This CVPR Workshop paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Gain-first or Exposure-first: Benchmark for Better Low-light Video Photography and Enhancement

Haiyang Jiang
The University of Tokyo
Tokyo, Japan
jiang-haiyang777@g.ecc.u-tokyo.ac.jp

Zhihang Zhong
Shanghai AI Laboratory
Shanghai, China
zhongzhihang@pjlab.org.cn

Yinqiang Zheng *
The University of Tokyo
Tokyo, Japan
yqzheng@ai.u-tokyo.ac.jp

Abstract

Acquiring visually pleasing videos under insufficient lighting has been an important and challenging task for both photographers and algorithm engineers. Current methods have evolved into two major paradigms: prioritizing camera gain, which induces higher level of noise, and prioritizing exposure time, which brings about undesirable motion blur. Though both paths can lead to satisfying outputs, there is still a lack of direct comparison between them under the context of fast evolving deep learning algorithms, which can be crucial to shed light on better ways for capturing and enhancement. In this paper, we present a thorough study using state-of-the-art image and video enhancement frameworks, comparing ***Gain-first****,* ***Exposure-first****, and* ***Mixed*** *strategies on a large dataset collected by a special optical system so that three strategies can compete fairly under controlled conditions. Experiment results across multiple camera gain levels and exposure time settings, as well as a theoretical analysis, show advantages of Gain-first strategy over Exposure-first one under relatively small ratio, and superiority of Mixed strategy under extreme low-light cases, providing a basis for optimal videography and enhancement algorithm designs in the future.*

1. Introduction

As digital imaging devices being widely used throughout past decades, there has been increasing demands on algorithms to help them adapt to insufficient lighting situations. Particularly, low-light imaging of dynamic scenes is of great significance than ever, playing crucial roles in mobile photography, visual surveillance, video conferencing [41], underwater exploration [27], microscopy [77] and autonomous driving [37] etc. Yet, hindered by the essence of low photon count in the dark and short exposure time limited by frame rates, such task of low-light video enhancement remains a challenging one in terms of generating visually satisfying and smooth results.

Without exclusively designed hardware modifications, current solutions for low-light video capture mainly fall into two categories: 1. raising gain value, and 2. elongating exposure time. The former elevates camera gain level, which readily gives rise to conspicuous noise patterns, while the latter faces inevitable obstacles of motion blur. Therefore, dark video enhancement tasks can either be formulated as denoising [13, 30, 48, 63, 73, 76, 79] in the perspective of increasing gain level, or as deblurring [38, 68, 86, 88] in the situation of using longer exposure.

Although efforts are being made to combine two aspects together in a form of joint enhancement for low-light videos [29, 89], these paths have been developing independently of each other. Several prior works [25, 46, 47] have tried to compare denoising and deblurring under the context of HDR or imaging systems, yet neither are they able to verify their conclusions on large scale real-world data, nor do they incorporate advanced deep learning enhancement algorithms in their analysis.

To tackle this conundrum, we started by collecting a large benchmark dataset that has paired video samples for **Gain-first**, **Exposure-first**, and **Mixed** strategies, as shown in Figure 1. The dataset, consisting of more than 720 videos (i.e. 54000 images) in total spanning 3 light intensity ratios and 60 scenes for each strategy, is shot by a synchronized optical system that can record image sequences of exactly identical scenes on multiple cameras. By varying camera settings, diverse training inputs and ground truths can be recorded simultaneously.

With this dataset, we managed to analyze different strategies under precisely controlled conditions. A massive number of experiments are carried out, incorporating 2 state-of-the-art video-oriented [7, 40], 2 latest image-oriented [75, 80], and 1 joint-denoising-deblurring [89] enhancement algorithms, to provide straightforward comparisons. Based on that, rigorous conclusions can be drawn on advantages

* Corresponding author.
For more resources, please visit our Project page.

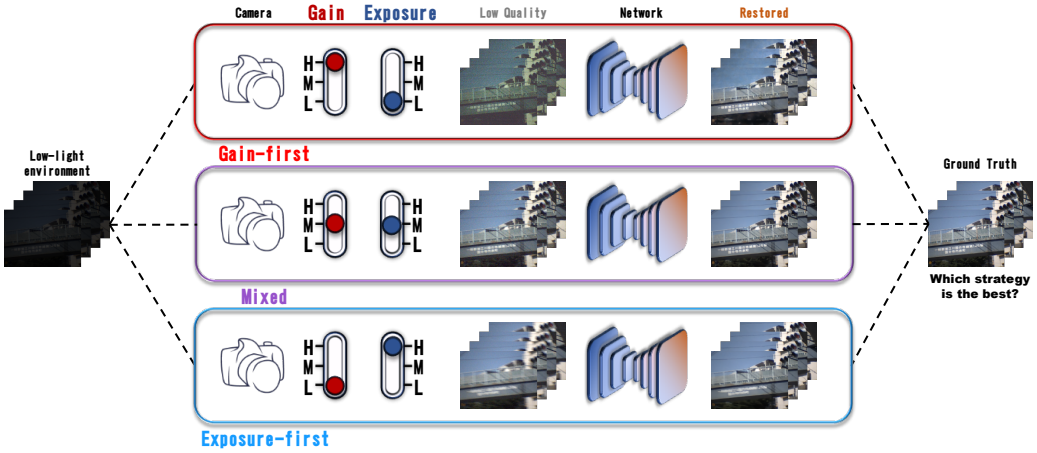

Figure 1. As raising camera gain and prolonging exposure time can both produce bright images from low-light environments, it is still unclear whether their unwanted artifacts, i.e. noise and motion blur, cause similar level of difficulties for enhancement networks. In this paper, we delve into **Gain-first**, **Exposure-first**, and **Mixed** strategies, comparing them under strictly controlled conditions, to investigate an optimal selection for low-light video photography and enhancement algorithm designs.

of Gain-first against Exposure-first under common light reduction ratios and supremacy of Mixed under a large ratio, offering heuristics for better low-light video photography and enhancement pipeline designs. We also provide a theoretical justification in our supplementary materials, incorporating the latest real-world noise [18] and motion blur modeling [4], proving the existence of a turning point where performances of different strategies switch places, pointing out a promising direction for future exploration under the complexity of this topic.

The main contributions of our work can be summarized as:

1. A newly designed coaxial optical system is proposed, capable of capturing image sequences with the same contents concurrently on multiple cameras, facilitating various training sample collection of dynamic scenes.
2. We build a large scale benchmark dataset containing over 54000 training images across multiple gain levels, exposure time, light intensity ratios, frame rates, making it possible to compare different camera setting strategies on identical contents for current and future algorithms.
3. Experiments are conducted, training state-of-the-art enhancement models on the collected dataset, demonstrating hindrances brought about by Gain-first and Exposure-first strategies respectively, drawing conclusions that are reinforced by a theoretical explanation, giving insights for better low-light videography and enhancement.

