UNIDEBLUR: UNIFIED MOTION DEBLURRING VIA FOURIER CONVOLUTION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Image deblurring is a fundamental task in computer vision, yet existing methods mainly specialize in isolated scenarios. Consequently, the deblurring performance of these models drops significantly across scenarios with different cameras and image contents. This motivates the development of a unified motion deblurring framework that can handle diverse blur patterns. To this end, we construct a multi-source dataset and develop a data pruning and quality control pipeline. Then, to restore blur images with various motion blur kernels across different scenarios, we propose a unified motion deblurring framework via Fourier Convolution Learning. Existing deblurring networks usually rely on pixel-level mappings and overlook the convolutional nature of blur. Thus, the network mostly fits the image content itself rather than learning the convolution process, making it difficult to handle multi-source blur images. Instead of relying on spatial mapping, we perform blur-adaptive global filter learning in Fourier space, which avoids the limited spatial support of kernel estimation. Extensive experiments demonstrate that our unified model effectively handles motion blurs across diverse sources and scenes, achieving competitive performance across both in-domain and OOD benchmarks. Code will be released.

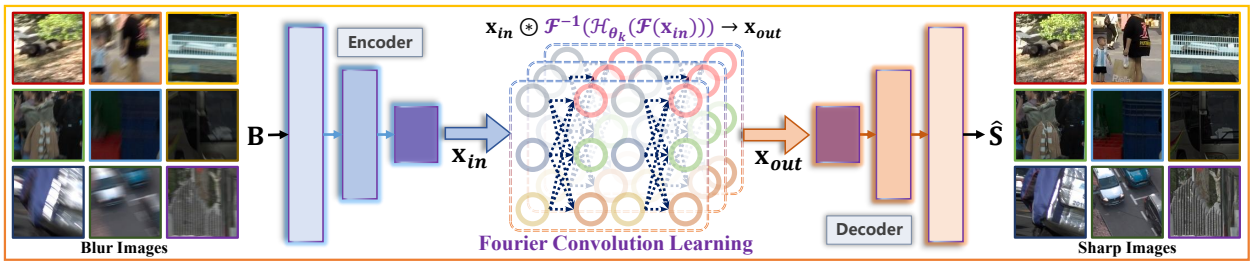


Figure 1: UniDeblur via Fourier Convolution Learning.

1 INTRODUCTION

Image deblurring, as a fundamental subtask of image restoration, aims to recover latent sharp images from blur inputs Nah et al. (2017); Kupyn et al. (2019). In real-world scenarios, the formation of a blurry image B is formulated as the spatial convolution of a sharp latent image S with a motion blur kernel k_{blur}, along with additive noise n Fergus et al. (2006); Pan et al. (2016):

$$\mathrm{B} = \mathrm{S} \circledast \mathrm{k}_{blur} + \mathrm{n}, \tag{1}$$

where $\circledast$ denotes the 2D convolution operator. This degradation process destroys high-frequency edge details and delicate textures, leading to visually unpleasant artifacts and severely impairing downstream vision tasks.

Recently, deep learning-based approaches have achieved remarkable improvements in image deblurring Zamir et al. (2021); Chen et al. (2022); Wang et al. (2022). Existing methods predominantly develop specialized frameworks on isolated benchmark datasets. However, as shown in Table 1, real-world motion blurs exhibit substantial divergence across datasets due to disparate cameras (GoPro, Sony, CHRONOS), distinct resolutions, and diverse scenes (outdoor streetscapes, pedestrians, night-time, low-light). Consequently, models trained on a single isolated scenario suffer from severe performance drops when deployed across unseen environments. To alleviate cross-domain degradation, recent All-in-One restoration models attempt to handle diverse corruptions within a unified framework Li et al. (2022); Potlapalli et al. (2023). Nonetheless, their performance on motion deblurring remains constrained. Denoising and deraining are predominantly additive corruptions, whereas motion blur is a convolutional degradation involving structural distortion and phase shifts; forcibly optimizing these tasks within a single model inevitably causes optimization

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 2: Visualization of blur images, deconvolution kernels and kernels by FCL on various datasets.

Table 1: Summary of various motion deblurring datasets.

Dataset	Acquisition	Images	Videos	Camera	Resolution	Content
GoPro Nah et al. (2017)	Averaged	3.2K	33	GoPro HERO4	1280 × 720	Outdoor Scenes
HIDE Shen et al. (2019)	Averaged	8.4K	31	GoPro HERO	1280 × 720	Pedestrian in Outdoor
REDS Nah et al. (2021)	Averaged	30K	300	GoPro HERO6	1280 × 720	Outdoor Scenes
RSBlur Rim et al. (2022)	Beam Splitter	13.3K	697	-	≈1915 × 1200	Outdoor Scenes
RealBlur Rim et al. (2020)	Beam Splitter	9.1K	232	Sony A7RM3	≈660 × 750	Static Object at Night
LOLBlur Zhou et al. (2022)	Simulated	12K	200	Sony RX10 IV	1120 × 640	Low-light Scenes
MIORe Ciubotariu et al. (2025)	Averaged	135K	333	CHRONOS 2.1-HD	1920 × 1080	Outdoor Scenes
FoundIR Li et al. (2025)	Slider	-	-	Different Cameras	1920 × 1080	Static Object
BSD Zhong et al. (2023)	Beam Splitter	33K	100	-	640 × 480	Outdoor Scenes

conflicts and feature interference. Moreover, these All-in-One approaches Potlapalli et al. (2023); Tang et al. (2026) heavily over-rely on the synthetic GoPro dataset Nah et al. (2017), which comprises only 3.2K blur-sharp pairs across 33 scenes (with merely 2.1K training pairs). Models trained on such limited distributions lack the rich semantic and kernel variations of the real world, leading to severe overfitting. To this end, developing a unified motion deblurring framework becomes highly imperative.

As shown in Fig. 1, we present a unified motion deblurring framework, termed **UniDeblur**, driven by **Fourier Convolution Learning** (FCL). On the data side, recognizing that high-quality datasets are essential for reliable training and evaluation, we design a multi-source training/evaluation protocol equipped with a dedicated data pruning and quality control pipeline. This pipeline removes samples with obvious quality issues, such as spatial misalignment and residual blur in the ground truth, to prevent the network from fitting erroneous degradation patterns. Consequently, it ensures clean supervision for model optimization and reduces the influence of obvious acquisition and annotation artifacts across both in-domain and out-of-domain (OOD) benchmarks.

On the framework side, mainstream networks are typically confined to spatial pixel-to-pixel black-box mapping, often overlooking the convolutional nature of motion blur. To overcome this limitation, we elevate the restoration paradigm to blur-adaptive convolution learning. By exploiting the Convolution Theorem, FCL reformulates spatial kernel estimation as element-wise multiplication in the Fourier space. As shown in Fig. 2, this formulation enables the network to predict blur-adaptive frequency filters and dynamically modulate the deep features. Thus, it captures degradation-aware kernels in diverse motion blur scenes with minimal computational overhead.

The main contributions are summarized as follows:

- We propose UniDeblur, a unified motion deblurring framework with a multi-source training and evaluation protocol, ensuring effective training supervision and evaluation.
- We design the Fourier Convolution Learning module, which reformulates complex spatial kernel prediction into efficient frequency-domain multiplication, enabling blur-adaptive filter learning to effectively capture diverse input-specific blurs.
- Extensive experiments on both in-domain and OOD benchmarks demonstrate that our method achieves competitive restoration performance and superior cross-domain generalization ability.

2 Related Work

Deep Image Deblurring In recent years, deep learning-based methods have achieved significant success in image deblurring Nah et al. (2017); Tao et al. (2018); Zamir et al. (2021). Pioneering approaches attempted to estimate the blur kernel using Convolutional Neural Networks (CNNs) and then restore the latent image via non-blind deconvolution Sun et al. (2015). However, these methods struggle with spatially-variant blurs in dynamic scenes and are sensitive to noise. Consequently, almost all recent methods have shifted towards end-to-end training paradigms, mapping blurry images directly to their sharp counterparts. To achieve better performance, various architectural improvements have been proposed, ranging from multi-stage CNNs (MPRNet Zamir et al. (2021), DMPHN Zhang et al. (2019)) to advanced Vision Transformers (Uformer Wang et al. (2022), Restormer Zamir et al. (2022)). Despite their success, these methods primarily focus on *pixel-level* restoration, merely learning the residual between blur and sharp image pairs while overlooking the convolutional nature of the blur process. To revisit kernel learning, some recent works introduce kernel prior-assisted models: UFPNet Fang et al. (2023) pre-trains a flow-based model to estimate spatially variant kernel maps, and PGDN Fang et al. (2025) introduces a parameterized motion kernel estimator. However, kernels learned in the spatial domain are often limited in size and capacity for convolution.

Frequency Domain Methods for Vision Tasks A growing literature corpus has demonstrated the effectiveness of extracting information from the frequency domain for various vision tasks Rippel et al. (2015); Qin et al. (2020); Rao et al. (2021). Since the Convolution Theorem states that convolution in the spatial domain is equivalent to element-wise multiplication in the frequency domain, Fourier space presents a tractable framework for handling large-scale spatial transformations. For example, LaMa Suvorov et al. (2022) applies Fast Fourier Convolution (FFC) to capture a wider receptive field for image inpainting. GFNet Rao et al. (2021) learns long-term spatial dependencies in the frequency domain for image classification. For image deblurring, DeepRFT Mao et al. (2023) introduces an additive paradigm of Res-FFT-ReLU to recognize blur patterns. LoFormer Mao et al. (2024) and FFTformer Kong et al. (2023) further explore frequency information for self-attention. SegDeblur Kim et al. (2024) and Self-Deblur Kim et al. (2025) try to predict the image-level deconvolution kernel in logarithmic Fourier space for reliable prior information. However, these approaches merely utilize frequency representations for additive feature aggregation or auxiliary prior, leaving the potential of convolution learning in the feature space largely underexplored for the deblurring task.

Unified Image Restoration Nowadays, almost all existing deblurring models are trained and evaluated on specific datasets (GoPro Nah et al. (2017) for global camera shake, LOLBlur Zhou et al. (2022) for night-time blur). This reliance on single-dataset training severely limits their generalization capability when deployed in complex, real-world open scenarios. Recently, All-in-One image restoration models Li et al. (2022); Potlapalli et al. (2023); Hu et al. (2025); Tang et al. (2026) attempted to solve multiple degradations (such as noise, rain, and blur) using a single unified network. However, denoising and deraining are additive degradations requiring local spatial aggregation, whereas motion blur is a convolutional degradation that demands large-scale deformation and phase shift correction, so that their performance on deblurring remains constrained.

DeMoE Feijoo et al. (2026) proposed a Mixture-of-Experts decoder for unified image deblurring, aiming to address both motion and defocus blur within a single model. However, compared to the diverse range of available motion blur benchmarks, existing defocus datasets Abuolaim & Brown (2020); Lee et al. (2021); Ruan et al. (2021); Li et al. (2023b) suffer from several inherent issues, such as color shifts, pixel misalignment, and limited scene diversity. Therefore, instead of mixing multiple degradations, we specifically target a Unified Motion Deblurring framework which learns the essence of blur degradation via Fourier Convolution Learning.

3 Unified Motion Deblurring Framework

Overall Architecture. As illustrated in Fig. 3(a), the overall architecture of our UniDeblurNet adopts a 3-level U-Net structure, inspired by the recent success of state-of-the-art architectures such as EVSSM Kong et al. (2025) and BioIR Cui et al. (2026). The restoration process can be formally described as follows:

$$\mathrm{B} \xrightarrow{\mathcal{E}_{\theta_{enc}}} \{\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3\} \xrightarrow{\mathcal{D}_{\theta_{dec}}} \{\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3\} \xrightarrow{\mathcal{T}_{\theta_{tail}}} \hat{\mathrm{P}} \xrightarrow{+\mathrm{B}} \hat{\mathrm{S}}, \tag{2}$$

where $\mathcal{E}_{\theta_{enc}}$, $\mathcal{D}_{\theta_{dec}}$ and $\mathcal{T}_{\theta_{tail}}$ denote the encoder, the decoder and the convolution tail (parameterized by θ_{enc}, θ_{dec} and θ_{tail}), $\{\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3\}$, $\{\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3\}$ and $\hat{\mathrm{P}}$ are the outputs of the encoder, the decoder and the convolution tail, $\hat{\mathrm{S}}$ is the restored image of the model.

