampling $\times 2$
	1×1 Conv	$[BT_{\mathrm{out}}, C_{\mathrm{fus}}, H, W]$	$[B, T_{\mathrm{out}}, C, H, W]$	$C_{\mathrm{fus}} \to C$	Shared over leads and reshape
Calibration	Monotone map	$P^0 : [B, T_{\mathrm{out}}, C, H, W]$; $X : [B, T_{\mathrm{in}}, C, H, W]$	$[B, T_{\mathrm{out}}, C, H, W]$	$C \to C$	Threshold-invariant

C.4 NORMAL–TANGENT REFINEMENT

The normal stage combines an 8×8 quaternion response with a learned 64×64 conservative correction. Its seven fine-grid input channels are the future field, local mean, local maximum, latest observation, temporal tendency, temporal curvature, and one coarse response channel. The two-channel readout predicts shape-flux and mass-reaction logits. The tangent stage reuses the conditional moment head applied during base calibration; no second set of tangent parameters is introduced. Let C_{n}, C_{m}, and C_{d} denote the real-valued hidden widths of the normal, moment, and fine-detail paths, respectively; their values are $C_{\mathrm{n}} = 16$, $C_{\mathrm{m}} = 24$, and $C_{\mathrm{d}} = 12$.

Table 6: Layer-wise specification of the normal transport and shared conditional moment/tangent head. Shapes with BT_{out} fold batch and forecast lead for 2-D convolution.

Module	Submodule	Input shape	Output shape	Channels	Configuration
Normal coarse	Pooling and packing	prediction and history	$[B, T_{\mathrm{out}}C, 4, 8, 8]$	$T_{\mathrm{out}}C$ Q-ch	Four physical components
	Quaternion Linear	$[B, T_{\mathrm{out}}C, 4, 8, 8]$	$[B, 4, 4, 8, 8]$	$T_{\mathrm{out}}C \to 4$	ModReLU
	Quaternion Linear	$[B, 4, 4, 8, 8]$	$[B, T_{\mathrm{out}}C, 4, 8, 8]$	$4 \to T_{\mathrm{out}}C$	Coarse readout
Normal fine	1×1 Conv	$[BT_{\mathrm{out}}, 7C, H_w, W_w]$	$[BT_{\mathrm{out}}, C_{\mathrm{n}}, H_w, W_w]$	$7C \to C_{\mathrm{n}}$	Feature projection
	DW ResBlock $\times 3$	$[BT_{\mathrm{out}}, C_{\mathrm{n}}, H_w, W_w]$	$[BT_{\mathrm{out}}, C_{\mathrm{n}}, H_w, W_w]$	$C_{\mathrm{n}} \to C_{\mathrm{n}}$	3×3, dilations $(1, 2, 3)$
	1×1 Conv	$[BT_{\mathrm{out}}, C_{\mathrm{n}}, H_w, W_w]$	$[BT_{\mathrm{out}}, 2C, H_w, W_w]$	$C_{\mathrm{n}} \to 2C$	Shape/mass logits
Normal output	Composition	coarse/fine responses	$P^n : [B, T_{\mathrm{out}}, C, H, W]$	$C \to C$	Zero-mean flux, mass reaction
Moment head	Concatenation	forecast and history	$[BT_{\mathrm{out}}, (10C + 4), H, W]$	$(10C + 4)$ real ch	Conditional features
	1×1 Conv	$[BT_{\mathrm{out}}, (10C + 4), H, W]$	$[BT_{\mathrm{out}}, C_{\mathrm{m}}, H, W]$	$(10C + 4) \to C_{\mathrm{m}}$	Group norm, SiLU
	DW ResBlock $\times 4$	$[BT_{\mathrm{out}}, C_{\mathrm{m}}, H, W]$	$[BT_{\mathrm{out}}, C_{\mathrm{m}}, H, W]$	$C_{\mathrm{m}} \to C_{\mathrm{m}}$	5×5, dilations $(1, 2, 4, 8)$
	1×1 Conv	$[BT_{\mathrm{out}}, C_{\mathrm{m}}, H, W]$	$[BT_{\mathrm{out}}, C, H, W]$	$C_{\mathrm{m}} \to C$	Moment readout
	Fine Conv path	$[BT_{\mathrm{out}}, (9C + 4), H, W]$	$[BT_{\mathrm{out}}, C, H, W]$	$(9C + 4) \to C_{\mathrm{d}} \to C$	5×5 and 1×1 Conv
Tangent output	Composition	P^n and calibrated field	$\widehat{Y} : [B, T_{\mathrm{out}}, C, H, W]$	$C \to C$	Interval-preserving map

C.5 PARAMETER ALLOCATION AND INACTIVE MODULES

Table 7 reports exact trainable parameter counts from the active implementation. Counts are mutually exclusive and sum to 185,433. Fixed Haar transforms, QFT/IQFT operations, tensor additions, and range maps contain no trainable parameters.

