BurstDeflicker: A Benchmark Dataset for Flicker Removal in Dynamic Scenes

Lishen Qu[1,3,5], **Zhihao Liu**[3], **Shihao Zhou**[1,3], **Yaqi Luo**[3]
Jie Liang[5], **Hui Zeng**[5], **Lei Zhang**[4,5], **Jufeng Yang**[1,2,3,*]
[1]Nankai International Advanced Research Institute (SHENZHEN·FUTIAN)
[2]Peng Cheng Laboratory [3]College of Computer Science, Nankai University
[4]The Hong Kong Polytechnic University [5]OPPO Research Institute
{qulishen, 2212602, zhoushihao96, 2323483}@mail.nankai.edu.cn
liang27jie@163.com, cshzeng@gmail.com, cslzhang@comp.polyu.edu.hk
yangjufeng@nankai.edu.cn
https://github.com/qulishen/BurstDeflicker

Abstract

Flicker artifacts in short-exposure images are caused by the interplay between the row-wise exposure mechanism of rolling shutter cameras and the temporal intensity variations of alternating current (AC)-powered lighting. These artifacts typically appear as non-uniform brightness distribution across the image, forming noticeable dark bands. Beyond compromising image quality, this structured noise also impacts high-level tasks, such as object detection and tracking, where reliable lighting is crucial. Despite the prevalence of flicker, the lack of a large-scale, realistic dataset has been a significant barrier to advancing research in flicker removal. To address this issue, we present *BurstDeflicker*, a scalable benchmark constructed using three complementary data acquisition strategies. First, we develop a Retinex-based synthesis pipeline that redefines the goal of flicker removal and enables controllable manipulation of key flicker-related attributes (e.g., intensity, area, and frequency), thereby facilitating the generation of diverse flicker patterns. Second, we capture 4,000 real-world flickering images from different scenes, which help the model better understand the spatial and temporal characteristics of real flicker artifacts and generalize more effectively to wild scenarios. Finally, due to the non-repeatable nature of dynamic scenes, we propose a green-screen method to incorporate motion into image pairs while preserving real flicker degradation. Comprehensive experiments demonstrate the effectiveness of our dataset and its potential to advance research in flicker removal.

1 Introduction

Flicker artifacts commonly arise in images captured under alternating current (AC)-powered light sources [1, 2, 3]. This phenomenon is especially prevalent in high-speed photography [4, 5, 6], high dynamic range (HDR) imaging [7, 8, 9], and slow-motion video recording [10, 11, 12], where short exposures are either required or frequently used. We provide a schematic overview of flicker formation as shown in Figure 1(a). Flicker artifacts arise primarily from two reasons. First, since the intensity of AC varies periodically, AC-powered light sources inherently exhibit periodic fluctuations in brightness [1, 3, 13]. Although these fluctuations are generally imperceptible to the human eye due to the persistence of vision [14], they become visible in images captured with short exposure times, which may sample only a narrow temporal slice of the flicker cycle. Second, most consumer-grade

*Corresponding Author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Track on Datasets and Benchmarks.

 https://doi.org/10.52202/085713-1821

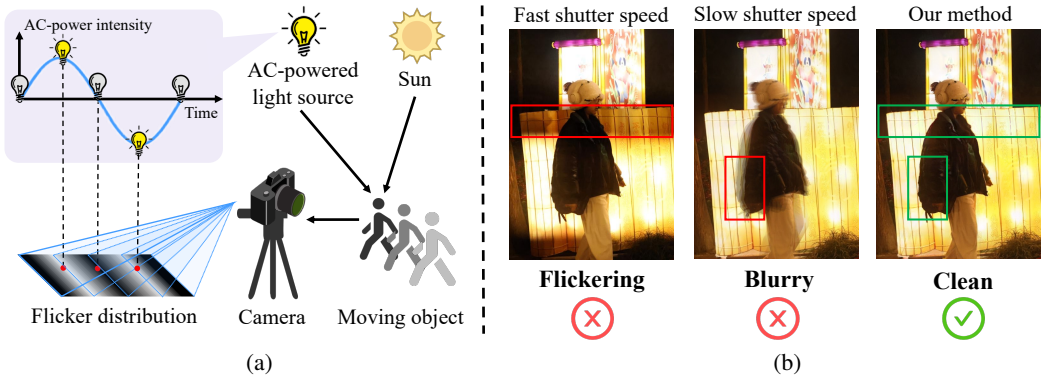

Figure 1: (a) Illustration of flicker formation. The moving object is illuminated by stable light and AC-powered flickering sources. The intensity of the flickering component changes over time (purple area), and each row is exposed at a slightly different moment, leading to a non-uniform brightness distribution across the captured image. (b) Capturing short-exposure images under artificial lighting often results in flicker degradation (red box in the left). Although increasing the exposure time can mitigate flicker artifacts, it introduces motion blur [17, 18, 19] (red box in the middle). Our method effectively removes flicker while preserving fine image details (green box in the right).

digital cameras utilize rolling shutter and line-scan exposure mechanisms [1, 15, 16], in which different rows of the image sensor are exposed at slightly different times. This temporal offset introduces spatial inconsistencies in illumination. Flicker artifacts degrade visual quality, potentially ruining valuable moments for users, especially in scenes with abundant artificial lighting, such as amusement parks, lantern festivals, and cinemas. In addition, flicker impairs the performance of downstream computer vision tasks such as detection, tracking, and recognition [20, 21, 22].

Traditional approaches for flicker mitigation primarily rely on sensor-level solutions, such as integrating flicker detection circuits into camera hardware [23]. Upon detecting flicker, these systems often extend the exposure time to mitigate its effects. However, this introduces motion blur [24, 25, 26, 27, 28], as presented in Figure 1(b). Owing to the adaptability and scalability of deep learning [29, 30], data-driven methods have become the dominant paradigm in image restoration tasks [31, 32, 33]. Their effectiveness heavily depends on the diversity and realism of training datasets. Wong *et al.* [1] propose simulating flicker by superimposing sinusoidal intensity variations onto static images to emulate periodic illumination fluctuations. However, their synthetic data is primarily intended for geo-tagging tasks, rather than for flicker removal. Besides, unlike globally consistent dimness in low-light scenarios [34], flicker artifacts are spatially localized and temporally dynamic. Therefore, single-image flicker removal (SIFR) methods [2, 35] struggle to differentiate flicker from similar dark regions (e.g., shadows), and cannot recover the severely degraded areas due to the lack of pixel context [36]. The intrinsic ambiguity of spatial-only observations results in unreliable restoration.

