

High-Quality and Efficient Turbulence Mitigation with Events

Xiaoran Zhang^{1†}, Jian Ding^{1†}, Yuxing Duan¹, Haoyue Liu^{1*}, Gang Chen², Yi Chang¹, Luxin Yan¹

¹ State Key Laboratory of Multispectral Information Intelligent Processing Technology
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology

² School of Computer Science and Engineering, Sun Yat-sen University

{xiaoranzhang, dingjian, liuhuy, yichang}@hust.edu.cn, cheng83@mail.sysu.edu.cn

Abstract

Turbulence mitigation (TM) is highly ill-posed due to the stochastic nature of atmospheric turbulence. Most methods rely on multiple frames recorded by conventional cameras to capture stable patterns in natural scenarios. However, they inevitably suffer from a trade-off between accuracy and efficiency: more frames enhance restoration at the cost of higher system latency and larger data overhead. Event cameras, equipped with microsecond temporal resolution and efficient sensing of dynamic changes, offer an opportunity to break the bottleneck. In this work, we present EHETM, a high-quality and efficient TM method inspired by the superiority of events to model motions in continuous sequences. We discover two key phenomena: (1) turbulence-induced events exhibit distinct polarity alternation correlated with sharp image gradients, providing structural cues for restoring scenes; and (2) dynamic objects form spatiotemporally coherent “event tubes” in contrast to irregular patterns within turbulent events, providing motion priors for disentangling objects from turbulence. Based on these insights, we design two complementary modules that respectively leverage polarity-weighted gradients for scene refinement and event-tube constraints for motion decoupling, achieving high-quality restoration with few frames. Furthermore, we construct two real-world event-frame turbulence datasets covering atmospheric and thermal cases. Experiments show that EHETM outperforms SOTA methods, especially under scenes with dynamic objects, while reducing data overhead and system latency by approximately 77.3% and 89.5%, respectively. Our code is available at <https://github.com/Xavier667/EHETM>.

1. Introduction

Atmospheric and thermal turbulence are major sources of degradation in long-range imaging, arising from random

[†]Equal contribution. ^{*}Corresponding author.

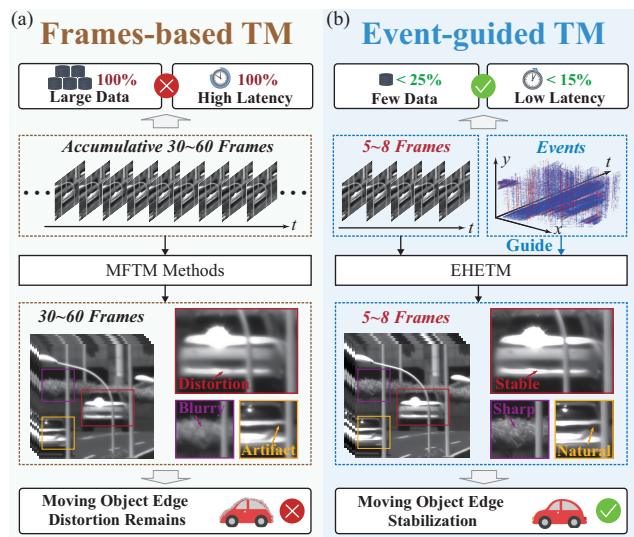


Figure 1. Workflow comparison between our EHETM and multi-frame TM methods (MFTM, exemplified by DATUM [8]). (a) DATUM [8] suffers from high system latency and large overhead due to multi-frame accumulation, while distortions at dynamic object edges remains. (b) Our EHETM uses events for guidance, producing higher-quality and efficient restoration with few frames.

fluctuations of air refractive index [1–5]. Such turbulence, introducing complex distortions characterized by geometric tilts and spatial-varying blurs, degrades image quality and hinders the performance of downstream vision tasks [6, 7].

To address this challenge, early classical methods based on pixel registration, lucky fusion, and deblurring, etc., have provided insights into turbulence mitigation (TM) [4, 9–15]. However, they are computationally expensive and struggling with dynamic scenes. Recently, deep learning-based methods have shown great potential, which can be categorized into single-frame [16–28] and multi-frame methods [1, 8, 12, 29–44]. Single-frame methods offer simplicity but are limited by the ill-posedness of turbulence degradation, undermining restoration quality. Multi-frame methods utilize sufficient temporal cues from a large number of frames

(typically 30–60) recorded by conventional cameras. However, as illustrated in Fig. 1 (a), they suffer a trade-off: more frames enhance restoration at the expense of higher system latency¹(about 1–2 s) and larger data overhead (about 30–60×), impeding their real-time applicability.

Beyond the efficiency–accuracy trade-off, conventional cameras suffer from low-frequency sampling, failing to offer adequate and stable motion information for accurate motion decoupling between object motion and turbulence. It raises a question: *with a high-speed imaging system capable of capturing transient turbulence dynamics, can we extract finer-grained temporal cues within a shorter duration to break the bottleneck in frames-only methods and achieve high-quality and efficient restoration?*

Event cameras [45–47] asynchronously sense per-pixel brightness changes, capturing motion cues with microsecond temporal resolution. Such properties make them a powerful complement to frames-based TM methods by offering temporally dense yet spatially efficient (edge-triggered) information. To intuitively illustrate advantages of high-temporal-resolution data, Fig. 2 compares temporal averaging from a high-speed (800 FPS) and a conventional (25 FPS) camera under thermal turbulence. By capturing finer-grained motion cues, high-temporal-resolution observations rapidly and accurately reveal the zero-mean characteristics of turbulence, producing higher quality within a shorter time than low-temporal-resolution ones. However, high-speed frame cameras incur prohibitively high recourse consumption, making long-duration operation impractical.

In this work, we propose an Event-guided High-quality and Efficient TM (EHETM) method. As shown in Fig. 1 (b), EHETM addresses two key limitations of frames-only methods: it **accelerates TM** by supplying high-temporal-resolution information from events, and it **enhances TM quality** by exploiting temporally dense events to accurately decouple object motion from turbulence. We discover two phenomena: (1) Turbulence-induced pixel displacement exhibits polarity alternation of events along sharp image gradients, offering structural cues for restoring scenes. (2) Objects in natural scenes exhibits continuous motion, which form spatiotemporally coherent “event tubes” in contrast to the irregular patterns of turbulent events, providing stable priors for decoupling object motion from turbulence. Motivated by these observations, we design two complementary modules in EHETM: the former encodes polarity alternation frequency into gradients to reduce edge distortions of scenes, while the latter stabilizes dynamic objects by modeling event tubes. They jointly learn stable patterns from turbulence for high-quality and efficient TM.

Besides, for comprehensive demonstration, we construct two real-world event-frame paired turbulence datasets:

¹System latency: the output for a given frame may depend on future inputs, which imposes an inherent delay independent of computation.

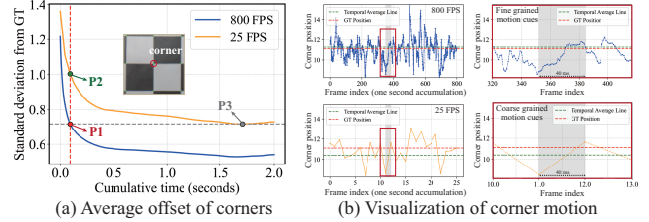


Figure 2. Merits of high-speed observations. (a) It requires less accumulation time to reach the same error (P1 vs. P3) and achieves lower error within the same time (P1 vs. P2). (b) It captures finer-grained motion cues compared to low-temporal-resolution ones.

