International Journal of Computer Vision (2026) 134:390
https://doi.org/10.1007/s11263-026-02977-y

FaverNet: All-in-One Video Restoration via Frequency-Discriminative Conditioning

Haiyu Zhao[1] · Yuanbiao Gou[1] · Boyun Li[1] · Xinyan Xiao[2] · Wangmeng Zuo[3] · Xi Peng[4,5]

Received: 27 August 2025 / Accepted: 16 July 2026 / Published online: 8 August 2026
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2026

Abstract

Traditional video restoration methods are typically designed for a single, known degradation type, while real-world videos often experience time-varying and unknown degradations. Although more general restoration methods have been proposed, they largely overlook the time-varying nature of real-world degradations. In this work, we focus on time-varying unknown degradations, which introduce frame-wise distortions and pose two significant challenges to video restoration. First, the unknown nature of degradations expands the input degradation space and introduces significant uncertainty into the restoration process. Second, time-varying degradations cause inconsistent distortions across frames, hindering the effective aggregation of temporal information from neighboring frames. To address these two challenges, we propose FaverNet, a Frequency-guided All-in-one VidEo Restoration Network, which incorporates a Frequency-discriminative Conditioning Mechanism (FCM) and a Prompt-guided Alignment Mechanism (PAM). Specifically, FCM enhances the degradation-awareness of the restoration process by conditioning the model with degradation cues extracted from frequency domain, where different degradation types are more separable than in the spatial domain. PAM further leverages these degradation cues to assist alignment between neighboring frames, enabling temporally coherent and degradation-robust information aggregation. Furthermore, we build two benchmarks covering both classic low-level distortions (e.g., noise, blur) and complex weather-related degradations (e.g., haze, rain). Experiments show that FaverNet offers an encouraging solution for practical video restoration and achieves superior performance.

Keywords Video restoration · All-in-one video restoration · Time-varying unknown degradations · Adverse weather conditions

1 Introduction

Video restoration aims to reconstruct high-quality videos from low-quality inputs degraded by factors such as noise, blur, haze or rain. In the recent few years, extensive studies (Zhou et al., 2019; Li et al., 2022; Wu et al., 2024; Lin et al., 2022b; Wang et al., 2023; Deng et al., 2021) have led to substantial progresses in this field. However, most existing methods are designed for specific types of degradation and presume the degradation is known beforehand an unrealistic expectation in real-world scenarios, where degradations are typically mixed and unknown.

Communicated by Joey Tianyi Zhou.

✉ Xi Peng
pengx.gm@gmail.com

Haiyu Zhao
haiyuzhao.gm@gmail.com

Yuanbiao Gou
gouyuanbiao@gmail.com

Boyun Li
liboyun.gm@gmail.com

Xinyan Xiao
xiaoxinyan@baidu.com

Wangmeng Zuo
cswmzuo@gmail.com

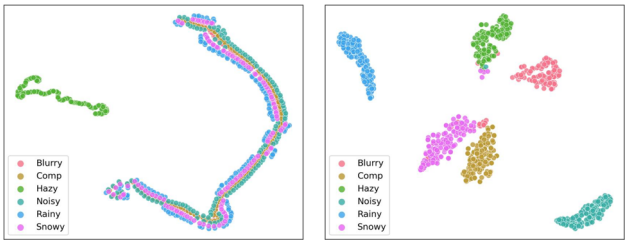
[1] College of Computer Science, Sichuan University, Sichuan, China

[2] Baidu Inc., Beijing, China

[3] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

[4] The School of Artificial Intelligence, Sichuan University, Sichuan, China

[5] Tianfu Jincheng Laboratory, Chengdu, China



(a) Spatial Domain (b) Frequency Domain

Fig. 1 Visualization of degraded frames in spatial and frequency domains using t-SNE. Points with distinct colors represent the same video frames with different degradation types. It can be observed that the frames with different degradations are more separable in the frequency domain than in the spatial domain

To address this limitation, a few studies (Cheng et al., 2023; Yang et al., 2023, 2024; Zhao et al., 2024) resort to all-in-one video restoration, which aims to handle multiple unknown degradations using a unified model. However, these methods typically assume that all frames in a video share the same, albeit unknown, degradation type. In real-world scenarios, this assumption is often impractical, as degradations (e.g., weather) usually change over time. For instance, motion blur may occur when a stationary object begins to move, while rain streaks can appear as weather conditions change. Compared to time-invariant degradations, time-varying degradations are not only more practical in real-world scenarios, but also significantly more challenging to handle.

In this work, we study such an underexplored yet practically important problem, named time-varying unknown degradations (TUD), which refers to scenarios where the degradation types are unknown and change across video frames over time. Specifically, TUD poses two challenges to video restoration. On the one hand, the unknown nature of degradations enlarges the input degradation space, forcing the model to handle a wide range of possible corruptions without explicit prior knowledge, such as degradations illustrated in Fig. 1a. This expansion introduces significant uncertainty into the restoration process, making the inverse mapping from degraded inputs to clean outputs more ambiguous and ill-posed. On the other hand, the temporal variation of degradations introduces inconsistencies in degradation types over frames, resulting in noticeable pixel-level shifts between adjacent frames, as illustrated in Figs. 2 and 3. Such shifts disrupt temporal coherence and pose significant challenges to effective aggregation of complementary information from neighboring frames.

To address these challenges, we introduce two key design principles. First, to reduce the uncertainty caused by multiple unknown degradations, we perform explicit degradation modeling to transform implicit degradation cues into a more identifiable form. This transition is facilitated in the frequency domain, where different types of degradation exhibit more separable patterns than in the spatial domain, as illustrated in Fig. 1. Second, to cope with the temporal variation of degradations, degradation cues inherently presented in each frame are leveraged to alleviate the pixel-level shifts. Incorporating these cues into cross-frame alignment helps reduce the effects of degradation inconsistency, enabling more coherent and robust temporal information aggregation.

Based on the above principles, we propose a Frequency-guided All-in-one VidEo Restoration Network (FaverNet), including a Frequency-discriminative Conditioning Mechanism (FCM) and a Prompt-guided Alignment Mechanism (PAM). As illustrated in Fig. 1, different degradations exhibit better separability in the frequency domain than in the spatial domain. This observation motivates FCM to perform explicit degradation modeling by exploiting the distinct frequency characteristics of various degradations. Specifically, FCM first transforms input video frames into the frequency domain to capture degradation-relevant cues, which are then utilized to guide the restoration process in a degradation-aware manner. To alleviate pixel shifts induced by time-varying degradations, PAM first leverages degradation cues to assist the estimation of pixel-level offsets between adjacent frames. These offsets are then used to align shifted pixels, enabling temporally coherent and degradation-robust information integration.

To summarize, the contributions of this work are as below:

- We propose FaverNet, a novel all-in-one video restoration network tailored for more realistic scenarios where degradation types and levels are both unknown and time-varying.
- We observe that degradations are more distinguishable in the frequency domain, which enables explicit modeling of unknown degradation patterns and their conversion into identifiable cues for all-in-one video restoration.
- We build two benchmarks targeting time-varying unknown degradations, including both classic low-level distortions (e.g., noise, blur) and complex weather-related degradations (e.g., haze, rain).

2 Related Works

This section presents an overview of related work, including all-in-one image restoration, general video restoration, and all-in-one video restoration. We also discuss how this work differs from prior efforts.

2.1 All-in-One Image Restoration

Recent studies in all-in-one image restoration (AIR) focus on recovering images corrupted by multiple unknown types of

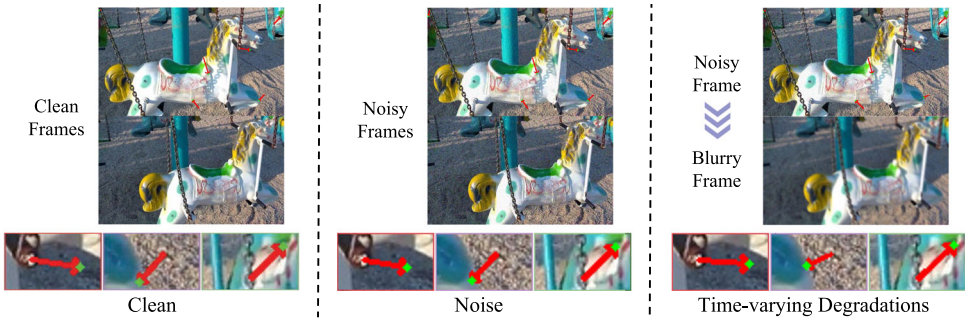


Fig. 2 Illustration of the pixel-level shifts. Optical flow is computed between two consecutive frames under three conditions: clean, time-invariant degradations, and time-varying degradations. A set of directional vectors, shown as red arrows, represent the estimated pixel-level alignments between consecutive frames. It can be observed that time-varying degradations result in less accurate alignment estimations compared to time-invariant degradations, leading to larger and more complex pixel shifts