2. Related Works

In this section, we give a summary of development in low-light image/video enhancement from aspects of denoising methods, deblurring techniques, and related datasets.

Denoising. In conventional image processing, noise reduction can be approached by optimizing pixel intensity distributions with image priors, e.g. sparsity [14, 16, 17, 43], self-similarity [22], total variation [55, 66]. In deep learning era, high performances can be achieved by uncomplicated synthetic noisy image data [10, 24, 42, 44, 72, 83], real world dark/noisy datasets [9, 11, 13, 28, 48, 67, 79], and unsupervised learning methods [3, 23, 34–36, 49] that largely alleviates difficulties of gathering paired training samples. Recently, complicated noise models [5, 18, 73, 76, 85] are put forward, modeling camera imaging process from a physics perspective integrating photon and electronic properties, synthesizing authentic data for training.

Deblurring. For decades, blind deconvolution for uniform motion deblurring [2, 6, 15, 20, 26, 45, 58, 78, 81] and more general non-uniform motion deblurring [21, 31–33, 51, 53, 54, 56, 57, 71] have made substantial progresses. Apart from single-image-based deblurring, numerous video-oriented models are also proposed [33, 59, 61, 64, 69, 70, 82] to model temporal complementary information. End-to-end deep-learning-based models prevail over recent years, including: a multi-scale network termed as MSCNN [50] restoring sharp images in a coarse-to-fine manner; a symmetric U-shape network design along with parameter sharing strategy proposed by Gao *et al*. [19]; and MT-RNN [60] that introduces recurrent U-Net strategy to sharpen images in an iterative fashion, while MIMO-UNet [12] processes multi-scale images simultaneously with a single network, achieving high quantitative performances and computational efficiency.

Although several works of deblurring directly applied to low-light video enhancement can be found [68, 86, 88], state-of-the-art deblurring networks are more prevalent in generic image/video restoration models, e.g. VRT [39]/RVRT [40], BasicVSR++ [7, 8], Restormer [80], Uformer [75], etc. Most of them are capable of reaching outstanding performance in both denoising, deblurring, and other tasks like super-resolution as well.

Beyond focusing on only one side of the story, latest works also show possibilities of achieving two goals of denoising and deblurring at the same time. Establishing a paired dataset between dark blurry images and bright sharp GTs, Zhou *et al*. [89] trained an effective network, LEDNet, to jointly restore clean images from intertwined degradation. Though demonstrating impressive results, the paper, however, is in lack of a discussion on the necessity of this joint method, *i.e*., it doesn't answer a question about when jointly denoising-deblurring would become an optimal solution, given that other denoising and deblurring algorithms might be powerful enough to deal with ordinary cases.

Dataset. For denoising, synthetic datasets [14, 22, 66] usually adopts additive, white, Gaussian noise model [76]. Overcoming the obstacles of data collection in real-world, SID [9], SMID [11], SMOID [28] are assembled raw image/video datasets for low-light enhancement. DND [62] and SIDD [1], produced with scrutiny, have become popular benchmarking datasets for denoising algorithms. These datasets contain sRGB and RAW format, enabling networks to overcome enhancement ratios up to 300 (*e.g*., SID [9], ELD [76], both in RAW format). Yet, cases of ratio over 100 are still challenging for sRGB.

For deblurring, DVD [70], GoPro [50], and REDS [52] are commonly used synthetic datasets, yet unable to generalize well to the real-world scenarios and even cannot remove the artifacts introduced by insufficient lighting. RealBlur [65] and Beam-Splitter Deblurring Dataset (BSD) [87], both have exclusively designed hardware systems, collect aligned blurry and sharp images in real world, elevating algorithm performances with captured nuances. Meanwhile, Zhou *et al*. [89] utilizes existing Zero-DCE [23] model to simulate low-light degradation, generating LOL-Blur dataset for joint denoising and deblurring. All of deblurring datasets are in RGB format. While some of them considers camera response function (CRF) or image signal processing (ISP) in their pipeline, all of them still land their network inside RGB domain. Typically, deblurring algorithms are often concerned with situations that have an exposure ratio between sharp and blurry frames close to or below 10, *e.g*., RealBlur [65], BSD [87].

Despite the prosperity of enhancement algorithms and datasets, which way, with the help of deep learning frameworks, is the most suitable for low-light videos remains unanswered. Prior works that compares denoising and deblurring [25, 46, 47] fail to verify their conclusions on real-world data, nor did they consider data-driven enhancements or joint methods. In this work, we gather a large dataset containing precisely aligned Gain-first, Exposure-first, and Mixed images for exactly identical scenes, which enables us to compare three strategies directly. Extensive number of experiments using latest models are carried out, accompanied by theoretical analysis of image formation, shedding light on better dark video capturing and enhancement.

3. Experiment Settings

In this section, we first introduce our optical system for data collection and parameter setting in each strategy, followed by detailed description of our dataset structure. We present experiment settings of networks in the last subsection.

3.1. Optical System

Illustrated in Figure 2 is our camera system, composed of 3 beam splitters to separate incident light beams identically in two levels, transmitting exact same photon signals towards four cameras independently. Taking advantage of ND filter's ability to attenuate light intensity without color shifting, camera 1, 2, and 3 in the picture can record dark images that are aligned with bright images recorded by camera 4. All optical components are carefully aligned by hand while inspecting the discrepancy images among four cameras of a reference pattern. Although tremendous efforts have been taken, geometric misalignment within a few pixels is impossible to be eliminated by hand, due to the complexity of our system. Therefore, we further estimate homographies and use them to correct images. All components are firmly fixed on a metal breadboard with plastic cushion to avoid shift due to vibration during capture. Signal generator that governs timing of shooting further enforces simultaneity among all cameras, guaranteeing spatial and temporal alignments.

The cameras we use are FLIR GS3-U3-23S6C Camera, equipped with a global shutter, adjustable analog gain from 0 to 30 dB, and exposure time from 30μs to 3.2s. It is commonly used for industrial utilities, containing a Sony IMX174 image sensor that can also be found in commercial cameras, ruling out irrelevant factors like rolling shutter effect, while maintaining genericness for our experiments. We choose 3 types of ND filters with transmittance of 1/10, 1/30, and 1/100, in order to emulate multiple levels of darkness, covering meaningful ratios for both denoising and deblurring tasks as mentioned in section 2, namely: 1/10 to cover common deblurring cases; 1/100 to push the limit of denoising methods; and 1/30 as a middle point. For a specific instance, if all mounted ND filters are of 1/10 transmittance ratio, and all 4 cameras using the same gain and exposure time, then camera 4 should produce a raw image that has pixel values roughly[1] 10 times of that of images shoot at the same time by camera 1, 2 and 3. We refer to the 3 darkness conditions as *ratio10*, *ratio30*, and *ratio100*, for brevity. The relationship between such ratio and camera's analog gain is: $AnalogGain = 20 \times log_{10}(ratio)$.