Table 7: Trainable parameter allocation of the default C_{hid}-channel QFSNet (with $C_{\mathrm{hid}} = 64$).

Active component	**Parameters**	**Share**
Fine-scale QFT encoder–translator–decoder	163,917	88.40%
LL predictor, temporal projector, and bounded gate	518	0.28%
Local non-linear phase transport	48	0.03%
Global and local phase-consistency projections	556	0.30%
Wavelet fusion decoder and fusion scale	2,634	1.42%
Threshold-invariant intensity calibrator	100	0.05%
Shared conditional moment/tangent head	16,134	8.70%
Normal shape/mass transport	1,526	0.82%
Total	**185,433**	**100.00%**

D Detailed Experimental Configuration

D.1 Datasets and Preprocessing

We evaluate QFSNet on four widely used benchmark datasets for precipitation nowcasting, covering diverse geographic regions, radar types, and weather patterns. A summary of the key statistics is provided in Table 8. To ensure a fair comparison and compatibility with our wavelet-quaternion operators, all input sequences are uniformly preprocessed to single-channel 128×128 spatial resolution and normalized to the $[0, 1]$ range.

Table 8: Summary of the four benchmark datasets used in this study.

Dataset	**Region**	**Physical Quantity**	**Native Resolution**	**Temporal Interval**	**Protocol**	**Train / Eval**
Shanghai	Shanghai, China	Composite reflectivity (dBZ)	$\sim$1 km	6 min	$5 \to 20$	1,534 / 526
SEVIR	Contiguous US	NEXRAD VIL mosaic	$\sim$1 km	5 min	$5 \to 20$	20,750 / 8,527
Meteo	France	Rain radar composite (CRF)	$\sim$1 km	5 min	$5 \to 20$	6,308 / 1,310
CIKM	Eastern China	Radar reflectivity (4 heights)	$\sim$1 km	6 min	$5 \to 10$	8,000 / 4,000

Shanghai Radar Dataset (Chen et al., 2020). The original data of Shanghai Radar dataset[1] spans from October 2015 to July 2018, containing 170,000 echo images generated by volume scans at approximately 6-minute intervals. The data is interpolated onto a $0.01^\circ \times 0.01^\circ$ longitude-latitude grid, covering a 510 km $\times$ 510 km region. Following DiffCastYu et al. (2024), we process the data into 25 frames per event. The maximum pixel value of each frame is clipped to 90, and the thresholds are set to [20, 30, 35, 40].

SEVIR Veillette et al. (2020). The Storm EVent ImagRy (SEVIR) dataset[2] is an annotated, curated, and spatio-temporally aligned collection encompassing five modalities, including visible and infrared satellite imagery, NEXRAD radar mosaics of Vertically Integrated Liquid (VIL), and ground lightning events. In this work, we focus on short-term weather forecasting and utilize the VIL radar mosaics as the primary data source. The dataset comprises 20,393 weather events recorded across the continental United States from 2017 to 2020. Each event consists of a 4-hour image sequence sampled at 5-minute intervals, covering a spatial extent of $384\,\mathrm{km} \times 384\,\mathrm{km}$. Following the prediction protocol of forecasting 20 future frames (100 min) given 5 observed frames (25 min), we adopt the strategy in Gao et al. (2022b) to sample 25 consecutive frames with a stride of 12 from each event. The dataset is chronologically split into training, validation, and test sets using January 1, 2019, and June 1, 2019, as the respective cutoff dates. Pixel values are rescaled to the range $[0, 255]$, and the binarization thresholds for computing CSI and POD are set to $[16, 74, 133, 160, 181, 219]$, consistent with the original SEVIR benchmark Veillette et al. (2020).

[1] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2GKMQJ

[2] https://nbviewer.org/github/MIT-AI-Accelerator/eie-sevir/blob/master/examples/SEVIR_Tutorial.ipynb

Meteo Dataset (Larvor & Berthomier, 2021). MeteoNet[3] is a large-scale multimodal meteorological dataset integrating full time series of satellite imagery, radar composites, numerical weather model outputs, and ground-based observations. It covers a geographic region of $550\,\text{km} \times 550\,\text{km}$ in the northwestern quarter of France, with observations recorded every 5 minutes from 2016 to 2018. Similar to the SEVIR preprocessing pipeline, we partition the radar sequences spanning 2016–2018 into training, validation, and test sets using January 1, 2018, and June 1, 2018, as temporal boundaries. We then apply a stride-20 sliding window filtering algorithm to extract precipitation events and suppress background noise, utilizing a mean pixel threshold T_{pixel} as the event activation criterion. The pixel intensity range is normalized to $[0, 70]$, and the evaluation thresholds for CSI and POD are set to $[12, 18, 24, 32]$ following Larvor & Berthomier (2021).