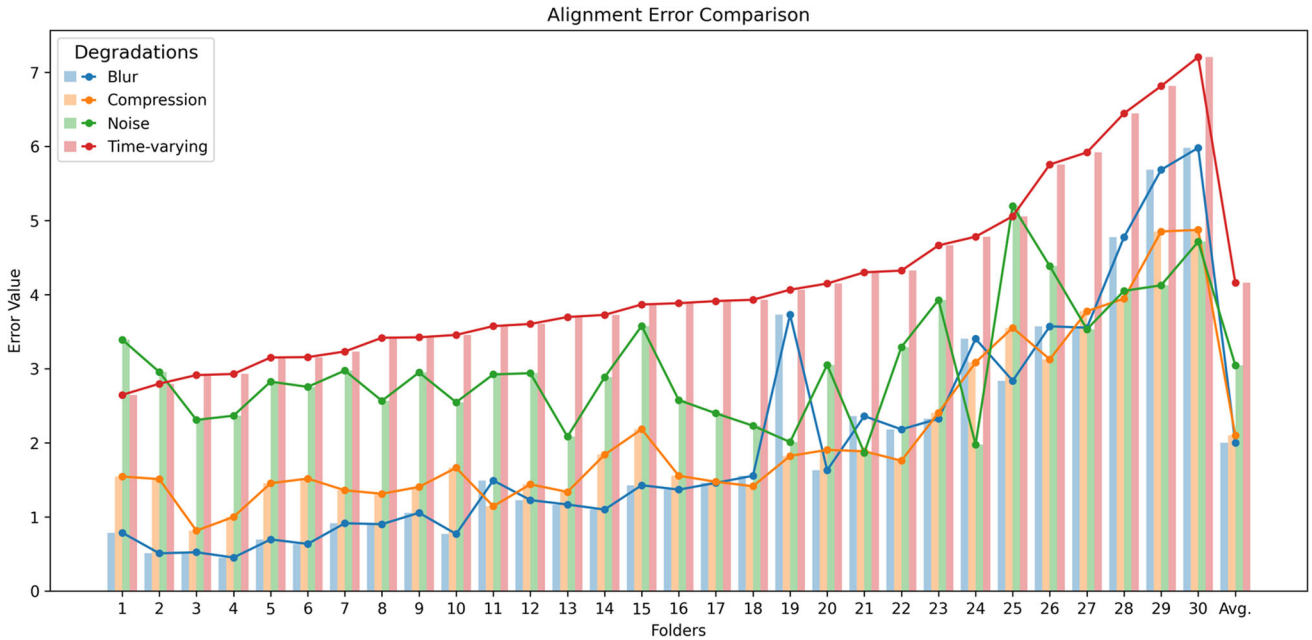


Fig. 3 Visualization of pixel-level shift statistics. We compute the optical flow for videos affected by both time-invariant and time-varying degradations. The estimated flows are compared against the flow computed on the ground-truth video to measure the pixel-level shifts, which are visualized as polylines and bars. One could observe that the time-varying degradations cause significantly larger shifts compared to time-invariant degradations

degradations using a unified model. These methods typically learn degradation-specific representations and leverage them to guide the restoration process.

For example, AirNet (Li et al., 2022) introduce contrastive learning to acquire degradation-aware representations, which are subsequently incorporated into the network for AIR. PromptIR (Potlapalli et al., 2023) introduces learnable prompts to encode degradation-specific information and leverages them to guide the restoration process. DPPD (Wu et al., 2025) assigns degradation representations to predefined prototypes and models prompts as distributions to generate dynamic prompts that guide the restoration process. ProRes (Ma et al., 2023) introduces additional visual prompts to incorporate task-specific information and utilize the prompts to guide the network for restoration. MPerceiver (Ai et al., 2024b) adopts a multimodal prompt-learning approach, utilizing generative priors from Stable Diffusion (Rombach et al., 2022) to achieve high-fidelity all-in-one restoration. AdaIR (Cui et al., 2024) explores frequency modulation by separately processing high-frequency and low-frequency regions to achieve all-in-one image restoration. CDN_MRF (He et al., 2018) presents a cascaded architecture with multiple receptive fields to enhance the restoration of fine image details. DRFIR (Liu et al., 2025)

extracts joint spatial-frequency embeddings and integrates them with image features to further boost reconstruction quality. AutoDIR (Jiang et al., 2023) integrates a CLIP-based module to detect unknown degradations and adopts a latent-diffusion-based module with a text prompt for restoration. Besides, recent work TAO (Gou et al., 2024) has explored a challenging task of open-set image restoration, which aims to handle unknown degradations that were unseen in the training phase.

In contrast to prior works specialized for image restoration, our method addresses the more complex task of all-in-one video restoration, which involves recovering videos corrupted by time-varying and unknown degradations through a single model.

2.2 General Video Restoration

In recent years, extensive studies have shown promising performance on video restoration tasks such as video denoising (Xu et al., 2023; Tassano et al., 2019, 2020), video deblurring (Shang et al., 2021; Ren et al., 2023; Lin et al., 2022a; Pan et al., 2020; Lin et al., 2020), video dehazing (Zhang et al., 2021; Ren et al., 2018; Li et al., 2018), and video deraining (Zhang et al., 2022; Sun et al., 2023; Li et al., 2018). These video restoration methods can be roughly divided into two categories: sliding window-based and recurrent models, each following a distinct temporal modeling strategy. In general, sliding window-based methods process a short temporal window and focus on restoring the center frame. By leveraging temporal information from adjacent frames, these methods have achieved competitive performance in video restoration. For example, EDVR (Wang et al., 2019) introduces a pyramid structure and temporal-spatial attention to fuse features across frames effectively. ShiftNet (Li et al., 2023) introduces an efficient framework that employs grouped spatial-temporal shift modules to implicitly aggregate information across frames. FANeRV (Li et al., 2025) utilizes wavelet transforms to decompose signals into high-frequency and low-frequency components, allowing for tailored enhancement for each part. To effectively exploit both spatial and temporal information in videos for deblurring, DSTNet (Pan et al., 2025) introduces a channel-wise gated dynamic network to adaptively model spatial features, and a discriminative temporal feature fusion module to aggregate temporal information. To better explore the redundent information in consecutive frames, CDVDTSPNL (Pan et al., 2023) proposes a convolutional network to integrate a temporal sharpness prior for removing blur in videos. Self-blind-VSR (Bai & Pan, 2024) introduces an effective self-supervised framework for video super-resolution, leveraging auxiliary training pairs synthesized directly from the original input sequences. DBVSR (Pan et al., 2021) introduces a blind video super-resolution framework that explicitly estimates blur kernels and restores sharp content through a dedicated image deconvolution step. Although these methods achieve impressive performance, they struggle to handle long video sequences and often incur high memory overhead.

Recurrent approaches accumulate temporal information by sequentially propagating hidden features from one frame to the next, enabling the restoration of subsequent frames based on prior frames. For instance, BasicVSR (Chan et al., 2021) proposes a concise and efficient network with a bidirectional propagation scheme to enhance temporal feature propagation. BasicVSR++ (Chan et al., 2022a) enhances the bidirectional propagation framework by introducing second-order grid connections, leading to more effective recurrent model. RVRT (Liang et al., 2022) proposes a transformer-based recurrent model that incorporates locally parallel designs to efficiently aggregate temporal information through the self-attention mechanism.

In contrast to previous methods that typically target a single, known degradation type, our approach is designed to handle multiple unknown and temporally varying degradations within a unified framework. Built upon a recurrent architecture, it extracts degradation-aware cues in the frequency domain and uses them to adaptively guide the restoration of each frame.

2.3 All-in-One Video Restoration

Different from most video restoration methods that focus on a specific type of degradation, recent works (Yang et al., 2023; Cheng et al., 2023; Yang et al., 2024; Zhao et al., 2024) explore all-in-one video restoration, which aims to address multiple unknown degradations within a unified model. For instance, ViWSNet (Yang et al., 2023) introduces degradation messenger tokens to encode degradation-specific information, which are subsequently used to guide the restoration process. CUDN (Cheng et al., 2023) adaptively estimates degradation representations and uses them to guide the restoration under multiple unknown degradations. Diff-TTA (Yang et al., 2024) introduces test-time adaptation to address multiple unknown degradations by aligning the test data distribution and updating the parameters of pre-trained models accordingly.

Different from these methods, our previous work (Zhao et al., 2024) was the first to focus on the time-varying nature of video degradations, and presented an all-in-one restoration approach, AverNet, to address time-varying unknown degradations. Specifically, AverNet incorporates prompt learning into a recurrent framework, where degradation-specific information is encoded as prompts and used to guide each frame restoration.

In this work, we reveal that degradations are more distinguishable in the frequency domain, enabling explicit

modeling of multiple unknown degradation patterns and transforming them into identifiable cues for all-in-one video restoration. Compared to the spatial prompts used in AverNet, the fundamental advantage of frequency-domain cues lies in the discriminative features of different degradations. Leveraging this discriminative property, our approach could effectively identify the underlying degradation type and conditionally restore the corrupted frames. As validated in experiments, our method outperforms AverNet by at least 0.42dB in PSNR, while requiring less than half parameters and runtime. Furthermore, we establish two comprehensive benchmarks for time-varying unknown degradations, providing a solid foundation for future research in this area.

3 Methods

In this section, we first introduce the problem formulation of all-in-one video restoration. Then, we present the overall network architecture, followed by detailed descriptions of the core components of the proposed method.

3.1 Problem Formulation

Video restoration aims to recover a clean video X from its degraded observation $Y = \mathcal{D}(X)$, where $\mathcal{D}(\cdot)$ represents the degradation operator. This operator can encompass diverse corruptions, including noise, blur, compression artifacts, or environmental factors such as rain, snow, and haze, depending on the specific restoration task. From a Bayesian perspective, the restored video $\hat{X}$ can be obtained by solving a Maximum A Posteriori (MAP) estimation problem:

$$\hat{X} = \arg\max_{X} \log p(X|Y). \tag{1}$$

However, most existing methods simplified the problem by introducing a restrictive assumption, i.e., the degradations are identical across all video frames. Under this assumption, the problem is formulated as follows:

$$\begin{aligned} &\hat{x}_i = \arg\max_{x_i} \log p(x_i|y_1, y_2, \ldots, y_n), \\ &\text{s.t.} \quad \begin{cases} y_i = \mathcal{D}_i(x_i), \\ \mathcal{D}_i = \mathcal{D}_j, \forall\, i,\ j, \end{cases} \end{aligned} \tag{2}$$

where $\hat{x}_i \in \hat{X}$ is the i-th restored frame, $x_i \in X$ and $y_i \in Y$ are the corresponding clean and degraded frames, respectively, n is the number of frames, and $\mathcal{D}_i$ denotes the degradation operator applied to x_i. Clearly, such an assumption of uniform degradation is unrealistic, as it overlooks the inherently time-varying nature of real-world video degradations. When this assumption is violated, most methods exhibit a noticeable drop in performance.