[1]Considering existence of noise, it is not precisely, but should be fairly close to, 10 times.

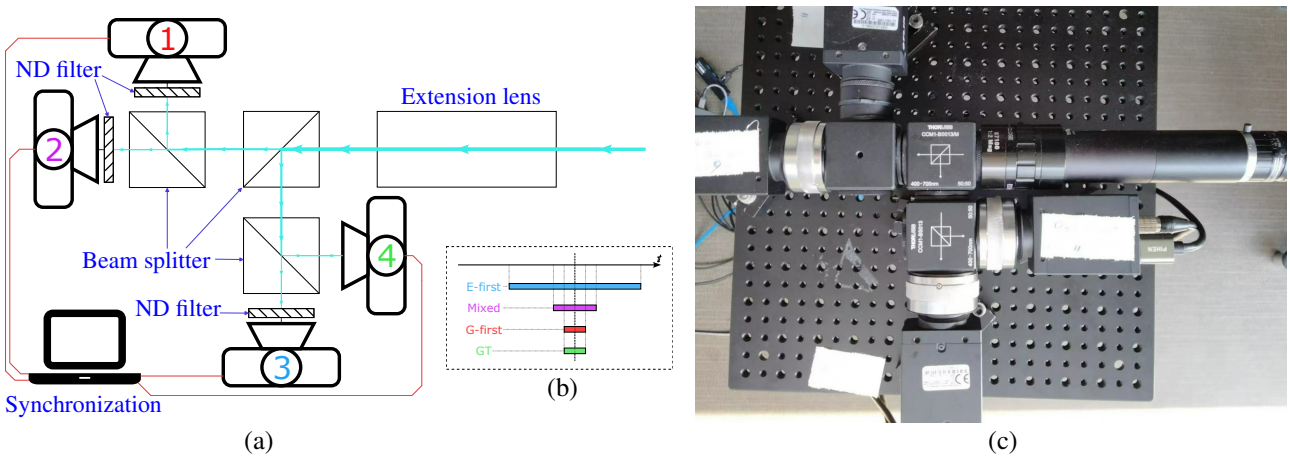

Figure 2. (a) Optical diagram of our camera system. (b) Illustration of center-aligned exposure timings for different lengths of exposure. (c) A top-view photo of our system. Exploiting beam splitter's equally dividing light property and ND filter's unbiased light intensity reduction, we are able to record 3 dark versions on camera 1, 2, 3 and 1 bright version on camera 4 of the exact same scenes.

Table 1. Camera parameter settings for ratio10, ratio30, and ratio100. ET, AG, FR means exposure time, analog gain, and frame rate, respectively. For each capture, parameters are fixed throughout the sequence.

Ratio	Param.	G-first	Mixed	E-first	GT
10	ET (ms)	2	6	20	2
	AG (dB)	20	10	0	0
	FR (fps)	30	30	30	30
30	ET (ms)	2	6	60	2
	AG (dB)	29	20	0	0
	FR (fps)	15	15	15	15
100	ET (ms)	2	6	20	2
	AG (dB)	30	30	20	0
	FR (fps)	30	30	30	30

3.2. Parameter Settings

Table 1 shows key parameters used in the process of data collection. In the table, G-first stands for Gain-first strategy, shot on camera 1, compensating attenuated brightness primarily by raised analog gain value. E-first stands for Exposure-first strategy, shot on camera 3, increasing its luminance primarily by longer exposure time. Mixed strategy is shot on camera 2, benefiting from both moderate gain values and slightly longer exposure time. Ground truth bright image sequences are shot on camera 4, with constant 0dB gain and 2ms exposure as a standard. All 4 cameras, when set with different length of exposure, are temporally center-aligned for exposure time as shown in Figure 2 (b), *i.e.*, as Exposure-first has the longest time, all other cameras have their respective starting and ending point of exposure being symmetric with respect to the exact center point of Exposure-first's exposure duration on time axis.

In ratio10 situation, all brightness gap can be covered by analog gain (20dB = 10-fold pixel intensity amplification) or exposure time (20ms vs. 2ms), producing perfect samples for controlled experiments that only vary single factor at a time. In ratio30's case, this can be achieved as well (29dB = 28.18-fold amplification; 60ms vs. 2ms), since we lowered frame rate of all cameras at the same time, making sure the longest required exposure time, 60ms, is still possible at a 15 fps frame rate. However, longer exposure time introduces stronger biases against Exposure-first strategy, as we saw in preliminary tests and can be seen from the results that will be presented in this paper later. If this pursuit of purity continues to isolate denoising and deblurring factors completely for two strategies, we will arrive at an extremely low frame rate for ratio100, which undoubtedly diminishes the necessity of further experimentation. Thus, to proceed in a more meaningful way, we adopt 30fps for ratio100, setting analog gain to the maximum possible value, 30dB (=31.62-fold amplification), for Gain-first, and 20ms exposure for Exposure-first. The remaining ratio gap for Gain-first are compensated by a digital gain, while settings for Exposure-first under ratio100 also come with a 20dB (=10-fold amplification) gain, so that starting points of all compared strategies are in a close range with each other, leading to fairer and more distinct comparisons.

3.3. Dataset Structure

A subset of our collected data can be seen in Figure 1, 3, and more in supplementary materials. For each ratio, we filmed 60 scenes with immense amount of objects, like buildings, vehicles and pedestrians, covering as much complicated motion as possible. Camera movements are diversified among all scenes, ranging from being completely still to fast panning or tilting movements, creating graded levels of difficulties for all strategies, reducing possible inclination towards any specific strategy. Each scene contains 75

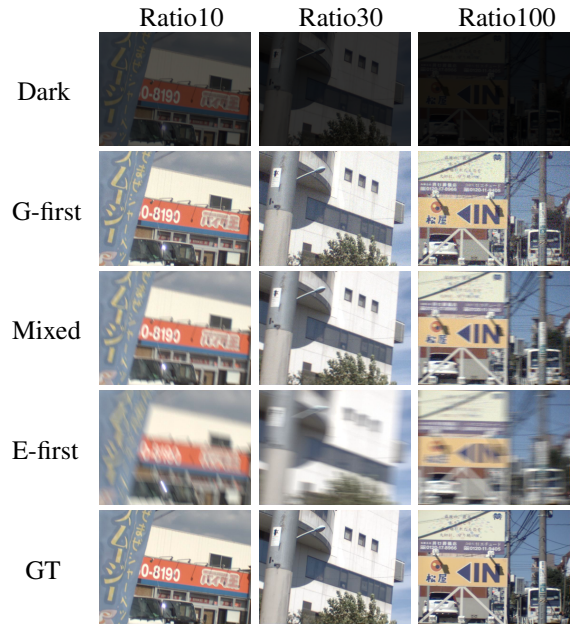

Figure 3. Dataset samples. Dark images are obtained by dividing GT images by amplification ratios, serving as a reference for what a camera would see without any help from gain or exposure. It is recommended to zoom in to see different levels of noise and blurry effects among pictures.