Close-range Thermal Turbulence Hybrid (CTTH) dataset and Long-range Atmospheric Turbulence Hybrid (LATH) dataset. CTTH contains dynamic objects with GTs, while LATH covers diverse scenes at varying shooting distances. In summary, we make the following primary contributions:

- We break the bottleneck induced by accuracy–efficiency trade-off in frames-based approaches by exploiting finer-grained motion cues from events within a short duration. With only 5 to 8 input frames and events, our method significantly reduces data overhead and system latency, while achieving high-quality restoration.
- We reveal distinct spatiotemporal characteristics of turbulent events via experimental and theoretical analysis—scenes exhibit polarity alternation while dynamic objects form event tubes. We accordingly propose EHETM that leverage polarity-weighted gradients for scene refinement and event-tube priors for motion decoupling, attaining high-quality and efficient TM with few frames.
- We bridge the data gap in TM by constructing the first long-range atmospheric turbulence event-frame dataset shooting at various distances, and the first thermal turbulence event-frame dataset with dynamic objects and GTs. Extensive experiments show that our method outperforms SOTA methods in both restoration quality and efficiency.

2. Related Works

Learning-based Turbulence Mitigation. Recent AI advances have pushed learning-based methods to the forefront of TM, covering both single- and multi-frame paradigms. Single-frame methods, including CNNs [23, 25, 26], GANs [17, 19, 22], physics-informed Transformers [16, 18], diffusion-based methods [21], and approaches exploiting narrowband priors [48] or the seg-then-res pipeline [28], etc., offer simplicity but lack temporal information, limiting restoration performance. Multi-frame methods exploit temporal cues, including video-based GANs [1, 34] and Transformers [37], network-incorporated traditional paradigms [8], the state-space model [41], and neural representations [40, 43, 44], etc., showing strong potential for TM. How-

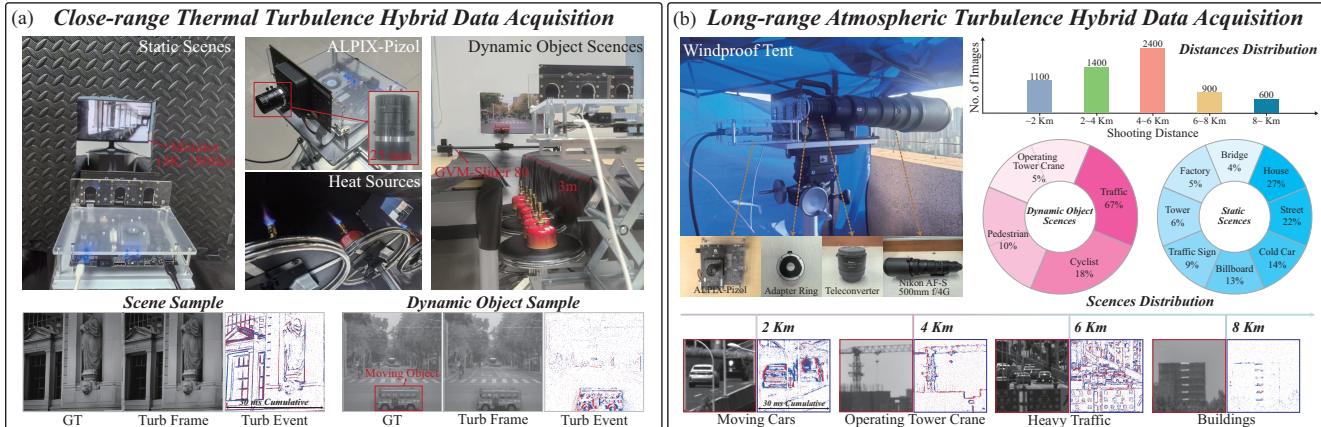


Figure 3. Our real-world event-frame turbulence datasets and acquisition setups. (a) CTTH dataset is captured under adjustable thermal turbulence, containing about 30k image-event pairs with dynamic objects and GTs. (b) LATH dataset is acquired using a 1000 mm lens covering multiple scenes at various shooting distances.

ever, this benefit comes at a cost: every extra frame brings additional system latency and data overhead. Moreover, at low frame rates, severe aliasing between object motion and turbulence are difficult to be accurately decoupled, degrading restoration quality. In contrast, our method overcomes these limitations by leveraging high-speed events to guide few frames for efficient and high-quality restoration.

Event-based Video Restoration. Event cameras can efficiently capture transient motion often missed by conventional cameras [49–53]. In video super-resolution and motion deblurring, events provide precise motion cues, improving edge recovery [54–60]. These results highlight that high-temporal-resolution events reveal motion patterns invisible to conventional cameras, offering new opportunities for TM. Recent works have explored event-based TM: Liu *et al.* [61] model tilt and blur effects in turbulent events, while Li *et al.* [62] are committed to identify “event-lucky regions”. Although insightful, these studies do not exploit events on objects, leaving motion-turbulence entanglement largely unaddressed. We address this gap by leveraging fine-grained temporal information to model object motion.

Turbulence Datasets. Current datasets primarily encompass three categories: physics-based simulation algorithms [1, 8, 30, 37], close-range thermal turbulence [1, 4, 18, 61–63], and long-range atmospheric turbulence [1, 9, 15, 30, 36, 64]. Most existing datasets consist of low-frame-rate image sequences, with only [61, 62] providing event-frame paired data for thermal turbulence. To our best knowledge, there is currently no event-frame paired turbulence dataset that includes dynamic objects with GTs, limiting the validation of event-based methods under dynamic object scenes. Moreover, the lack of real-world hybrid atmospheric turbulence datasets hinders the evaluation of event-based methods across diverse turbulence conditions. Thus, we construct CTTH and LATH datasets to bridge the data gap.

3. Real-World Hybrid Turbulence Datasets

Capturing real-world event-frame paired turbulence data is very valuable, as high-speed events can precisely characterize the spatiotemporal dynamics of turbulence, providing finer-gained motion cues than conventional frames.

CTTH Dataset. The CTTH setup is shown in Fig. 3 (a). Inspired by [1], we use a 4K 180Hz monitor and an ALPIX-pizol hybrid camera with a 25mm lens. The camera’s split-focal-plane design ensures hardware-level synchronization of events and images, eliminating spatiotemporal misalignment. Our base data source comprises over 300 groups of clean images (512×512) from TMT dataset [37], spanning indoor, outdoor, textual, and facial categories, which are super-resolved to 2048×2048 via Real-ESRGAN [65]. During acquisition, the synchronized monitor-camera setup captured 100 frames per scene at 25 fps, producing approximately 30000 turbulent and turbulence-free image-event pairs (512×512). Turbulence is generated via multiple gas sources, with the camera placed 3 meters from the monitor. For dynamic object scenes, the setup incorporates multiple posters, a programmable slider (GVM-Slider), mechanical fixtures, and the ALPIX-pizol camera. Various toy vehicles in controlled experiments across different poster backgrounds. Precise regulation of slider motion parameters ensures spatial consistency, with 200 frames captured per scene at 25 fps, producing approximately 3000 turbulent and turbulence-free image-event pairs (512×512).

LATH Dataset. As shown in Fig. 3 (b), LATH dataset is captured using the ALPIX-pizol hybrid camera configured with a 1000 mm telephoto setup (500 mm lens + $2\times$ teleconverter) at 25 fps across multiple distances and scenes. To ensure stability, all recordings are conducted inside a windproof tent with a professional tripod to eliminate vibrations. Each sequence lasts about 8 seconds, from which

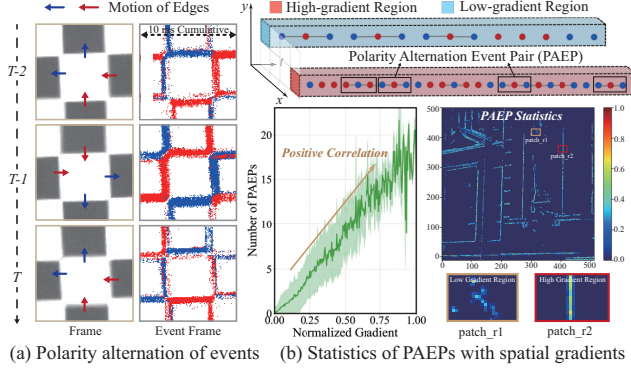


Figure 4. Event responses of scenes caused by turbulence. (a) Turbulence induces event polarity alternation along structural edges. (b) The PAEP numbers correlate with sharp image gradients, providing structural cues for suppressing edge distortions.

the stable 5 seconds segment is retained, yielding roughly 6500 curated image-event pairs at 512×512 resolution. LATH dataset covers shooting distances from 1-8 km and diverse long-range scenes, such as moving cars, operating tower cranes, heavy traffic, and buildings, etc.