To address this limitation, we reformulate the video restoration problem in Eq. (1) by explicitly modeling non-uniform degradations. The revised formulation is as follows:

$$\begin{aligned} &\hat{x}_i = \arg\max_{x_i} \log p(x_i|y_1, y_2, \ldots, y_n), \\ &\text{s.t.} \quad \begin{cases} y_i = \mathcal{D}_i(x_i), \\ \mathcal{D}_i \neq \mathcal{D}_j, \exists\, i,\ j, \end{cases} \end{aligned} \tag{3}$$

where each frame y_i is degraded by an unknown degradation D_i, which may differ from that of other frames.

3.2 Architecture of FaverNet

To address this problem, we propose an all-in-one video restoration network called FaverNet, which consists of two core components: a Frequency-discriminative Conditioning Mechanism (FCM) for degradation modeling and a Prompt-guided Alignment Mechanism (PAM) for temporal alignment. We first present the overall architecture, followed by detailed descriptions of the two components.

As illustrated in Fig. 4a, FaverNet consists of three phases. In feature extraction phase, shallow features are extracted from each degraded video frame using a *ConvBlock*. In propagation phase, the features are propagated twice sequentially across the video sequence to incorporate temporal information through *Propagation Block*s. Specifically, the first propagation is performed backward, where features are propagated from the last frame to the first. The second propagation is forward, where features are propagated from the first frame to the last. This bidirectional propagation allows the features of each frame to incorporate complementary information from entire video sequence for restoration. After propagation, the features are passed through an *FCM* and a subsequent *Upsample* to generate the restored output for each frame.

3.3 Frequency-Discriminative Conditioning Mechanism

To address unknown degradations, we propose a Frequency-discriminative Conditioning Mechanism (FCM), inspired by the observation that different degradation types exhibit more separable and interpretable patterns in the frequency domain than in the spatial domain. Leveraging this property, FCM extracts degradation-relevant cues from the frequency domain and uses them to condition the restoration process under time-varying unknown degradations. Consequently, the problem in Eq. (3) is reformulated as follows:

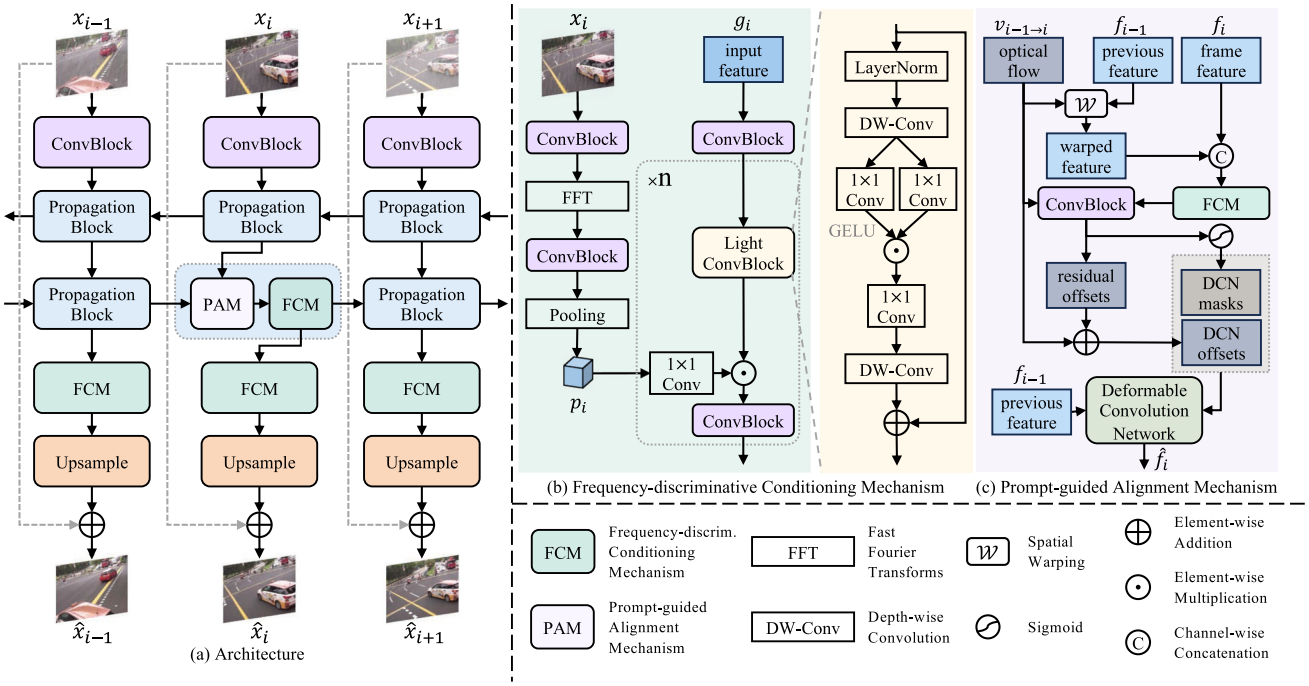


Fig. 4 Architecture overview. **a** Overall architecture of the proposed method, which is mainly composed of propagation blocks. Each propagation block consists of a **b** Frequency-discriminative Conditioning Mechanism (FCM) for conditioning the restoration with frequency-based degradation cues and a (c) Prompt-guided Alignment Mechanism (PAM) for spatially aligning features across frames with time-varying degradations

$$\hat{x}_i = \arg\max_{x_i} \log p(x_i | y_i, \mathcal{F}(y_i)),$$
$$\text{s.t.} \quad \begin{cases} y_i = \mathcal{D}_i(x_i), \\ \mathcal{D}_i \neq \mathcal{D}_j, \exists\, i,\ j, \end{cases} \tag{4}$$

where $\mathcal{F}(y_i)$ denotes the frequency-based degradation cues extracted from the degraded frame y_i. In the following, we describe how these cues are obtained and integrated into our network for all-in-one video restoration.

As illustrated in Fig. 4b, FCM conditions the restoration process in two steps: first, extracting degradation cues to generate prompts; and second, integrating the prompts into network to guide the restoration. To extract degradation cues, FCM transforms the input video into the frequency domain via *ConvBlock* followed by Fast Fourier Transforms (FFT). The resulting frequency representations are then further processed by *ConvBlock* and pooling layer to generate degradation-aware prompts. Note that these prompts are generated only once and reused throughout the restoration process. Formally, given the input frame $y_i \in \mathbb{R}^{3\times h\times w}$, where h, w are the spatial dimensions, the prompt generation process is formulated as follows:

$$p_i = GAP(\mathcal{C}_2(F(\mathcal{C}_1(y_i)))), \tag{5}$$

where $p_i \in \mathbb{R}^c$ represents the prompts for the i-th frame, $\mathcal{C}_1, \mathcal{C}_2$ are convolutional blocks, F denotes the FFT, and GAP is the global average pooling. Note that the FFT operation outputs only the magnitude of the frequency representations, and we use solely these magnitudes for prompt generation.

After generating the degradation-aware prompts, FCM integrates them into the network to dynamically modulate the frame features based on the underlying degradation characteristics. This integration is realized through convolutional blocks and Light Convolution Blocks (LCB). Specifically, given the input feature g_i, FCM first processes it through a convolution block as follows:

$$\tilde{g}_i = \mathcal{C}_3(g_i), \tag{6}$$

where $\mathcal{C}_3$ is the convolution block. Then, FCM progressively modulate the intermediate features through the prompts p_i:

$$\bar{g}_i = \{\mathcal{C}_4(\text{Conv}_{1\times1}(p_i) * LCB(\tilde{g}_i))\}_{\times n}, \tag{7}$$

where $\mathcal{C}_4$ denotes the convolution block, $\text{Conv}_{1\times1}$ is a point-wise convolution layer, $\times n$ indicates the number of repetitions, and $*$ denotes point-wise multiplication.

To efficiently enhance features, LCB adopts a combination of point-wise and depth-wise convolutions. Specifically, given the input feature $\tilde{g}_i$, LCB first applies:

$$g_i' = \Theta_1(LN(\tilde{g}_i)), \tag{8}$$

where Θ_1 denotes a depth-wise convolution, and LN represents layer normalization. Then, it projects g_i' into a high-dimensional, non-linear space (Ma et al., 2024) through:

$$g_i'' = \sigma(\phi_1(g_i')) * \phi_2(g_i'), \tag{9}$$

where σ is GELU activation function, and ϕ_1, ϕ_2 are point-wise convolutions. At last, the features are further refined with a residual connection:

$$\hat{g}_i = \Theta_2(g_i'') + \tilde{g}_i, \tag{10}$$

where Θ_2 is the depth-wise convolution, and $\hat{g}_i$ is the output of LCB.