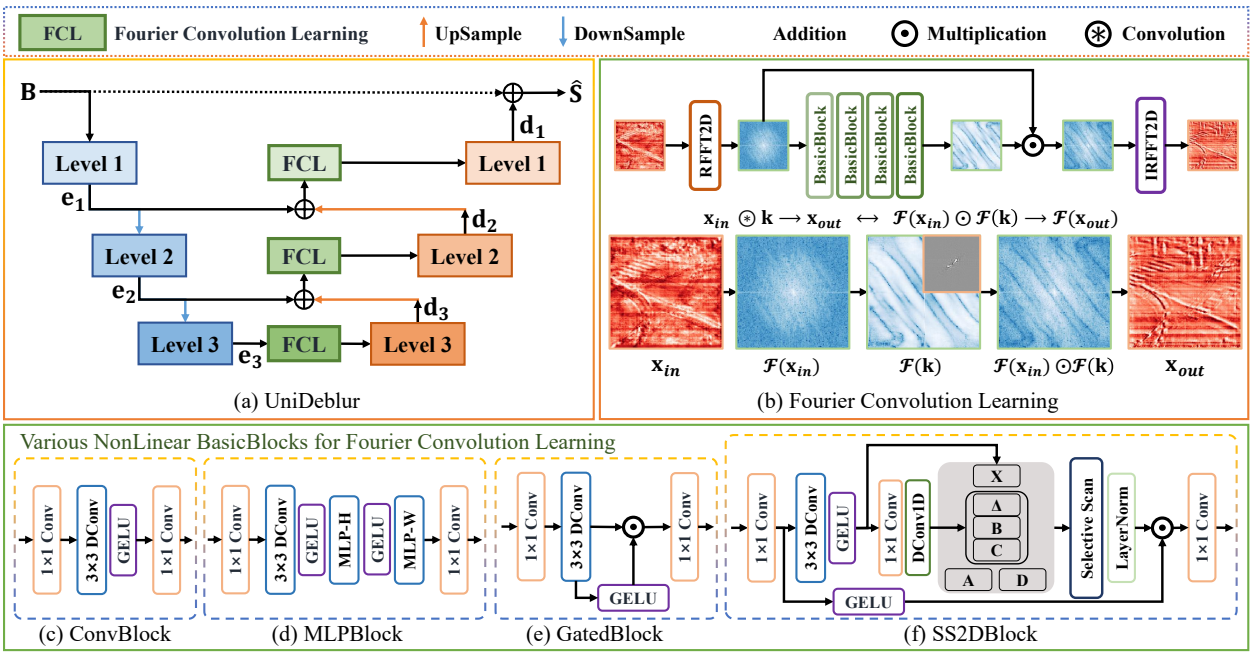


Figure 3: Architecture of UniDeblur, Fourier Convolution Learning and Basicblocks.

3.1 Fourier Convolution Learning: Blur-Adaptive Filter

From Spatial Convolution to Frequency Filtering. As formulated in Eq. 1, motion blur is fundamentally a convolution process. However, most existing deep networks rely on a direct pixel-to-pixel mapping strategy $\mathrm{B} \rightarrow \hat{\mathrm{S}}$ in the spatial domain. This black-box mapping merely attempts to fit spatial differences while overlooking the convolutional nature of the blur degradation. Forcing standard CNNs to implicitly learn complex, large-scale inverse convolutions inevitably leads to severe receptive field bottlenecks and optimization difficulties. To this end, we shift the convolution learning process to the Fourier space. According to the Convolution Theorem, spatial convolution is mathematically equivalent to element-wise multiplication in the frequency domain:

$$\mathcal{F}(\mathrm{B}) = \mathcal{F}(\mathrm{S}) \odot \mathcal{F}(\mathrm{k}_{blur}) \rightarrow \mathrm{S} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathrm{B}) \odot \mathcal{F}(\mathrm{k}_{deblur})\right), \tag{3}$$

where k_{deblur} is the inverse kernel of k_{blur}, $\mathcal{F}$ denotes the 2D Real Fast Fourier Transform (RFFT2D), $\mathcal{F}^{-1}$ is the Inverse RFFT2D, and $\odot$ is the Hadamard product. This conversion simply reformulates the intractable spatial kernel estimation into a simpler multiplication.

Fourier Convolution Learning (FCL). Existing frequency-based architectures fail to fully exploit this property. For instance, DeepRFT Mao et al. (2023) adopts an additive paradigm:

$$\mathrm{x}_{out} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathrm{x}_{in}) + \mathcal{H}_{\theta}(\mathcal{F}(\mathrm{x}_{in}))\right), \tag{4}$$

which treats frequency representations merely as additive residual features rather than convolution kernels. Conversely, while GFNet Rao et al. (2021) utilizes frequency-domain multiplication:

$$\mathrm{x}_{out} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathrm{x}_{in}) \odot \mathrm{W}\right), \tag{5}$$

it learns a static global filter W that acts merely as a degradation-agnostic prompt, lacking the adaptability to handle varying, image-specific motion blurs.

Inspired by Eq. 3, we propose the FCL module (Fig. 3(b)) to instantiate a blur-adaptive convolution process:

$$\mathrm{x}_{out} = \mathcal{F}^{-1}\left(\mathcal{F}(\mathrm{x}_{in}) \odot \mathcal{H}_{\theta_k}(\mathcal{F}(\mathrm{x}_{in}))\right). \tag{6}$$

Specifically, FCL transforms x_{in} from spatial domain to frequency domain, where a lightweight non-linear network $\mathcal{H}_{\theta_k}$ dynamically predicts blur-adaptive global filters $\mathcal{F}(\mathrm{k})$. These blur-adaptive filters are then multiplied with $\mathcal{F}(\mathrm{x}_{in})$ before applying the IRFFT2D. This design enables the model to perform blur-adaptive convolution in Fourier space.

Multi-Scale FCL. In standard U-Net architectures, encoder-decoder transitions are typically achieved via additive mechanisms (skip-addition): $\hat{\mathrm{x}}_i = \mathrm{e}_i + \mathrm{d}^{\uparrow}_{i-1}$. While effectively preserving high-frequency structures, this straightforward spatial combination simultaneously propagates redundant, degradation-corrupted features from the encoder. These degraded features can interfere with the decoder when reconstructing sharp details.



(a) BSD (blur) (b) MIORe (noise & defocus) (c) VAR-MIORe (noise & defocus) (d) FoundIR (misalign)

Figure 4: Various flaws in OOD datasets.

Table 2: Summary of training and testing datasets.

Dataset	Train			Test			
	Images	Patches	Resolution	Images	Removed	Resolution	Step
GoPro Nah et al. (2017)	2,103	16,824	512 × 512	1,111	0	1280 × 720	1
HIDE Shen et al. (2019)	6,397	12,794	720 × 720	203	0	1280 × 720	10
REDS Nah et al. (2021)	24,000	48,000	720 × 720	299	1	1280 × 720	10
RSBlur Rim et al. (2022)	8,878	53,268	720 × 720	336	0	≈1915 × 1200	10
RealBlur Rim et al. (2020)	7,516	15,032	≈660 × 750	1,960	0	640 × 640	1
LOLBlur Zhou et al. (2022)	10,200	20,400	640 × 640	600	0	1120 × 640	3
BSD1(1ms-8ms) Zhong et al. (2023)	-	-	-	500	0	640 × 480	6
BSD2(2ms-16ms) Zhong et al. (2023)	-	-	-	471	29	640 × 480	6
BSD3(3ms-24ms) Zhong et al. (2023)	-	-	-	401	99	640 × 480	6
MIORe Ciubotariu et al. (2025)	-	-	-	458	328	1920 × 1080	10
VAR-MIORe Ciubotariu et al. (2025)	-	-	-	203	230	1920 × 1080	3
FoundIR Li et al. (2025)	-	-	-	150	0	1920 × 1080	1

To address this issue, we introduce Multi-Scale FCL at all encoder-decoder transitions:

$$\mathrm{x}_i = \mathcal{F}^{-1}\left(\mathcal{F}(\hat{\mathrm{x}}_i) \odot \mathcal{H}_{\theta_k}(\mathcal{F}(\hat{\mathrm{x}}_i))\right), \tag{7}$$

where the feature $\hat{\mathrm{x}}_i$ is refined into x_i before entering the decoder. By executing blur-adaptive filtering, the network effectively modulates the encoded representations and filters out degradation-irrelevant information. Consequently, Multi-Scale FCL transforms conventional skip connections into blur-adaptive information transmission, ensuring the decoder receives purified representations tailored to the specific blur pattern of the current input.

3.2 Multi-Source Dataset Construction and Evaluation Protocol

To rigorously train and evaluate our unified model, we establish a comprehensive multi-source dataset, categorized into in-domain training sets and out-of-distribution (OOD) test sets.

In-Domain Training Datasets. We construct our training corpus using six high-quality datasets: GoPro Nah et al. (2017), HIDE Shen et al. (2019), REDS Nah et al. (2021), RSBlur Rim et al. (2022), RealBlur Rim et al. (2020), and LOLBlur Zhou et al. (2022). These datasets feature pristine ground truths (GTs) and cover diverse scenarios (synthetic/real blurs, low-light, and dynamic motions), providing a robust foundation for unified deblurring.

OOD Testing and Quality Control. To evaluate cross-domain generalization, we select six challenging OOD test sets: BSD1/BSD2/BSD3 Zhong et al. (2023), MIORe/VAR-MIORe Ciubotariu et al. (2025), and FoundIR Li et al. (2025). As shown in Fig. 4, these datasets inherently suffer from acquisition flaws, such as mixed degradations in MIORe, residual GT blur in BSD, and pixel misalignments in FoundIR. To ensure reliable and efficient evaluation, we systematically prune the data (Table 2). Specifically, we subsample the test sets to a manageable scale and strictly filter out flawed GT-blur pairs. Additionally, high-resolution training images are cropped into patches for better I/O.

Evaluation Protocol. To mitigate evaluation bias from inconsistent GT quality, we adopt a hybrid protocol. For in-domain datasets, we use PSNR and SSIM. For OOD datasets, we rely mainly on robust no-reference metrics (CLIP-IQA Wang et al. (2023) and MUSIQ Ke et al. (2021)) to assess perceptual quality.

Table 3: Comparisons on in-domain datasets.

Method	GoPro			HIDE			REDS			RSBlur			RealBlur		LOLBlur	
	PSNR	SSIM	MUSIQ	PSNR	SSIM	MUSIQ	PSNR	SSIM	MUSIQ	PSNR	SSIM	MUSIQ	PSNR	SSIM	PSNR	SSIM
GT	-	1.000	47.97	-	1.000	60.43	-	1.000	66.37	-	1.000	37.31	-	1.000	-	1.000
Blur	25.64	0.793	24.87	23.94	0.759	26.28	26.34	0.772	39.01	28.74	0.850	21.90	30.08	0.874	33.82	0.904
MIMOUNet	30.47	0.906	39.91	28.93	0.888	47.22	32.14	0.908	62.08	32.30	0.917	30.57	33.01	0.925	41.70	0.977
DeepRFT	31.92	0.932	41.11	30.84	0.921	51.70	32.36	0.912	61.21	33.45	0.937	32.40	33.79	0.938	43.38	0.982
NAFNet	32.81	0.940	43.61	31.11	0.924	54.99	33.66	0.931	64.50	33.89	0.941	34.62	33.93	0.939	44.22	0.985
LoFormer	32.33	0.936	42.28	30.98	0.922	53.09	33.62	0.934	64.34	33.46	0.937	32.95	33.86	0.939	44.12	0.985
PromptIR	31.94	0.931	42.86	30.95	0.922	54.85	32.98	0.922	64.70	33.65	0.939	33.89	33.92	0.939	43.84	0.984
MoCEIR	32.36	0.936	43.54	31.19	0.925	55.53	33.24	0.926	65.14	33.48	0.937	34.06	33.93	0.939	44.23	0.985
BioIR	33.12	0.943	44.77	31.77	0.933	57.34	34.48	0.942	66.00	33.87	0.941	35.22	34.05	0.940	45.11	0.987
EVSSM	33.26	0.945	44.45	31.70	0.932	56.66	34.68	0.944	65.81	33.91	0.942	34.95	34.04	0.940	44.94	0.987
UniDeblur-B	33.48	0.945	44.42	32.01	0.936	57.12	34.41	0.943	65.69	34.13	0.944	35.16	34.08	0.941	44.82	0.987
UniDeblur-L	34.02	0.953	45.27	32.49	0.941	58.10	35.20	0.951	66.31	34.37	0.947	35.90	34.17	0.942	45.57	0.988

Table 4: Comparisons on OOD datasets.