4. Turbulent Event Signal Characteristics

4.1. Turb-induced Scene Event Response

In image domain, turbulence mainly distorts gradient-rich regions, causing geometric tilts and edge blurring. In event domain, it remains concentrated in these areas, manifesting as high-frequency polarity alternation induced by local brightness oscillations, as presented in Fig. 4 (a).

To quantify this relationship, we conduct statistical experiments under varying thermal turbulence conditions. The analysis of polarity alternation event pair (PAEP) counts versus sharp image gradients reveals a strong positive correlation across different turbulence levels (mean Pearson’s $r = 0.82$, $p < 0.001$), as presented in the line graph in Fig. 4 (b). This demonstrates that regions with higher gradients experience more intense turbulence-induced brightness fluctuations, resulting in more frequent polarity alternations. Notably, we discover that edges generated by high frequency PAEPs at high-gradient locations tend to be sharper, since these regions are more sensitive to abrupt intensity variations and produce denser responses along structural boundaries, leading to better preservation of fine details. We attribute this phenomenon to the zero-mean property of turbulence, which causes perturbations to predominantly oscillate around their original positions [15].

4.2. Dynamic Object Events: Turb vs. No Turb

From both intuitive and theoretical perspectives, we analyze the response differences between the object and turbulence in events, and reveal how turbulence affects event responses

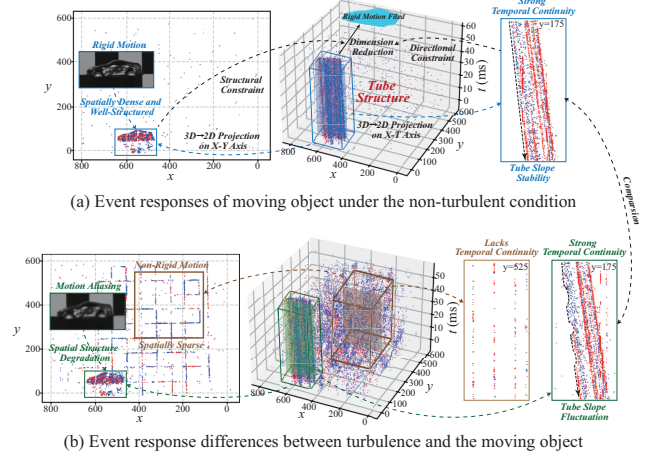


Figure 5. Event responses of the dynamic object under turbulence. (a) Turbulence-free: the object triggers temporally continuous and spatially dense events, forming a stable event tube. (b) Under turbulence: the event tube is structurally degraded, while turbulence produces temporally discontinuous and spatially sparse events.

of the object. Event generation process [66] is described by:

$$\log I(x, y, t) - \log I(x, y, t - \Delta t) = pC, \quad (1)$$

where $\log I(x, y, t)$ is the logarithmic brightness, Δt is the interval, $p \in \{-1, 1\}$ is the polarity, and C is the threshold.

The spatial displacement of a dynamic object generates continuous brightness changes, forming a coherent event stream, as shown in Fig. 5 (a). Without turbulence, object events exhibit strong spatiotemporal consistency, marked by well-structured spatial patterns and sustained temporal continuity. Within a short time window (e.g., 40ms), this forms a spatiotemporal coherent 3D event tube. This phenomenon benefits from the superiority of events to model continuous motion. It also holds under long-range observations, where small angular velocities ensure the object’s geometry and motion state remain nearly unchanged over short periods [67]. Furthermore, we observe that the event tube can be effectively represented by its 2D motion fields, providing a compact and informative description of objects.

We adopt the notion of “particle” from Lagrangian particle tracking [68], where each particle is represented by a 3D coordinate (x, t) . Let the trajectory of a particle located at (x_0, t_0) over a temporal window of length $\Delta > 0$ be given by a function $T_{x_0}^{t_0} : [t_0 - \Delta, t_0 + \Delta] \rightarrow \mathbb{R}^2$ which maps time to its spatial position, e.g., $T_{x_0}^{t_0}(t_0) = x_0$. The trajectory can be written as the point set $\{(T_{x_0}^{t_0}(t), t) : t \in [t_0 - \Delta, t_0 + \Delta]\}$. When Δ is small, the trajectory can be approximated by the first-order Taylor expansion at t_0 :

$$\begin{aligned} T_{x_0}^{t_0}(t) &= T_{x_0}^{t_0}(t_0) + (t - t_0) \cdot T_{x_0}^{t_0 \prime}(t_0) + o(|t - t_0|) \\ &= x_0 + (t - t_0) \cdot T_{x_0}^{t_0 \prime}(t_0) + o(\Delta) \\ &\approx x_0 + (t - t_0) \cdot T_{x_0}^{t_0 \prime}(t_0), \end{aligned} \quad (2)$$

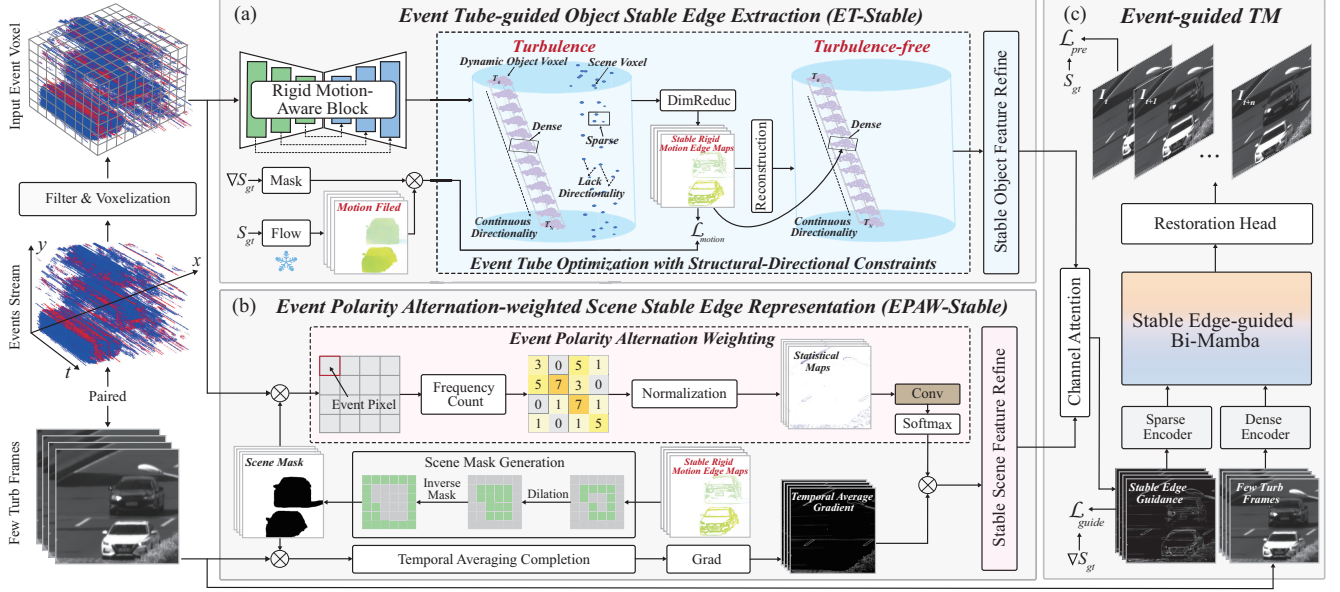


Figure 6. Overall Architecture of the EHETM. (a) ET-Stable stabilizes object motion via the rigid motion-aware block and event tube optimization, leveraging dense spatiotemporal cues to efficiently decouple object motion from turbulence. (b) EPAW-Stable generates masks to isolate turbulence-affected scene regions, and encodes the frequency of event polarity alternation as adaptive weights on temporally averaged gradients to suppress scene distortions. (c) ET-Stable and EPAW-Stable extract stable patterns from turbulence to guide a light-weight Mamba-based video restoration network, enabling high-quality and efficient restoration with few input frames.

where $T_{x_0}^{t_0'}(t_0) \in \mathbb{R}^2$ is the first-order derivative. Hence the trajectory is fully determined by this motion vector.