Modern handheld devices typically capture multiple frames in a single shot, inherently providing rich temporal cues and inter-frame correlations [37]. These cues are highly beneficial for accurate flicker localization and removal, effectively reducing the reliance on explicit priors such as masks in deshadowing [38, 39]. Leveraging this property, multi-frame flicker removal (MFFR) emerges as a more promising and practical solution, particularly under dynamic conditions. However, building a large-scale and high-quality MFFR dataset for dynamic scenes remains highly challenging. On the one hand, capturing a large number of real-world flickering image pairs is labor-intensive and time-consuming [40]. Even with sufficient manpower, the quantity and diversity of flickering patterns that can be collected through real-world capture remain limited. On the other hand, the non-repeatable nature of dynamic scenes makes it nearly impossible to acquire aligned flickering and flicker-free pairs with motion. The lack of such dynamic paired data leads to a critical issue: models tend to misinterpret motion-induced pixel variations as flicker artifacts.

To address these challenges, we construct a comprehensive dataset from three complementary perspectives. First, we propose a flicker synthesis method grounded in Retinex theory [41], capable of generating an unlimited number of flickering images across different scenes. The method explicitly models the interplay between ambient lighting and flickering light sources and supports diverse flicker

https://doi.org/10.52202/085713-1821

patterns caused by common light sources with different rectification modes. Second, to bridge the domain gap between synthetic and real-world datasets, we collect real-world flickering sequences from various static scenes. These sequences serve as a foundation for understanding the spatial and temporal characteristics of real flicker artifacts. Finally, to overcome the inherent non-repeatability of dynamic scenes, we introduce a novel green-screen compositing method. Specifically, we extract foreground subjects with motion from green-screen footage and composite them onto the previously captured flickering backgrounds. This results in a set of flickering image pairs that preserve real flicker degradation while introducing realistic motion dynamics. By integrating these three complementary data sources, we introduce the first MFFR dataset, named *BurstDeflicker*. It comprises an unlimited number of synthetic images, 4,000 real-world flickering image pairs across diverse scenes, and 3,690 green-screen dynamic image pairs generated from the real images.

Our main contributions are summarized as follows: (1) We present BurstDeflicker, the first dataset for multi-frame flicker removal (MFFR), consisting of synthetic, real-world captured, and manually constructed dynamic data. (2) We propose a Retinex-based flicker synthesis method that jointly models ambient and flickering illumination. This enables the scalable generation of synthetic images with diverse and realistic flicker patterns. To overcome the difficulty of acquiring dynamic flickering image pairs, we introduce a green-screen compositing method, which helps mitigate motion ghosting artifacts in multi-frame restoration. (3) Extensive quantitative and qualitative experiments validate the effectiveness of our proposed MFFR dataset. We believe that it will provide a strong foundation for future research in flicker removal.

2 Related work

Hardware-based methods. Early efforts in flicker removal commonly rely on specialized hardware components, such as photodiodes, to monitor periodic brightness fluctuations in the environment. These systems dynamically adjust camera exposure settings in response, thereby reducing the visibility of flicker artifacts [42]. For instance, Park *et al.* [43] proposed a basic flicker detection and avoidance strategy using a lookup table combined with PID control to modulate exposure timing. Poplin [23] introduced an automatic flicker detection approach optimized for embedded camera systems, aiming for real-time deployment. More recently, neuromorphic vision sensors [44, 45, 46] have opened up new avenues in flicker detection by capturing event-based brightness changes, offering significantly higher temporal resolution than traditional frame-based acquisition. For example, Wang *et al.* [47] proposed a linear comb filter that effectively exploits the high temporal resolution of event-based sensors for flicker removal. Despite their technical advantages, the high cost, limited accessibility, and complex calibration requirements of these methods hinder deployment in consumer-level applications.

Single image flicker removal. The term "flicker" is sometimes used to describe brightness discontinuities across video frames [48, 49], such as those found in old films [50]. Restoring such flicker in videos, typically referred to as photometric stabilization [51], is a distinct task from ours. In this paper, we focus on image degradation caused by flicker of AC-powered lights. Several methods [52, 53, 54] have been proposed to suppress flicker, assuming prior knowledge of the lighting conditions. However, the unavailability of such prior information in most real-world scenarios significantly limits their practicality. Yoo *et al.* [35] presented a wide dynamic range system for flicker removal, leveraging long-exposure frames to recover flicker degradation in short-exposure frames. Kim *et al.* [55] introduced a multiplicative model that estimates spatial illumination variations from uniform reflectance areas, which experiences a significant drop in performance when dealing with complex backgrounds. More recently, Lin *et al.* [2] have introduced DeflickerCycleGAN, the first learning-based approach for single-image flicker removal (SIFR). By designing tailored loss functions, Flicker Loss and Gradient Loss, they effectively harness the translation capabilities of CycleGAN [56] to suppress flicker artifacts.

Flicker removal dataset. The success of learning-based flicker removal models is heavily dependent on the availability of high-quality paired datasets [57, 58, 59]. However, acquiring a large-scale dataset consisting of both flicker-corrupted and flicker-free image pairs is challenging and labor-intensive, especially for dynamic scenes. Wong *et al.* [1] proposed a flicker synthesis method based on electric network frequency and the rolling shutter mechanism of cameras, primarily for geo-tagging purposes. Then, Lin *et al.* [2] directly used it to construct a flicker removal dataset. As this synthesis approach is originally intended for the geo-tagging task rather than flicker removal, the resulting flickering images lack sufficient variability and realism. Moreover, these synthetic pairs model

 https://doi.org/10.52202/085713-1821

flicker removal as the complete elimination of flicker-induced illumination, whereas in reality, the illumination should be adjusted to its effective value rather than entirely removed. As a result, models trained on such data struggle to generalize well to real-world flicker removal tasks.

To address these limitations, we propose a Retinex-based flicker synthesis method tailored for flicker removal. Our method models the interaction between flickering and stable light sources and covers diverse flicker patterns and intensities, better reflecting the variability found in real-world scenarios. We also collect many real-world flickering images and employ a green-screen compositing technique to address the lack of dynamic paired data.

3 BurstDeflicker

To the best of our knowledge, there is no publicly available dataset specifically designed for flicker removal, which presents a major obstacle for training and evaluating deep-learning models. To tackle this issue, we present the BurstDeflicker dataset, which is composed of three subsets: synthetic data, static data captured from the real world, and dynamic green-screen data derived from real static data. These three subsets respectively address the challenges of acquiring large-scale, realistic, and dynamic data, as illustrated in Figure 2. We believe this dataset will serve as a valuable benchmark for the multi-frame flicker removal (MFFR) task and foster further research in this underexplored yet practically important domain.