frames in total, except for ratio10 where each scene has 150 frames. Doubling the number of images allows us to conduct an extra experiment to test methods under ratio10 with 15fps, simply by extracting 75 frames out of 150 with a two-frame temporal stride. Without hardware setting modifications, this extra experiment serves as an examination for validity on our choices of divergent frame rates. Original ratio10 30fps experiment uses first 75 consecutive frames from each scene. Basically, our experiments cover 4 categories: ratio10 15fps, ratio10 30fps, ratio30 15fps, and ration100 30fps, each containing 60 scenes with 75 distinct frames per scene for 3 strategies, bringing the total number of unique images in our dataset to $4 \times 60 \times 75 \times 3 = 54000$. Images are stored in RGB and RAW format, with spatial dimension of 480×640. In consideration of the fact that existing deblurring methods focus solely on RGB formats, we prioritized the usage of RGB images in following experiments, while RAW counterparts will be released as well, promoting deeper exploration to the topic.

The 60 scenes under each category are partitioned into 3 parts: 51 for training, 3 for validation, and 6 for test. On 3 selected scenes for validation set, we manually examined and adjusted the selections to include 2 scenes with camera movements and 1 scene with none. Likewise, test set is also balanced in the sense of camera motion, *i.e.*, 3 moving and 3 still, for the reason that Exposure-first strategy is more motion-sensitive than others, demanding a closer inspection on performances for still and moving scenes respectively rather than merely presenting results on test set as a whole.

3.4. Networks

We selected 5 video/image enhancement algorithms based on their state-of-the-art performances on denoising and deblurring tasks at the time being. 2 video-oriented works: RVRT [40] and BasicVSR++ [7, 8]. 2 image-oriented works: Restormer [80] and Uformer [75]. 1 joint enhancement model: LEDNet [89]. All models are trained on Gain-first, Exposure-first, and Mixed images paired with clean GTs for all 4 categories of intensity ratios and frame rates. Hence the total number of experiments is $4 \times 3 \times 5 = 60$, taking more than 150 equivalent days on 4 RTX-3090 GPUs.

We maintained all of these models' default structures, and make sure that under one category (*e.g.*, Ratio10 15fps) all hyperparameters are exactly the same for 3 strategies, guaranteeing impartial competitions. Specific modifications for each model are detailed in our supplementary materials.

4. Results

4.1. Quantitative Evaluations

All trained models are evaluated by PSNR↑, SSIM↑ [74], LPIPS↓ [84]. Comprehensive results can be found in Ta-

Table 2. Overall quantitative results, averaged over entire test set for each experiment respectively. Colors denote video-oriented, image-oriented, joint methods.

PSNR/SSIM/LPIPS	Ratio10 15fps G-first	Ratio10 15fps Mixed	Ratio10 15fps E-first
RVRT	37.1/0.958/0.0127	36.1/0.950/0.0147	30.9/0.891/0.0426
BasicVSR++	36.9/0.960/0.0130	35.6/0.949/0.0158	32.3/0.915/0.0327
Restormer	37.7/0.962/0.0112	36.5/0.958/0.0132	34.3/0.940/0.0233
Uformer	34.0/0.956/0.0188	33.6/0.950/0.0192	32.7/0.927/0.0268
LEDNet	36.3/0.959/0.0150	34.2/0.948/0.0197	30.9/0.910/0.0371
PSNR/SSIM/LPIPS	Ratio10 30fps G-first	Ratio10 30fps Mixed	Ratio10 30fps E-first
RVRT	36.9/0.959/0.0133	36.0/0.950/0.0152	33.3/0.925/0.0278
BasicVSR++	36.7/0.958/0.0150	35.9/0.954/0.0149	31.5/0.924/0.0301
Restormer	37.8/0.963/0.0113	36.5/0.958/0.0138	33.9/0.941/0.0240
Uformer	36.4/0.961/0.0138	35.1/0.952/0.0161	32.4/0.923/0.0284
LEDNet	35.9/0.959/0.0155	34.5/0.950/0.0197	30.3/0.904/0.0425
PSNR/SSIM/LPIPS	Ratio30 15fps G-first	Ratio30 15fps Mixed	Ratio30 15fps E-first
RVRT	32.8/0.938/0.0235	34.3/0.949/0.0206	24.5/0.788/0.1271
BasicVSR++	34.0/0.952/0.0204	34.5/0.954/0.0193	28.0/0.868/0.0669
Restormer	34.0/0.946/0.0210	34.7/0.952/0.0200	26.9/0.853/0.0820
Uformer	33.1/0.944/0.0219	32.7/0.949/0.0214	26.4/0.843/0.0859
LEDNet	33.1/0.949/0.0243	32.5/0.945/0.0292	25.0/0.809/0.1040
PSNR/SSIM/LPIPS	Ratio100 30fps G-first[2]	Ratio100 30fps Mixed	Ratio100 30fps E-first
RVRT	32.1/0.925/0.0271	33.5/0.942/0.0202	31.7/0.904/0.0340
BasicVSR++	32.6/0.931/0.0250	32.8/0.939/0.0229	28.8/0.869/0.0658
Restormer	31.6/0.913/0.0315	32.5/0.929/0.0244	32.5/0.925/0.0275
Uformer	31.5/0.913/0.0299	32.5/0.927/0.0254	32.0/0.920/0.0309
LEDNet	30.7/0.903/0.0361	31.2/0.922/0.0303	30.4/0.903/0.0421

[2]Tainted background color here is a reminder that, at Ratio100, G-first is with digital gain; E-first with analog gain, neither of which is purely enhanced by a single factor.

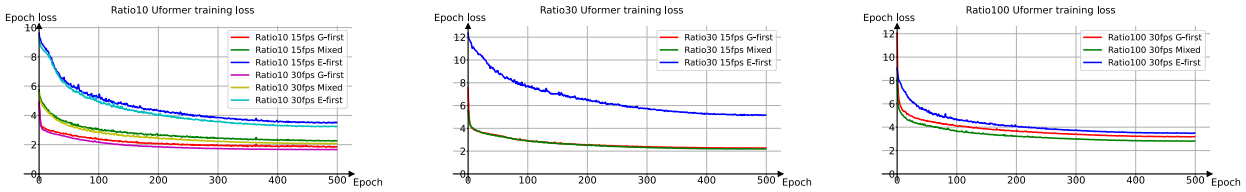

Figure 4. Epoch losses recorded during Uformer training, delivering an intuitive illustration for initial degree of difficulties and convergence speeds for different settings.

Table 3. Quantitative results on test videos with no camera movement. Colors denote video-oriented, image-oriented, joint methods.