Under turbulence, as shown in Fig. 5 (b), the spatial structure is distorted. The observed trajectory become

$$\tilde{T}_{x_0}^{t_0}(t) = T_{x_0}^{t_0}(t) + \delta(t), \quad (3)$$

where $\delta : \mathbb{R} \rightarrow \mathbb{R}^2$ is turbulence and cannot be approximated linearly in the given time interval Δ due to the short coherent time of turbulence. Now consider the trajectories of all particles on a uniform $h \times w$ grid at time t_0 :

$$T^{t_0} = \{T_{\mathbf{x}}^{t_0} : \mathbf{x} \in [h] \times [w]\}, \quad (4)$$

$$\tilde{T}^{t_0} = \{T_{\mathbf{x}}^{t_0} + \delta_{\mathbf{x}} : \mathbf{x} \in [h] \times [w]\}, \quad (5)$$

where $[n] = \{1, 2, \dots, n\}$ for any positive integer n , T^{t_0} is the set of clean trajectories and \tilde{T}^{t_0} is the turbulent version.

Since each clean trajectory is linearized by Eq. (2), the entire set T^{t_0} can be compactly represented by a tensor $T^{t_0'}(t_0) \in \mathbb{R}^{h \times w \times 2}$. In contrast, \tilde{T}^{t_0} lacks such a low-dimensional representation. This contrast highlights the intrinsic reducibility of clean 3D tubes, which provides a strong structural prior for recovering turbulence-free trajectories from distorted observations.

5. Event-guided Turbulence Mitigation

5.1. Event Tube-guided Object Edge Extraction

Dynamic objects generate spatiotemporally coherent event tubes, offering dense temporal cues absent in frame-based

methods. Building on this finding, we design the ET-Stable module, which exploits event spatiotemporal continuity to decouple object motion from turbulence and generate stable guidance of objects for TM, as illustrated in Fig. 6 (a).

Rigid Motion-Aware Block. While conceptually intuitive, directly modeling distinctive event tubes of dynamic objects is challenging due to the randomness of turbulence. Consequently, we propose a lightweight rigid motion-aware block (RMAB) integrating 3D residual convolutions (3D-ResConvs) with 3D channel attention units (3D-CAUs) to extract spatiotemporally consistent dense motion features from events. Specifically, multi-scale learning strategy is employed on 3D-ResConvs to capture spatiotemporal patterns across scales, while 3D-CAUs adaptively enhance and decode deep features along spatiotemporal dimensions. Skip connections enable effective multi-scale fusion following classic U-Net [69]. To balance capability and efficiency, we employ only three layers of each 3D-ResConvs and 3D-CAUs for lightweight spatiotemporal feature extraction.

Event Tube Optimization. Dimensionality-reduced event tube optimization (ETO) aims to constrain object motion by exploiting the high-temporal-resolution information of events to project the 3D spatiotemporal volume onto a 2D space, effectively decoupling object motion from turbulence while also improving the computational efficiency. Building upon the dense spatiotemporal cues extracted by the rigid motion-aware block, this projection enforces motion consistency along event tubes. We find that T^{t_0} in Eq. (4)

Table 1. Quantitative comparisons on CTTH, LATH and UDET datasets. The best results are in **bold** and the second-best are underlined.

Methods	Task / Input Type	CTTH / UDET Datasets			LATH Dataset	Efficiency (512 × 512 Patches)				
		PSNR↑	SSIM↑	LPIPS↓	NIQE↓	#Params (M)↓	FPS↑	FLOPs (G)↓	Latency (ms)↓	Data Size (%)↓
TurbNet [18]	TM / Single-frame	28.17 / 24.12	0.8315 / 0.7784	0.2210 / 0.2679	16.15	26.6	5.52	686.7	–	2.5
VRT [70]	VR / Multi-frame	32.55 / 26.58	0.8986 / 0.8139	0.1613 / 0.2334	14.37	18.3	0.31	2509.8	600	40.0
TSRWGAN [1]	TM / Multi-frame	30.19 / 26.72	0.8569 / 0.8027	0.1812 / 0.2358	13.87	46.3	1.33	1280.3	<u>280</u>	37.5
TMT [37]	TM / Multi-frame	32.56 / 27.73	0.9046 / 0.8449	0.1743 / 0.1975	13.09	26.1	0.11	463.9	760	50.0
DATUM [8]	TM / Multi-frame	<u>34.89</u> / 28.85	<u>0.9340</u> / <u>0.8688</u>	0.1448 / 0.1565	<u>10.87</u>	<u>5.8</u>	12.1	191.5	1560	100.0
MambaTM [41]	TM / Multi-frame	34.38 / <u>28.91</u>	0.9219 / 0.8610	<u>0.1290</u> / 0.1678	11.49	6.9	<u>23.5</u>	62.3	760	75.0
Ev-DeblurVSR [71]	Deblur / Hybrid	31.82 / 25.11	0.9188 / 0.7961	0.1897 / 0.2376	15.48	8.3	2.73	448.8	560	68.1
EGVSR [72]	VSR / Hybrid	30.15 / 24.27	0.8838 / 0.7882	0.2189 / 0.2539	14.19	10.2	1.06	273.8	360	33.6
EHETM [Ours]	TM / Hybrid	35.17 / 29.33	0.9425 / 0.8717	0.1281 / 0.1517	10.12	5.6	29.5	<u>181.7</u>	160	<u>22.7</u>

Note. The number of inference frames for each multi-frame method follows the configuration reported in their original paper. FLOPs are computed by averaging the total operations over 100 inference frames. Data Size indicates the total data overhead, including both image and event inputs.

can be represented by the clean optical flow, which can model the ideal object motion. Accordingly, we use a pre-trained RAFT model [73] to derive clean motion fields that serve as references for optimizing consistent event tubes, enabling the model to preserve true object structures while effectively stabilizing the turbulence. Besides, we employ clean gradient masks to clean motion fields, yielding stable edge-aware motion maps. This manner enhances the event voxel alignment and accelerate convergence by reducing redundancy in flat regions. The resulting stable edge motion fields are reconstructed into structured 3D representations and refined through an encoder, yielding stable dynamic object edge guidance with structural-directional constraints.

5.2. Event-weighted Scene Edge Representation

Building on ET-Stable for dynamic object stabilization, we further design the EPAW-Stable to handle turbulence in scenes, as shown in Fig. 6 (b). EPAW-Stable leverages stable object edge maps from ET-Stable as a foundation to generate scene masks, which are applied to both event voxels and frame sequences to isolate areas affected solely by turbulence. Then, we perform temporal averaging completion [28] and gradient computation, resulting in initial gradient maps. Simultaneously, event polarity alternation weighting (EPAW) encodes counts of event polarity alternations as adaptive weights for initial gradient maps, sharpening the edges of scenes, while an encoder further refines features to ensure consistent representations. ET-Stable and EPAW-Stable form a complementary framework that exploits high-speed events to stabilize both dynamic objects and scenes, accelerating TM while enhancing restoration quality.