CIKM AnalytiCup 2017 (Yao & Li, 2017). The CIKM dataset[4] originates from the CIKM AnalytiCup 2017 Competition and records precipitation events within a $101\,\text{km} \times 101\,\text{km}$ region in Guangdong Province, China. Each sample contains 15 consecutive radar echo maps with a 6-minute temporal interval. Following Yu et al. (2024), we pad each echo map to a spatial resolution of 128×128 and adopt the original competition split for training, validation, and testing. Pixel values are converted to reflectivity in the range of $[0, 76]$ dBZ, and the evaluation thresholds are set to $[20, 30, 35, 40]$ for CSI and POD computation.

D.2 Metrics and Thresholds

We evaluate the models using five standard metrics. For threshold-based event detection, let H, M, and F denote the number of hit, miss, and false alarm pixels at a specific intensity threshold. The Critical Success Index (CSI) and Probability of Detection (POD) are computed as:

$$\text{CSI} = \frac{H}{H + M + F}, \qquad \text{POD} = \frac{H}{H + M}. \tag{19}$$

For continuous intensity fidelity, we use Structural Similarity (SSIM), Mean Absolute Error (MAE), and Continuous Ranked Probability Score (CRPS). Let x and y be the predicted and ground-truth image patches, with local means μ_x, μ_y, variances σ_x^2, σ_y^2, and covariance σ_{xy}:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{20}$$

where c_1, c_2 are stability constants to avoid division by zero. The MAE is defined as:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|, \tag{21}$$

where N is the total number of pixels. Finally, the CRPS evaluates the predictive cumulative distribution function (CDF) F against the Heaviside step function $H(\cdot)$ of the observation y:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} \left(F(x) - H(x - y)\right)^2 dx. \tag{22}$$

For the deterministic point predictions in this work, the CDF F collapses to a step function, and CRPS strictly simplifies to the normalized MAE. The specific thresholds used are $(20, 30, 35, 40)$ for Shanghai and CIKM, $(16, 74, 133, 160, 181, 219)$ for SEVIR, and $(12, 18, 24, 32)$ for Meteo.

E Additional Ablation Studies

E.1 Component and Branch Controls

We extend the ablations in Sec. 5.3 to examine three aspects of QFSNet: backbone replacement, single-branch forecasting, and the contribution of exact local convolution. Table 9 reports these controls on Shanghai. Unlike the leave-one-branch-out experiments in Table 2, the single-branch variants retain only one of the fine-scale spectral, large-scale LL, or local phase-transport branches.

[3] https://meteofrance.github.io/meteonet/english/data/rain-radar/

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

Table 9: Additional component controls on Shanghai.

Variant	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓
Full QFSNet	**0.4252**	**0.3882**	**0.2869**	**0.6020**	**0.8040**	**0.0192**	**1.7280**
Wavelet-domain conv backbone	0.4146	0.3845	0.2816	0.5836	0.8011	0.0198	1.7318
Fine only (w/o LL & phase)	0.4178	0.3817	0.2816	0.5893	0.7993	0.0199	1.7873
LL only (w/o fine & phase)	0.3832	0.3428	0.2478	0.5553	0.7867	0.0218	1.9607
Phase only (w/o fine & LL)	0.2973	0.2248	0.0880	0.3978	0.7710	0.0225	2.0274
w/o exact local convolution	0.4011	0.3609	0.2636	0.5557	0.7962	0.0202	1.8198

Wavelet-domain convolutional backbone. To assess the contribution of the proposed spectral backbone beyond wavelet-domain convolution, we replace the complete frequency backbone with a wavelet-domain convolutional backbone while retaining the LL, phase, fusion, normal, and tangent modules. As shown in Table 9, this replacement reduces CSI-M from 0.4252 to 0.4146 and POD from 0.6020 to 0.5836. CSI-35 and CSI-40 also decrease from 0.3882 and 0.2869 to 0.3845 and 0.2816, respectively. Reconstruction quality follows the same direction: SSIM decreases from 0.8040 to 0.8011, while CRPS and MAE increase from 0.0192 and 1.7280 to 0.0198 and 1.7318. Although the MAE difference is small, the proposed spectral backbone achieves better results across all reported metrics, improving precipitation-event detection without sacrificing reconstruction quality. These results support retaining the spectral backbone within QFSNet, rather than replacing it with convolutional processing of the wavelet representation. Since the surrounding modules are retained and the backbone is replaced as a whole, this comparison evaluates the overall backbone design rather than the quaternion representation in isolation.