Method	MIORe		VAR-MIORe		BSD1		BSD2		BSD3		FoundIR		FLOPs (G)	Mem (GB)
	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ		
GT	0.3076	55.70	0.3045	56.81	0.2695	55.05	0.2919	51.42	0.2658	49.70	0.3207	61.42	-	-
Blur	0.2120	36.75	0.1953	26.85	0.2097	40.35	0.1854	34.15	0.1740	30.89	0.1477	23.76	-	-
MIMOUNet	0.2422	44.41	0.2147	32.97	0.2415	49.08	0.2306	42.24	0.2111	41.53	0.2756	48.21	67	12
DeepRFT	0.2797	52.01	0.2299	39.91	0.2656	52.70	0.2787	48.23	0.2395	47.76	0.3146	53.90	80	24
NAFNet	0.2704	51.00	0.2224	37.88	0.2886	53.74	0.2923	49.28	0.2649	48.02	0.3102	53.82	63	35
LoFormer	0.2766	51.95	0.2272	37.17	0.2708	55.01	0.2615	49.48	0.2362	48.21	0.3142	54.49	146	179
PromptIR	0.2636	49.26	0.2197	35.13	0.2867	53.97	0.2811	48.41	0.2492	42.15	0.3099	53.55	158	212
MoCEIR	0.2601	48.69	0.2232	35.80	0.2801	54.10	0.2830	49.35	0.2513	47.27	0.3122	54.48	87	136
BioIR	0.2689	50.49	0.2265	37.67	0.2990	54.82	0.2943	49.13	0.2747	49.10	0.3239	55.83	126	233
EVSSM	0.2624	48.52	0.2248	36.46	0.2916	53.41	0.2856	47.88	0.2617	47.93	0.3172	55.62	122	236
UniDeblur-B	0.2784	51.70	0.2317	38.61	0.3006	55.62	0.3031	51.17	0.2749	49.09	0.3231	56.70	80	103
UniDeblur-L	0.2820	52.23	0.2367	40.44	0.3017	55.11	0.3010	50.08	0.2781	48.79	0.3319	57.72	130	265

4 EXPERIMENT

4.1 EXPERIMENTAL SETUP

UniDeblur Architecture. We introduce two variants, UniDeblur-B and UniDeblur-L. As shown in Fig. 3(a), UniDeblur adopts a 3-scale U-Net architecture, with detailed configurations summarized in Table 5.

Comparative Methods. We selected representative end-to-end methods MIMO-UNet Cho et al. (2021), DeepRFT Mao et al. (2023), NAFNet Chen et al. (2022), LoFormer Mao et al. (2024), EVSSM Kong et al. (2025), PromptIR Potlapalli et al. (2023), MoCEIR Zamfir et al. (2025) and BioIR Cui et al. (2026) from both motion deblurring and All-in-One image restoration. We retrain all baselines under the unified training protocol. In addition, we calculate the computational cost (FLOPs) and the total training memory (Mem) of each model for comparison.

Configuration. All models are trained for 300K iterations using Adam Kingma & Ba (2014) ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 1×10^{-2}), initial learning rate 5×10^{-4}. The learning rate is steadily decreased to 1×10^{-6} with the cosine annealing schedule. The batch size is set to 16, and the patch size is 256×256.

Loss Function. All of the models are supervised exclusively by deblurring loss, defined as the sum of the L1 loss (ℓ_1) and the Frequency Reconstruction loss (ℓ_{freq}): $\ell_{total} = \ell_1 + 0.01\ell_{freq}$.

4.2 COMPARISONS WITH REPRESENTATIVE METHODS

Results on In-Domain Datasets. As shown in Table 3, recent models (PromptIR, MoCEIR) achieve suboptimal performance (31.94 dB and 32.36 dB on GoPro) on complex motion deblurring. This limitation primarily arises because their spatial feature routing lacks an explicit mechanism to model the convolutional nature of blur. While methods like DeepRFT demonstrate the effectiveness of frequency-domain learning over the spatial baseline MIMO-UNet, they typically impose severe training memory burdens due to dense spatial-frequency transformations. By introducing the

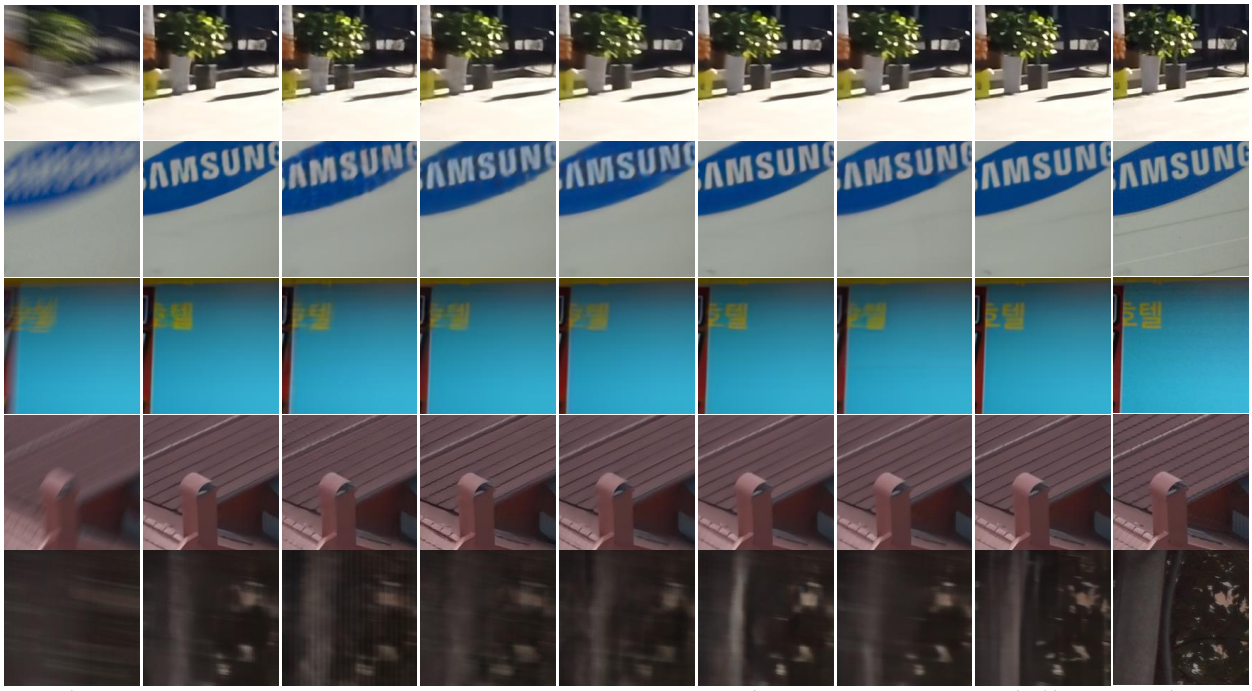


Figure 5: Visual comparison on GoPro, RSBlur, RealBlur, MIORe and VAR-MIORe.

Fourier Convolution Learning (FCL) module exclusively at the encoder-decoder transitions, UniDeblur-L boosts the GoPro PSNR from the strong baseline EVSSM's 33.26 dB to 34.02 dB. This 0.76 dB gain incurs a small computational overhead of only 8G FLOPs (from 122G to 130G), demonstrating that our FCL effectively models the convolutional degradation of blur.

Results on Out-of-Distribution Datasets. For OOD evaluation (Table 4), real-world GT quality inevitably degrades as exposure time increases (GT MUSIQ drops from 55.05 to 49.70 on BSD), justifying our reliance on robust no-reference metrics (CLIPIQA and MUSIQ). Equipped with FCL, our UniDeblur-L (adopting the identical EVSBlock and depth as EVSSM) delivers marked perceptual improvements across unseen scenarios, notably outperforming EVSSM on MIORe (52.23 vs. 48.52 MUSIQ) and VAR-MIORe (0.2367 vs. 0.2248 CLIPIQA). Furthermore, to validate structural efficiency, UniDeblur-B integrates the same BioBlock as BioIR into a shallower backbone. Despite requiring significantly lower computation and memory (80G FLOPs / 103GB Mem vs. BioIR's 126G / 233GB), UniDeblur-B surpasses BioIR across nearly all OOD metrics, trailing only marginally on BSD3 MUSIQ (49.09 vs. 49.10) and FoundIR CLIPIQA (0.3231 vs. 0.3239). These results indicate that FCL improves cross-domain performance while introducing limited computational overhead.

Visual Comparisons. As shown in Fig. 5 and Fig. 9 $\sim$ Fig. 16, while existing methods often leave noticeable residual blur, UniDeblur-L consistently recovers crisp edges and fine textures. Across both in-domain and unseen OOD scenarios, our UniDeblur effectively removes blur artifacts to reconstruct accurate and visually pleasing details.

4.3 Analysis and Discussion

To fully understand the contribution of each component in our proposed method, we conduct extensive ablation studies and visualization analysis. As summarized in Table 5, we systematically evaluate the effectiveness of Fourier Convolution Learning, the impact of multi-scale integration, the choice of frequency-domain fusion operators, the basicblock for backbone (NAFBlock Chen et al. (2022), BioBlock Cui et al. (2026), EVSBlock Kong et al. (2025)), and the basicblock for FCL (ConvBlock Fig. 3(c), MLPBlock Fig. 3(d), GatedBlock Fig. 3(e), SS2DBlock Fig. 3(f)).

Effectiveness of Fourier Convolution Learning. We first evaluate the fundamental impact of integrating FCL into the baseline network. As shown in Table 5, the baseline model (Exp 1) without FCL achieves 31.96 dB on the GoPro dataset with 14.20G FLOPs and 21GB of memory. By introducing a single-scale FCL module at the encoder-decoder transition, regardless of the lightweight non-linear basicblocks used (ConvBlock in Exp 3, MLPBlock in Exp 4, GatedBlock in Exp 6, or SS2DBlock in Exp 7), the network obtains significant performance gains across all general

Table 5: Ablation for Fourier Convolution Learning. Exp is the index of each experiment. Dim is the channel number of Level 1. Num means the number of basicblock of FCL. Mode means get $\mathcal{F}(\mathrm{x}_{out})$ via $\odot$ (multiplication) or $\oplus$ (addition). MS denotes applying FCL in multi-scale or not. Exp 6 and 13 are trained without auxiliary supervision.