5.3. Event-guided Video Restoration

Sequence models with linear complexity, such as Mamba [41, 74–76], have shown remarkable efficiency in TM and other vision tasks. Thus, we propose a stable edge-guided Bi-Mamba, as shown in Fig. 6 (c), which, together with few frames, leverages fused outputs from ET-Stable and EPAW-Stable as full-scene edge guidance. A dual-encoder extracts sparse edge and dense texture features, with shallow features processed via local Hilbert scanning [77] be-

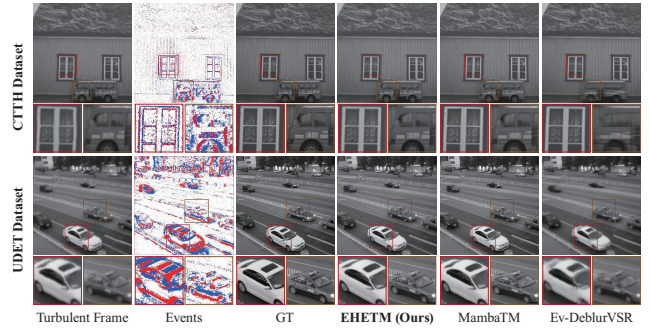


Figure 7. Visual comparisons on CTTH and UDET datasets. Our EHETM yields more visual pleasing results, especially around regions with moving objects.

fore deep feature extraction. Edge priors guide feature evolution, achieving faster and higher-quality restoration. Network details are provided in the supplementary material.

5.4. Loss Functions

The overall loss function of the EHETM is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{motion}} + \mathcal{L}_{\text{guide}} + \mathcal{L}_{\text{pre}}. \quad (6)$$

Motion consistency loss $\mathcal{L}_{\text{motion}} = \mathcal{L}_{\text{char}}(\hat{\mathbf{M}}, \mathbf{M})$ ensures the extraction of stable edge motion fields of objects, where $\hat{\mathbf{M}}$ and \mathbf{M} denote the predicted and RAFT-generated motion fields, respectively. Prior loss $\mathcal{L}_{\text{guide}} = \mathcal{L}_{\text{char}}(\hat{\mathbf{G}}, \mathbf{G})$ supervises the ET-Stable and EPAW-Stable to generate stable priors, where $\hat{\mathbf{G}}$ and \mathbf{G} represent the predicted and clean gradients. Restoration loss $\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{char}}(\hat{\mathbf{J}}, \mathbf{J}) + \alpha_p \mathcal{L}_p(\hat{\mathbf{J}}, \mathbf{J})$ ensures the quality of outputs, where $\hat{\mathbf{J}}$ and \mathbf{J} denote restoration results and GTs. $\mathcal{L}_{\text{char}}$ represents the Charbonnier loss [78] and \mathcal{L}_p represents the perceptual loss [79], and the weight coefficient α_p is set to 0.3.

6. Experiments

6.1. Experimental Settings

Training Settings. Our network is implemented in PyTorch [80] and optimized with Adam [81] using an initial learning rate of 2×10^{-4} , scheduled by cosine annealing. Training is

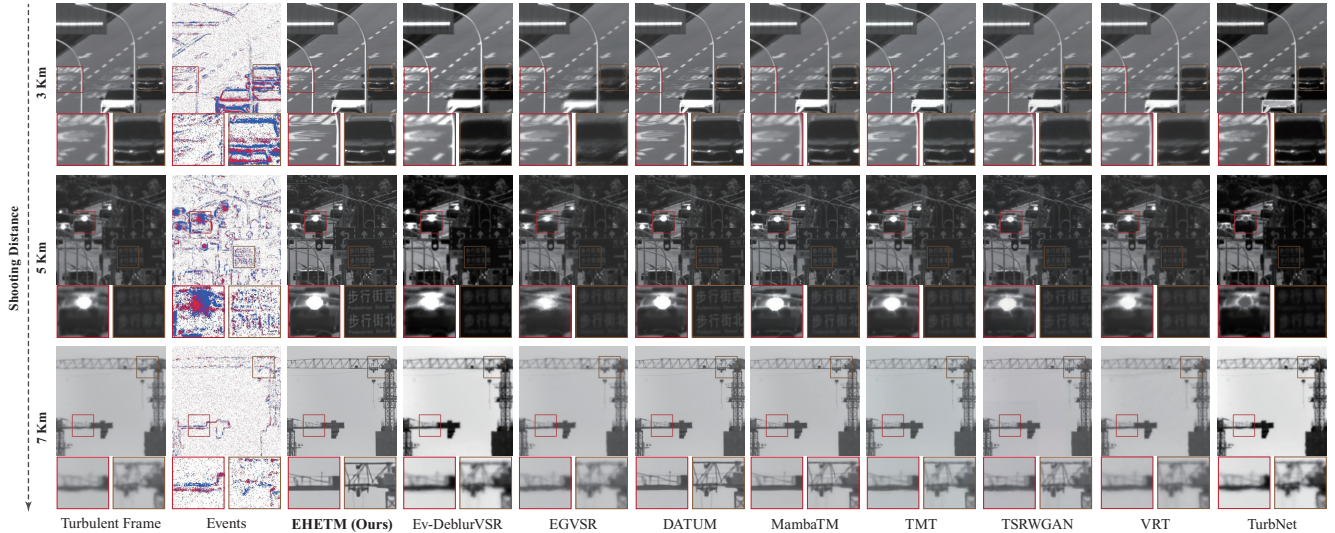


Figure 8. Visual comparisons on LATH dataset. EHETM produces more satisfactory visual results and achieves outstanding generalization under different real-world scenes. The number of inference frames from EHETM to VRT are as follows: 8, 15, 10, 40, 20, 20, 15, 16.

conducted on 8 NVIDIA RTX 3090 GPUs, and the detailed training scheme is provided in the supplementary material.

Comparison Methods. We select eight representative competing methods, including TurbNet [18], VRT [70], TSRWGAN [1], TMT [37], DATUM [8], MambaTM [41], Ev-DeblurVSR [71], and EGVSR [72]. The last two are hybrid-based deblurring and video super-resolution (VSR) methods, while the others are frame-based TM or video restoration (VR) approaches. For the fair comparison, all methods use their released pretrained weights and are additionally re-trained on our CTTH dataset.

Datasets. To comprehensively evaluate our method, we further adopt the UA-DETRAC dataset [82] in addition to our CTTH and LATH datasets. Following [8], different levels of turbulence are simulated, and the corresponding events are synthesized with the v2e simulator [83], yielding the UA-DETRAC-EvTurb (UDET) dataset.

Evaluation Metrics. For CTTH and UDET datasets, we evaluate the restoration performance using PSNR [84], SSIM [85], and LPIPS [79], and report efficiency including #Params, FPS, data size, FLOPs and the system latency. For LATH dataset, which lacks GTs, we adopt natural image quality evaluator (NIQE) [86] to assess visual quality of restored images in a no-reference manner.

6.2. Quantitative Evaluation

Table 1 demonstrates that all multi-frame methods outperform the single-frame one, highlighting the merits of temporal cues in mitigating turbulence distortions. Notably, guided by high-temporal-resolution events, our EHETM achieves the best restoration (+0.38 PSNR, +0.0057 SSIM, -0.0108 LPIPS, -0.75 NIQE vs. the second best: DATUM [8]) by suppressing turbulence-induced distortions and sta-

Table 2. Ablation studies of EHETM

RMAB	ETO	EPAW	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
✓			31.13	0.8962	0.2261
	✓		33.85	0.9241	0.1736
✓	✓		34.19	0.9309	0.1563
✓	✓	✓	35.17	0.9425	0.1281

bilizing object edges. Furthermore, with only 5.6 M parameters, 181.7 FLOPs and 5 input frames with corresponding events (22.7% data size), EHETM attains the highest 29.5 FPS and lowest 160 ms system latency, highlighting its efficiency and practical applicability.