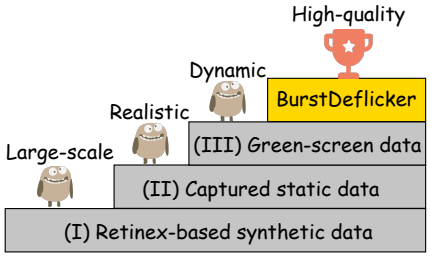

Figure 2: A visual illustration of the three-stage growth of the BurstDeflicker dataset.

3.1 Retinex-based synthetic flicker

The previous flicker synthesis theory [2] treats the brightness caused by flickering light as harmful. The purpose of this method is to remove the brightness changes caused by flickering illumination. We argue that the flickering illumination should not be completely removed. Instead, the instantaneous flicker illumination should be adjusted to an effective value.

The light in a scene can be categorized into two components: flickering light and ambient light [35], as shown in Figure 1(a). According to the Retinex theory [41], the flickering image $I_{flicker} \in \mathbb{R}^{H\times W\times 3}$ can be expressed as:

$$I_{flicker} = R \odot (L_a + L_f) \tag{1}$$

where $\odot$ is the element-wise multiplication. $R \in \mathbb{R}^{H\times W\times 3}$ represents the reflectance, and $L_a \in \mathbb{R}^{H\times W}$ and $L_f \in \mathbb{R}^{H\times W}$ are the illumination maps of the ambient light and the flickering light, respectively.

The goal of flicker removal is to obtain a flicker-free image $I_{clean} \in \mathbb{R}^{H\times W\times 3}$. In contrast to previous work [2], which model $I_{clean} = R \odot L_a$, we propose that the correct formulation should be $I_{clean} = R \odot (L_a + \overline{L_f})$, where $\overline{L_f}$ denotes the effective value of the flickering light illumination. To derive the relationship between the flickering and flicker-free images, we rewrite Equation (1) as follows:

$$I_{flicker} = R \odot (L_a + \overline{L_f} - \overline{L_f} + L_f) \tag{2}$$

We assume that the ambient light intensity is k times the flickering light intensity. The flickering images can be derived:

$$I_{flicker} = (k+1)R \odot \overline{L_f} \cdot \left(1 + \frac{L_f/\overline{L_f} - 1}{k+1}\right) = I_{clean} \cdot \left(1 + \frac{L_f/\overline{L_f} - 1}{k+1}\right) \tag{3}$$

By changing the background, flicker modes, and the ratio of ambient light to flickering light, we can synthesize flickering images with diverse patterns.

Light source rectification modes. In the real world, different light sources have different rectification modes, including full-wave rectified fluorescent lights, half-wave rectified incandescent bulbs, and PWM-rectified LED lights [60, 61]. The sinusoidal alternating current, after undergoing different

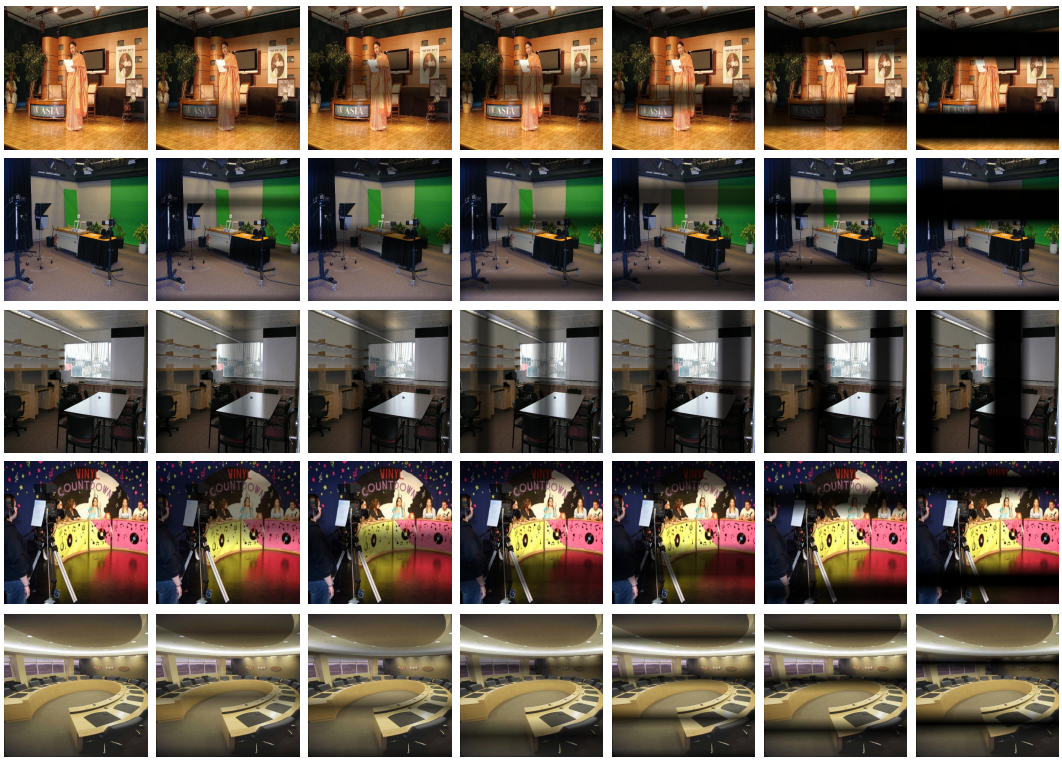

Figure 3: Flicker synthesis results based on the proposed Retinex-based method. Background images are sourced from the indoorCVPR dataset [63]. A pre-training is conducted using the synthetic data, providing a strong initialization for subsequent training on real data.

rectification methods, results in distinct patterns. The flicker caused by light with different rectification modes can be denoted as:

$$
\begin{aligned}
L_{f\sim full} &= A^c \left| \cos\left(2\pi f_{\text{enf}} \frac{y}{f_{\text{row}}} + \varphi\right) \right| \\
L_{f\sim half} &= A^c \max\left(0, \cos\left(2\pi f_{\text{enf}} \frac{y}{f_{\text{row}}} + \varphi\right)\right) \\
L_{f\sim pwm} &= \begin{cases} A^c, & \text{if } \cos\left(2\pi f_{\text{enf}} \frac{y}{f_{\text{row}}} + \varphi\right) > \cos(\pi D) \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
\tag{4}
$$

where $L_{f\sim full}$, $L_{f\sim half}$ and $L_{f\sim pwm}$ represent the flicker intensity distributions under full-wave, half-wave, and PWM rectification modes, respectively. f_{enf} denotes the electric network frequency and f_{row} represents the row scanning frequency of the camera. y represents the y-th row of pixel points along the scanning direction of the sensor. φ is the initial phase when capturing images. D is the duty cycle of the PWM rectification mode, with a range of values from 0 to 1. A^c represents the intensity of each RGB channel, which depends on the spectra of the light source [62]. Specific solutions for L_f in Equation 3 including full-wave rectification, half-wave rectification, and PWM, which are derived in Equation 4 using $L_{f\sim full}$, $L_{f\sim half}$, and $L_{f\sim pwd}$, respectively.