3.4 Prompt-Guided Alignment Mechanism

To alleviate pixel-level shift issue induced by time-varying degradations, PAM employs deformable convolution network (DCN) (Dai et al., 2017) to align spatial pixels between adjacent frames. DCN relies on two parameters: offsets and masks, which represent spatial coordinate shifts and weights, respectively. Typically, the offsets and masks are estimated from the frame features and optical flow (Fischer et al., 2015) through convolutions. However, time-varying degradations can impact the estimation process, leading to inaccurate offsets and masks, which deteriorate spatial alignment. Therefore, we introduce degradation-aware prompts that encode degradation cues to assist the estimation. The graphical illustration is shown in Fig. 4c. In the following, we will detail the alignment procedure in the forward propagation, which is similar in the backward propagation.

Given the previous feature f_{i-1} from $(i-1)$-th frame, and the optical flow $v_{(i-1)\to i}$ from the $(i-1)$-th to i-th frame, we first warp f_{i-1} through $v_{(i-1)\to i}$:

$$\tilde{f}_{i-1} = \mathcal{W}(f_{i-1}, v_{(i-1)\to i}), \tag{11}$$

where $\mathcal{W}$ is the spatial warping function.

Due to the temporal variations in degradations, estimating offsets and masks through a simple network can be challenging and easily result in alignment errors. To address this, we utilize degradation-aware prompts in FCM to assist in the estimation process. Specifically, PAM incorporates the degradation information by integrating the prompts $[p_{i-1}, p_i]$ from the frames to be aligned:

$$h_i = \text{FCM}([\tilde{f}_{i-1}, f_i]; [p_{i-1}, p_i]), \tag{12}$$

where h_i is the intermediate features to compute DCN parameters and $\tilde{f}_{i-1}$ is the warped feature from $(i-1)$-th frame. Note that the prompts p_{i-1} and p_i are generated by FCM from frames y_{i-1} and y_i, respectively. Finally, the DCN offsets $o_{(i-1)\to i}$ and modulation mask $m_{(i-1)\to i}$ are calculated as:

$$\begin{aligned} o_{(i-1)\to i} &= v_{(i-1)\to i} + \mathcal{C}^o([h_i, v_{(i-1)\to i}]), \\ m_{(i-1)\to i} &= \sigma(\mathcal{C}^m([h_i, v_{(i-1)\to i}])), \end{aligned} \tag{13}$$

where $\mathcal{C}^{\{o,m\}}$ denotes a stack of convolutions, and σ denotes the sigmoid function. The DCN is then applied to the previous feature f_{i-1} for spatial alignment.

$$\hat{f}_i = \mathcal{D}(f_{i-1}; o_{(i-1)\to i}, m_{(i-1)\to i}), \tag{14}$$

where $\mathcal{D}$ denotes a deformable convolution network.

4 Experiments

In this section, we conduct experiments to assess the proposed method. First, we introduce the datasets that are used to study time-varying unknown degradations. We then provide the details on experimental settings and present quantitative and qualitative results. Finally, we conduct ablation analyses to demonstrate the effectiveness of the key modules.

4.1 Datasets

To study time-varying unknown degradations (TUD), we construct two video datasets, i.e., TUD-Common and TUD-Weather. Both datasets are synthesized by applying degradations to videos sourced from the base datasets listed in Table 1. As shown in the table, TUD-Common covers classic degradations (e.g., noise, blur, compression), consisting of 90 training videos (6.2 k frames) and 38 test videos (4.5 k frames). TUD-Weather focuses on weather-related degradations (e.g., rain, snow, haze), comprising 43 training videos (2.7 k frames) and 23 test videos (2.4 k frames).

Table 1 Synthesized datasets for studying time-varying unknown degradations

	TUD-common	TUD-weather
Base datasets	DAVIS, Set8	Rain motion, KITTI snow
Degradations	Noise, blur, compression	Rain, snow, haze
Training videos/frames	90/6.2k	43/2.7k
Testing videos/frames	38/4.5k	23/2.4k

TUD-Common includes classic degradations, while TUD-Weather contains weather-related degradations

Video Synthesis Since collecting TUD data with ground-truth is challenging or even infeasible, we synthesize TUD-Common and TUD-Weather by adding degradations to clean videos. To be specific, TUD-Common is generated through the video synthesis approach in AverNet (Zhao et al., 2024). A clean video is first divided into multiple clips with a fixed interval t. For each clip, degradations are randomly sampled from seven candidate types, including Poisson noise, Gaussian noise, speckle noise, resizing blur, Gaussian blur, JPEG compression, and video compression, with a sampling probability of 0.55. Detailed configurations for each degradation are provided in the supplementary material. The sampled degradations are then shuffled and sequentially applied to the clip to produce its corrupted version. Finally, these corrupted clips are assembled into a video containing time-varying unknown degradations.

For TUD-Weather, we adopt a similar procedure. Since weather degradations typically evolve slowly in real-world scenarios, we ensure that weather conditions change only 1 to 3 times per video. Specifically, a clean video is divided into 2 to 4 clips, and for each clip, degradations from rain, snow, and haze are randomly sampled with a probability of 0.55 and then combined before being applied. We simulate these degradations using Adobe After Effects, employing the CC Rainfall and CC Snowfall presets for rain and snow, respectively. Haze is simulated by generating an additive haze layer using Fractal Noise combined with blur, which is then overlaid onto the clean video. To ensure smooth transitions between different degradations, we gradually adjust the degradation intensity across six frames from one clip to the next. Detailed configurations for each degradation and the adjustment strategy are provided in the supplementary material. After applying the combined weather degradations to each clip, the clips are concatenated to form a video with time-varying weather degradations.

4.2 Implementation

Architecture Details We adopt the same settings for all experiments. Specifically, the number of channels is 96 and the number of convolution blocks in feature extraction is 3. The hyperparameter n for the FCM is set to 7 during propagation and 5 during the reconstruction. For optical flow estimation, we employ the pre-trained SPyNet (Ranjan & Black, 2017), whose parameters and runtime accounted for as part of our method. Note that the flow estimation network is initialized with pre-trained weights and jointly optimized together with the restoration network during training.

Training Details The experiments are conducted in PyTorch (Paszke, 2019) framework with NVIDIA GeForce RTX 3090 GPUs and NVIDIA A800. Note that our model is trained on two NVIDIA GeForce RTX 3090 GPUs. For training, we use Charbonnier loss (Charbonnier et al., 1994) and Adam optimizer (Kingma & Ba, 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is set to $2e^{-4}$ for the TUD-Common dataset and $1e^{-4}$ for the TUD-Weather dataset. In both cases, the learning rate is gradually decreased to $1e^{-7}$ through cosine annealing strategy (Loshchilov & Hutter, 2016). The number and resolution of input frames are set to 12 and 256×256, respectively. The networks are trained with a batch size of 2 for 600K iterations. The optical flow network is frozen during the first 5K iterations.

Table 2 Comparisons of parameters, FLOPs and runtime

Type	Method	#Params	FLOPs	Runtime
Image Restoration	WDiffusion	80.0 M	73.26 T	3956.50 s
	TransWeather	38.1 M	38.88 G	3.32 s
	AirNet	8.9 M	1.89 T	12.98 s
	PromptIR	35.6 M	991.47 G	10.81 s
	LoRA-IR	68.87 M	398.97 G	5.45 s
	PromptRestorer	24.45 M	66.62 G	46.93 s
	AdaIR	28.78 M	922.76 G	27.13 s
Video restoration	EDVR	23.6 M	3.30 T	6.78 s
	BasicVSR++	7.4 M	2.00 T	5.09 s
	Shift-Net	12.9 M	708.88 G	45.54 s
	RVRT	13.6 M	3.61 T	15.07 s
All-in-one video restoration	ViWS-Net	57.8 M	548.88 G	12.20 s
	AverNet	41.3 M	290.18 G	6.93 s
	FaverNet (ours)	13.4 M	240.25 G	**3.04 s**

The lowest runtime is highlighted in bold

Table 3 Quantitative results compared with state-of-the-art methods on test sets with different degradation change intervals, where t denotes the interval at which degradations change

Method	DAVIS-test						Set8						Average	
	t = 6		t = 12		t = 24		t = 6		t = 12		t = 24			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
WDiffusion	31.74	0.8768	31.79	0.8784	31.92	0.8809	30.31	0.8784	30.02	0.8716	30.82	0.8746	31.10	0.8768
TransWeather	31.11	0.8694	31.13	0.8699	31.26	0.8741	29.24	0.8662	28.95	0.8565	29.15	0.8632	30.14	0.8647
AirNet	32.46	0.8873	32.46	0.8887	32.75	0.8928	30.71	0.8874	30.40	0.8806	31.16	0.8825	31.66	0.8866
PromptIR	32.18	0.8843	32.19	0.8867	32.45	0.8900	30.79	0.8903	30.43	0.8821	31.19	0.8847	31.54	0.8864
LoRA-IR	27.12	0.6696	26.75	0.6518	26.76	0.6596	24.90	0.6427	24.43	0.6188	24.98	0.6464	25.82	0.6482
PromptRestorer	31.38	0.8707	31.46	0.8695	31.53	0.8737	29.87	0.8663	29.50	0.8549	30.79	0.8733	30.76	0.8681
AdaIR	33.12	0.8998	33.08	0.9016	33.15	0.9028	31.05	0.8943	30.86	0.8904	32.10	0.8994	32.23	0.8981
EDVR	28.70	0.7224	28.37	0.6991	29.07	0.7289	26.75	0.7259	26.94	0.7382	28.71	0.7675	28.09	0.7303
BasicVSR++	33.22	0.9204	33.07	0.9180	33.32	0.9210	30.90	0.9048	30.52	0.8965	31.35	0.9011	32.06	0.9103
Shift-Net	33.09	0.9096	33.10	0.9113	33.34	0.9133	31.15	0.9027	30.82	0.8947	31.88	0.9000	32.23	0.9053
RVRT	33.99	0.9314	33.98	0.9311	34.10	0.9315	31.73	0.9192	31.39	0.9113	32.47	0.9178	32.94	0.9237
ViWSNet	16.47	0.5243	16.48	0.5224	16.47	0.5242	13.83	0.3585	13.81	0.3598	13.81	0.3574	15.15	0.4411
AverNet	34.07	0.9333	34.09	0.9339	34.28	0.9356	31.73	**0.9219**	31.47	0.9145	32.45	0.9189	33.01	0.9263
FaverNet (ours)	**34.60**	**0.9354**	**34.64**	**0.9356**	**34.68**	**0.9366**	**31.82**	0.9214	**31.60**	**0.9166**	**33.24**	**0.9271**	**33.43**	**0.9288**