	Backbone			**Fourier Convolution Learning**					**GoPro**	**HIDE**	**RSBlur**	**LOLBlur**	FLOPs	Mem
Exp	Basicblock	Level1 $\sim$ 3	Dim	Basicblock	mode	MS	Num	Dim	PSNR	PSNR	PSNR	PSNR	(G)	(GB)
1	NAFBlock	[1, 4, 8]	32	-	-	-	-	-	31.96	30.27	33.26	43.32	14.20	21
2	NAFBlock	[1, 4, 8]	32	ConvBlock	$\oplus$	$\times$	4	48	32.08	30.41	33.35	43.49	15.20	22
3	NAFBlock	[1, 4, 8]	32	ConvBlock	$\odot$	$\times$	4	48	32.23	30.56	33.51	43.62	15.20	22
4	NAFBlock	[1, 4, 8]	32	MLPBlock	$\odot$	$\times$	4	48	32.27	30.59	33.52	43.61	15.36	22
5	NAFBlock	[1, 4, 8]	32	GatedBlock	$\oplus$	$\times$	4	48	32.23	30.57	33.51	43.63	15.52	22
6	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	$\times$	4	48	32.21	30.53	33.53	43.65	15.52	22
6	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	$\times$	4	48	32.26	30.59	33.56	43.68	15.52	22
7	NAFBlock	[1, 4, 8]	32	SS2DBlock	$\odot$	$\times$	4	48	32.26	30.62	33.56	43.64	16.47	24
8	NAFBlock	[1, 4, 8]	32	ConvBlock	$\oplus$	✓	4	48	32.59	31.09	33.64	43.83	17.19	27
9	NAFBlock	[1, 4, 8]	32	ConvBlock	$\odot$	✓	4	48	32.74	31.23	33.75	43.94	17.19	27
10	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	✓	1	48	32.49	31.06	33.64	43.86	16.07	25
11	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	✓	4	32	32.71	31.24	33.78	43.93	16.49	27
12	NAFBlock	[1, 4, 8]	32	GatedBlock	$\oplus$	✓	4	48	32.70	31.24	33.72	43.96	18.31	28
13	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	✓	4	48	32.66	31.18	33.75	43.97	18.31	28
13	NAFBlock	[1, 4, 8]	32	GatedBlock	$\odot$	✓	4	48	32.74	31.28	33.78	44.04	18.31	28
14	NAFBlock	[1, 4, 8]	32	SS2DBlock	$\odot$	✓	1	48	32.56	31.08	33.65	43.88	16.79	25
15	NAFBlock	[1, 4, 8]	32	SS2DBlock	$\odot$	✓	4	48	32.85	31.41	33.83	44.10	21.19	40
16	NAFBlock	[1, 4, 8]	32	SS2DBlock	$\odot$	✓	8	32	32.90	31.45	33.88	44.12	20.50	46
17	BioBlock	[1, 4, 12]	48	-	-	-	-	-	32.52	31.14	33.57	44.28	75	95
18	BioBlock	[1, 4, 12]	48	GatedBlock	$\odot$	$\times$	4	48	33.24	31.72	34.02	44.78	77	96
19	UniDeblur-B	[1, 4, 12]	48	GatedBlock	$\odot$	✓	4	48	33.48	32.01	34.13	44.82	80	103
20	EVSBlock	[6, 6, 12]	48	-	-	-	-	-	33.26	31.70	33.91	44.94	122	236
21	EVSBlock	[6, 6, 12]	48	SS2DBlock	$\odot$	$\times$	4	48	33.73	32.20	34.26	45.42	125	248
22	UniDeblur-L	[6, 6, 12]	48	SS2DBlock	$\odot$	✓	4	48	34.02	32.49	34.37	45.57	130	265

motion deblurring datasets. In particular, Exp 6 improves the GoPro PSNR to 32.26 dB with only a 1.32G FLOPs and 1GB memory increase. This demonstrates that the FCL module can efficiently empower the model to understand the convolutional degradation process with minimal computational overhead.

Effectiveness of Multi-Scale FCL. Building on the success of single-scale, we further extend FCL to multi-scale (MS) FCL in Table 5. Comparing Exp 6 (single-scale) with Exp 13 (multi-scale) reveals a comprehensive performance improvement (GoPro PSNR increases from 32.26 dB to 32.74 dB), demonstrating the benefit of multi-scale FCL. Furthermore, we investigate the depth (Num) and width (Dim) of the non-linear basicblocks within the FCL module (Exp 10, 11, 13 and Exp 14, 15, 16). The results indicate that deeper and wider blocks generally yield better gains. For example, increasing the number of SS2DBlocks from 1 (Exp 14) to 4 (Exp 15) improves the GoPro PSNR from 32.56 dB to 32.85 dB. However, this comes with a severe memory trade-off: pushing the depth to 8 (Exp 16) brings only small gains (32.90 dB) but causes the memory footprint to surge to 46GB.

Convolution *v.s.* Addition. We compare feature integration in Fourier space via addition ($\oplus$) and multiplication ($\odot$) in Table 5. Across all single- and multi-scale settings (Exp 2 vs. Exp 3, Exp 8 vs. Exp 9), $\odot$ consistently outperforms $\oplus$. This is consistent with our formulation: multiplication in the frequency domain corresponds to convolution in the spatial domain, explicitly encouraging the network to learn the blur degradation. Using $\odot$ for the final integration also performs better when the backbone contains multiplicative gating (GatedBlock).

Auxiliary Supervision or Not. Following existing multi-scale designs (MIMO-UNet Cho et al. (2021), DeepRFT Mao et al. (2023), ConvIR Cui et al. (2024)), we project the intermediate FCL output feature into the image space via simple convolutional layers, which are supervised by downsampled ground-truth images. These projection layers are completely discarded during inference, ensuring no additional computational overhead. As shown in Table 5, although the performance gains from this supervision are marginal in Exp 6 and Exp 13, we retain this cost-free auxiliary multi-scale guidance in our final framework.

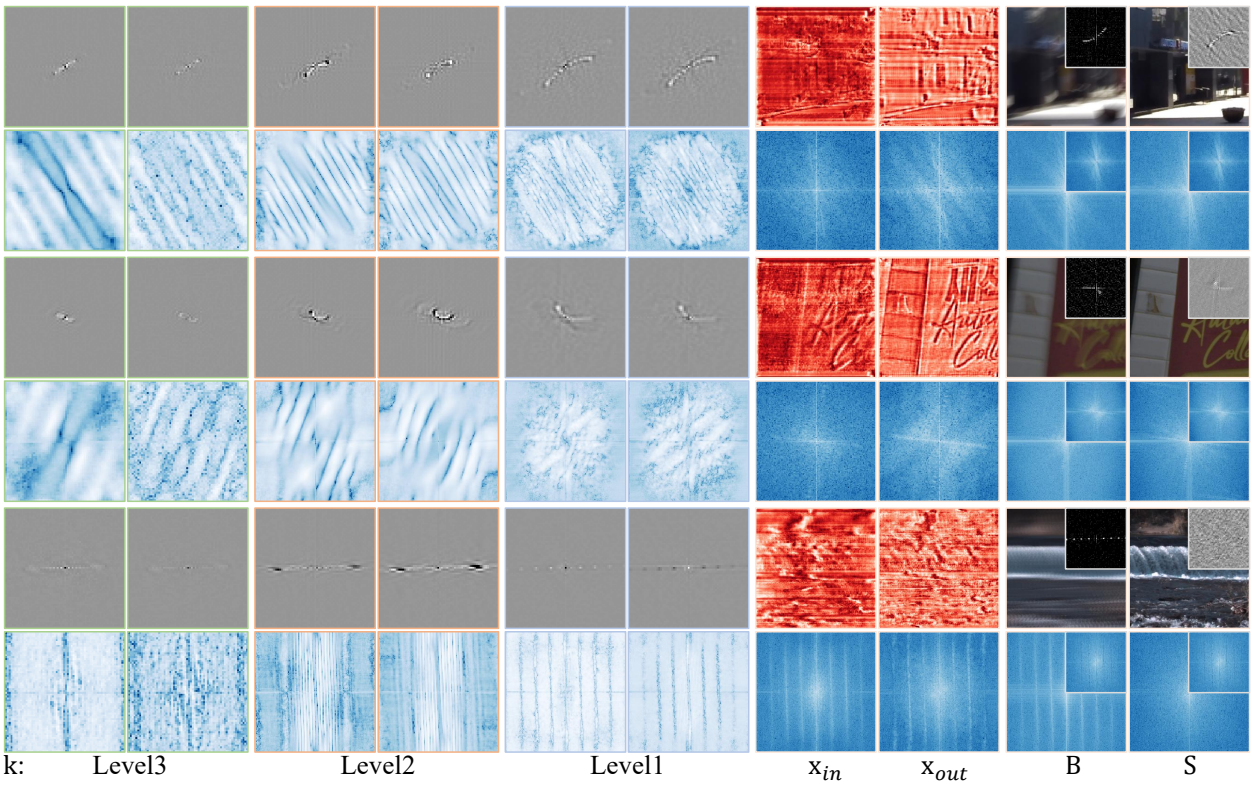


Figure 6: Visualization of the frequency filters $\mathcal{F}(\mathrm{k})$ and corresponding spatial kernels k at Levels $1 \sim 3$, together with the averaged x_{in} / x_{out} and the corresponding B / S / k_{blur} / k_{deblur} on GoPro, RSBlur, and VAR-MIORe.

Universality of Fourier Convolution Learning. To evaluate whether FCL generalizes across different backbone architectures, we further evaluate its effectiveness on other distinct backbones (Exp 17-22). As shown in Table 5, beyond the lightweight NAFBlock, FCL exhibits strong generalization on both the BioBlock Cui et al. (2026) and the EVSBlock Kong et al. (2025). For instance, the baseline BioBlock model (Exp 17) achieves 32.52 dB in PSNR on the GoPro dataset with 75G FLOPs. Integrating single-scale (Exp 18) and multi-scale FCL (Exp 19, our UniDeblur-B) significantly boosts the PSNR to 33.24 dB and 33.48 dB, respectively, while incurring a small computational overhead of only 2G to 5G FLOPs. This consistent scaling trend is similarly observed on the more advanced and heavier EVSBlock backbone ((Exp 20-22). These results indicate that FCL can be integrated into different backbone architectures with limited additional computation.

Visualization of Fourier Convolution Learning. To gain deeper insights into our proposed Fourier Convolution Learning, we visualize the global filters ($\mathcal{F}(\mathrm{k})$) and their spatial convolution kernels (k) in different U-Net scales (Level $1 \sim 3$) alongside the corresponding feature transformations (x_{in} to x_{out}) in Fig. 6 and Fig. 7. The learned filters show a clear relationship with the underlying blur degradation. As observed in the spatial domain (obtained via IFFT), the learned multi-scale kernels are similar to the theoretical deconvolution kernels (k_{deblur}). In the frequency domain, they adaptively capture the directional frequency responses corresponding to specific motion blurs. Furthermore, by comparing the frequency spectrums of the averaged x_{in} and x_{out}, it is evident that the element-wise multiplication with these blur-adaptive filters achieves an effective full-frequency enhancement. This visualization provides evidence that our FCL successfully executes input-dependent convolution learning.

5 Conclusion

In this paper, we proposed UniDeblur, a unified framework for motion deblurring based on Fourier Convolution Learning (FCL). FCL performs blur-adaptive filtering in the Fourier space, allowing the network to explicitly model the convolutional nature of motion blur. We also constructed a multi-source training and evaluation protocol with data quality control. Experiments on in-domain and OOD datasets show that FCL improves deblurring performance and generalization across different blur sources, with moderate computational overhead. These results demonstrate the effectiveness of frequency filtering learning for unified motion deblurring.

AI USE STATEMENT

In this work, we used generative AI tools for paper polishing. We have not used generative AI tools for writing code, developing ideas, conducting experiments, or creating figures and tables. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All datasets utilized in this work are publicly available and freely accessible without any application restrictions. To guarantee full reproducibility, our code, test datasets and experimental results will be systematically organized and released.

REFERENCES

Abdullah Abuolaim and Michael S. Brown. Defocus deblurring using dual-pixel data. In *Proc. ECCV*, pp. 111–126, 2020.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proc. ECCV*, pp. 17–33, 2022.

Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proc. ICCV*, pp. 4641–4650, 2021.