PSNR/SSIM/LPIPS	Ratio10 15fps G-first	Ratio10 15fps Mixed	Ratio10 15fps E-first
RVRT	36.3/0.955/0.0173	37.1/0.956/0.0156	34.4/0.934/0.0267
BasicVSR++	36.5/0.960/0.0163	37.0/0.955/0.0165	34.9/0.941/0.0237
Restormer	37.1/0.960/0.0145	37.5/0.961/0.0150	36.7/0.956/0.0190
Uformer	35.2/0.956/0.0188	35.3/0.959/0.0186	35.4/0.954/0.0213
LEDNet	35.6/0.956/0.0194	35.3/0.956/0.0209	34.0/0.945/0.0276
PSNR/SSIM/LPIPS	Ratio10 30fps G-first	Ratio10 30fps Mixed	Ratio10 30fps E-first
RVRT	36.2/0.956/0.0175	37.1/0.954/0.0170	35.2/0.942/0.0226
BasicVSR++	35.6/0.950/0.0211	37.0/0.955/0.0171	34.2/0.934/0.0276
Restormer	37.4/0.961/0.0144	37.5/0.960/0.0162	35.7/0.951/0.0229
Uformer	35.9/0.957/0.0173	36.7/0.960/0.0175	34.8/0.948/0.0248
LEDNet	35.2/0.955/0.0202	35.6/0.955/0.0212	33.0/0.934/0.0368
PSNR/SSIM/LPIPS	Ratio30 15fps G-first	Ratio30 15fps Mixed	Ratio30 15fps E-first
RVRT	31.1/0.928/0.0274	34.6/0.956/0.0166	28.6/0.893/0.0575
BasicVSR++	32.8/0.947/0.0229	33.9/0.956/0.0183	30.3/0.917/0.0394
Restormer	32.1/0.935/0.0246	34.6/0.956/0.0178	29.5/0.909/0.0501
Uformer	31.6/0.933/0.0256	33.1/0.954/0.0189	29.7/0.916/0.0458
LEDNet	31.8/0.942/0.0268	32.8/0.953/0.0231	28.1/0.903/0.0580
PSNR/SSIM/LPIPS	Ratio100 30fps G-first	Ratio100 30fps Mixed	Ratio100 30fps E-first
RVRT	31.9/0.931/0.0288	33.3/0.950/0.0207	33.7/0.944/0.0221
BasicVSR++	32.2/0.940/0.0266	32.9/0.950/0.0238	32.4/0.947/0.0302
Restormer	31.1/0.915/0.0351	32.5/0.940/0.0245	33.8/0.955/0.0202
Uformer	31.3/0.917/0.0323	32.4/0.936/0.0254	33.5/0.952/0.0223
LEDNet	30.2/0.905/0.0390	31.4/0.938/0.0289	32.3/0.952/0.0292

Table 4. Quantitative results on test videos with camera movements. Colors denote video-oriented, image-oriented, joint methods.

PSNR/SSIM/LPIPS	Ratio10 15fps G-first	Ratio10 15fps Mixed	Ratio10 15fps E-first
RVRT	37.9/0.961/0.0082	35.1/0.944/0.0138	27.3/0.848/0.0586
BasicVSR++	37.3/0.960/0.0097	34.2/0.942/0.0151	29.7/0.888/0.0418
Restormer	38.2/0.965/0.0080	35.5/0.955/0.0115	31.9/0.924/0.0276
Uformer	32.8/0.956/0.0189	31.9/0.940/0.0197	30.1/0.900/0.0324
LEDNet	37.0/0.962/0.0107	33.1/0.941/0.0186	27.9/0.875/0.0467
PSNR/SSIM/LPIPS	Ratio10 30fps G-first	Ratio10 30fps Mixed	Ratio10 30fps E-first
RVRT	37.5/0.963/0.0091	34.9/0.946/0.0134	31.3/0.908/0.0329
BasicVSR++	37.9/0.965/0.0088	34.8/0.953/0.0127	28.8/0.914/0.0325
Restormer	38.3/0.966/0.0081	35.5/0.957/0.0113	32.1/0.931/0.0251
Uformer	36.9/0.964/0.0104	33.5/0.945/0.0146	30.0/0.897/0.0320
LEDNet	36.7/0.964/0.0107	33.5/0.944/0.0181	27.6/0.874/0.0481
PSNR/SSIM/LPIPS	Ratio30 15fps G-first	Ratio30 15fps Mixed	Ratio30 15fps E-first
RVRT	34.5/0.948/0.0197	34.0/0.943/0.0246	20.4/0.683/0.1967
BasicVSR++	35.3/0.956/0.0180	35.0/0.952/0.0203	25.6/0.820/0.0944
Restormer	35.9/0.958/0.0174	34.7/0.949/0.0221	24.4/0.796/0.1139
Uformer	34.7/0.956/0.0181	32.3/0.944/0.0239	23.1/0.769/0.1260
LEDNet	34.4/0.956/0.0218	32.2/0.937/0.0354	21.9/0.715/0.1501
PSNR/SSIM/LPIPS	Ratio100 30fps G-first	Ratio100 30fps Mixed	Ratio100 30fps E-first
RVRT	32.4/0.920/0.0255	33.6/0.933/0.0196	29.7/0.863/0.0459
BasicVSR++	32.9/0.923/0.0234	32.7/0.928/0.0219	25.2/0.790/0.1015
Restormer	32.1/0.910/0.0279	32.6/0.919/0.0243	31.2/0.895/0.0348
Uformer	31.8/0.910/0.0276	32.5/0.918/0.0253	30.6/0.888/0.0396
LEDNet	31.2/0.902/0.0331	31.1/0.907/0.0318	28.6/0.854/0.0549

ble 2, showing overall tendencies among all settings. We also included measurements obtained by testing on subset of static shot and subset of moving shot respectively in Table 3 and 4, to further portray characteristics of test results. A visualization in the form of heatmap for Table 2, 3, 4 are presented in our supplementary document, which can serve as a visual aid for descriptions and conclusions in the following paragraphs.

Superiority of Gain-first. It can be easily noticed that Gain-first strategy has predominant advantages against Exposure-first in ratio10 and ratio30. Gain-first overpowers Mixed strategy in all ratio10 settings, showing non-inferiority in ratio30 against Mixed as well. Additionally, Gain-first results demonstrate robustness in dynamic scenes, while others crippled by distortions, especially in the case of the longest exposure for ratio30. This proves the usefulness of camera gain over exposure, as when increasing gain and prolonging exposure are both available under these ratios, noisy videos have more potential to be restored by deep learning algorithms than blurry ones.