6.3. Qualitative Evaluation

Evaluation on CTTH and UDET Datasets. Fig. 7 depicts a visual comparison between EHETM and two typical methods. EHETM produces sharper edges and more coherent structures, particularly around dynamic objects. In contrast, MambaTM [41] and Ev-DeblurVSR [71] exhibit blurry edges and residual distortions. These results highlight that using temporal cues from high-speed events is crucial to TM, leading to higher-quality restoration.

Evaluation on LATH Dataset. We compare EHETM with eight typical methods for generalization evaluation, as shown in Fig. 8. TurbNet [18] struggles with distortions as it lacks of temporal modeling for TM. VRT [70], TSRWGAN [1], and TMT [37] partially reduce turbulence effects, but some blur and noise remain unsolved. Ev-DeblurVSR [71] alleviates blur at the cost of degraded contrast, and EGVSR [72] introduces unexpected artifacts. While DATUM [8] and MambaTM [41] perform reasonably well, severe aliasing between object motion and turbulence still causes deviations. In contrast, our EHETM consistently produces supe-

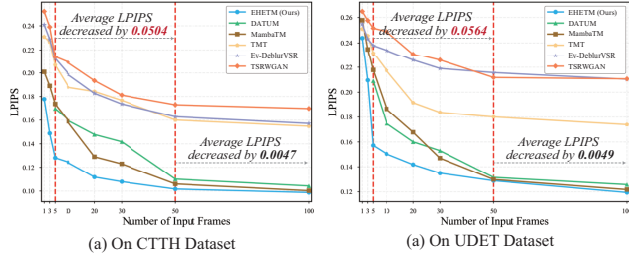


Figure 9. Impact of the inference data size. With only few frames and events, our method demonstrates comparable or even higher quality restoration compared to other multi-frame methods.

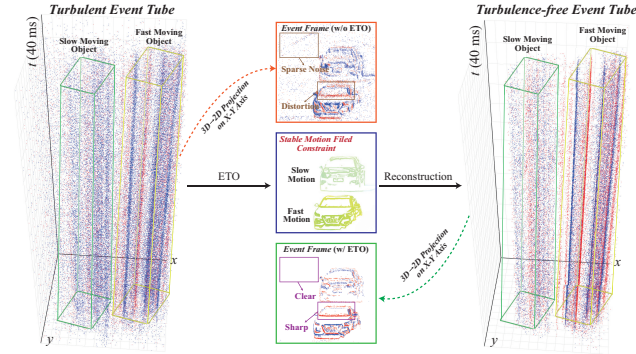


Figure 10. Effectiveness of ETO. It extracts clean event tubes triggered by objects with various motions from turbulent events.

rior results across different scenes, suppressing turbulence distortions while stabilizing dynamic objects.

6.4. Ablation Study and Discussion

Impact of the Inference Data Size. We investigate the impact of inference data size on TM, as shown in Fig. 9. With very limited inputs (1 to 5 frames), EHETM excels by using the high-speed events to capture the zero-mean characteristics of turbulence. As data size grows (5 to 50 frames), all methods improve substantially (average LPIPS reduction > 0.5), yet EHETM retains a clear advantage due to finer-grained temporal cues. For larger data size (50 to 100 frames), gains plateau (average LPIPS reduction < 0.05) while computational costs rise sharply, highlighting the importance of efficiently mining temporal information for TM. **Effectiveness of Event Tube Optimization.** As shown in Fig. 10, using 40 ms accumulated event voxels, we find that w/o ETO, turbulence-corrupted event tubes produce sparse noise and distorted edges when projected to the event frame. By applying 3D-to-2D constraint, ETO suppresses these distortions while maintaining sharp edges and scene clarity, yielding spatiotemporally consistent event tubes. It is noteworthy that ETO effectively constrains objects even when their motion speeds differ. Table 2 further confirms the effectiveness of ETO, and when integrated with RMAB, it enables effective object edge stabilization and significantly contributes to high-quality restoration.

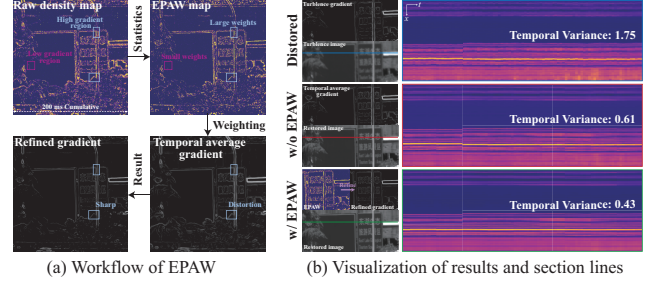


Figure 11. Effectiveness of EPAW. (a) How does EPAW work. (b) Visualization of selected section lines and the variance computed along the t axis.

Effectiveness of Event Polarity Alternation Weighting.

EPAW refines distorted edges within the temporal average gradient to guide accurate scene restoration. In Fig. 11 (a), we illustrate the workflow of EPAW. By computing the EPAW map, sharp structures in high-gradient regions are assigned greater weights, thereby optimizing distorted gradients and producing sharper edge representations. As shown by the selected section lines in Fig. 11 (b), w/ EPAW produces smoother results, and the temporal variance along these lines quantitatively verifies the improvement in stability. Moreover, Table 2 verifies the improvement in overall restoration brought by EPAW. In summary, w/ EPAW enables the model to learn sharp edges during end-to-end training, resulting in higher-quality scene restoration.

Limitation and Future Work. EHETM exhibits high efficiency but faces challenges in ultra-long-range scenes (e.g. > 30 km) and under severe platform motion (e.g. fast object tracking). Ultra-long-range atmospheric sparsifies event responses, lacking information for reliable restoration. Severe platform motion entangles turbulence, scenes and objects, making motion disentanglement extremely challenging. In future, we plan to introduce inertial measurement unit and ultra-telephoto optics to overcome these problems.

7. Conclusion

In this work, we present an event-guided high-quality and efficient TM method. By mining fine-grained motion cues from events within a short duration, we overcome the accuracy–efficiency bottleneck in frames-based methods. With experimental and theoretical analysis, we reveal distinct spatiotemporal characteristics of turbulent events—scenes exhibit polarity alternation while objects form event tubes. Based on these insights, we leverage polarity-weighted gradients for scene refinement and event-tube priors for motion decoupling. Moreover, for comprehensive demonstration, we construct two real-world event-frame paired turbulence datasets: CTTH and LATH. Our method achieves superior restoration quality and efficiency over SOTAs, and we believe it can bring new perspectives to the community.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant U24B20139 and 62371203, the Open Fund of the State Key Laboratory of Dynamic Optical Imaging and Measurement E41302A1L1, the Hubei Provincial Natural Science Foundation under Grant JCZRQT202600099.