The best outcomes are highlighted in bold

Table 4 Quantitative results compared with state-of-the-art methods under three degradation combinations, namely noise & blur, noise & compression, and blur & compression

Method	DAVIS-test						Set8						Average	
	Noise & Blur		Noise & Comp.		Blur & Comp.		Noise & Blur		Noise & Comp.		Blur & Comp.			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
WDiffusion	32.70	0.8990	33.52	0.9124	33.76	0.9142	31.64	0.8943	30.88	0.8968	31.22	0.8978	32.29	0.9024
TransWeather	31.74	0.8863	32.53	0.9062	32.18	0.9017	29.74	0.8714	29.93	0.8886	29.61	0.8830	30.95	0.8895
AirNet	33.41	0.9078	34.23	0.9184	34.59	0.9224	32.15	0.9065	31.27	0.9019	31.60	0.9027	32.88	0.9100
PromptIR	33.69	0.9128	34.18	0.9213	33.87	0.9179	32.10	0.9033	31.54	0.9106	31.53	0.9047	32.82	0.9118
LoRA-IR	25.79	0.6231	26.85	0.6571	29.27	0.7579	23.92	0.6056	23.68	0.6037	27.08	0.7163	26.10	0.6606
PromptRestorer	32.31	0.8893	33.54	0.9127	34.12	0.9101	31.03	0.8683	31.10	0.8994	30.69	0.8760	32.13	0.8926
AdaIR	34.35	0.9212	35.31	0.9331	35.57	0.9317	32.65	0.9097	32.13	0.9166	32.22	0.9071	33.71	0.9199
EDVR	28.00	0.6809	29.58	0.7036	34.17	0.9082	27.82	0.7268	27.23	0.7245	32.15	0.8845	29.83	0.7714
BasicVSR++	33.89	0.9324	34.72	0.9391	34.82	0.9392	31.88	0.9189	31.42	0.9152	31.35	0.9146	33.01	0.9266
Shift-Net	34.00	0.9277	34.91	0.9390	35.26	0.9376	32.66	0.9159	31.86	0.9184	32.08	0.9164	33.46	0.9258
RVRT	34.67	0.9438	35.69	0.9504	35.94	0.9503	32.70	0.9291	32.40	0.9297	32.38	0.9291	33.96	0.9387
ViWSNet	16.48	0.5237	16.45	0.5226	16.49	0.5233	13.81	0.3562	13.73	0.3521	13.76	0.3563	15.12	0.4390
AverNet	35.01	**0.9468**	35.89	**0.9531**	35.87	0.9506	**32.94**	**0.9326**	32.33	0.9309	32.37	**0.9306**	34.07	**0.9408**
FaverNet (ours)	**35.41**	0.9462	**36.67**	0.9526	**37.14**	**0.9540**	32.78	0.9294	**32.66**	**0.9335**	**32.48**	0.9278	**34.52**	0.9406

The best outcomes are highlighted in bold

Table 5 Quantitative results compared to state-of-the-art methods on TUD-Weather

Method	Rain Motion		KITTI Snow		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
WDiffusion	27.79	0.9272	27.24	0.8863	27.52	0.9068
TransWeather	27.84	0.8972	24.88	0.8375	26.36	0.8674
AirNet	28.56	0.9298	27.31	0.8945	27.94	0.9122
PromptIR	30.75	0.9351	27.58	0.8981	29.17	0.9166
LoRA-IR	5.82	0.0530	10.57	0.3060	8.20	0.1795
PromptRestorer	21.47	0.8229	16.65	0.6600	19.06	0.7415
AdaIR	32.59	0.9493	28.86	0.9124	30.73	0.9309
EDVR	23.94	0.7758	20.14	0.6957	22.04	0.7358
BasicVSR++	31.58	0.9607	28.68	0.9235	30.13	0.9421
Shift-Net	31.58	0.9580	**29.76**	0.9275	30.67	0.9428
RVRT	29.58	0.9388	27.60	0.9018	28.59	0.9203
ViWSNet	11.99	0.3089	13.54	0.4635	12.77	0.3862
AverNet	29.70	0.9290	27.21	0.8924	28.46	0.9107
FaverNet (ours)	**33.37**	**0.9638**	29.26	**0.9315**	**31.32**	**0.9477**

4.3 Comparison with State-of-the-Arts

In this section, we compare FaverNet with state-of-the-art methods, including all-in-one image restoration, conventional video restoration, and all-in-one video restoration approaches. To be specific, all-in-one image restoration methods are WDiffusion (Özdenizci & Legenstein, 2023), TransWeather (Valanarasu et al., 2022), AirNet (Li et al., 2022), PromptIR (Potlapalli et al., 2023), LoRA-IR (Ai et al., 2024a), PromptRestorer (Wang et al., 2023), and AdaIR (Cui et al., 2024). Conventional video restoration methods are EDVR (Wang et al., 2019), BasicVSR++ (Chan et al., 2022a), Shift-Net (Li et al., 2023), and RVRT (Liang et al., 2022). All-in-one video restoration methods are ViWS-Net (Yang et al., 2023) and AverNet (Zhao et al., 2024). All methods were trained from scratch on the training sets of TUD-Common and TUD-Weather, following the training settings reported in their original papers. Note that image restoration methods were trained and evaluated on individual video frames. The parameters, FLOPs and runtime of our and the compared methods are summarized in Table 2. Specifically, runtime is measured on a 48-frame video from the DAVIS-test set, while FLOPs are calculated on a single frame with a resolution of 856×480, ensuring a fair comparison between image and video restoration methods.

Evaluation on TUD-Common To comprehensively evaluate the effectiveness in handling time-varying common degradations, we test all methods on TUD-Common and several of its variants. Specifically, the variants are constructed by adjusting the settings of the video synthesis process to assess performance under different degradation change intervals and degradation combinations.

We evaluate on test sets with three degradation change intervals, namely $t = 6$, $t = 12$, and $t = 24$, where t denotes the number of frames after which the degradation type and intensity are changed. The results, shown in Table 3, demonstrate that our method nearly outperforms others across all test sets, achieving the highest average PSNR and SSIM scores. Our method surpasses AverNet by at least 0.40 dB in PSNR on DAVIS-test, and outperforms RVRT by up to 0.66 dB on DAVIS-test and 0.77 dB on Set8. Averaged over all test sets, our method exceeds the second-best approach AverNet by more than 0.42 dB in PSNR and 0.0025 in SSIM. Compared to all-in-one image restoration methods, our method achieves significantly better results, revealing the limitations of frame-wise approaches when dealing with time-varying unknown degradations, and highlighting the necessity of considering frame-to-frame variations. Specifically, our FaverNet exceeds AdaIR by 1.20 dB in PSNR and 0.0307 in SSIM on average. Furthermore, according to Table 2, our method achieves superior performance while maintaining competitive efficiency among the compared methods. Notably, its runtime is less than half that of AverNet and approximately one-fifth that of RVRT. Meanwhile, our model size is significantly smaller, requiring only 32% of AverNet's parameters (Fig. 5).

To assess the robustness under different degradation combinations, we construct three types of combinations by removing one degradation from the original set, namely noise and blur, noise and compression, blur and compression. In these test sets, both the degradation types and intensities are time-varying, with changes occurring every 6 frames. Across both Table 2 and Table 4, it can be observed that our method achieves comparable or even superior performance while being significantly more efficient. For instance, under the

Table 6 Quantitative results compared to state-of-the-art methods under three weather degradation combinations, i.e., rain & snow, rain & haze, and snow & haze