Following Lin *et al.* [2], we select the images from the indoorCVPR dataset [63] as background images. We synthesize flicker with the same pattern on each burst sequence, only changing φ of the AC power. The range of the intensity ratio k between ambient light and flicker light is set from 0 to 1. According to [1], f_{enf} is 50 or 60 Hz. We use the same f_{row} between 100 kHz and 160 kHz as in [2] for 512×512 resolution. Representative visualizations of the synthetic data are provided in Figure 3.

https://doi.org/10.52202/085713-1821

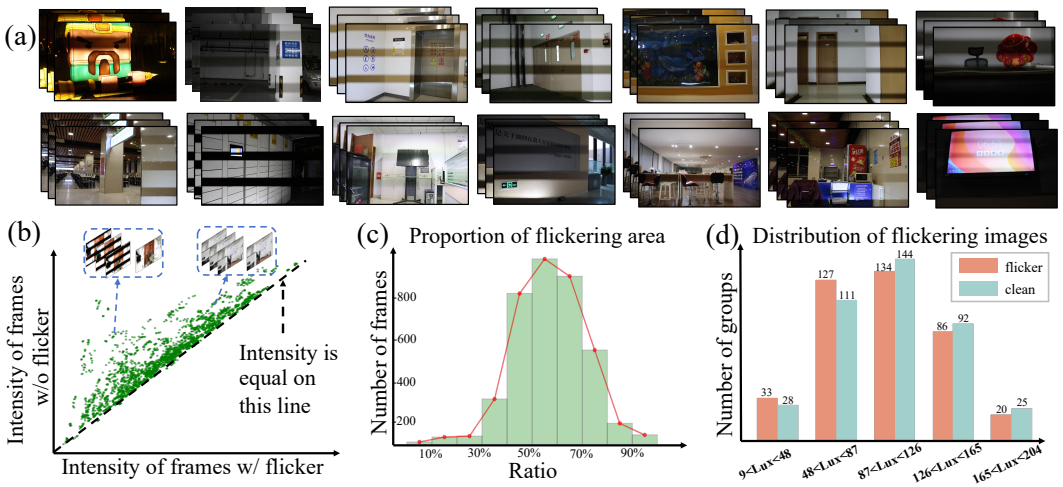

Figure 4: Illustration of the real captured dataset. (a) Example images from our dataset, which include a diverse range of common artificial lighting scenarios. (b) The intensity distributions of flicker and non-flicker frames. (c) The area ratio of flicker degradation per image. (c) The luminance distribution of flickering and clean images across different scenes.

3.2 Collection of real-world flickering image pairs

To construct a high-quality dataset containing real-world flickering images, we design a data acquisition pipeline that ensures spatial alignment, high resolution, and scene diversity. A key challenge lies in accurately capturing flicker artifacts while minimizing misalignment between flickering frames and their corresponding ground truth (GT) references.

To address this, we employ a Canon EOS R7 with a 15–150mm f/2.8 lens and a Canon EOS R6 Mark II paired with a 24–105mm f/4–7.1 STM lens, both securely mounted on tripods. To minimize vibration, a remote shutter release is used, and the cameras operate in electronic shutter burst mode to eliminate mechanical jitter during high-speed capture. All recordings are made in manual mode to ensure consistent imaging conditions. First, we reduce the shutter speed to 1/1000-1/2000 seconds to capture flicker artifacts in static scenes. These short exposures capture the high-frequency luminance variations caused by artificial light sources. Second, we capture a clean, flicker-free image using a slow shutter speed of 1/50 or 1/60 seconds, depending on the local electric network frequency [1]. The ISO is adjusted accordingly to ensure that the clean long-exposure image receives the same amount of light as flickering images. This long exposure integrates multiple flicker cycles, effectively suppressing temporal luminance variations and producing an ideal illumination reference.

We capture 10 consecutive frames in a single burst sequence, preserving visible flickering artifacts with precise spatial alignment. We collect flickering sequences across 369 different real-world scenes, including indoor (e.g., offices, supermarkets, subway stations) and outdoor (e.g., LED billboards, parking lots) environments. Since all the flickering images in this subset are captured under static scenes, we refer to this set of 4,000 images as BurstDeflicker-S. Its distribution characteristics and representative sequences are illustrated in Figure 4.

3.3 Green-screen compositing flickering image

Since dynamic motion in the real world cannot be exactly replicated, real-world flickering image pairs can only be captured in static scenes. Insufficient exposure to dynamic data may lead the model to misinterpret motion-related pixel variations as flicker, thereby introducing ghosts or misaligned artifacts in the restored images. To address this challenge, we draw inspiration from green-screen compositing techniques widely used in the film industry and propose a novel green-screen approach to simulate realistic motion in flickering sequences. We present a schematic illustration of the motion synthesis process, as shown in Figure 5(a). The synthesis phase takes a group of real-world flickering image pairs as background, and overlays green-screen foregrounds selected from the VideoMatte240K dataset [64]. For each scene, we manually select foreground clips that are semantically and spatially

https://doi.org/10.52202/085713-1821

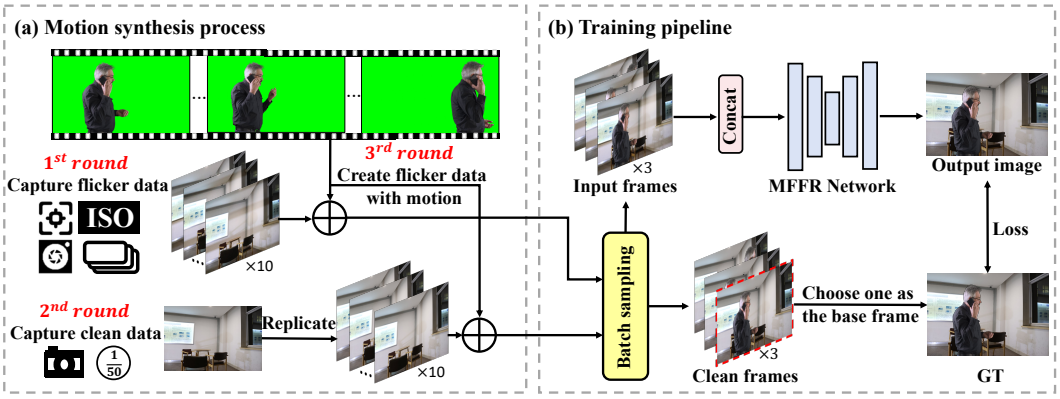

Figure 5: The motion synthesis process and training pipeline. (a) The green-screen footage is selected from the VideoMatte240K dataset [64]. The compositing of green-screen foregrounds with the backgrounds is manually performed using Adobe After Effects. (b) Given a sequence of flickering frames, we select three frames as input to form a training batch, with the target being a single clean reference image. The synthetic dataset and BurstDeflicker-S also follow this pipeline.