Single-branch forecasting. Among the single-branch variants, the fine-only configuration performs best across all reported metrics, reaching a CSI-M of 0.4178. The LL-only and phase-only configurations obtain 0.3832 and 0.2973, respectively. The phase-only configuration also yields a CSI-40 of 0.0880, consistent with the phase branch's intended role as a motion-residual correction rather than a complete forecasting pathway. Nevertheless, full QFSNet outperforms the fine-only configuration on every reported metric: CSI-M increases from 0.4178 to 0.4252, POD from 0.5893 to 0.6020, and MAE decreases from 1.7873 to 1.7280. These results support combining fine-scale spectral evolution, large-scale LL prediction, and local phase transport instead of relying on a single forecasting branch.

Exact local convolution. Removing the exact bi-sided Hamilton convolution at the translator bottleneck decreases CSI-M from 0.4252 to 0.4011 and CSI-40 from 0.2869 to 0.2636, while increasing MAE from 1.7280 to 1.8198. This degradation in both event detection and continuous-intensity accuracy supports retaining the local operator alongside the global radial mixer, providing evidence for the benefit of complementary global and local processing within the proposed spectral backbone.

E.2 Event-loss Weight Sensitivity

We vary the weight λ_e of the boundary-aware asymmetric event loss while retaining the complete architecture and all other training settings. Setting $\lambda_e = 0$ removes the event term and leaves only the continuous reconstruction objective. Table 10 reports the resulting threshold detection and continuous-intensity metrics.

Table 10: Sensitivity to the event-loss weight λ_e on Shanghai. The selected $\lambda_e = 0.1$ row uses the QFSNet result reported in the main paper; all other rows are obtained from the controlled weight sweep.

λ_e	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓
0.0	0.3130	0.2352	0.1474	0.3565	0.8017	**0.0188**	**1.6878**
0.1 (selected)	**0.4252**	**0.3882**	**0.2869**	**0.6020**	**0.8040**	0.0192	1.7280
0.2	0.4144	0.3775	0.2782	0.5977	0.7976	0.0205	1.8457
0.3	0.4150	0.3784	0.2789	0.6057	0.7968	0.0209	1.8820
0.4	0.4154	0.3791	0.2796	0.6084	0.7961	0.0213	1.9140
0.5	0.4145	0.3789	0.2785	0.6142	0.7947	0.0220	1.9785
0.6	0.4143	0.3788	0.2781	0.6147	0.7944	0.0223	2.0076
0.7	0.4137	0.3783	0.2771	0.6154	0.7938	0.0226	2.0341
0.8	0.4131	0.3780	0.2760	0.6157	0.7934	0.0228	2.0499
0.9	0.4119	0.3776	0.2721	0.6166	0.7931	0.0229	2.0627
1.0	0.4152	0.3796	0.2787	0.6115	0.7939	0.0228	2.0507

Without the event term, the model attains low continuous error but loses the ability to recover threshold exceedances, reducing CSI-M from 0.4252 to 0.3130 and POD from 0.6020 to 0.3565. Larger weights restore event recall, but the gain saturates and increasingly biases optimization away from amplitude fidelity: from $\lambda_e = 0.2$ to 1.0, SSIM decreases while CRPS and MAE increase markedly. We therefore use $\lambda_e = 0.1$, which provides the best overall balance between threshold detection and continuous reconstruction.

E.3 Random-seed Robustness

We assess optimization stability by training the complete model from scratch with three random seeds while keeping the architecture, data split, and all other optimization settings fixed.

Table 11: Random-seed robustness of QFSNet on Shanghai with $\lambda_e = 0.1$. The final row reports the mean $\pm$ sample standard deviation over three independent training runs, computed from the displayed per-seed values.

Seed	CSI-M↑	CSI-35↑	CSI-40↑	POD↑	SSIM↑	CRPS↓	MAE↓
42	0.4252	0.3882	0.2869	0.6020	0.8040	0.0192	1.7280
43	0.4231	0.3862	0.2868	0.5879	0.8039	0.0192	1.7314
44	0.4241	0.3872	0.2868	0.6015	0.8037	0.0192	1.7259
Mean $\pm$ std	0.4241 ± 0.0011	0.3872 ± 0.0010	0.2868 ± 0.0001	0.5971 ± 0.0080	0.8039 ± 0.0002	0.0192 ± 0.0000	1.7284 ± 0.0028

As shown in Table 11, all three independently trained models attain closely matched final performance. CSI-M ranges from 0.4231 to 0.4252, with a mean and sample standard deviation of 0.4241 ± 0.0011, while CSI-40 remains between 0.2868 and 0.2869. Continuous-intensity accuracy is also consistent, with SSIM of 0.8039 ± 0.0002 and MAE of 1.7284 ± 0.0028. Although POD exhibits more variation across seeds (0.5971 ± 0.0080), changing the seed does not lead to a substantial deterioration in overall predictive quality in these runs. Together, these results provide empirical evidence of stable optimization across the tested seeds and are consistent with convergence to solutions of similar predictive quality, rather than reliance on a single favorable initialization.