References

- [1] Darui Jin, Ying Chen, Yi Lu, Junzhang Chen, Peng Wang, Zichao Liu, Sheng Guo, and Xiangzhi Bai. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. *Nature Machine Intelligence*, 3(10):876–884, 2021. 1, 2, 3, 6, 7
- [2] Yuduo Guo, Yuhan Hao, Sen Wan, Hao Zhang, Laiyu Zhu, Yi Zhang, Jiamin Wu, Qionghai Dai, and Lu Fang. Direct observation of atmospheric turbulence with a video-rate wide-field wavefront sensor. *Nature Photonics*, 18(9):935–943, 2024.
- [3] Paul Hill, Nantheera Anantrasirichai, Alin Achim, and David Bull. Deep learning techniques for atmospheric turbulence removal: a review. *Artificial Intelligence Review*, 58(4):101, 2025.
- [4] Nantheera Anantrasirichai, Alin Achim, Nick G Kingsbury, and David R Bull. Atmospheric turbulence mitigation using complex wavelet-based fusion. *IEEE Transactions on Image Processing*, 22(6):2398–2408, 2013. 1, 3
- [5] Stanley H Chan. Tilt-then-blur or blur-then-tilt? clarifying the atmospheric turbulence model. *IEEE Signal Processing Letters*, 29:1833–1837, 2022. 1
- [6] David Cornett, Joel Brogan, Nell Barber, Deniz Aykac, Seth Baird, Nicholas Burchfield, Carl Dukes, Andrew Duncan, Regina Ferrell, Jim Goddard, et al. Expanding accurate person recognition to new altitudes and ranges: The briar dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 593–602, 2023. 1
- [7] Xingguang Zhang and Chih-Hsien Chou. Source-free domain adaptation for video object detection under adverse image conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5010–5019, 2024. 1
- [8] Xingguang Zhang, Nicholas Chimitt, Yiheng Chi, Zhiyuan Mao, and Stanley H Chan. Spatio-temporal turbulence mitigation: A translational perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2889–2899, 2024. 1, 2, 3, 6, 7
- [9] Michael Hirsch, Suvrit Sra, Bernhard Schölkopf, and Stefan Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 607–614. IEEE, 2010. 1, 3
- [10] Akshay Patel, Dippal Israni, Nerella Arun Mani Kumar, and Chintan Bhatt. An adaptive image registration technique to remove atmospheric turbulence. *Statistics, Optimization & Information Computing*, 7(2):439–446, 2019.
- [11] Tufan Caliskan and Nafiz Arica. Atmospheric turbulence mitigation using optical flow. In *2014 22nd International Conference on Pattern Recognition*, pages 883–888. Ieee, 2014.
- [12] Hamidreza Fazlali, Shahram Shirani, Michael Bradford, and Thia Kirubarajan. Atmospheric turbulence removal in long-range imaging using a data-driven-based approach. *International Journal of Computer Vision*, 130(4):1031–1049, 2022. 1
- [13] Xiang Zhu and Peyman Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):157–170, 2012.
- [14] Dehao Qin, Ripon Kumar Saha, Woojeh Chung, Suren Jayasuriya, Jinwei Ye, and Nianyi Li. Unsupervised moving object segmentation with atmospheric turbulence. In *European Conference on Computer Vision*, pages 18–37. Springer, 2024.
- [15] Shengqi Xu, Run Sun, Yi Chang, Shuning Cao, Xueyao Xiao, and Luxin Yan. Long-range turbulence mitigation: a large-scale dataset and a coarse-to-fine framework. In *European Conference on Computer Vision*, pages 311–329. Springer, 2024. 1, 3, 4
- [16] Ajay Jaiswal, Xingguang Zhang, Stanley H Chan, and Zhangyang Wang. Physics-driven turbulence image restoration with stochastic refinement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12170–12181, 2023. 1, 2
- [17] Chun Pong Lau, Carlos D Castillo, and Rama Chellappa. Atfacegan: Single face semantic aware image restoration and recognition from atmospheric turbulence. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2):240–251, 2021. 2
- [18] Zhiyuan Mao, Ajay Jaiswal, Zhangyang Wang, and Stanley H Chan. Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. In *European conference on computer vision*, pages 430–446. Springer, 2022. 2, 3, 6, 7
- [19] Kangfu Mei and Vishal M Patel. Ltt-gan: Looking through turbulence by inverting gans. *IEEE Journal of Selected Topics in Signal Processing*, 17(3):587–598, 2023. 2
- [20] Nithin Gopalakrishnan Nair and Vishal M Patel. Confidence guided network for atmospheric turbulence mitigation. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1359–1363. IEEE, 2021.
- [21] Nithin Gopalakrishnan Nair, Kangfu Mei, and Vishal M Patel. At-ddpm: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3434–3443, 2023. 2
- [22] Shyam Nandan Rai and CV Jawahar. Removing atmospheric turbulence via deep adversarial learning. *IEEE Transactions on Image Processing*, 31:2633–2646, 2022. 2
- [23] Rajeev Yasarla and Vishal M Patel. Cnn-based restoration of a single face image degraded by atmospheric turbulence. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2):222–233, 2022. 2
- [24] Chun Pong Lau, Amit Kumar, and Rama Chellappa. Semi-supervised landmark-guided restoration of atmospheric tur-

- bulent images. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):204–215, 2021.
- [25] Gongping Chen, Zhisheng Gao, Qiaolu Wang, and Qingqing Luo. U-net like deep autoencoders for deblurring atmospheric turbulence. *Journal of Electronic Imaging*, 28(5):053024–053024, 2019. 2
- [26] Rajeev Yasarla and Vishal M Patel. Learning to restore images degraded by atmospheric turbulence using uncertainty. In *2021 IEEE international conference on image processing (ICIP)*, pages 1694–1698. IEEE, 2021. 2
- [27] Yubo Wu, Kuanhong Cheng, Ting Cao, Dong Zhao, and Junhui Li. Semi-supervised correction model for turbulence-distorted images. *Optics Express*, 32(12):21160–21174, 2024.
- [28] Ripon Kumar Saha, Dehao Qin, Nianyi Li, Jinwei Ye, and Suren Jayasuriya. Turb-seg-res: A segment-then-restore pipeline for dynamic videos with atmospheric turbulence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25286–25296, 2024. 1, 2, 6
- [29] Nantheera Anantrasirichai. Atmospheric turbulence removal with complex-valued convolutional neural network. *Pattern Recognition Letters*, 171:69–75, 2023. 1
- [30] Yadong Wang, Darui Jin, Junzhang Chen, and Xiangzhi Bai. Revelation of hidden 2d atmospheric turbulence strength fields from turbulence effects in infrared imaging. *Nature Computational Science*, 3(8):687–699, 2023. 3
- [31] Jing Gao, Nantheera Anantrasirichai, and David Bull. Atmospheric turbulence removal using convolutional neural network. *arXiv preprint arXiv:1912.11350*, 2019.
- [32] Wai Ho Chak, Chun Pong Lau, and Lok Ming Lui. Subsampled turbulence removal network. *arXiv preprint arXiv:1807.04418*, 2018.
- [33] Yifei Xia, Chu Zhou, Chengxuan Zhu, Chao Xu, and Boxin Shi. Planet: Learning to mitigate atmospheric turbulence in planetary images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 8584–8592, 2025.
- [34] Brandon Y Feng, Mingyang Xie, and Christopher A Metzler. Turbugan: An adversarial learning approach to spatially-varying multiframe blind deconvolution with applications to imaging through turbulence. *IEEE Journal on Selected Areas in Information Theory*, 3(3):543–556, 2023. 2
- [35] Xiangqing Liu, Gang Li, Zhenyang Zhao, Qi Cao, Zijun Zhang, Shaoan Yan, Jianbin Xie, and Minghua Tang. Eaf-wgan: Enhanced alignment fusion-wasserstein generative adversarial network for turbulent image restoration. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10):5605–5616, 2023.
- [36] Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Image reconstruction of static and dynamic scenes through anisoplanatic turbulence. *IEEE Transactions on Computational Imaging*, 6:1415–1428, 2020. 3
- [37] Xingguang Zhang, Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Imaging through the atmosphere using turbulence mitigation transformer. *IEEE Transactions on Computational Imaging*, 10:115–128, 2024. 2, 3, 6, 7
- [38] Zhicheng Zou and Nantheera Anantrasirichai. Deturb: atmospheric turbulence mitigation with deformable 3d convolutions and 3d swin transformers. In *Proceedings of the Asian Conference on Computer Vision*, pages 904–921, 2024.
- [39] Xijun Wang, Santiago López-Tapia, and Aggelos K Katsaggelos. Real-world atmospheric turbulence correction via domain adaptation. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 1466–1472. IEEE, 2024.
- [40] Haoming Cai, Jingxi Chen, Brandon Feng, Weiyun Jiang, Mingyang Xie, Kevin Zhang, Cornelia Fermuller, Yiannis Aloimonos, Ashok Veeraraghavan, and Chris Metzler. Temporally consistent atmospheric turbulence mitigation with neural representations. *Advances in Neural Information Processing Systems*, 37:44554–44574, 2024. 2
- [41] Xingguang Zhang, Nicholas Chimitt, Xijun Wang, Yu Yuan, and Stanley H Chan. Learning phase distortion with selective state space models for video turbulence mitigation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2127–2138, 2025. 2, 6, 7
- [42] Ziran Zhang, Yuhang Tang, Zhigang Wang, Yueting Chen, and Bin Zhao. Physical prior guided cooperative learning framework for joint turbulence degradation estimation and infrared video restoration. *arXiv preprint arXiv:2408.04227*, 2024.
- [43] Nianyi Li, Simron Thapa, Cameron Whyte, Albert W Reed, Suren Jayasuriya, and Jinwei Ye. Unsupervised non-rigid image distortion removal via grid deformation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2522–2532, 2021. 2
- [44] Weiyun Jiang, Vivek Boominathan, and Ashok Veeraraghavan. Nert: Implicit neural representations for unsupervised atmospheric turbulence mitigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4236–4243, 2023. 1, 2
- [45] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conrath, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 2
- [46] Haoyue Liu, Shihan Peng, Lin Zhu, Yi Chang, Hanyu Zhou, and Luxin Yan. Seeing motion at nighttime with an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25648–25658, 2024.
- [47] Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, and Boxin Shi. Eventps: Real-time photometric stereo using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9602–9611, 2024. 2
- [48] Yifei Xia, Chu Zhou, Chengxuan Zhu, Minggui Teng, Chao Xu, and Boxin Shi. Nb-gtr: narrow-band guided turbulence removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24934–24943, 2024. 2
- [49] Haoyue Liu, Jinghan Xu, Yi Chang, Hanyu Zhou, Haozhi Zhao, Lin Wang, and Luxin Yan. Timetracker: Event-based continuous point tracking for video frame interpolation with