compatible with the background, and use the corresponding alpha masks to perform high-quality compositing. Since the clean GT image is a single frame, we replicate it ten times and overlay the same ten foreground clips used in the flickering sequence. This compositing strategy allows us to simulate realistic dynamic content while preserving authentic flicker artifacts in the background, effectively addressing the lack of dynamic scenes in the dataset. Using this approach, we construct a semi-synthetic green-screen dataset consisting of 3,690 images with motion, which we denote as BurstDeflicker-G. It serves as a crucial step toward improving the robustness and generalization of flicker removal models in dynamic, real-world scenarios. More details can be found in the supplementary materials.

We obtain 4,000 real-world static flickering images (referred to as BurstDeflicker-S) along with their corresponding dynamic green-screen image pairs (referred to as BurstDeflicker-G). These datasets are partitioned into training and testing splits using an 8/2 ratio. To simulate natural handheld motion, we introduce synthetic camera shake by applying random rotations in the range of $[-3°, 3°]$ and translations within $[-5, 5]$ pixels to the burst sequences. Besides, to facilitate model training, input images are often cropped into small patches in burst image super resolution [36, 65, 66]. However, flicker artifacts are localized and exhibit periodic patterns along the line-scan direction. To preserve this periodicity, we resize the images during training instead of cropping them.

Training pipeline. We present the MFFR training pipeline in Figure 5(b). For each training iteration, we randomly sample three frames at intervals of 1 to 3 to simulate different varying camera capture rates. These three frames are augmented and concatenated, then fed into different MFFR networks for restoration. Note that the final output is a single image, corresponding to one of the three flickering frames. Owing to the significant expense and effort involved in collecting paired MFFR data, training a model from scratch with a large real dataset is impractical [67]. Following the previous burst super-resolution work [36], we use the proposed Retinex-based synthesis method to generate a large dataset for pre-training the network. The pre-trained model acts as a strong initialization, then undergoes fine-tuning on our BurstDeflicker dataset for real-world flicker removal.

4 Experiments

4.1 Comparison with previous work

Experimental settings. Low-light enhancement is similar to the flicker removal task, so we use the pre-trained model of the representative Retinexformer [68] for flicker removal. We also test the performance of the state-of-the-art SIFR method, DeflickerCycleGAN, proposed by Lin *et al.* [2]. We conduct a benchmark using three baselines trained on our dataset: Burstormer [69] for burst image restoration, HDRTransformer [70] for HDR imaging, and Restormer [71] for single-image restoration. Specifically, we change Restormer's input channels from 3 to 9 and concatenate the input

https://doi.org/10.52202/085713-1821

(a) Input (b) Retinexformer [68] (c) Lin *et al.* [2] (d) Ours (e) Reference

Figure 6: Visual results of flicker removal on the static test data. Our results are obtained by training the Restormer on our dataset. It demonstrates the best performance across diverse flickering scenes.

frames along the channel dimension to achieve multi-frame fusion. Due to the memory limitation (24 GB) of the RTX 3090, we reduce the parameters of Restormer [71]. Specifically, we reduce the number of refinement blocks in Restormer [71] from 4 to 2, and set the feature channel dimension to 32 instead of the default 48. Besides, since Burstformer [69] originally takes 8-frame inputs while Restormer [71] is designed for single-frame inputs, we modify both models to take 3-frame inputs for consistency. More experimental details, settings, and qualitative comparisons are presented in the supplementary materials.

Qualitative comparison. We show the visual comparison of flicker removal on the test data of BurstDeflicker-S in Figure 6. The low-light image enhancement method, Retinexformer [68], globally brightens the image and introduces color shifts, which is undesirable for the goal of flicker removal. The SIFR method of Lin *et al.* [2] has little effect on flicker removal, especially in cases of severe flicker and insufficient lighting. Restormer, trained on our dataset, demonstrates effective flicker removal in dealing with mild indoor flicker and strong nighttime flicker.

Quantitative comparison. We evaluate the performance of different methods on the static test set using three full-reference metrics, including PSNR, SSIM [72], and LPIPS [73]. Since the test set and training set come from the same domain, the overfitted model may achieve better results. Therefore, we capture 50 dynamic test sequences using various mobile devices and consumer cameras to evaluate the model's performance and robustness in real-world dynamic scenarios. Since the ground truths of real dynamic images are unavailable, we employ MUSIQ [74], BRISQUE [75], and PIQE [76] as no-reference evaluation metrics.

The flicker removal results of different methods are shown in Table 1. Retinexformer performs poorly in flicker removal, indicating that training on a low-light dataset alone is insufficient to

Table 1: Quantitative results of different methods. Note that '*' indicates models that are not trained on our dataset but directly tested using their original versions. To ensure a fair comparison of dataset effectiveness, we retrain Retinexformer [68] and the model of Lin *et al.* [2] on our dataset using single-frame input, following the same setup as in their original work.

Method	Flops (G)	Params (M)	Static test data			Dynamic test data		
			PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	PIQE ↓	BRISQUE ↓
*Retinexformer [68]	69.23	1.61	15.704	0.707	0.213	53.596	50.269	30.242
*Lin *et al.* [2]	509.80	92.08	20.358	0.838	0.134	55.228	43.875	25.121
Lin *et al.* [2]	509.80	92.08	26.408	0.875	0.102	58.131	35.710	22.102
Retinexformer [68]	69.23	1.61	27.212	0.885	0.081	58.249	35.942	21.648
Burstormer [69]	141.05	0.17	29.439	0.910	0.056	58.527	37.014	20.451
HDRTransformer [70]	272.12	1.04	30.031	0.914	0.054	59.069	37.292	21.588
Restormer [71]	149.01	7.92	**30.634**	**0.918**	**0.045**	**59.097**	**34.896**	**19.324**

Table 2: Ablation study of Restormer trained on different parts of the dataset. The synthetic dataset enhances the model's robustness and overall performance. In particular, the green-screen data (BurstDeflicker-G) significantly improves the model's effectiveness in real-world dynamic scenarios.