Method	Rain motion						KITTI snow						Average	
	Rain & snow		Rain & haze		Snow & haze		Rain & snow		Rain & haze		Snow & haze			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
WDiffusion	30.48	0.9358	27.66	0.9446	27.78	0.9578	27.89	0.8995	28.42	0.9211	29.39	0.9434	28.60	0.9337
TransWeather	27.32	0.8832	28.83	0.9200	29.09	0.9418	24.86	0.8392	25.45	0.8764	26.07	0.9072	26.94	0.8946
AirNet	32.21	0.9410	28.40	0.9450	28.58	0.0603	28.39	0.9126	28.64	0.9265	29.50	0.9466	29.29	0.9387
PromptIR	31.59	0.9359	30.99	0.9502	30.87	0.9654	27.71	0.9062	28.31	0.9264	29.20	0.9476	29.78	0.9386
LoRA-IR	5.95	0.0549	5.78	0.0529	5.65	0.0492	14.26	0.3902	9.52	0.2269	11.35	0.3542	8.75	0.1881
PromptRestorer	26.14	0.7990	17.37	0.7845	19.50	0.8447	24.22	0.7604	16.72	0.6474	19.38	0.7796	20.56	0.7693
AdaIR	32.29	0.9439	31.24	0.9595	32.38	0.9734	28.26	0.9115	29.35	0.9336	30.25	0.9531	30.63	0.9458
EDVR	24.61	0.7791	23.37	0.7751	23.90	0.7856	20.60	0.7191	20.31	0.7167	20.38	0.7281	22.20	0.7506
BasicVSR++	33.99	0.9616	27.12	0.9332	31.06	0.9712	29.08	0.9316	29.18	0.9416	29.99	0.9590	30.07	0.9497
Shift-Net	33.60	0.9539	31.34	0.9607	31.98	0.9735	29.29	0.9256	**29.95**	0.9414	**30.99**	0.9603	31.19	0.9526
RVRT	31.02	0.9378	29.02	0.9496	29.28	0.9633	27.77	0.9068	28.31	0.9292	28.89	0.9497	29.05	0.9394
ViWSNet	11.96	0.3113	12.07	0.3111	11.93	0.3066	13.52	0.4706	13.47	0.4665	13.38	0.4683	12.72	0.3891
AverNet	30.62	0.9285	29.24	0.9446	29.39	0.9592	27.42	0.8989	28.13	0.9239	28.83	0.9438	28.94	0.9332
FaverNet (ours)	**34.41**	**0.9617**	**31.66**	**0.9659**	**32.93**	**0.9765**	**29.68**	**0.9385**	29.92	**0.9474**	30.54	**0.9616**	**31.52**	**0.9586**

The best outcomes are highlighted in bold

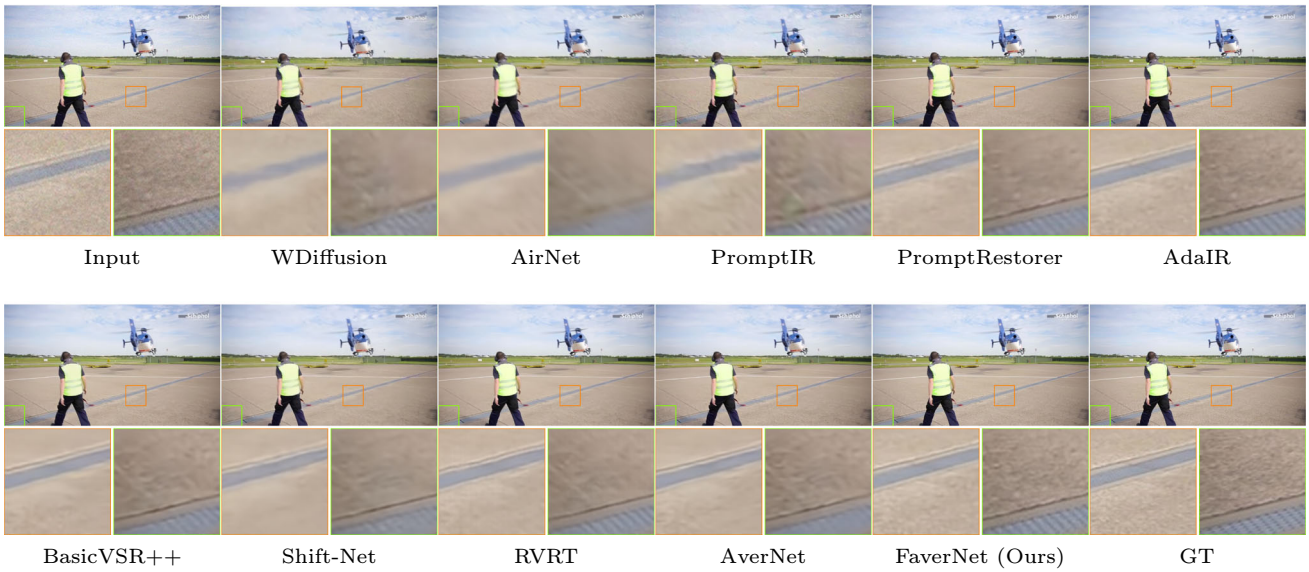


Fig. 5 Qualitative results on the “helicopter” video from DAVIS-test (t=24), from which one could observe that existing methods leave artifacts and lose fine texture details. In contrast, our method obtains clearer results that are closer to GT

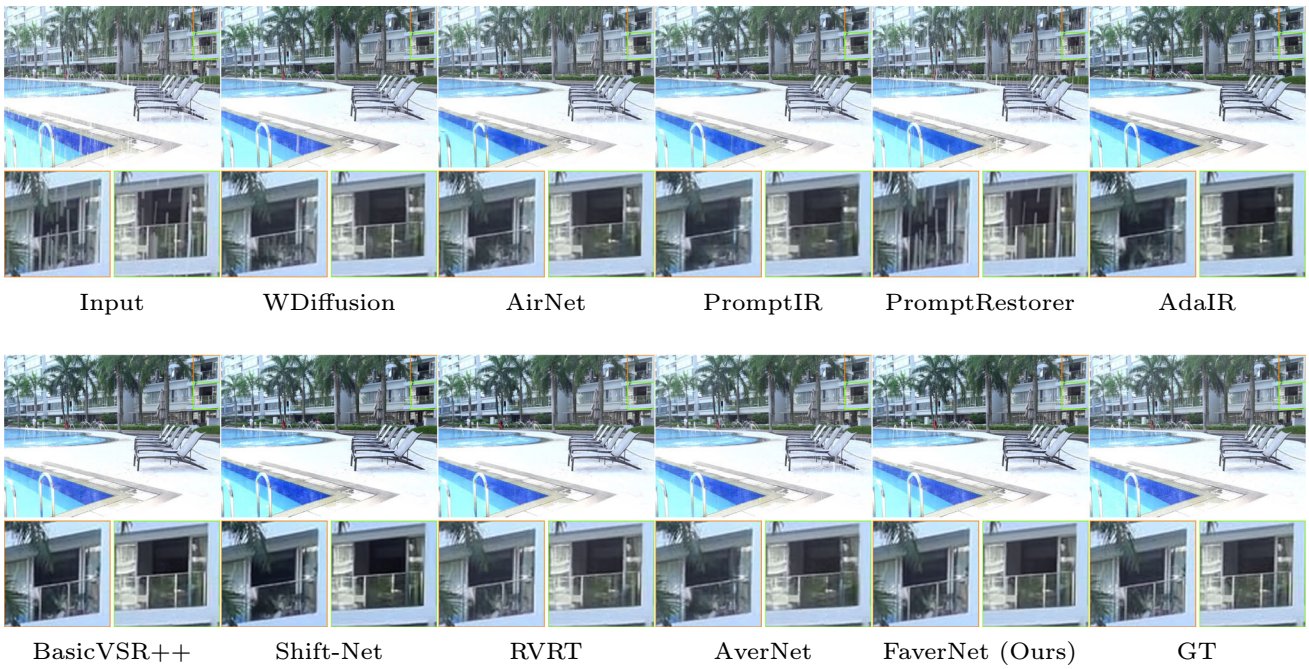


Fig. 6 Qualitative results on the "a2" video from Rain Motion, where it can be observed that existing methods leave noticeable residual artifacts and lose critical structural details, such as the railings. In contrast, our method produces clearer results that are closer to the ground truth

noise and compression combination, our method surpasses AverNet by 0.78 dB on DAVIS-test and 0.33 dB on Set8 in terms of PSNR, while requiring less than half of AverNet’s processing time. Moreover, our FaverNet outperforms RVRT by at least 0.74 dB in PSNR on DAVIS-test. On average across all six test sets, our method achieves the highest PSNR score, surpassing the second-best approach, AverNet, by more than 0.45 dB.

Evaluation on TUD-Weather To evaluate the effectiveness of our method against weather-related degradations, we conduct experiments on test sets, where each video is affected by time-varying weather conditions. As shown in Table 5, our method nearly outperforms other methods on both Rain Motion and KITTI Snow, achieving the highest average scores in PSNR and SSIM. Specifically, it surpasses AverNet by at least 2.05 dB in PSNR and 0.0348 in SSIM.

Table 7 Quantitative results compared to state-of-the-art methods on the realistic dataset VideoLQ

Dataset	VideoLQ	
Metric	NIQE ↓	BRISQUE ↓
WDiffusion	4.59	34.17
TransWeather	7.33	30.42
AirNet	4.70	36.11
PromptIR	4.03	16.33
LoRA-IR	4.02	13.50
PromptRestorer	4.05	13.52
AdaIR	4.70	35.08
EDVR	4.79	33.04
BasicVSR++	4.97	34.15
Shift-Net	4.71	35.11
RVRT	4.65	34.07
ViWSNet	5.01	29.34
AverNet	4.60	34.01
FaverNet (ours)	**3.95**	**12.87**

On average across two test sets, our approach exceeds the second-best method, Shift-Net, by more than 0.65 dB in PSNR and 0.0049 in SSIM.

To verify robustness under various degradation combinations, we evaluate the models under different weather conditions, namely rain and snow, rain and haze, haze and snow. As shown in Table 6, our method demonstrates superior performance across most scenes. For instance, our FaverNet outperforms AverNet by at least 2.42 dB on Rain Motion and 1.71 dB on KITTI Snow. Furthermore, it achieves at least a 0.0119 improvement in SSIM compared to RVRT on both test sets. On average across six test sets, our method achieves the highest PSNR and SSIM scores, surpassing the second-best approach, Shift-Net, by more than 0.33 dB and 0.0060 (Fig. 6).