Loss tendencies. With the help of gathered loss information from Uformer training processes in Figure 4, we are able to visualize traits of each experiment more clearly. At the beginning of each training, Exposure-first images add the most difficulties to tasks. Even under a balanced circumstance like ratio100 that has proximate starting points for all three, Exposure-first training exhibits relatively lower convergence. Another remarkable point from the loss plots is that, although quantitative measurements don't show a constant performance gain when comparing ratio10 15fps to 30fps, higher frame rate indeed lowers training losses for all strategies, substantiating our choice of preserving higher

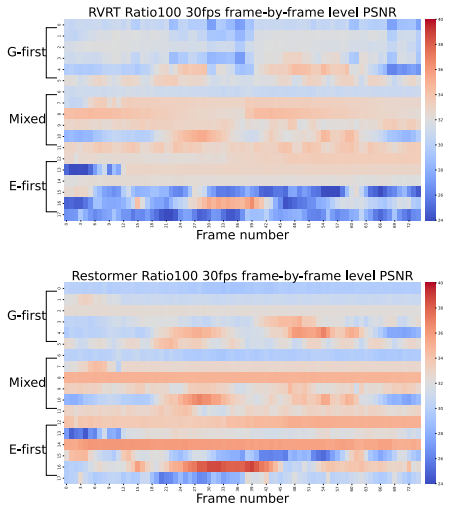

Figure 5. Heatmap depicted PSNR of each frame in RVRT [40] and Restormer [80] results. Each row represents PSNR values calculated on each frame of a restored video. Each strategy (G-first, Mixed, E-first), contains 6 videos that correspond to the same GTs, with upper 3 rows being static shot videos, lower 3 being moving shot ones.

frame rates at ratio100 to intensify competition, rather than leaving it at a similar landslide.

Fine-grained measurements. As is noticeable in previous tables, all strategies achieves comparable results under ratio100, since they are all implemented in a mixed fashion at this condition. Nonetheless, static-shot results and moving-shot results at this ratio show contrary relations between Gain-first and Exposure-first, urging a closer examination. Therefore, we draw heatmaps in Figure 5 to display PSNR distributions of 2 networks at ratio100 on a frame-by-frame level. Gain-first and Mixed have gradual color gradients, while Exposure-first has oscillating high and low values. Following this observation, we further inspected fluctuation levels of each strategy's numerical performances. In Table 5, there exist palpable differences between Gain-first and Exposure-first, as the latter can fluctuate more than two times of the former, raising the question on Exposure-first's robustness.

4.2. Qualitative Results

We pieced together sequential sample frames from restored results of 3 methods under 3 ratio conditions. Figure 6 and 7 are results from BasicVSR++ under ratio10 15fps, and LEDNet under ratio30 15fps, both of which are moving-shot videos that focus on showcasing Exposure-first's inability to recover sharp edges under large scene movements. Figure 8 juxtaposes result images from RVRT under ratio100 30fps. With this scene being a static-shot video, Exposure-first can recover frames to a favorable level, while

Table 5. Standard deviation of PSNR, SSIM, LPIPS calculated among restored frames. Colors denote video-oriented, image-oriented, joint methods. Reaffirming that tainted background color for strategy names here is a reminder that, at Ratio100, none of them is purely enhanced by a single factor.

Std of PSNR/SSIM/LPIPS	Among all Ratio100 30fps test scenes		
	G-first	Mixed	E-first
RVRT	1.74/0.031/0.014	1.96/0.024/0.010	3.03/0.053/0.021
BasicVSR++	1.32/0.026/0.010	1.81/0.024/0.010	4.68/0.098/0.050
Restormer	1.50/0.032/0.012	1.84/0.028/0.011	2.91/0.044/0.016
Uformer	1.53/0.029/0.012	1.83/0.027/0.012	2.67/0.045/0.017
LEDNet	1.32/0.038/0.012	1.14/0.031/0.011	2.99/0.066/0.024
Std of PSNR/SSIM/LPIPS	Among static shot test scenes		
	G-first	Mixed	E-first
RVRT	1.43/0.016/0.012	2.21/0.011/0.010	1.63/0.017/0.011
BasicVSR++	0.74/0.008/0.010	1.93/0.011/0.010	2.16/0.024/0.018
Restormer	1.13/0.009/0.010	2.08/0.010/0.011	2.24/0.016/0.012
Uformer	1.18/0.007/0.012	2.09/0.011/0.012	1.72/0.015/0.010
LEDNet	0.47/0.023/0.010	0.75/0.010/0.009	1.82/0.020/0.014
Std of PSNR/SSIM/LPIPS	Among moving shot test scenes		
	G-first	Mixed	E-first
RVRT	1.99/0.039/0.015	1.66/0.030/0.010	2.80/0.045/0.022
BasicVSR++	1.64/0.034/0.010	1.69/0.028/0.010	3.68/0.079/0.046
Restormer	1.64/0.044/0.012	1.58/0.035/0.011	2.96/0.042/0.016
Uformer	1.78/0.041/0.011	1.53/0.034/0.012	2.63/0.041/0.017
LEDNet	1.64/0.049/0.013	1.41/0.037/0.012	2.80/0.061/0.024

Gain-first cannot properly remove all noises, leaving visible noise patterns on the zoomed-in green box. More video results and visualizations are included in our supplementary materials, corroborating conclusions drawn from quantitative measurements.

5. Discussion

Based on above experiment results, it can be inferred that Gain-first strategy is a more preferable choice for video enhancement than Mixed and Exposure-first under acceptable light attenuation levels. This gives us a new perspective on deblurring long exposure videos shot in dark, where larger gain might preserve more information that is easier to restore if increasing gain level is accessible. When the surrounding light reduction ratio approaches maximum value of camera gain amplification, Gain-first reduces to be comparable with Mixed, but still better than Exposure-first. When going beyond camera gain level, even with state-of-the-art denoising models, noise pattern persists, leaving chances for Mixed and Exposure-first. This answers the question of when jointly denoising-deblurring methods are called upon, a vital discussion on its necessity that is absent in [89]. Extreme dark environments require mixed scheme to concurrently suppress noise pattern and motion blur distortion.

Theoretical interpretation. To further support our conclusions, a theoretical reasoning based on noise [18] and motion blur modeling [4] coupled with synthetic data test is detailed in our supplementary material, demonstrating clear

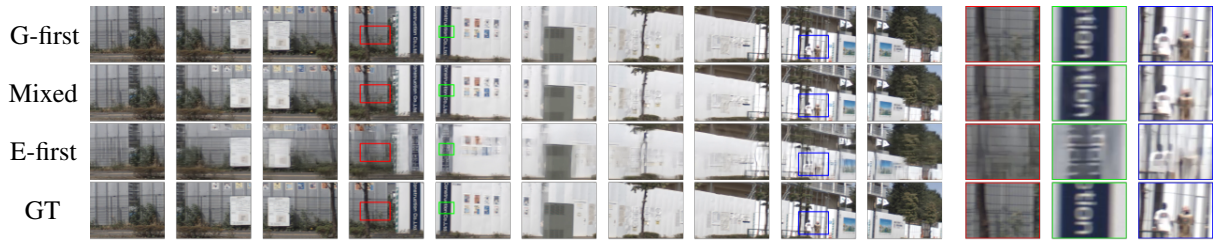

Figure 6. Consecutive restored frames from BasicVSR++ [7] of ratio10 15fps, with enlarged patches on the right.