Training data			Static test data			Dynamic test data		
Synthetic data	BurstDeflicker-S	BurstDeflicker-G	PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	PIQE ↓	BRISQUE ↓
✓			24.483	0.862	0.122	57.096	39.726	23.755
	✓	✓	30.481	0.915	0.053	58.011	37.498	21.523
✓	✓		**30.645**	0.916	0.052	58.431	36.259	20.338
✓	✓	✓	30.634	**0.918**	**0.045**	**59.097**	**34.896**	**19.324**

(a) Input (b) w/o BurstDeflicker-G (c) Full data

Figure 7: The visual comparison on the dynamic test data of Restormer trained with/without BurstDeflicker-G. The introduction of BurstDeflicker-G helps reduce motion ghosting artifacts (red boxes) without compromising the flicker removal performance (green boxes).

address the flicker problem. We re-train Retinexformer [68] and Lin [2]'s network on our dataset, achieving significant improvements, which demonstrates the effectiveness of the BurstDeflicker dataset. Additionally, we train three representative networks using a burst size of three, all of which demonstrate strong flicker removal performance, with Restormer achieving the best results.

4.2 Ablation study

Parts of BurstDeflicker. Synthetic data helps reduce the risk of overfitting and enhances the model's robustness, as demonstrated in the second and fourth rows of Table 2. The BurstDeflicker-G subset helps improve the model's performance on the dynamic test set, as shown in the third and fourth rows of Table 2. To intuitively explain the improvement caused by green-screen data, we present a visual comparison of flicker removal on the handheld-captured data in Figure 7. The model without green-screen data may introduce motion ghosts. The essence of the green-screen method is to allow the model to "see" flickering image pairs with motion. By training on such data, the model learns to distinguish whether pixel value changes are caused by motion or flicker.

https://doi.org/10.52202/085713-1821

Table 3: The test results with different numbers of burst image inputs. Multi-frame flicker removal achieves better performance compared to single-image restoration. Due to the redundancy between adjacent frames, the performance gain from increasing the input from two to three frames is smaller than that from one to two frames, which indicates the marginal effect of adding more input frames.

Input	Flops (G)	Static test data			Dynamic test data		
		PSNR ↑	SSIM ↑	LPIPS ↓	MUSIQ ↑	PIQE ↓	BRISQUE ↓
Single image	148.55	27.310 (+0.000)	0.891 (+0.000)	0.069 (-0.000)	58.664	35.821	20.010
Burst-2	148.78	30.264 (+2.594)	0.915 (+0.024)	0.048 (-0.021)	58.786	35.277	19.798
Burst-3	149.01	30.634 (+3.324)	0.918 (+0.027)	0.045 (-0.024)	**59.097**	**34.896**	**19.324**

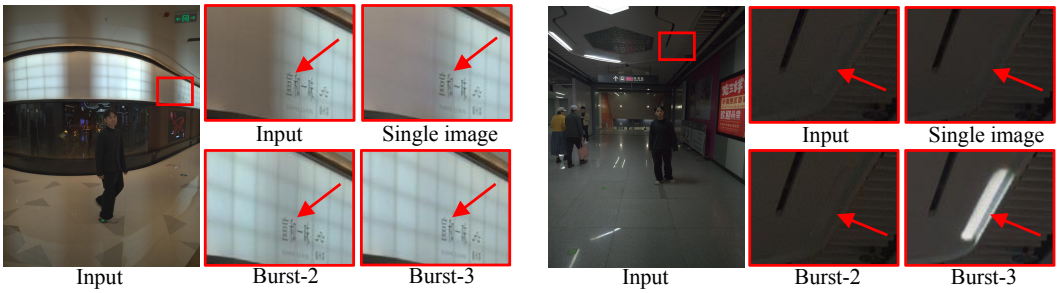

Figure 8: The visual comparison of flicker removal using different numbers of input frames. As the number of input frames increases, the model exhibits progressively better restoration performance for banding artifacts (left). Moreover, when the number of input frames reaches three, the model successfully restores the originally extinguished light source (right).

Number of input frames. Following previous multi-frame restoration tasks [36, 57], we train Restormer with varying numbers of input frames to validate the effectiveness of the MFFR strategy. We present the performance of Restormer [71] under varying numbers of input frames, as shown in Table 3. The PSNR improvement from 2 to 3 input frames (0.730 dB) is less significant than that from 1 to 2 input frames (2.594 dB), which can be attributed to the overlapping clean regions among multiple input frames. We provide a visual comparison of restoration results using different numbers of input frames, as depicted in Figure 8.

5 Conclusion

In this paper, we introduced the first multi-frame flicker removal (MFFR) dataset, BurstDeflicker, which consists of large-scale synthetic images, 4,000 real-world captured image pairs, and 3,690 manually created image pairs with motion. The synthetic method simulated the interaction between ambient light and flicker light, considering various flicker patterns for pretraining the flicker removal model. The real-world captured dataset was used to fine-tune the model for improved performance. Furthermore, since paired motion images are difficult to capture, we proposed a motion embedding method based on the green-screen technique, which helped mitigate motion ghosting issues in multi-frame fusion. Comprehensive experiments demonstrated the effectiveness of our MFFR method and dataset, which can facilitate future research in flicker removal.

Limitation. When an AC-powered light source serves as the sole illumination source, the resulting flicker degradation can be particularly severe, potentially leading to noticeable color shifts in the restoration results, as illustrated in the fourth row of Figure 6. Although MFFR methods can leverage multi-frame information for more accurate restoration, they still struggle when the available frames lack complete scene content. Besides, when there is significant misalignment between frames (e.g., strong handheld jitter), the performance of multi-frame restoration may degrade to that of single-frame methods. To address these challenges, future work could focus on designing more efficient network architectures that are specifically tailored to the unique features and priors of flicker.

Acknowledgement. This work was supported by Shenzhen Science and Technology Program (No. JCYJ20240813114229039), National Natural Science Foundation of China (No. 624B2072), Natural Science Foundation of Tianjin (No.24JCZXJC00040), Supercomputing Center of Nankai University, and OPPO Research Fund.

References

[1] Chau-Wai Wong, Adi Hajj-Ahmad, and Min Wu. Invisible geo-location signature in a single image. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1987–1991, 2018.

[2] Xiaodan Lin, Yangfu Li, Jianqing Zhu, and Huanqiang Zeng. Deflickercyclegan: Learning to detect and remove flickers in a single image. *IEEE Transactions on Image Processing*, 32:709–720, 2023.

[3] Pawan Kumar, Bhim Singh, Ambrish Chandra, and Kamal Al-Haddad. Flicker detection, measurement and means of mitigation: A review. *Journal of The Institution of Engineers (India): Series B*, 95(2):109–119, 2014.