George Ciubotariu, Zhuyun Zhou, Zongwei Wu, and Radu Timofte. Miore & var-miore: Benchmarks to push the boundaries of restoration. In *Proc. ICCV*, pp. 19784–19793, 2025.

Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Revitalizing convolutional network for image restoration. *IEEE TPAMI*, 46(12):9423–9438, 2024.

Yuning Cui, Wenqi Ren, and Alois Knoll. Bio-inspired image restoration. In *Proc. NeurIPS*, pp. 80452–80481, 2026.

Zhenxuan Fang, Fangfang Wu, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Self-supervised non-uniform kernel estimation with flow-based motion prior for blind image deblurring. In *Proc. CVPR*, 2023.

Zhenxuan Fang, Fangfang Wu, Tao Huang, Le Dong, Weisheng Dong, Xin Li, and Guangming Shi. Parameterized blur kernel prior learning for local motion deblurring. In *Proc. CVPR*, pp. 23006–23015, 2025.

Daniel Feijoo, Paula Garrido-Mellado, Jaesung Rim, Álvaro García, and Marcos V Conde. Towards unified image deblurring using a mixture-of-experts decoder. In *Proc. CVPR*, pp. 4946–4956, 2026.

Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. *ACM TOG*, 25(3):787–794, 2006.

JiaKui Hu, Lujia Jin, Zhengjian Yao, and Yanye Lu. Universal image restoration pre-training via degradation classification. In *Proc. ICLR*, pp. 64647–64668, 2025.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proc. ICCV*, pp. 5128–5137, 2021.

Insoo Kim, Jae Seok Choi, Geonseok Seo, Kinam Kwon, Jinwoo Shin, and Hyong-Euk Lee. Real-world efficient blind motion deblurring via blur pixel discretization. In *Proc. CVPR*, pp. 25879–25888, 2024.

Insoo Kim, Geonseok Seo, Hyong-Euk Lee, and Jinwoo Shin. Refdeblur: Blind motion deblurring with self-generated reference image. *Transactions on Machine Learning Research*, 2025.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proc. CVPR*, pp. 5886–5895, 2023.

Lingshun Kong, Jiangxin Dong, Jinhui Tang, Ming-Hsuan Yang, and Jinshan Pan. Efficient visual state space model for image deblurring. In *Proc. CVPR*, pp. 12710–12719, 2025.

Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proc. ICCV*, pp. 8878–8887, 2019.

Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In *Proc. CVPR*, pp. 2034–2042, June 2021.

Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proc. CVPR*, pp. 17431–17441, 2022.

Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *Proc. ICCV*, pp. 12626–12636, 2025.

Haoying Li, Ziran Zhang, Tingting Jiang, Peng Luo, Huajun Feng, and Zhihai Xu. Real-world deep local motion deblurring. In *Proc. AAAI*, pp. 1314–1322, 2023a.

Yu Li, Dongwei Ren, Xinya Shu, and Wangmeng Zuo. Learning single image defocus deblurring with misaligned training pairs. In *Proc. AAAI*, pp. 1495–1503, 2023b.

Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *Proc. AAAI*, pp. 1905–1913, 2023.

Xintian Mao, Jiansheng Wang, Xingran Xie, Qingli Li, and Yan Wang. Loformer: Local frequency transformer for image deblurring. In *Proc. ACM MM*, pp. 10382–10391, 2024.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. CVPR*, pp. 3883–3891, 2017.

Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, and Kyoung Mu Lee. NTIRE 2021 challenge on image deblurring. In *Proc. CVPR Workshop*, 2021.

Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *Proc. CVPR*, 2016.

Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *Proc. NeurIPS*, pp. 71275–71293, 2023.

Zequn Qin, Pengyi Zhang, Fei Wu, and Xi Li. Fcanet: Frequency channel attention networks. In *Proc. ICCV*, pp. 783–792, 2020.

Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. In *Proc. NeurIPS*, 2021.

Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Proc. ECCV*, pp. 184–201, 2020.

Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *Proc. ECCV*, pp. 487–503, 2022.

Oren Rippel, Jasper Snoek, and Ryan P. Adams. Spectral representations for convolutional neural networks. In *Proc. NeurIPS*, 2015.

Lingyan Ruan, Bin Chen, Jizhou Li, and Miu-Ling Lam. Aifnet: All-in-focus image restoration network using a light field-based dataset. *IEEE Transactions on Computational Imaging*, 7:675–688, 2021.

Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *Proc. ICCV*, 2019.

Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proc. CVPR*, pp. 769–777, 2015.

Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proc. WACV*, pp. 2149–2159, 2022.

Xiaole Tang, Xiaoyi He, Jiayi Xu, Xiang Gu, and Jian Sun. Learning continuous wasserstein barycenter space for generalized all-in-one image restoration. *IEEE TPAMI*, 2026.

Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proc. CVPR*, pp. 8174–8182, 2018.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proc. AAAI*, pp. 2555–2563, 2023.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. In *Proc. CVPR*, pp. 17683–17693, 2022.

Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, and Radu Timofte. Complexity experts are task-discriminative learners for any image restoration. In *Proc. CVPR*, pp. 12753–12763, 2025.

Syed Waqas Zamir, Aditya Arora, Salman H. Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proc. CVPR*, pp. 14821–14831, 2021.

Syed Waqas Zamir, Aditya Arora, Salman H. Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. CVPR*, pp. 5728–5739, 2022.

Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proc. CVPR*, pp. 5978–5986, 2019.

Zhihang Zhong, Ye Gao, Yinqiang Zheng, Bo Zheng, and Imari Sato. Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. *IJCV*, 131(1):284–301, 2023.

Shangchen Zhou, Chongyi Li, and Chen Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *Proc. ECCV*, pp. 573–589, 2022.

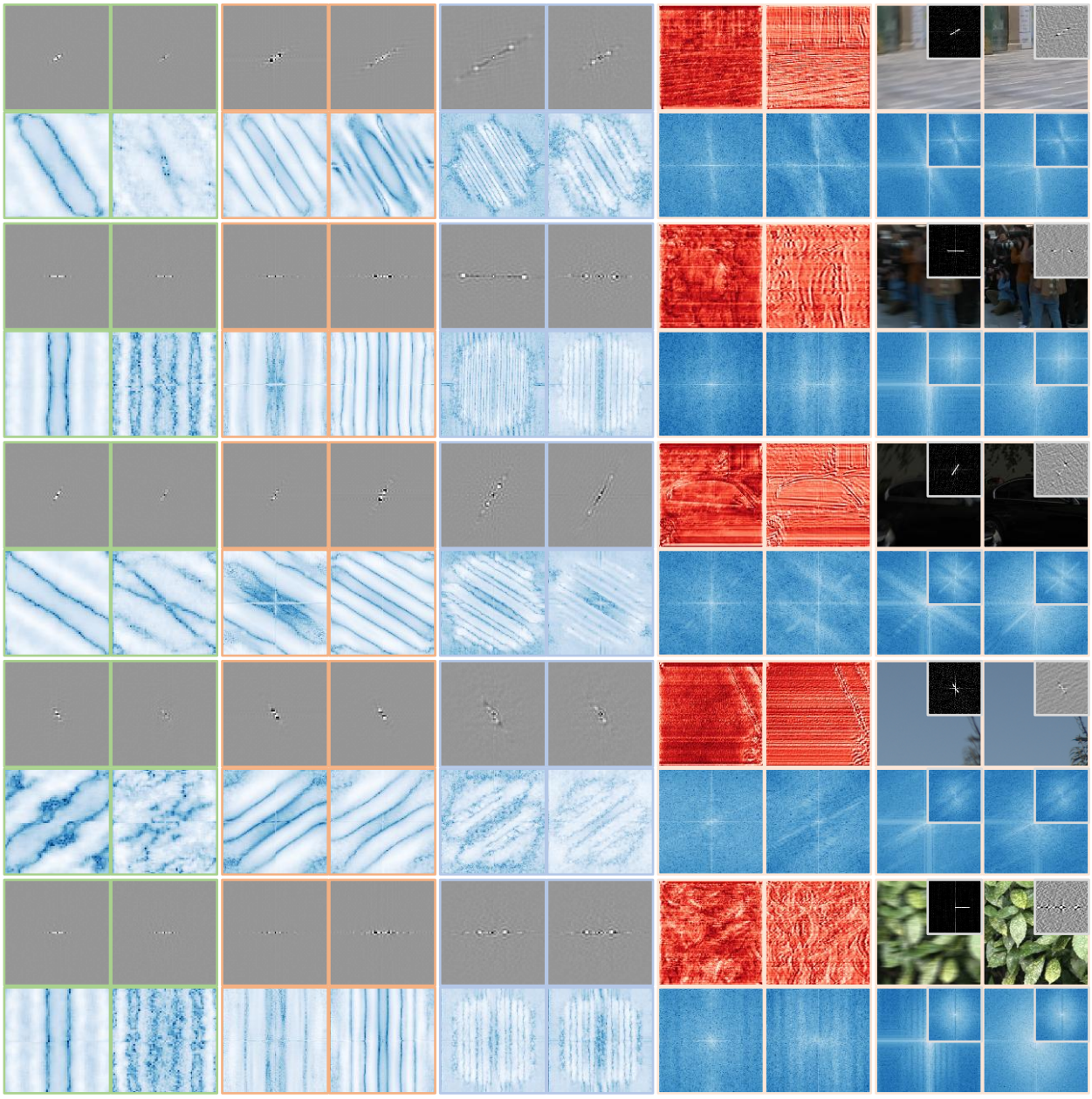


Figure 7: Visualization of the frequency filters $\mathcal{F}(\mathrm{k})$ and corresponding spatial kernels k at Levels $1 \sim 3$, together with the averaged x_{in} / x_{out} and the corresponding B / S / k_{blur} / k_{deblur} on HIDE, REDS, LOLBlur, BSD and FoundIR.

A Appendix

A.1 Visualization of Fourier Convolution Learning

Extended FCL Visualizations and Dataset Analysis. In Fig. 7, we provide additional visualizations of the FCL's input/output features (x_{in}, x_{out}) and the learned kernels (k) across different hierarchical levels (Level $1 \sim 3$) on both in-domain and OOD test sets (including HIDE, REDS, LOLBlur, BSD, and FoundIR). These visual results clearly demonstrate that our FCL module effectively learns blur-adaptive global filters that reflect the distinct motion blur degradation. Furthermore, we visualize the theoretical blur kernel (k_{blur}) in the top-right corner of the blur images (column B). Notably, in the FoundIR example (bottom row), the theoretical kernel displays a distinct off-center horizontal shift. This observation visually confirms a severe left-right spatial misalignment between the blur



Figure 8: Color shift and residual blur of ReLoBlur dataset.

Table 6: Comparisons on ReLoBlur and RealLOL.

Dataset	GT		Blur		PromptIR		MoCEIR		EVSSM		UniDeblur-L	
	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ	CLIPIQA	MUSIQ
ReLoBlur	0.2103	40.14	0.1889	37.20	0.2546	41.30	0.2169	39.85	0.2841	43.61	0.2913	43.74
RealLOL	-	-	0.1370	26.12	0.1979	29.74	0.2028	29.69	0.1995	30.04	0.2088	30.28

inputs and the ground truths in the FoundIR dataset, directly corroborating the data quality concerns discussed in the main manuscript.

A.2 Unused Datasets

ReLoBlur and RealLOL. As shown in Table 6 (provided only as diagnostic references and not used to support the quantitative benchmark claim), our proposed UniDeblur-L demonstrates promising cross-domain generalization on both the ReLoBlur Li et al. (2023a) and RealLOL Zhou et al. (2022) datasets. Despite these superior results, we deliberately exclude both datasets from our main experimental comparisons due to their inherent data defects, which could severely compromise evaluation integrity.

Specifically, the ReLoBlur dataset, captured via a dual-camera beam-splitting system, exhibits noticeable color shifts between the blur inputs and the ground truths (GTs). More severely, the majority of its GT images still contain perceptible motion blur artifacts (shown in Fig. 8). This flaw is quantitatively evidenced in Table 6, where the no-reference scores of the original blurry inputs (CLIPIQA: 0.1889, MUSIQ: 37.20) and the so-called GT images (CLIPIQA: 0.2103, MUSIQ: 40.14) are exceptionally close. Comparing restoration models on such a dataset is unreliable.