Evaluation on Real Datasets To further evaluate the effectiveness of our approach in real-world scenarios, we test the models on the realistic VideoLQ (Chan et al., 2022b) dataset. The results are presented in Table 7, showing that our method achieves the best overall performance. Specifically, our approach yields a 0.65 lower NIQE compared with AverNet and is the only video restoration method to achieve a BRISQUE value below 20. Notably, it is also the only method that attains both a NIQE lower than 4 and a BRISQUE lower than 13.

4.4 Analysis Experiments

In this section, we conduct evaluations on extended test sets to further analyze the effectiveness of the models in handling time-varying unknown degradations. In addition, we perform ablation studies on the core components, namely FCM and PAM, to validate their contributions.

To further demonstrate the capability of the proposed FaverNet, we conduct evaluations on test sets with time-worsening unknown degradations. Specifically, each video is synthesized by sequentially introducing multiple unknown types of degradations, with the intensities progressively increasing over time. Taking noise as an example, Poisson noise, Speckle noise, and Gaussian noise are introduced sequentially and cumulatively, with their noise levels progressively increasing during its introduction period. The introduction order of different degradation types is randomized for each video to ensure diversity.

As shown in Table 8, our method outperforms existing methods on test sets with increasing blur and compression, achieving the highest average scores and thus demonstrating superior capability in handling time-worsening degradations. Specifically, on the increasing blur test set, our method outperforms RVRT by 0.35 dB in PSNR and 0.0122 in SSIM on DAVIS-test, and by 0.34 dB in PSNR and 0.0147 in SSIM on Set8. When averaged over all test sets, our method surpasses the second-best approach, AverNet, by 0.15 dB in PSNR and 0.0061 in SSIM. Although AverNet achieves slightly better results on the increasing noise test set, our method provides a superior overall balance across different degradation types, along with significantly higher efficiency.

To validate the effectiveness of FCM and PAM, we conduct ablation studies by individually removing them from the network and evaluating the performance. The results are presented in the Table 9, showing that each component significantly improves performance.

Specifically, removing FCM leads to PSNR drops of 0.70 dB on DAVIS-test and 0.47 dB on Set8, indicating that without degradation prompts, the model faces challenges in handling unknown degradations. Similarly, the removal of PAM causes PSNR drops of 1.05 dB on DAVIS-test and 0.49 dB on Set8, respectively. This suggests that without PAM, the network struggles to effectively align frame features under time-varying degradations, which results in performance degradation. Furthermore, when both FCM and PAM are removed, the model shows significant PSNR drops of 2.22 dB and 2.52 dB on DAVIS-test and Set8, respectively, highlighting the crucial role of both frequency-based conditioning and prompt-guided alignment in achieving all-in-one video restoration.

To further evaluate the effectiveness of our frequency-based conditioning and demonstrate its advantages, we conduct additional comparisons by replacing frequency prompts with the spatial prompts used in AverNet. As shown in Table 10, using spatial prompts results in a model size increase of over 30% compared to our method. Meanwhile, our frequency prompts achieve superior performance, with a 0.14 dB gain in PSNR and a 0.0024 improvement in

Table 8 Quantitative results compared to state-of-the-art methods on test sets with increasing degradation intensity

Method	DAVIS-test						Set8						Average	
	Increasing noise		Increasing blur		Increasing comp.		Increasing noise		Increasing blur		Increasing comp.			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
WDiffusion	34.02	0.9162	29.17	0.8271	28.49	0.8298	32.41	0.9162	27.29	0.7953	27.41	0.8349	29.80	0.8533
TransWeather	33.02	0.9114	28.69	0.8150	28.76	0.8456	30.77	0.9007	26.80	0.7768	27.63	0.8478	29.28	0.8496
AirNet	34.99	0.9271	29.20	0.8274	29.22	0.8604	33.12	0.9302	27.31	0.7940	28.31	0.8599	30.36	0.8665
PromptIR	35.41	0.9364	29.30	0.8321	29.13	0.8643	33.21	0.9287	27.32	0.7966	28.35	0.8676	30.45	0.8710
LoRA-IR	23.94	0.5590	25.99	0.6505	25.68	0.6865	20.61	0.5027	24.49	0.5998	24.26	0.6640	24.16	0.6104
PromptRestorer	33.79	0.9141	28.23	0.7999	28.53	0.8494	32.19	0.9095	26.54	0.7619	27.62	0.8538	29.48	0.8481
AdaIR	35.57	0.9364	29.19	0.8271	29.43	0.8701	33.41	0.9298	27.29	0.7912	28.48	0.8698	30.56	0.8707
EDVR	24.82	0.7707	23.99	0.7224	23.82	0.7507	22.22	0.6178	22.31	0.6396	22.30	0.6780	23.24	0.6965
BasicVSR++	35.33	0.9432	28.56	0.8168	24.32	0.7664	32.80	0.9332	26.64	0.7702	24.91	0.7841	28.76	0.8357
Shift-Net	35.66	0.9459	29.00	0.8238	29.35	0.8655	33.60	0.9394	27.14	0.7863	28.49	0.8743	30.54	0.8725
RVRT	36.27	0.9553	28.97	0.8279	29.14	0.8618	33.80	0.9477	26.99	0.7851	28.67	0.8796	30.64	0.8762
ViWSNet	16.51	0.5230	16.59	0.5327	16.22	0.4834	13.83	0.3538	13.93	0.3783	13.78	0.3426	15.14	0.4356
AverNet	**36.76**	**0.9583**	28.93	0.8279	29.36	0.8695	**33.97**	**0.9496**	26.86	0.7827	28.68	0.8789	30.76	0.8778
FaverNet (ours)	36.40	0.9555	**29.32**	**0.8401**	**29.72**	**0.8759**	33.79	0.9471	**27.33**	**0.7998**	**28.91**	**0.8851**	**30.91**	**0.8839**

The best outcomes are highlighted in bold

Table 9 Ablation studies

	#Params	FCM	PAM	DAVIS-test		Set8	
				PSNR	SSIM	PSNR	SSIM
(A)	6.0 M			32.42	0.8938	29.08	0.8488
(B)	10.4 M		✓	33.94	0.9242	31.13	0.9065
(C)	8.7 M	✓		33.59	0.9158	31.11	0.9039
(D)	13.4 M	✓	✓	34.64	0.9356	31.60	0.9166

Each module contributes to improvements in PSNR and SSIM, validating their effectiveness

Table 10 Comparisons between frequency-based prompts and spatial prompts

Strategy	#Params	PSNR	SSIM
Spatial Prompts	17.6M	34.50	0.9332
Frequency Prompts	**13.4M**	**34.64**	**0.9356**

The lowest number of parameters and the best performance outcomes are highlighted in bold

SSIM, demonstrating both the efficiency and effectiveness of frequency-based guidance over spatial prompts.

5 Conclusions

In this paper, we introduced FaverNet, a novel all-in-one video restoration network tailored for realistic scenarios characterized by time-varying unknown degradations. FaverNet leverages the key observation that degradations exhibit greater separability in the frequency domain, enabling explicit modeling of unknown degradation patterns and their transformation into identifiable cues to guide the restoration process. To this end, FaverNet incorporates two core components: an FCM to extract degradation cues from the frequency domain, and a PAM to enhance temporal coherence by alleviating pixel-level shifts across frames. To support evaluations, we also constructed two benchmarks targeting time-varying unknown degradations, covering both classic low-level degradations and complex weather-related degradations. Extensive experiments demonstrate that FaverNet not only achieves superior performance across diverse degradation scenarios but also offers a more efficient solution for practical video restoration.

Acknowledgements This work was supported in part by Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (Grant No. JYB2025XDXM610); in part by Postdoctoral Fellowship Program (Grade C) of China Postdoctoral Science Foundation under Grant GZC20251052; in part by Fundamental Research Funds for the Central Universities under Grant CJ202303; in part by Sichuan Science and Technology Planning Project under Grant 2024NSFTD0038; in part by TCL Science and Technology Innovation Fund.

Data Availability The datasets used in this study are publicly available. Specifically, DAVIS and Set8 can be accessed at https://davischallenge.org and https://drive.google.com/drive/folders/11chLkbcX-oKGLOLONuDpXZM2-vujn_KD, respectively. Rain Motion and KITTI Snow are available at https://drive.google.com/file/d/1905B_e2RgQGnyfHd5xpjB4lTLYoq0Jm4/view and https://drive.google.com/file/d/1_1IsyT5nTvYjrCwNgP4LDOh_rXnPP_LE/view. In addition, part of the synthesized data with time-varying degradations is available at https://drive.google.com/drive/folders/1NIV4YSMhBJfjQu2SNQgvuluTAZDeFXH6, and the remaining data will be released together with the source code upon acceptance.

References

Ai, Y., Huang, H.,& He, R. (2024a). Lora-ir: Taming low-rank experts for efficient all-in-one image restoration. arXiv preprint arXiv:2410.15385

Ai, Y., Huang, H., Zhou, X., Wang, J.,& He, R. (2024b). Multimodal prompt perceiver: Empower adaptiveness generalizability and fidelity for all-in-one image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 25432–25444).

Bai, H., & Pan, J. (2024). Self-supervised deep blind video super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(7), 4641–4653.

Chan, K. C. K., Wang, X., Yu, K., Dong, C., & Loy, C. C. (2021). Basicvsr: The search for essential components in video super-resolution and beyond. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4947–4956).

Chan, K.C.K., Zhou, S., Xu, X., Loy, C.C. (2022a). Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5972–5981).

Chan, K. C. K., Zhou, S., Xu, X., Loy, C. C. (2022b). Investigating tradeoffs in real-world video super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5962–5971).