[4] Liang Gao, Jinyang Liang, Chi Li, and Lihong V. Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014.

[5] Yajing Zheng, Lingxiao Zheng, Zhaofei Yu, Boxin Shi, Yonghong Tian, and Tiejun Huang. High-speed image reconstruction through short-term plasticity for spiking cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6358–6367, 2021.

[6] Yunhao Zou, Yinqiang Zheng, Tsuyoshi Takatani, and Ying Fu. Learning to reconstruct high speed and high dynamic range videos from events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2024–2033, 2021.

[7] Yiheng Chi, Xingguang Zhang, and Stanley H Chan. Hdr imaging with spatially varying signal-to-noise ratios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5724–5734, 2023.

[8] Zhiyuan Pu, Peiyao Guo, M Salman Asif, and Zhan Ma. Robust high dynamic range (hdr) imaging with complex motion and parallax. In *Proceedings of the Asian Conference on Computer Vision*, pages 134–149, 2020.

[9] Yuang Meng, Xin Jin, Lina Lei, Chun-Le Guo, and Chongyi Li. Ultraled: Learning to see everything in ultra-high dynamic range scenes. *arXiv preprint arXiv:2510.07741*, 2025.

[10] Meiguang Jin, Zhe Hu, and Paolo Favaro. Learning to extract flawless slow motion from blurry videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8112–8121, 2019.

[11] Xiaoyu Xiang, Yapeng Tian, Yulun Zhang, Yun Fu, Jan P Allebach, and Chenliang Xu. Zooming slow-mo: Fast and accurate one-stage space-time video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3370–3379, 2020.

[12] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 1000 fps hdr video with a spike-rgb hybrid camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22180–22190, 2023.

[13] Mark Sheinin, Yoav Y Schechner, and Kiriakos N Kutulakos. Computational imaging on the electric grid. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6437–6446, 2017.

[14] Joseph Anderson and Barbara Anderson. The myth of persistence of vision revisited. *Journal of Film and Video*, pages 3–12, 1993.

[15] Mark Sheinin, Yoav Y Schechner, and Kiriakos N Kutulakos. Rolling shutter imaging on the electric grid. In *IEEE International Conference on Computational Photography*, pages 1–12, 2018.

[16] Luc Oth, Paul Furgale, Laurent Kneip, and Roland Siegwart. Rolling shutter camera calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1360–1367, 2013.

 https://doi.org/10.52202/085713-1821

[17] Jiangxin Dong, Jinshan Pan, Zhongbao Yang, and Jinhui Tang. Multi-scale residual low-pass filter network for image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12345–12354, 2023.

[18] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023.

[19] Duosheng Chen, Shihao Zhou, Jinshan Pan, Jinglei Shi, Lishen Qu, and Jufeng Yang. A polarization-aided transformer for image deblurring via motion vector decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28061–28070, 2025.

[20] Jiacheng Wang, Hongyang Du, Dusit Niyato, Mu Zhou, Jiawen Kang, Zehui Xiong, and Abbas Jamalipour. Through the wall detection and localization of autonomous mobile device in indoor scenario. *IEEE Journal on Selected Areas in Communications*, 42(1):161–176, 2023.

[21] Tanvir Ahmed, Toon Calders, Hua Lu, and Torben Bach Pedersen. Risk detection and prediction from indoor tracking data. *SIGSPATIAL Special*, 9(2):11–18, 2017.

[22] Chuanxin Song and Xin Ma. Srrm: Semantic region relation model for indoor scene recognition. In *International Joint Conference on Neural Networks*, pages 01–08, 2023.

[23] D. Poplin. An automatic flicker detection method for embedded camera systems. *IEEE Transactions on Consumer Electronics*, 52(2):308–311, 2006.

[24] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023.

[25] Chengxu Liu, Xuan Wang, Xiangyu Xu, Ruhao Tian, Shuai Li, Xueming Qian, and Ming-Hsuan Yang. Motion-adaptive separable collaborative filters for blind motion deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25595–25605, 2024.

[26] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, Kyoung Mu Lee, Liangyu Chen, Jie Zhang, Xin Lu, Xiaojie Chu, Chengpeng Chen, et al. NTIRE 2021 challenge on image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 149–165, 2021.

[27] George Ciubotariu, Florin-Alexandru Vasluianu, Zhuyun Zhou, Nancy Mehta, Radu Timofte, Ke Wu, Long Sun, Lingshun Kong, Zhongbao Yang, Jinshan Pan, et al. Aim 2025 challenge on high fps motion deblurring: Methods and results. *arXiv preprint arXiv:2509.06793*, 2025.

[28] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, 2011.

[29] Shihao Zhou, Jinshan Pan, Jinglei Shi, Duosheng Chen, Lishen Qu, and Jufeng Yang. Seeing the unseen: A frequency prompt guided transformer for image restoration. In *European Conference on Computer Vision*, pages 246–264, 2024.

[30] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019.

[31] Kun Zhou, Xinyu Lin, Zhonghang Liu, Xiaoguang Han, and Jiangbo Lu. Ups: Unified projection sharing for lightweight single-image super-resolution and beyond. In *Advances in Neural Information Processing Systems*, pages 81105–81129, 2024.

[32] Dongwoo Lee, JoonKyu Park, and Kyoung Mu Lee. Gs-blur: A 3d scene-based dataset for realistic image deblurring. In *Advances in Neural Information Processing Systems*, pages 125394–125415, 2024.

[33] Jinjin Gu, Xianzheng Ma, Xiangtao Kong, Yu Qiao, and Chao Dong. Networks are slacking off: Understanding generalization problem in image deraining. In *Advances in Neural Information Processing Systems*, pages 28565–28584, 2023.

[34] Paras Maharjan, Li Li, Zhu Li, Ning Xu, Chongyang Ma, and Yue Li. Improving extreme low-light image denoising via residual learning. In *IEEE international conference on multimedia and expo*, pages 916–921, 2019.

[35] Yoonjong Yoo, Jaehyun Im, and Joonki Paik. Flicker removal for cmos wide dynamic range imaging based on alternating current component analysis. *IEEE Transactions on Consumer Electronics*, 60(3):294–301, 2014.

[36] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9209–9218, 2021.

[37] Hao Liu, Zhen Xu, Yifan Wei, Kai Han, and Xin Peng. Multispectral non-line-of-sight imaging via deep fusion photography. *Science China Information Sciences*, 68(4):1–19, 2025.