Similarly, the RealLOL dataset lacks corresponding GT images entirely. Furthermore, it consists predominantly of extreme low-light scenarios. Since no-reference perceptual metrics (CLIPIQA and MUSIQ) are sensitive to global image brightness and contrast, applying them to such dark scenes might inaccurately reflect illumination variations rather than the actual deblurring quality. So we ultimately omitted RealLOL from our core benchmark comparisons.

Table 7: Performance of NAFNet32 trained by different datasets .

	GoPro	HIDE	RealBlur	LOLBlur	REDS	RSBlur
train	PSNR	PSNR	PSNR	PSNR	PSNR	PSNR
GoPro	32.60	30.44	30.04	35.24	27.08	26.22
GoPro+HIDE	32.88	31.11	29.99	34.94	27.15	28.62
GoPro+HIDE+RealBlur	32.53	30.80	33.67	37.60	27.81	30.10
GoPro+HIDE+LOLBlur	32.63	30.93	30.63	43.46	28.58	28.68
GoPro+HIDE+RealBlur+LOLBlur	32.45	30.75	33.72	43.28	29.62	30.31

A.3 Impact of Multi-Source Training Dataset

To further investigate the impact of training on combined datasets, we evaluate NAFNet32 Chen et al. (2022) using different training data mixtures, as shown in Table 7. When combining datasets that share similar camera devices and generation pipelines (GoPro and HIDE), the model effectively aggregates homogeneous information, leading to mutual enhancement on these specific in-domain datasets. For instance, adding HIDE to the training set boosts the PSNR on GoPro from 32.60 dB to 32.88 dB, and on HIDE from 30.44 dB to 31.11 dB. However, this combination provides limited improvement on unseen scenarios.

Conversely, as we introduce more diverse, multi-source datasets (such as RealBlur and LOLBlur) into the training mixture, a characteristic trade-off emerges. The inclusion of heterogeneous blur patterns prevents the network from overfitting to a specific dataset bias, which results in a slight performance degradation on individual in-domain test sets (in the fully combined setting, the PSNR on GoPro and HIDE drops to 32.45 dB and 30.75 dB, respectively). Crucially, however, this diversification significantly enhances the model's cross-domain generalization capability. As shown in the rightmost columns of Table 7, the OOD performance steadily increases with the addition of diverse datasets, ultimately achieving the highest PSNR scores of 29.62 dB on REDS and 30.31 dB on RSBlur. This justifies our motivation for constructing a unified, multi-source training framework to tackle real-world motion blurs.

A.4 Visualization Comparisons

As shown in Fig. 9 $\sim$ Fig. 13, we provide detailed visual comparisons across multiple in-domain datasets, including GoPro, HIDE, REDS, RSBlur, RealBlur, and LOLBlur. For instance, on the GoPro (Fig. 9) and RSBlur (Fig. 12) datasets, UniDeblur-L restores legible sign numbers and poster letters. On the HIDE (Fig. 10) and REDS (Fig. 11) datasets, the leg contours of pedestrians and the textures of striped clothing are also well recovered.

Furthermore, Fig. 14 $\sim$ Fig. 17 illustrate the model's generalization ability on OOD test sets, such as MIORe, VAR-MIORe, BSD, FoundIR, and ReLoBlur. Specifically, in the localized close-ups from BSD (Fig. 15) and FoundIR (Fig. 16), UniDeblur-L recovers clear texts and geometric structures, yielding better visual results than the EVSSM.

A.5 Visualization of Removed Image Pairs

To provide a more intuitive understanding of our data pruning process, we visualize several examples of the excluded image pairs in Fig. 18 $\sim$ Fig. 22. As clearly demonstrated in these visualizations, the removed pairs suffer from severe acquisition flaws, such as residual motion blur in the ground truths and mixed degradations. This visual evidence strictly corroborates the data quality concerns discussed in the main manuscript.

It is worth reiterating that since all models are evaluated on the exact same cleaned subsets, this unified quality control avoids evaluation bias and ensures a fair, rigorous benchmark for assessing true generalization capabilities.

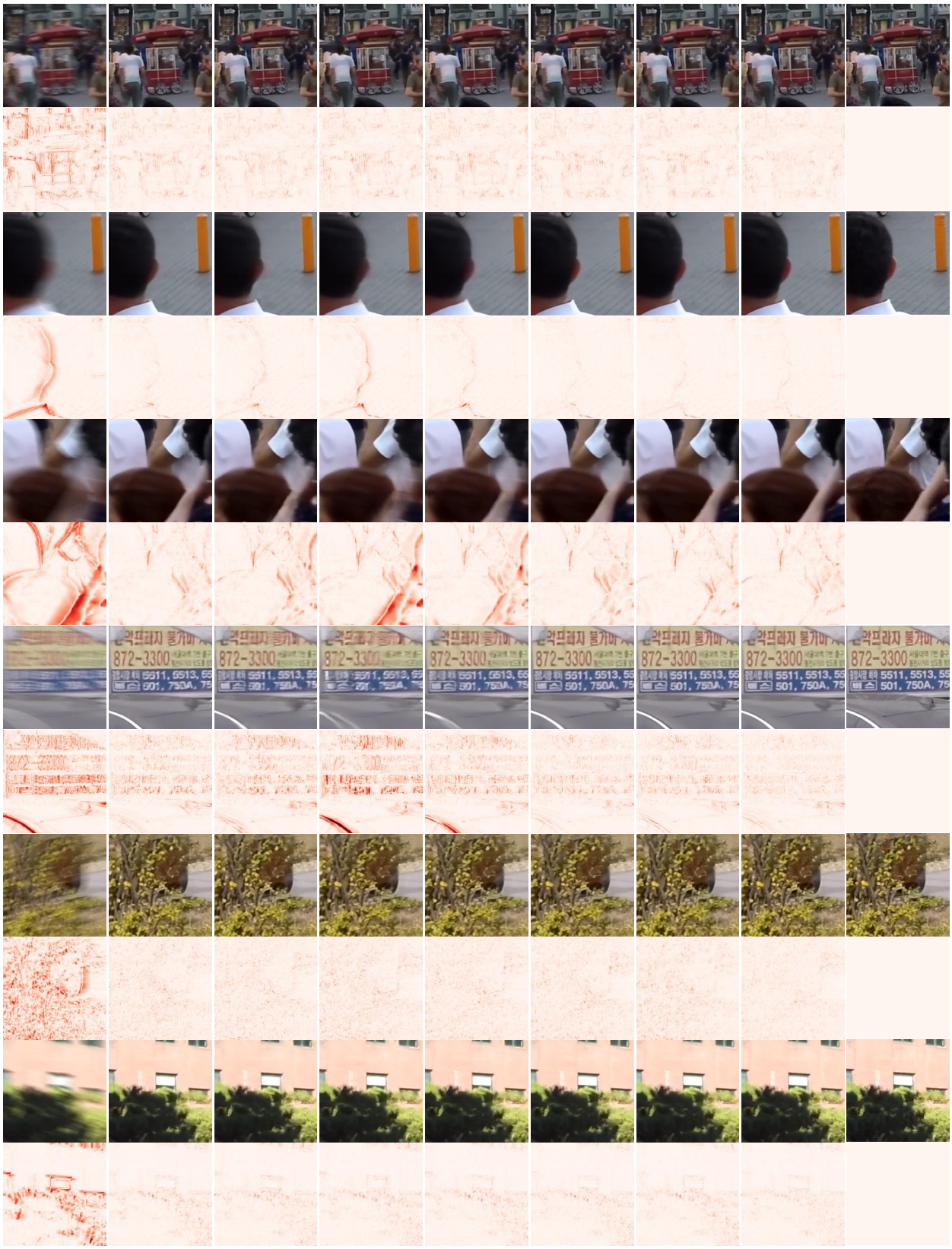


Figure 9: Visualization comparisons on GoPro dataset.

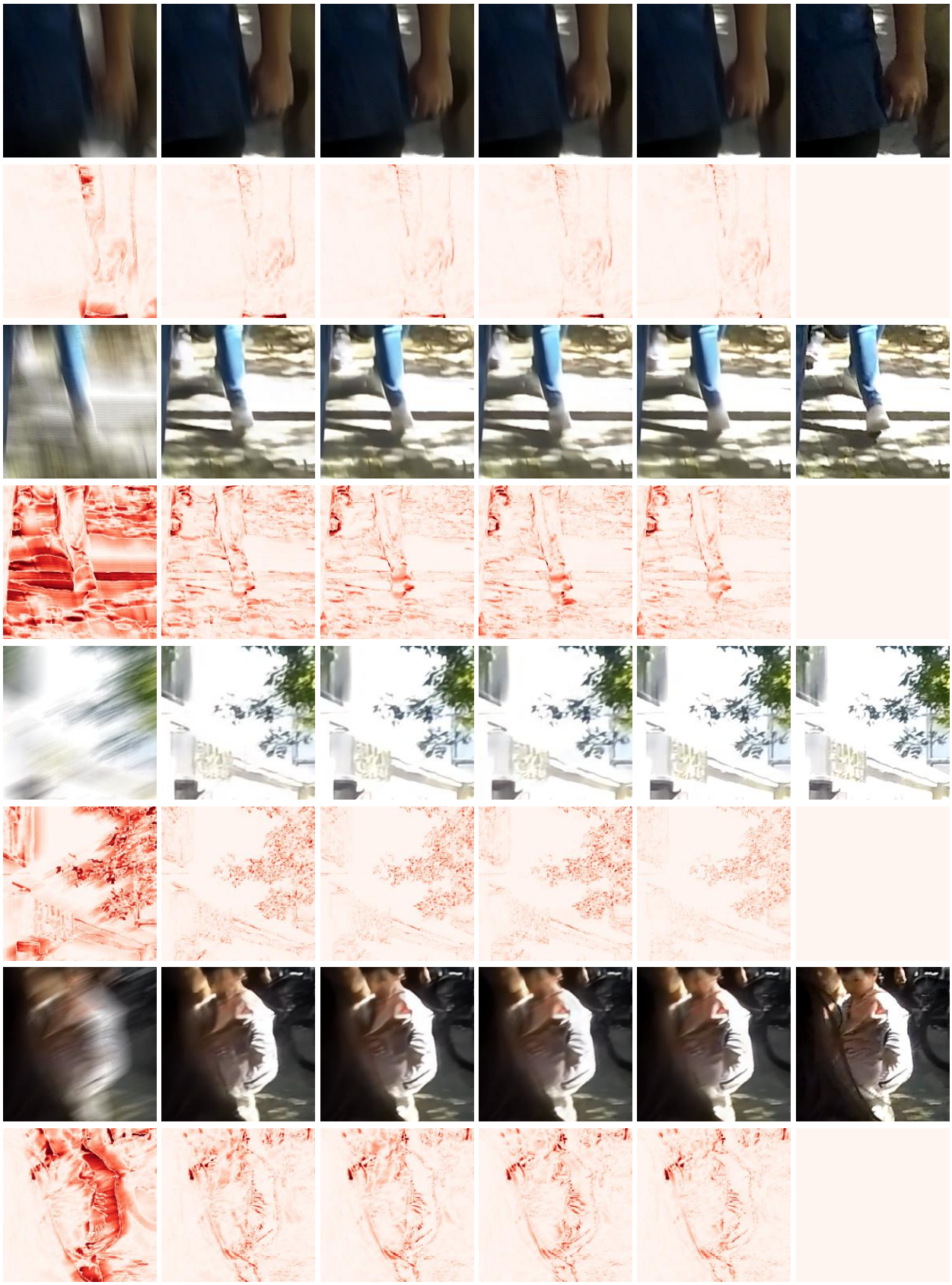


Figure 10: Visualization comparisons on HIDE dataset.