E.4 Capacity Ablation Study

We examine whether the predictive gains of QFSNet arise simply from model capacity by independently varying the hidden quaternion width H_q, the number of Q-Spectral translator blocks D_q, and the number of radial basis functions M. Each sweep changes one factor while fixing the other two at the default configuration $(H_q, D_q, M) = (64, 3, 8)$. All variants use the same Shanghai split and optimization protocol. Parameter counts include all trainable modules.

Table 12 reports predictive accuracy and computational cost jointly. The QFSNet reference row reproduces the result reported in the main comparison and is listed only once. Consequently, the repeated default points $H_q = 64$, $D_q = 3$, and $M = 8$ from the three new sweeps are omitted.

Table 12: Capacity ablations on Shanghai. H_q, D_q, and M denote hidden quaternion width, translator depth, and radial basis count, respectively. The QFSNet row uses the result reported in the main paper; each remaining row changes only the indicated capacity factor.

Configuration	H_q	D_q	M	CSI-M↑	CSI-35↑	CSI-40↑	MAE↓	Params (M)↓	FLOPs (G)↓	Latency (ms)↓
QFSNet	64	3	8	**0.4252**	**0.3882**	**0.2869**	**1.7280**	0.185	8.04	20.21
Width sweep	32	3	8	0.4086	0.3708	0.2711	1.8254	0.069	4.49	19.28
	128	3	8	0.4180	0.3828	0.2822	1.7779	0.624	22.76	25.13
	256	3	8	0.4170	0.3807	0.2794	1.7528	2.324	78.60	34.53
Depth sweep	64	2	8	0.4125	0.3757	0.2765	1.7905	0.144	6.95	20.20
	64	4	8	0.4121	0.3751	0.2751	1.7892	0.227	9.72	22.79
	64	5	8	0.4139	0.3779	0.2786	1.7678	0.268	11.10	24.67
	64	6	8	0.4137	0.3766	0.2779	1.7717	0.309	12.48	25.37
	64	7	8	0.4144	0.3779	0.2779	1.7764	0.351	13.86	26.64
	64	8	8	0.4106	0.3741	0.2756	1.7855	0.392	15.25	27.92
Basis sweep	64	3	4	0.4121	0.3746	0.2754	1.7952	0.118	6.08	20.01
	64	3	12	0.4092	0.3707	0.2709	1.7919	0.253	10.58	21.70
	64	3	16	0.4120	0.3753	0.2749	1.7817	0.321	12.83	22.65
	64	3	20	0.4079	0.3698	0.2706	1.7942	0.388	15.08	23.53
	64	3	24	0.4076	0.3691	0.2698	1.7729	0.456	17.33	24.41

The width sweep shows that increasing H_q from 32 to 128 improves CSI-M, but further expansion to 256 provides no additional event-detection gain while raising the cost to 2.324M parameters, 78.60G FLOPs, and 34.53ms. Increasing translator depth produces similarly limited returns: CSI-M remains between 0.4106 and 0.4144 for depths 2–8, whereas FLOPs and latency grow steadily. The radial-basis sweep also saturates early; expanding M beyond 8 increases spectral mixing cost without a consistent improvement in CSI.

F More Qualitative Results

Figures 4, 5, 6, and 7 present visualization examples of the prediction results from the baselines and QSFNet on the Shanghai, SEVIR, Meteo, and CIKM datasets, respectively. Convolution- and Transformer-based models, such as ConvLSTM and Earthformer, tend to produce blurred prediction results. Optical flow-based methods, such as NowcastNet, suffer from severe artifacts in their predictions. Spectral domain-based methods, lacking directional displacement modeling, are prone to predicting erroneous motion trajectories in rapidly evolving precipitation systems. Generative model-based methods incur high computational complexity and lack explicit motion information modeling, leading to deviations in the predicted motion direction. Benefiting from wavelet quaternion modeling and global receptive field processing in the quaternion spectral domain, our QSFNet can efficiently and accurately predict the future motion trends of precipitation.