[38] Yimin Xu, Mingbao Lin, Hong Yang, Fei Chao, and Rongrong Ji. Shadow-aware dynamic convolution for shadow removal. *Pattern Recognition*, 146:109969, 2024.

[39] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: global context helps shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 710–718, 2023.

[40] Xiang Chen, Jinshan Pan, Jiangxin Dong, and Jinhui Tang. Towards unified deep image deraining: A survey and a new benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):5414–5433, 2025.

[41] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[42] Nicolai Behmann and Holger Blume. Real-time LED flicker detection and mitigation: Architecture and fpga-implementation. In *IEEE International Conference on Electronics, Circuits and Systems*, pages 657–658, 2018.

[43] SangHyun Park, GyuWon Kim, and JaeWook Jeon. The method of auto exposure control for low-end digital camera. In *International Conference on Advanced Communication Technology*, pages 1712–1714, 2009.

[44] Giacomo Indiveri and Rodney Douglas. Neuromorphic vision sensors. *Science*, 288(5469):1189–1190, 2000.

[45] Qian-Bing Zhu, Bo Li, Dan-Dan Yang, Chi Liu, Shun Feng, Mao-Lin Chen, Yun Sun, Ya-Nan Tian, Xin Su, Xiao-Mu Wang, Song Qiu, Qing-Wen Li, Xiao-Ming Li, Hai-Bo Zeng, Hui-Ming Cheng, and Dong-Ming Sun. A flexible ultrasensitive optoelectronic sensor array for neuromorphic vision systems. *Nature Communications*, 12:1798, 2021.

[46] Yin Bi, Aaron Chadha, Alhabib Abbas, Eirina Bourtsoulatze, and Yiannis Andreopoulos. Graph-based object classification for neuromorphic vision sensing, 2019.

[47] Ziwei Wang, Dingran Yuan, Yonhon Ng, and Robert Mahony. A linear comb filter for event flicker removal. In *International Conference on Robotics and Automation*, pages 398–404, 2022.

[48] Pengpeng Ni, Ragnhild Eg, Alexander Eichhorn, Carsten Griwodz, and Pål Halvorsen. Flicker effects in adaptive video streaming to handheld devices. In *Proceedings of the ACM International Conference on Multimedia*, pages 463–472, 2011.

[49] Chenyang Lei, Xuanchi Ren, Zhaoxiang Zhang, and Qifeng Chen. Blind video deflickering by neural filtering with a flawed atlas. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10439–10448, 2023.

[50] Valery Naranjo and Antonio Albiol. Flicker reduction in old films. In *International Conference on Image Processing*, pages 657–659, 2000.

 https://doi.org/10.52202/085713-1821

[51] Xuaner Zhang, Joon-Young Lee, Kalyan Sunkavalli, and Zhaowen Wang. Photometric stabilization for fast-forward videos. In *Computer Graphics Forum*, pages 105–113, 2017.

[52] Ehsan Nadernejad, Claire Mantel, Nino Burini, and Søren Forchhammer. Flicker reduction in LED-LCDs with local backlight. In *International Workshop on Multimedia Signal Processing*, pages 312–316, 2013.

[53] Hyun-A Ahn, Seong-Kwan Hong, and Oh-Kyong Kwon. A highly accurate current LED lamp driver with removal of low-frequency flicker using average current control method. *IEEE Transactions on Power Electronics*, 33(10):8741–8753, 2017.

[54] Ignacio Castro, Aitor Vazquez, Manuel Arias, Diego G Lamar, Marta M Hernando, and Javier Sebastian. A review on flicker-free AC–DC LED drivers for single-phase and three-phase AC power grids. *IEEE Transactions on Power Electronics*, 34(10):10035–10057, 2019.

[55] Minwoong Kim, Kurt Bengtson, Lisa Li, and Jan P Allebach. Reducing flicker due to ambient illumination in camera captured images. In *Color Imaging XVIII: Displaying, Processing, Hardcopy, and Applications*, pages 42–51, 2013.

[56] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2223–2232, 2017.

[57] Pengxu Wei, Yujing Sun, Xingbei Guo, Chang Liu, Guanbin Li, Jie Chen, Xiangyang Ji, and Liang Lin. Towards real-world burst image super-resolution: Benchmark and method. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13233–13242, 2023.

[58] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[59] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9700–9709, 2021.

[60] IEEE Power Electronics Society. IEEE recommended practices for modulating current in high-brightness LEDs for mitigating health risks to viewers. *No. S1789-2015*, 2015.

[61] Jennifer A Veitch, Patricia Van Roon, Amedeo D'Angiulli, Arnold Wilkins, Brad Lehman, Greg J Burns, and E Erhan Dikel. Effects of temporal light modulation on cognitive performance, eye movements, and brain function. *Leukos*, 20(1):67–106, 2024.

[62] Saffet Vatansever, Ahmet Emir Dirik, and Nasir Memon. Factors affecting enf based time-of-recording estimation for video. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2497–2501, 2019.

[63] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 413–420, 2009.

[64] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian L Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Real-time high-resolution background matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8762–8771, 2021.

[65] Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Adaptive feature consolidation network for burst super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1279–1286, 2022.

[66] Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 998–1008, 2022.

[67] Siqi Li, Shaoyi Du, Jun-Hai Yong, and Yue Gao. Event-enhanced synthetic aperture imaging. *Science China Information Sciences*, 68(3):134101, 2025.

[68] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023.

[69] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5703–5712, 2023.

[70] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on Computer Vision*, pages 344–360, 2022.

[71] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.

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

[73] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[74] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021.

[75] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[76] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *National Conference on Communications*, pages 1–6, 2015.

 https://doi.org/10.52202/085713-1821

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Refer to Section 1 for details.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in the Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Theoretical result is implemented in Section 3.1.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In Section 3, we clearly describe the steps taken to make our datasets reproducible, including a detailed data acquisition pipeline and data processing methods. In Section 4, training details are clearly provided to make the results verifiable.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

https://doi.org/10.52202/085713-1821

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Our dataset and code are provided in the supplementary material.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The training and test details are specified at Section 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: Error bars or other statistical significance are unnecessary in our experiment.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The main model parameters and computational complexity are in Section 4. More computer resources are provided in the supplementary material.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Every aspect of the experimental paper complies with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The societal impacts are discussed in the supplementary material.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

https://doi.org/10.52202/085713-1821

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This question is unrelated to the research topic of this article.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The existing technologies used have clear references and mentions.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The details about training and limitations for our newly-introduced dataset are described in the paper at Section 3 and Section 5 respectively. The license and other documentation are provided alongside the dataset in the supplementary materials.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This question is unrelated to the research topic of this article.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: Yes.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/085713-1821

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This question is unrelated to the research topic of this article.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.