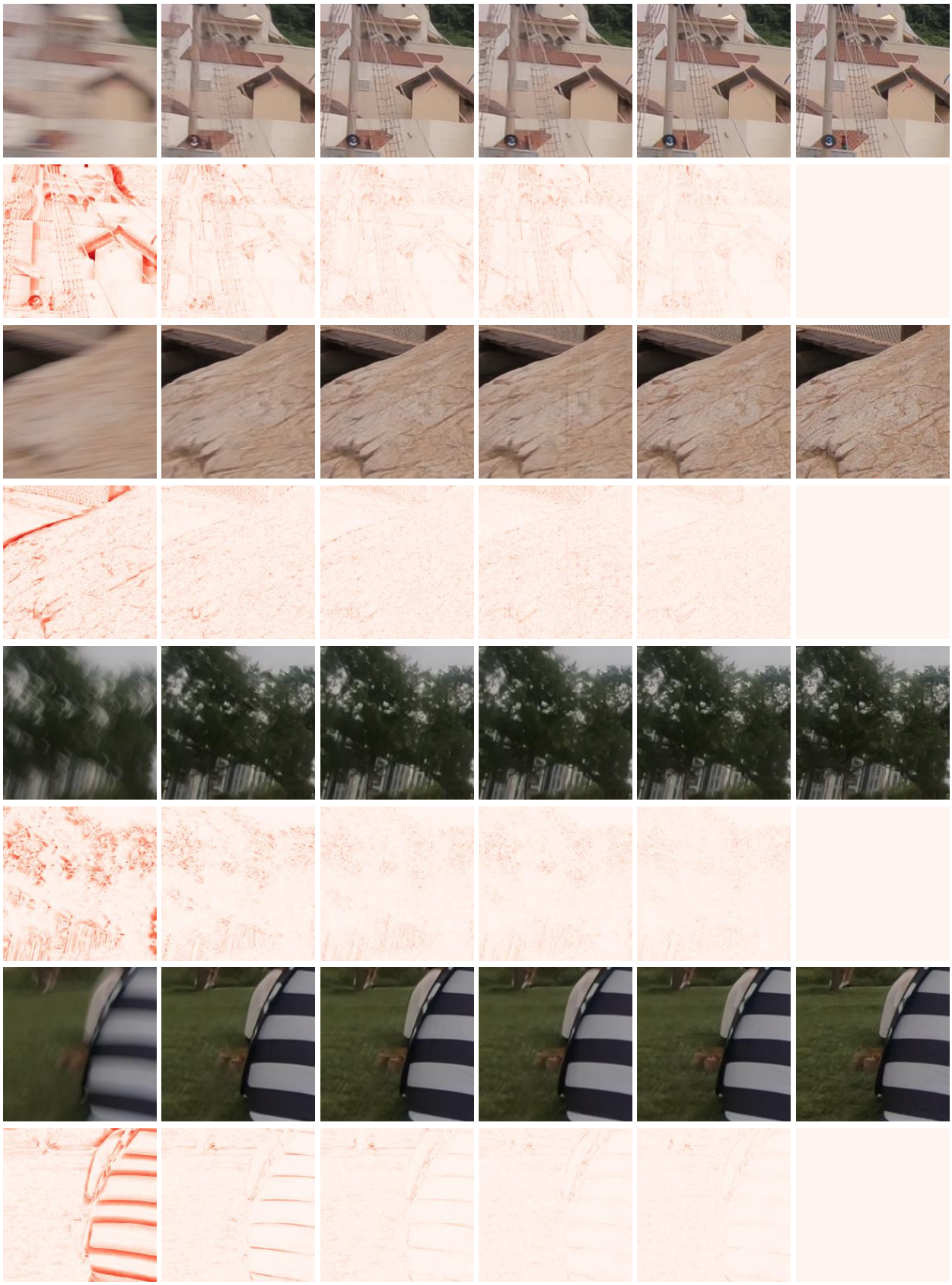


Figure 11: Visualization comparisons on REDS dataset.

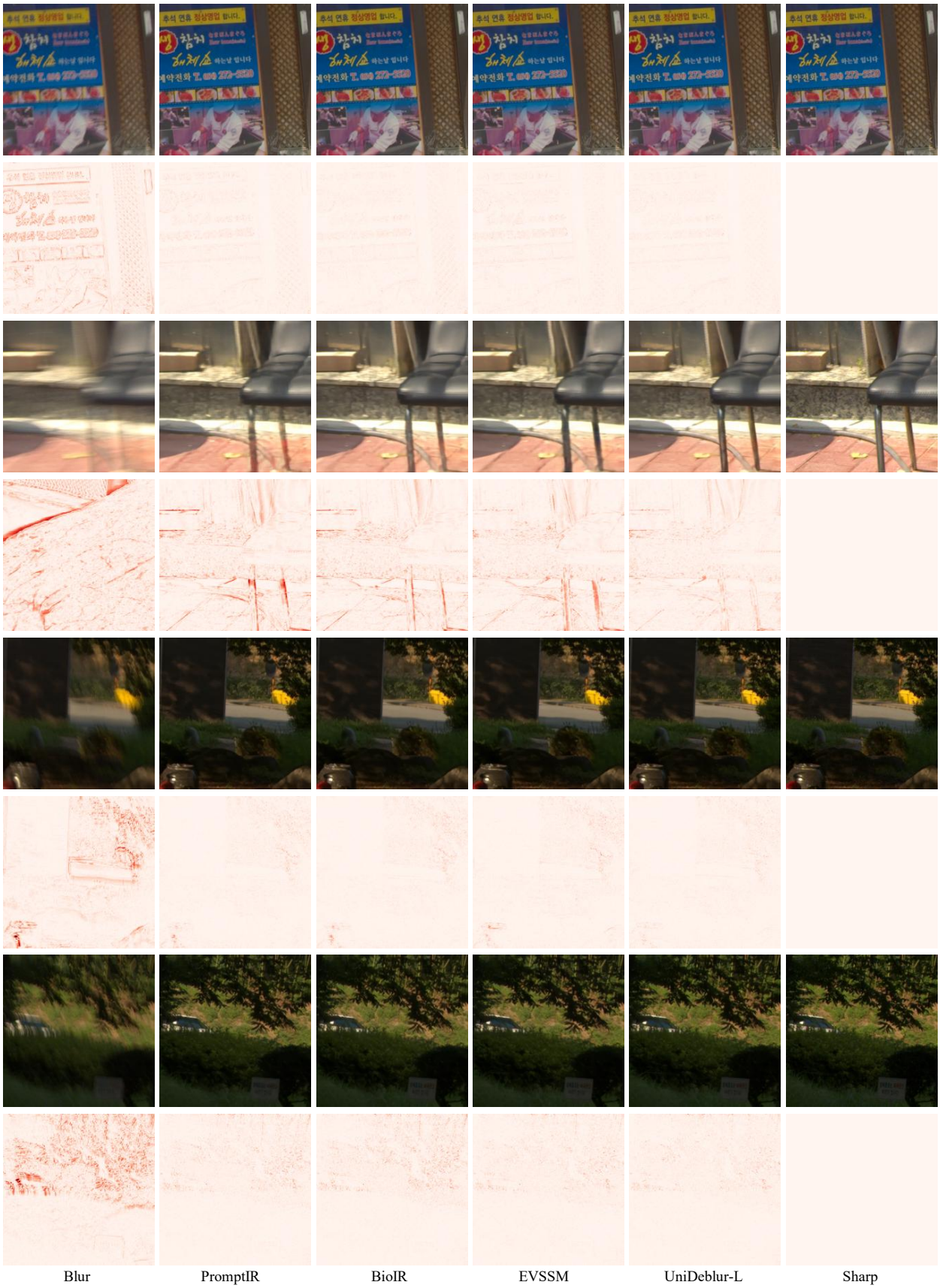


Figure 12: Visualization comparisons on RSBlur dataset.

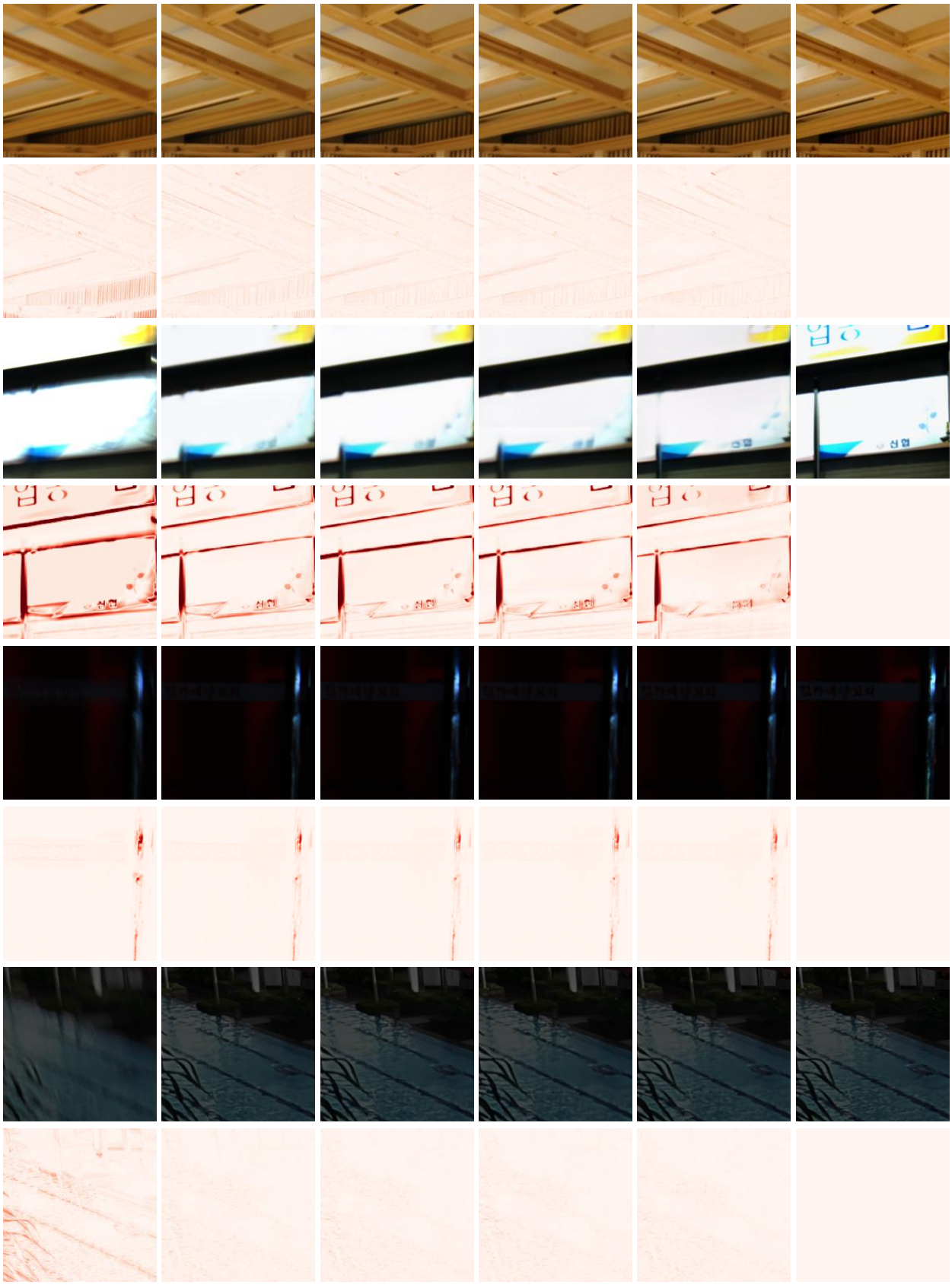


Figure 13: Visualization comparisons on RealBlur and LOLBlur dataset.



Figure 14: Visualization comparisons on MIORe and VAR-MIORe dataset.