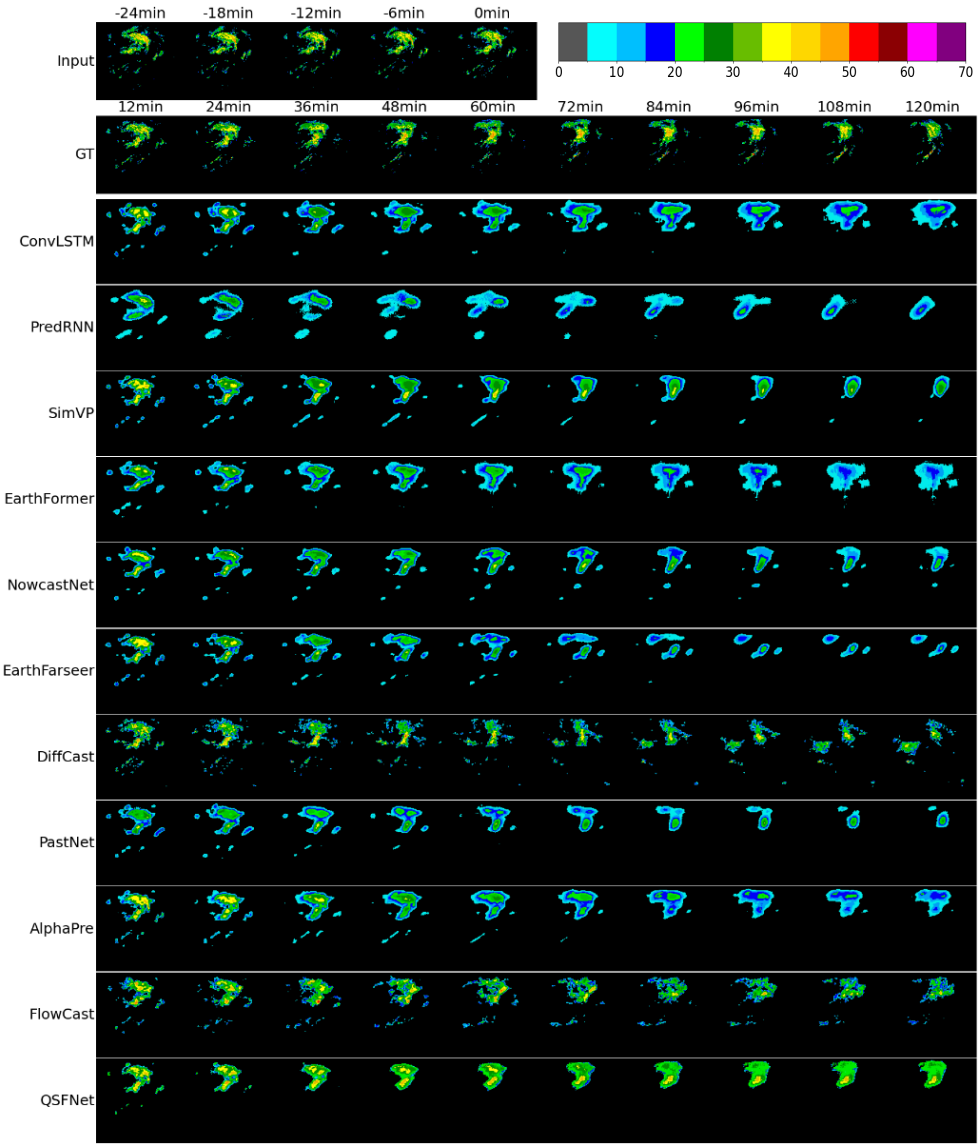

Figure 4: Qualitative comparison of QFSNet with baselines on the Shanghai dataset.