Charbonnier, P., Blanc-Feraud, L., Aubert, G., & Barlaud, M. (1994). Two deterministic half-quadratic regularization algorithms for computed imaging. *Proceedings of 1st international conference on image processing* (pp. 168–172). IEEE.

Cheng, Y., Shao, M., Wan, Y., Qiao, Y., Zuo, W., & Meng, D. (2023). Cross-consistent deep unfolding network for adaptive all-in-one video restoration arXiv:2309.01627 arXiv preprint.

Cui, Y., Zamir, S.W., Khan, S., Knoll, A., Shah, M.,& Shahbaz Khan, F. (2024). Adair: Adaptive all-in-one image restoration via frequency mining and modulation. arXiv preprint arXiv:2403.14614

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 764–773).

Deng, S., Ren, W., Yan, Y., Wang, T., Song, F., & Cao, X. (2021). Multi-scale separable network for ultra-high-definition video deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 14030–14039).

Fischer, P., Dosovitskiy, A., Ilg, E., Häusser, P., Hazırbaş, C., Golkov, V., van der Smagt, P., Cremers, D., Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2758–2766).

Gou, Y., Zhao, H., Li, B., Xiao, X., & Peng, X. (2024). Test-time degradation adaptation for open-set image restoration. *Proceedings of the 41st international conference on machine learning* (pp. 16167–16177).

He, Z., Tang, S., Yang, J., Cao, Y., Yang, M. Y., & Cao, Y. (2018). Cascaded deep networks with multiple receptive fields for infrared image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology, 29*(8), 2310–2322.

Jiang, Y., Zhang, Z., Xue, T., & Gu, J. (2023). Autodir: Automatic all-in-one image restoration with latent diffusion arXiv:2310.10123 arXiv preprint.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization arXiv:1412.6980 arXiv preprint.

Li, Y., Li, Z., Yao, C., Xiao, J., & Gabbouj, M. (2025). Fanerv: Frequency separation and augmentation based neural representation for video. In *Expert systems with applications* (pp. 129935).

Li, B., Liu, X., Hu, P., Wu, Z., Lv, J., & Peng, X. (2022a). All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 17452–17462).

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2018a). End-to-end united video dehazing and detection. In *Proceedings of the AAAI conference on artificial intelligence*.

Li, D., Shi, X., Zhang, Y., Cheung, K. C., See, S., Wang, X.,& Li, H. (2023). A simple baseline for video restoration with grouped spatial-temporal shift. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9822–9832).

Li, J., Wu, X., Niu, Z., & Zuo, W. (2022b). Unidirectional video denoising by mimicking backward recurrent modules with look-ahead forward ones. In *European conference on computer vision* (pp. 592–609). Springer.

Li, M., Xie, Q., Zhao, Q., Wei, W., Gu, S., Tao, J., & Meng, D. (2018b). Video rain streak removal by multiscale convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6644–6653).

Liang, J., Fan, Y., Xiang, X., Ranjan, R., Ilg, E., Green, S., & Gool, L. V. (2022). Recurrent video restoration transformer with guided deformable attention. *Advances in Neural Information Processing Systems, 35*, 378–393.

Lin, J., Cai, Y., Hu, X., Wang, H., Yan, Y., Zou, X.,& Van Gool, L. (2022a). Flow-guided sparse transformer for video deblurring. arXiv preprint arXiv:2201.01893

Lin, J., Hu, X., Cai, Y., Wang, H., Yan, Y., Zou, X.,& Van Gool, L. (2022b). Unsupervised flow-aligned sequence-to-sequence learning for video restoration. In *International conference on machine learning* (pp. 13394–13404). PMLR.

Lin, S., Zhang, J., Pan, J., Jiang, Z., Zou, D., Wang, Y.,& Ren, J. (2020). Learning event-driven video deblurring and interpolation. In *Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, part VIII 16* (pp. 695–710). Springer.

Liu, M., Cui, Y., Liu, X., Strand, L., Yin, H., & Knoll, A. (2025). Drfir: A dimensionality reduction framework for all-in-one image restoration in spatial and frequency domains. In *Expert systems with applications* (pp. 128959).

Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts arXiv:1608.03983 arXiv preprint.

Ma, J., Cheng, T., Wang, G., Zhang, Q., Wang, X., & Zhang, L. (2023) Prores: Exploring degradation-aware visual prompt for universal image restoration. arXiv preprint arXiv:2306.13653

Ma, X., Dai, X., Bai, Y., Wang, Y., & Fu, Y. (2024). Rewrite the stars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5694–5703).

Özdenizci, O., & Legenstein, R. (2023). Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*, 10346–10357.

Pan, J., Bai, H., & Tang, J. (2020). Cascaded deep video deblurring using temporal sharpness prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3043–3051).

Pan, J., Bai, H., Dong, J., Zhang, J., & Tang, J. (2021). Deep blind video super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4811–4820).

Pan, J., Sun, L., Xu, B., Dong, J., & Tang, J. (2025). Learning efficient deep discriminative spatial and temporal networks for video deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Pan, J., Xu, B., Bai, H., Tang, J., & Yang, M. H. (2023). Cascaded deep video deblurring using temporal sharpness prior and non-local spatial-temporal similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(8), 9411–9425.

Paszke, A. (2019). Pytorch: An imperative style, high-performance deep learning library arXiv:1912.01703 arXiv preprint.

Potlapalli, V., Zamir, S. W., Khan, S. H., & Shahbaz Khan, F. (2023). Promptir: Prompting for all-in-one blind image restoration. arXiv preprint arXiv:2306.13090

Ranjan, A., & Black, M. J. (2017). Optical flow estimation using a spatial pyramid network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4161–4170).

Ren, D., Shang, W., Yang, Y., & Zuo, W. (2023). Aggregating long-term sharp features via hybrid transformers for video deblurring arXiv:2309.07054 arXiv preprint.

Ren, W., Zhang, J., Xu, X., Ma, L., Cao, X., Meng, G., & Liu, W. (2018). Deep video dehazing with semantic segmentation. *IEEE Transactions on Image Processing, 28*(4), 1895–1908.

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

Shang, W., Ren, D., Zou, D., Ren, J. S., Luo, P., & Zuo, W. et al (2021). Bringing events into video deblurring with non-consecutively blurry frames. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4531–4540).

Sun, S., Ren, W., Li, J., Zhang, K., Liang, M., & Cao, X. (2023). Event-aware video deraining via multi-patch progressive learning. *IEEE Transactions on Image Processing, 32*, 3040–3053.

Tassano, M., Delon, J., & Veit, T. (2019). Dvdnet: A fast network for deep video denoising. In *2019 IEEE international conference on image processing (ICIP)* (pp. 1805–1809). IEEE.

Tassano, M., Delon, J., & Veit, T. (2020). Fastdvdnet: Towards real-time deep video denoising without flow estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1354–1363).

Valanarasu, J. M. J., Yasarla, R., & Patel, V. M. (2022). Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2353–2363).

Wang, X., Chan, K. C., Yu, K., Dong, C., & Change Loy, C. (2019). Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.

Wang, Z., Zhang, Y., Zhang, D., & Fu, Y. (2023b). Recurrent self-supervised video denoising with denser receptive field. In *Proceedings of the 31st ACM international conference on multimedia* (pp. 7363–7372).

Wang, C., Pan, J., Wang, W., Dong, J., Wang, M., Ju, Y., & Chen, J. (2023). Promptrestorer: A prompting image restoration method with degradation perception. *Advances in Neural Information Processing Systems, 36*, 8898–8912.

Wu, G., Jiang, J., Jiang, K., Liu, X., & Nie, L. (2025). Learning dynamic prompts for all-in-one image restoration. *IEEE Transactions on Image Processing*.

Wu, R., Zhang, Z., Zhang, S., Gou, L., Chen, H., Zhang, L.,& Zuo, W. (2024). Self-supervised video desmoking for laparoscopic

surgery. In *European conference on computer vision* (pp. 307–324). Springer.

Xu, X., Yu, Y., Jiang, N., Lu, J., Yu, B.,& Jia, J. (2023). Practical real video denoising with realistic degradation model. In *International conference on learning representations (ICLR)*.

Yang, Y., Aviles-Rivero, A. I., Fu, H., Liu, Y., Wang, W., & Zhu, L. (2023). Video adverse-weather-component suppression network via weather messenger and adversarial backpropagation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 13200–13210).

Yang, Y., Wu, H., Aviles-Rivero, A. I., Zhang, Y., Qin, J., & Zhu, L. (2024). Genuine knowledge from practice: Diffusion test-time adaptation for video adverse weather removal arXiv:2403.07684 arXiv preprint.

Zhang, X., Dong, H., Pan, J., Zhu, C., Tai, Y., Wang, C., & Wang, F. (2021). Learning to restore hazy video: A new real-world dataset and a new method. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9239–9248).

Zhang, K., Li, D., Luo, W., Ren, W., & Liu, W. (2022). Enhanced spatio-temporal interaction learning for video deraining: Faster and better. *IEEE Transactions on Pattern Analysis and Machine Intelligence,45*(1), 1287–1293.

Zhao, H., Tian, L., Xiao, X., Hu, P., Gou, Y., & Peng, X. (2024). Avernet: All-in-one video restoration for time-varying unknown degradations. In A. Globerson, L. Mackey, & D. Belgrave (Eds.), *Advances in neural information processing systems* (Vol. 37, pp. 127296–127316). Curran Associates Inc.

Zhou, S., Zhang, J., Pan, J., Xie, H., Zuo, W., & Ren, J. (2019). Spatio-temporal filter adaptive network for video deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2482–2491).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.