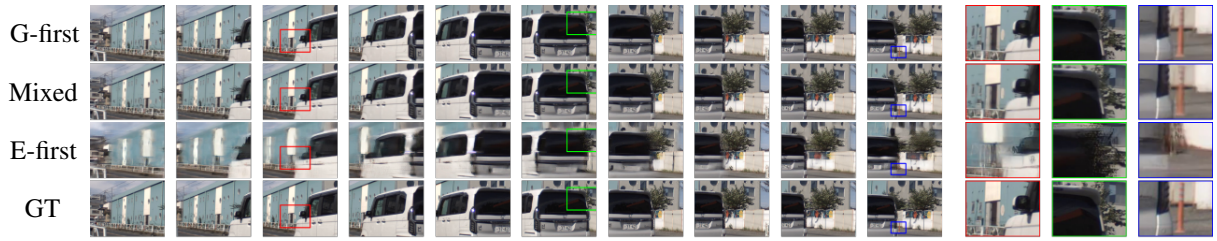

Figure 7. Consecutive restored frames from LEDNet [89] of ratio30 15fps, with enlarged patches on the right.

Figure 8. Consecutive restored frames from RVRT [40] of ratio100 30fps, with enlarged patches on the right.

existence of a critical turning point between Gain-first and Exposure-first, indicating possible optimal strategies with respect to ratio and motion speed.

Alignment with prior works. Though our conclusion is partly in alignment with former comparisons between denoising and deblurring [25, 47], that denoising/Gain-first is predominantly better than deblurring/Exposure-first, our work is unique in 3 aspects. Firstly, all previous analysis are done theoretically only, without real world data verification, while our benchmark dataset is collected with carefully designed optical system to ensure rigorous conclusions. Secondly, prior works only involve traditional algorithms, not concerned with deep learning methods, while we carry out our experiments with latest networks, clarifying the relationship between different strategies under the uncertainty introduced by data-driven algorithms. Thirdly, none of the prior comparisons considers Mixed strategy that was proposed only in recent years, showing an inspiring result of Mixed strategy from our experiments for future low-light video capture and enhancement algorithm design.

Limitations and future work. Since we use RGB formatted images, paths for low-light video enhancement through raw sensor data is left out of the picture for now. But such experiments can be carried out easily with the release of our dataset containing all RGB and RAW images. We will continue to explore authentic noise modeling, searching for a theoretical calculation for aforementioned critical points rather than finding in a discrete fashion. Such equilibrium points may offer connections between calibrated noise level and blurriness, uncovering new possible directions for both tasks and related areas, *e.g.*, multi-exposure HDR capture.

6. Conclusion

In this paper, we collected a large benchmark dataset for investigating optimal method for low-light video photography and enhancement among Gain-first, Exposure-first, and Mixed strategies. The dataset is built using newly designed optical system capable of recording aligned image sequences on multiple different cameras simultaneously. Large amount of experiments utilizing latest enhancement models are carried out, showing abundant evidence for Gain-first's advantage under relatively small brightness reduction ratio. In darker environments, Mixed strategy is a better choice to exceed limits of camera gain value. A theoretical explanation is also given in the supplementary materials, further supporting conclusions from empirical experiments.

Acknowledgment

This research was supported in part by JSPS KAKENHI Grant Numbers 22H00529, 20H05951, JST-Mirai Program JPMJMI23G1.

References

[1] A. Abdelhamed, S. Lin, and M. S. Brown. A high-quality denoising dataset for smartphone cameras. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1692–1700, Los Alamitos, CA, USA, 2018. IEEE Computer Society. 3

[2] Yuval Bahat, Netalee Efrat, and Michal Irani. Non-uniform blind deblurring by reblurring. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3306–3314, 2017. 2

[3] Coleman Broaddus, Alexander Krull, Martin Weigert, Uwe Schmidt, and Eugene Myers. Removing structured noise with self-supervised blind-spot networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020)*. IEEE, 2020. 2

[4] Mingdeng Cao, Zhihang Zhong, Yanbo Fan, Jiahao Wang, Yong Zhang, Jue Wang, Yujiu Yang, and Yinqiang Zheng. Towards real-world video deblurring by exploring blur formation process, 2022. 2, 7

[5] Yue Cao, Ming Liu, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Physics-guided iso-dependent sensor noise modeling for extreme low-light photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5744–5753, 2023. 2

[6] Ayan Chakrabarti. A neural approach to blind motion deblurring. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, pages 221–235. Springer, 2016. 2

[7] Kelvin C.K. Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. BasicVSR++: Improving video super-resolution with enhanced propagation and alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 1, 2, 5, 8

[8] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. On the generalization of BasicVSR++ to video deblurring and denoising. *arXiv preprint arXiv:2204.05308*, 2022. 2, 5

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3

[10] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 2

[11] Chen Chen, Qifeng Chen, Minh N. Do, and Vladlen Koltun. Seeing motion in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2, 3

[12] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2

[13] Michele Claus and Jan van Gemert. Videnn: Deep blind video denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 1, 2

[14] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 2, 3

[15] Jiangxin Dong, Jinshan Pan, Zhixun Su, and Ming-Hsuan Yang. Blind image deblurring with outlier handling. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2497–2505, 2017. 2

[16] Weisheng Dong, Xin Li, Lei Zhang, and Guangming Shi. Sparsity-based image denoising via dictionary learning and structural clustering. In *CVPR 2011*, pages 457–464, 2011. 2

[17] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 2

[18] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 1436–1444, 2022. 2, 7

[19] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3843–3851, 2019. 2

[20] Dong Gong, Mingkui Tan, Yanning Zhang, Anton Van Den Hengel, and Qinfeng Shi. Blind image deconvolution by automatic gradient activation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1827–1836, 2016. 2

[21] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3806–3815, 2017. 2

[22] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014. 2, 3

[23] Chunle Guo Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1780–1789, 2020. 2, 3

[24] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[25] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. In *2010 IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition, pages 553–560, 2010. 1, 3, 8

[26] Michael Hirsch, Christian J. Schuler, Stefan Harmeling, and Bernhard Schölkopf. Fast removal of non-uniform camera shake. In *2011 International Conference on Computer Vision*, pages 463–470, 2011. 2

[27] Muwei Jian, Xiangyu Liu, Hanjiang Luo, Xiangwei Lu, Hui Yu, and Junyu Dong. Underwater image processing and analysis: a review. *Signal Processing: Image Communication*, 91, 2021. 1

[28] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2, 3

[29] Efklidis Katsaros, Piotr K. Ostrowski, Daniel Wesierski, and Anna Jezierska. Concurrent video denoising and deblurring for dynamic scenes. *IEEE Access*, 9:157437–157446, 2021. 1

[30] Minjae Kim, Dubok Park, David K. Han, and Hanseok Ko. A novel approach for denoising and enhancement of extremely low-light video. *IEEE Transactions on Consumer Electronics*, 61(1):72–80, 2015. 1

[31] Tae Hyun Kim and Kyoung Mu Lee. Segmentation-free dynamic scene deblurring. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2766–2773, 2014. 2

[32] Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee. Dynamic scene deblurring. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013.