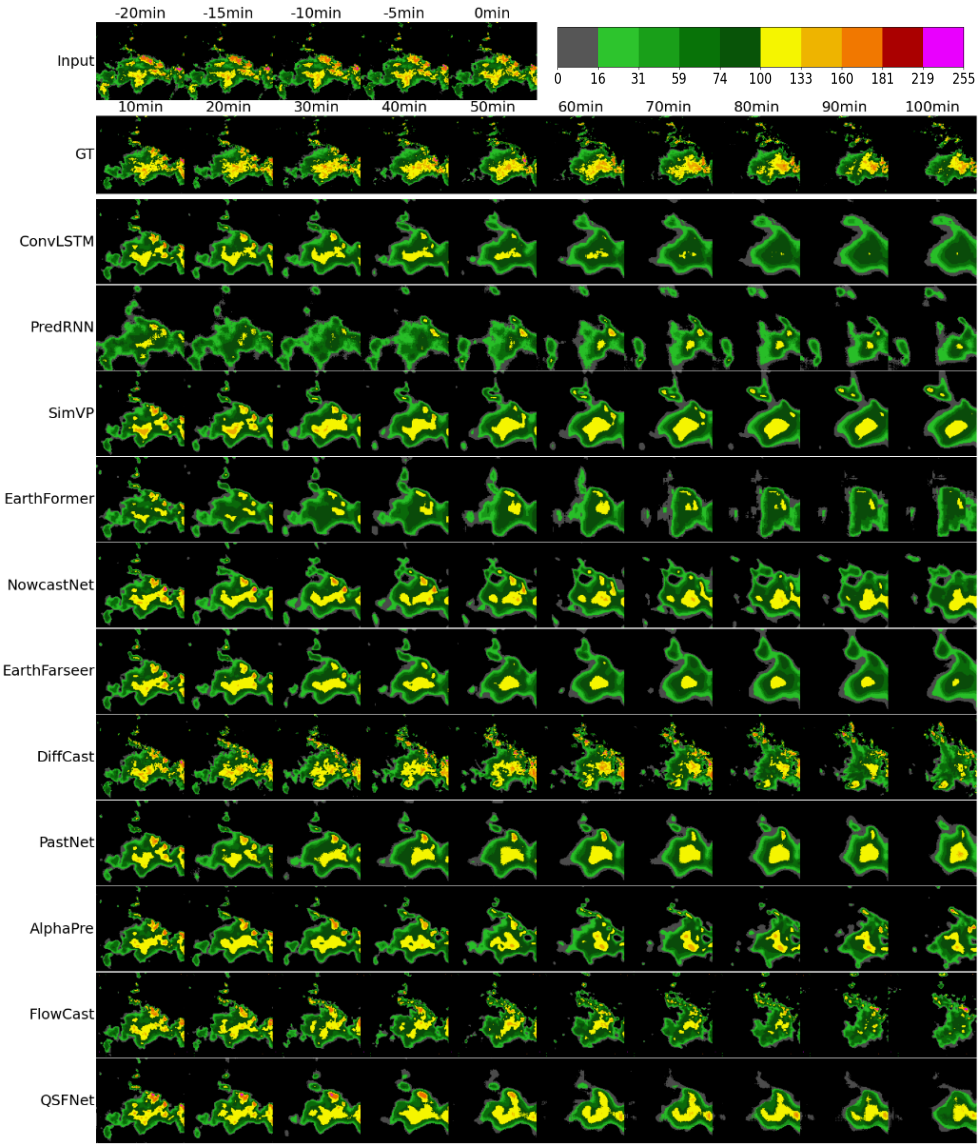

Figure 5: Qualitative comparison of QFSNet with baselines on the SEVIR dataset.