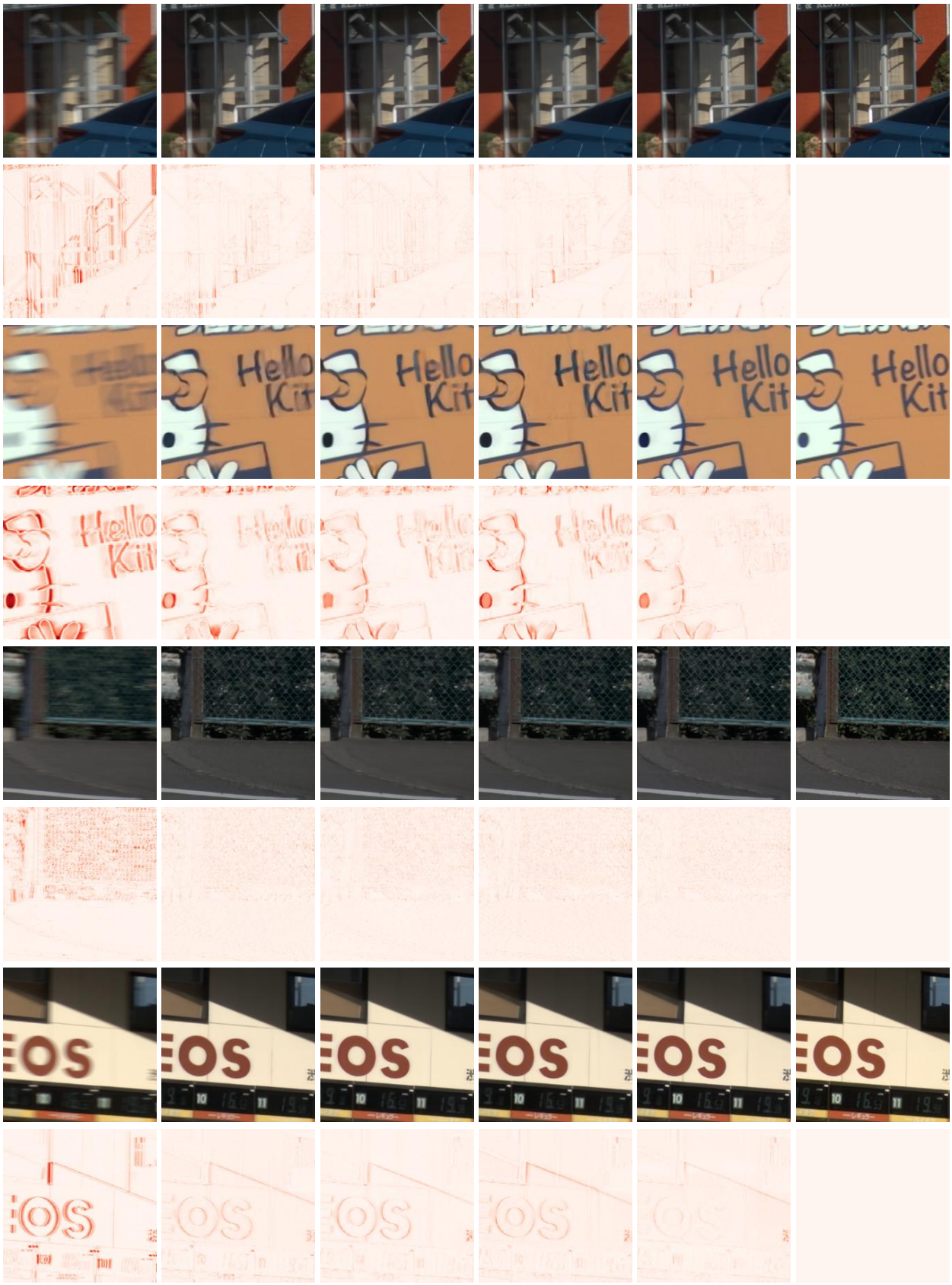


Figure 15: Visualization comparisons on BSD dataset.

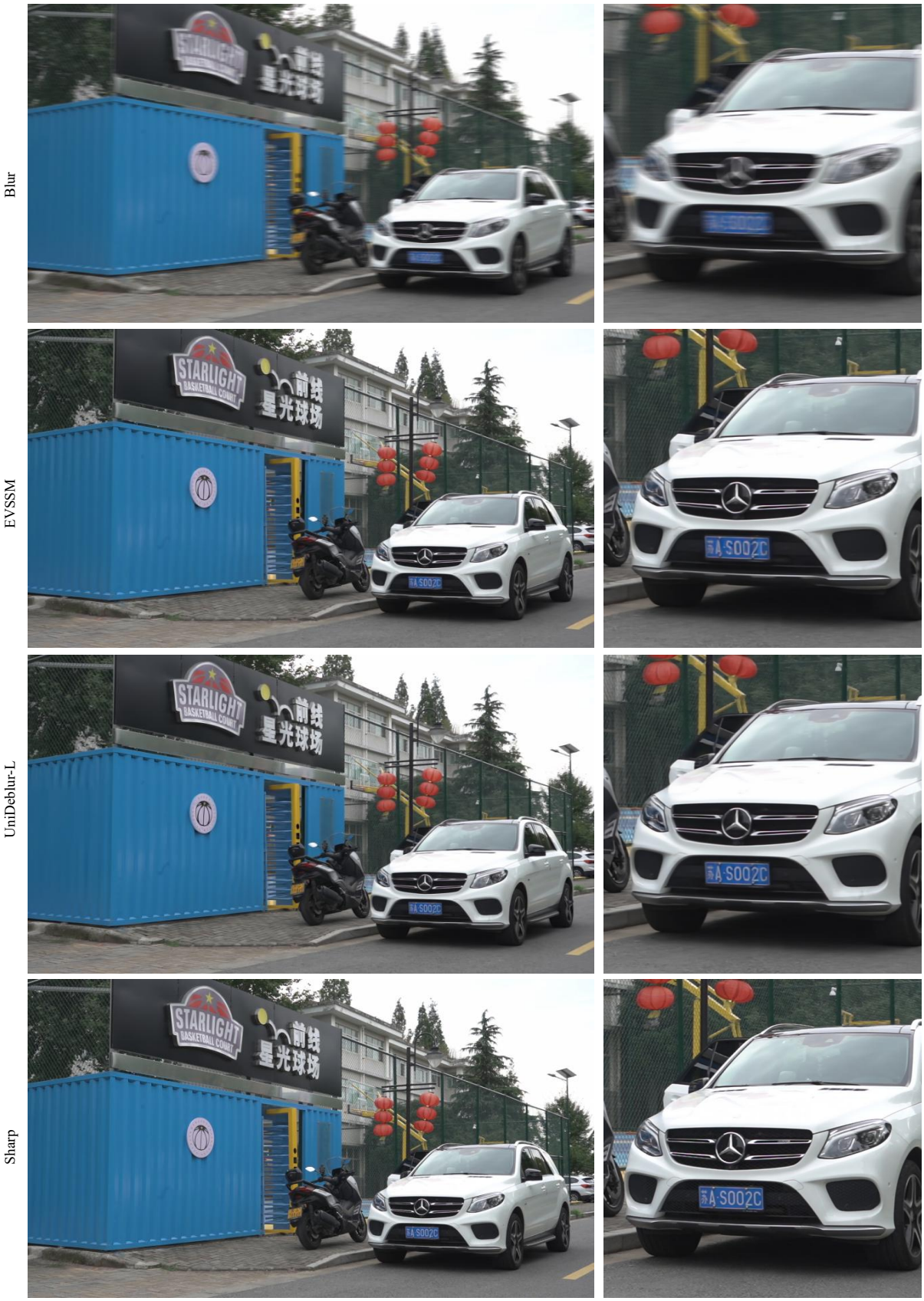


Figure 16: Visualization comparisons on FoundIR dataset.



Blur EVSSM



UniDeblur-L Sharp



Blur EVSSM UniDeblur-L Sharp

Figure 17: Visualization comparisons on ReLoBlur dataset.



Blur Sharp

Figure 18: Deleted image from REDS dataset.



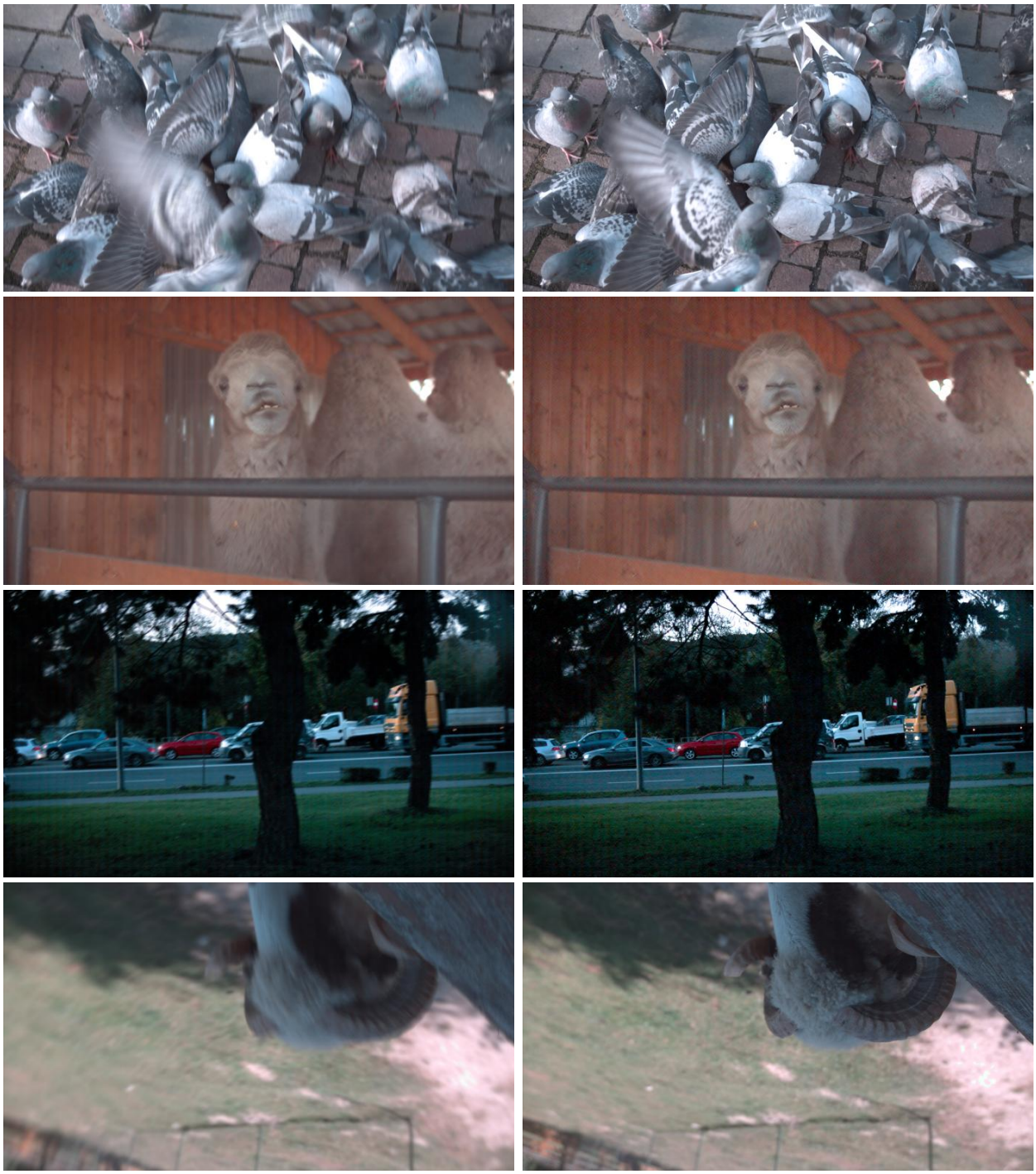
Blur Sharp

Figure 19: Deleted image from BSD2 dataset.



Blur Sharp

Figure 20: Deleted image from BSD3 dataset.



Blur Sharp

Figure 21: Deleted image from MIORe dataset.



Blur Sharp

Figure 22: Deleted image from VAR-MIORe dataset.