[33] Tae Hyun Kim, Kyoung Mu Lee, Bernhard Schölkopf, and Michael Hirsch. Online video deblurring via dynamic temporal blending network. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4058–4067, 2017. 2

[34] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019. 2

[35] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2965–2974. PMLR, 2018.

[36] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2022. 2

[37] Guofa Li, Yifan Yang, Xingda Qu, Dongpu Cao, and Keqiang Li. A deep learning based image enhancement approach for autonomous driving at night. *Knowledge-Based Systems*, 213:106617, 2021. 1

[38] Yawei Li, Hong Zhang, Yujie Wu, and Ding Yuan. Dynamic scene video deblurring using robust incremental weighted fourier aggregation. *IEEE Signal Processing Letters*, 28: 1565–1569, 2021. 1

[39] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *arXiv preprint arXiv:2201.12288*, 2022. 2

[40] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc Van Gool. Recurrent video restoration transformer with guided deformable attention. *arXiv preprint arXiv:2206.02146*, 2022. 1, 2, 5, 7, 8

[41] Zicheng Liu, Cha Zhang, and Zhengyou Zhang. Learning-based perceptual image quality improvement for video conferencing. In *2007 IEEE International Conference on Multimedia and Expo*, pages 1035–1038, 2007. 1

[42] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[43] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008. 2

[44] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2016. 2

[45] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In *Computer Vision – ECCV 2014*, pages 783–798, Cham, 2014. Springer International Publishing. 2

[46] Kaushik Mitra, Oliver Cossairt, and Ashok Veeraraghavan. To denoise or deblur: parameter optimization for imaging systems. In *Digital Photography X*, page 90230G. International Society for Optics and Photonics, SPIE, 2014. 1, 3

[47] Kaushik Mitra, Oliver S. Cossairt, and Ashok Veeraraghavan. A framework for analysis of computational imaging systems: Role of signal prior, sensor noise and multiplexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10):1909–1921, 2014. 1, 3, 8

[48] Kristina Monakhova, Stephan R. Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: Video denoising in starlight. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16241–16251, 2022. 1, 2

[49] Weiwen Mu, Huixiang Liu, Wenbai Chen, and Yiqun Wang. A more effective zero-dce variant: Zero-dce tiny. *Electronics*, 11(17), 2022. 2

[50] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 3

[51] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[52] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *CVPR Workshops*, 2019. 3

[53] T M Nimisha, Akash Kumar Singh, and A N Rajagopalan. Blur-invariant deep learning for blind-deblurring. In *2017*

IEEE International Conference on Computer Vision (ICCV), pages 4762–4770, 2017. 2

[54] Mehdi Noroozi, Paramanand Chandramouli, and Paolo Favaro. Motion deblurring in the wild. In *Pattern Recognition*, pages 65–77, Cham, 2017. Springer International Publishing. 2

[55] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005. 2

[56] Jinshan Pan, Zhe Hu, Zhixun Su, Hsin-Ying Lee, and Ming-Hsuan Yang. Soft-segmentation guided object motion deblurring. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 459–468, 2016. 2

[57] Jinshan Pan, Jiangxin Dong, Yu-Wing Tai, Zhixun Su, and Ming-Hsuan Yang. Learning discriminative data fitting functions for blind image deblurring. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1077–1085, 2017. 2

[58] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(10):2315–2328, 2018. 2

[59] Liyuan Pan, Yuchao Dai, Miaomiao Liu, and Fatih Porikli. Simultaneous stereo video deblurring and scene flow estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[60] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In *Computer Vision – ECCV 2020*, pages 327–343, Cham, 2020. Springer International Publishing. 2

[61] Haesol Park and Kyoung Mu Lee. Joint estimation of camera pose, depth, deblurring, and super-resolution from a blurred image sequence. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[62] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3

[63] Yuefei Qu, Ji Zhou, Song Qiu, Wei Xu, and Qingli Li. Recursive video denoising algorithm for low light surveillance applications. In *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–7, 2021. 1

[64] Wenqi Ren, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Video deblurring via semantic segmentation and pixel-wise non-linear kernel. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[65] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Computer Vision – ECCV 2020*, pages 184–201, Cham, 2020. Springer International Publishing. 3

[66] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. 2, 3

[67] Eli Schwartz, Raja Giryes, and Alex M. Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline. *Trans. Img. Proc.*, 28(2):912–923, 2019. 2

[68] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. Motion adaptive deblurring with single-photon cameras. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1945–1954, 2021. 1, 2

[69] Anita Sellent, Carsten Rother, and Stefan Roth. Stereo video deblurring. In *Computer Vision – ECCV 2016*, pages 558–575, Cham, 2016. Springer International Publishing. 2

[70] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 237–246, 2017. 2, 3

[71] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2

[72] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[73] Wei Wang, Xin Chen, Cheng Yang, Xiang Li, Xuemei Hu, and Tao Yue. Enhancing low light videos by exploring high sensitivity camera noise. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 2

[74] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 5

[75] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17683–17693, 2022. 1, 2, 5

[76] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8520–8537, 2022. 1, 2, 3

[77] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, 2018. 1

[78] Yanyang Yan, Wenqi Ren, Yuanfang Guo, Rui Wang, and Xiaochun Cao. Image deblurring via extreme channels prior. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6978–6986, 2017. 2

[79] Huanjing Yue, Cong Cao, Lei Liao, Ronghe Chu, and Jingyu Yang. Supervised raw video denoising with a benchmark

dataset on dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2

[80] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 1, 2, 5, 7

[81] Haichao Zhang and David Wipf. Non-uniform camera shake removal using a spatially-adaptive sparse penalty. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2013. 2

[82] Haichao Zhang and Jianchao Yang. Intra-frame deblurring by leveraging inter-frame camera motion. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4036–4044, 2015. 2

[83] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

[84] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 5

[85] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4593–4601, 2021. 2

[86] Yangyang Zhong, Zunjie Zhu, and Zhefeng Xu. Video deblurring network based on dark light enhancement. In *5th International Conference on Computer Information Science and Application Technology (CISAT 2022)*, page 124511Z. International Society for Optics and Photonics, SPIE, 2022. 1, 2

[87] Zhihang Zhong, Ye Gao, Yinqiang Zheng, Bo Zheng, and Imari Sato. Real-world video deblurring: A benchmark dataset and an efficient recurrent neural network. *International Journal of Computer Vision*, 2022. 3

[88] Chu Zhou, Minggui Teng, Jin Han, Chao Xu, and Boxin Shi. Delieve-net: Deblurring low-light images with light streaks and local events. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1155–1164, 2021. 1, 2

[89] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *ECCV*, 2022. 1, 3, 5, 7, 8