- non-linear motion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17649–17659, 2025. 3
- [50] Xinan Xie, Qing Zhang, and Wei-Shi Zheng. Diffusion-based event generation for high-quality image deblurring. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2194–2203, 2025.
- [51] Zeyu Xiao and Xinchao Wang. Event-based video super-resolution via state space models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12564–12574, 2025.
- [52] Yuxing Duan. Led: A large-scale real-world paired dataset for event camera denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25637–25647, 2024.
- [53] Hanyu Zhou, Yi Chang, and Zhiwei Shi. Bring event into rgb and lidar: Hierarchical visual-motion fusion for scene flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26477–26486, 2024. 3
- [54] Zeyu Xiao, Dachun Kai, Yueyi Zhang, Zheng-Jun Zha, Xiaoyan Sun, and Zhiwei Xiong. Event-adapted video super-resolution. In *European Conference on Computer Vision*, pages 217–235. Springer, 2024. 3
- [55] Yongcheng Jing, Yiding Yang, Xinchao Wang, Mingli Song, and Dacheng Tao. Turning frequency to resolution: Video super-resolution via event cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7772–7781, 2021.
- [56] Dachun Kai, Yueyi Zhang, and Xiaoyan Sun. Video super-resolution via event-driven temporal alignment. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 2950–2954. IEEE, 2023.
- [57] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6820–6829, 2019.
- [58] Wei Yu, Jianing Li, Shengping Zhang, and Xiangyang Ji. Learning scale-aware spatio-temporal implicit representation for event-based motion deblurring. In *Forty-first International Conference on Machine Learning*, 2024.
- [59] Xiang Zhang, Lei Yu, Wen Yang, Jianzhuang Liu, and Gui-Song Xia. Generalizing event-based motion deblurring in real-world scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10734–10744, 2023.
- [60] Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Cmta: Cross-modal temporal alignment for event-guided video deblurring. In *European Conference on Computer Vision*, pages 1–19. Springer, 2024. 3
- [61] Yixing Liu, Minggui Teng, Yifei Xia, Peiqi Duan, and Boxin Shi. Evturb: Event camera guided turbulence removal. *arXiv preprint arXiv:2508.10582*, 2025. 3
- [62] Huanan Li, Rui Fan, Juntao Guan, Weidong Hao, Lai Rui, Tong Wu, Yikai Wang, and Lin Gu. Egtm: Event-guided efficient turbulence mitigation. *arXiv preprint arXiv:2509.03808*, 2025. 3
- [63] UG2+ challenge at CVPR 2023: Bridging the gap between computational photography and visual recognition. http://cvpr2023.ug2challenge.org/dataset23_t2.html, 2023. 3
- [64] Jérôme Gilles and Nicholas B Ferrante. Open turbulent image set (otis). *Pattern Recognition Letters*, 86:38–41, 2017. 3
- [65] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*. 3
- [66] Federico Paredes-Vallés and Guido CHE De Croon. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3446–3455, 2021. 4
- [67] Guillermo Gallego and Davide Scaramuzza. Accurate angular velocity estimation with an event camera. *IEEE Robotics and Automation Letters*, 2(2):632–639, 2017. 4
- [68] Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–6. IEEE, 2007. 4
- [69] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 5
- [70] Jingyun Liang, Jiezhong Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *IEEE Transactions on Image Processing*, 33:2171–2182, 2024. 6, 7
- [71] Dachun Kai, Yueyi Zhang, Jin Wang, Zeyu Xiao, Zhiwei Xiong, and Xiaoyan Sun. Event-enhanced blurry video super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 4175–4183, 2025. 6, 7
- [72] Yunfan Lu, Zipeng Wang, Minjie Liu, Hongjian Wang, and Lin Wang. Learning spatial-temporal implicit neural representations for event-guided video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1557–1567, 2023. 6, 7
- [73] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision*, pages 402–419. Springer, 2020. 6
- [74] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. 6
- [75] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pages 222–241. Springer, 2024.
- [76] Lianghai Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024. 6
- [77] Hongtao Wu, Yijun Yang, Huihui Xu, Weiming Wang, Jinni Zhou, and Lei Zhu. Rainmamba: Enhanced locality learning

- with state space models for video deraining. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 7881–7890, 2024. [6](#)
- [78] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on image processing*, 6(2):298–311, 1997. [6](#)
- [79] 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 conference on computer vision and pattern recognition*, pages 586–595, 2018. [6](#), [7](#)
- [80] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. [6](#)
- [81] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [6](#)
- [82] Longyin Wen, Dawei Du, Zhaowei Cai, Zhen Lei, Ming-Ching Chang, Honggang Qi, Jongwoo Lim, Ming-Hsuan Yang, and Siwei Lyu. Ua-detrac: A new benchmark and protocol for multi-object detection and tracking. *Computer Vision and Image Understanding*, 193:102907, 2020. [7](#)
- [83] Yuhuang Hu, Shih-Chii Liu, and Tobi Delbruck. v2e: From video frames to realistic dvs events. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1312–1321, 2021. [7](#)
- [84] Zhou Wang and Alan Conrad Bovik. Modern image quality assessment. 2006. [7](#)
- [85] 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. [7](#)
- [86] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. [7](#)