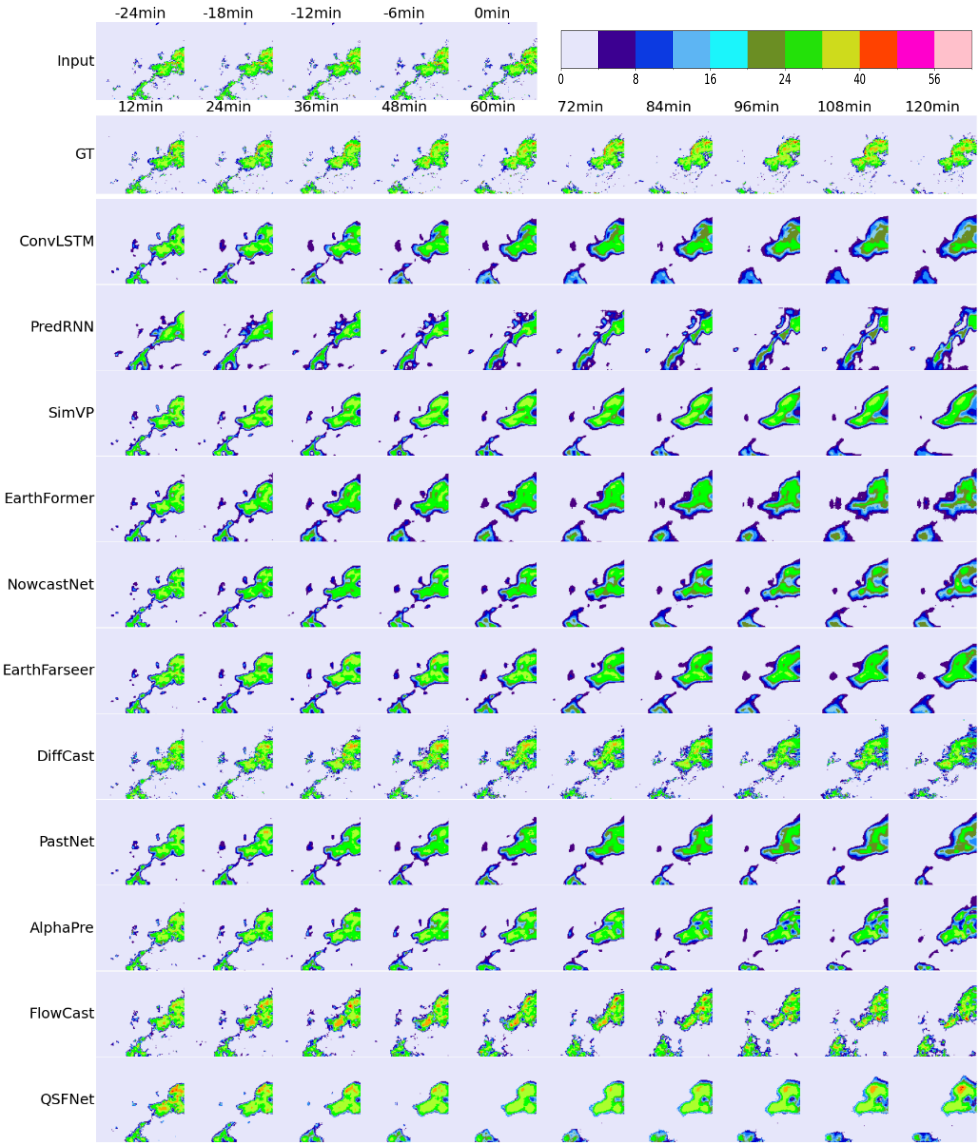

Figure 6: Qualitative comparison of QFSNet with baselines on the Meteo dataset.

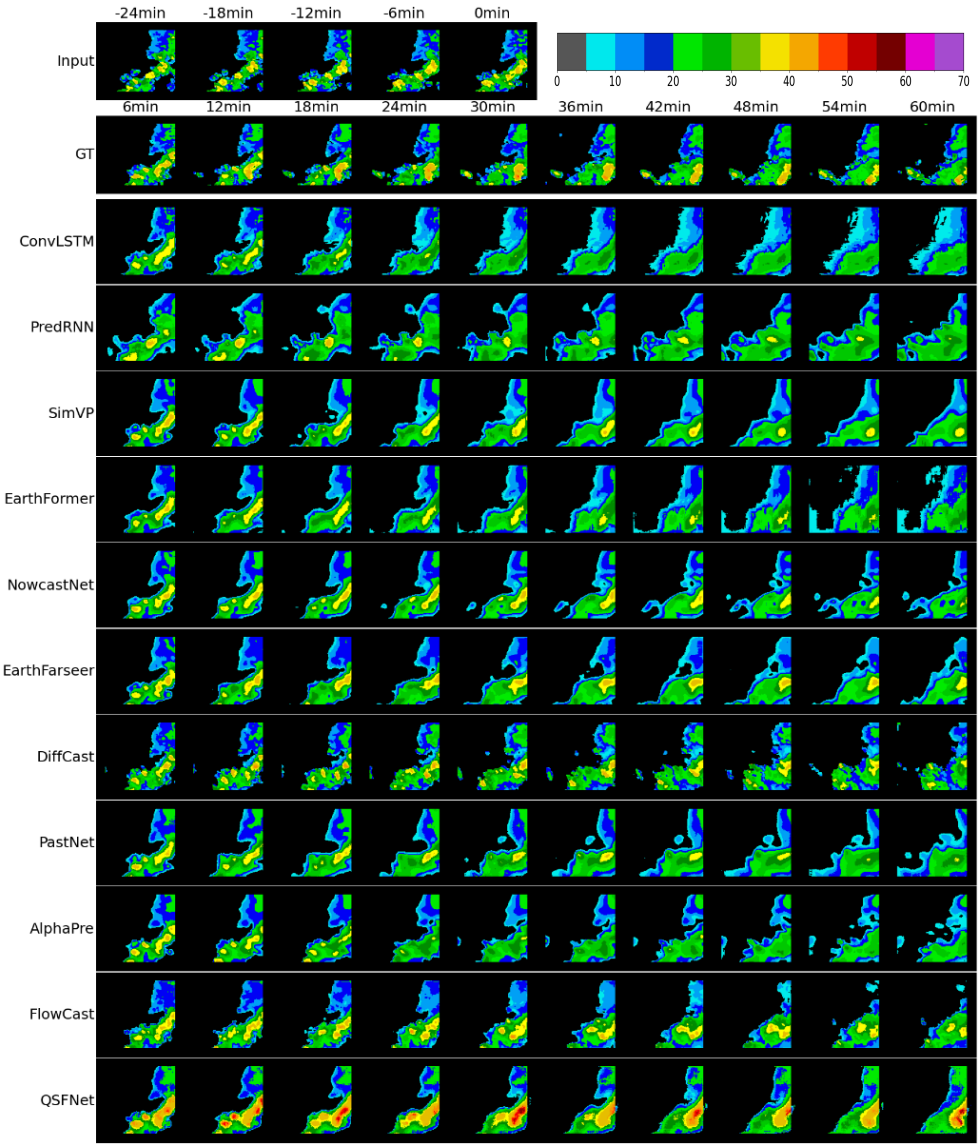

Figure 7: Qualitative comparison of QFSNet with baselines on the CIKM dataset.