A. Details of the Proposed ET-Turb Dataset

A.1. Visualization of the Turbulence Degradation Process

Our data synthesis method decomposes atmospheric turbulence degradation into two components: tilt and blur [10, 11, 25, 27, 45]. This decomposition enables quantitative analysis and provides diverse training data with controllable degradation parameters for mitigation algorithms. In this section, we visualize the synthesis pipeline and demonstrate how configuration parameters affect image quality.

Tilt Component: Fig. 10b illustrates tilt effects from low-order wavefront aberrations, causing statistically correlated pixel displacements that vary non-uniformly across the image. The color-coded displacement field indicates the magnitude of pixel displacements.

Blur Component: Fig. 10c shows the exposure-dependent blur applied to the tilt image. The blur intensity depends on exposure-time τ, with longer exposures accumulating more turbulence. By integrating our ET-MTF formulation with spatially varying blur kernels, we simulate realistic, temporally and spatially varying turbulence blur.

Complete Turbulence Synthesis: Fig. 10d presents the final turbulence-degraded images from the combined tilt and blur components, providing high-quality training data for deep learning-based mitigation algorithms.

The synthesis pipeline is demonstrated across six distinct parameter configurations, as shown in Tab. 8, highlighting the flexibility of our approach in generating a wide range of turbulence conditions. All parameters are selected randomly, and the complete synthesis process is presented, further demonstrating the broad applicability of our method in simulating real-world scenarios.

A.2. Generation of Turbulent Images under Varying Intensities and Exposure-Time

To further demonstrate the flexibility and physical fidelity of our data synthesis framework, Fig. 11 shows turbulence-degraded images of a resolution target under varying turbulence intensities and exposure-time. As both exposure-time and turbulence strength increase, the degree of blur becomes progressively more pronounced, reflecting the cumulative effect of turbulence states. For this visualization, we fix the distance to 500 m, the focal length to 300 mm, the F-number to 8, the height to 50 m, and the wind speed to 5 m/s, corresponding to a practically relevant ground-to-ground imaging scenario commonly encountered in long-range surveillance and outdoor observation. These results illustrate that our pipeline reproduces the expected interplay between exposure duration, turbulence strength, and image degradation across a broad range of operating conditions, providing qualitative evidence for the physical plausibility and robustness of our turbulence synthesis approach.

A.3. Analysis of ET-Turb-Real

ET-Turb-Real is a real-world turbulence dataset, compiled from publicly available atmospheric turbulence datasets [2, 23, 50]. This dataset is used to evaluate the cross-domain generalization of models trained on synthetic data, providing diverse real-world turbulence-degraded video sequences captured under various environmental and optical conditions.

The dataset spans a wide range of scene categories recorded in different settings, including urban (streets, buildings), recreational (sports fields), industrial (factories), transportation (vehicles), and commercial (billboards). Data acquisition was performed using professional imaging systems such as the Canon EOS-1D Mark IV, Nikon D750, and Nikon Coolpix P1000. These datasets were collected with consistent camera parameters, and we consolidated them to maximize scene diversity while maintaining uniform imaging conditions, ensuring a comprehensive representation of real-world turbulence.

Fig. 7 shows the dataset composition with stacked bar charts, illustrating the distribution of camera models and scene categories. These charts visualize the proportion of different scene types captured by various imaging devices, ensuring the dataset covers turbulence effects across different scales and environments, essential for evaluating the robustness and generalization of turbulence mitigation models.

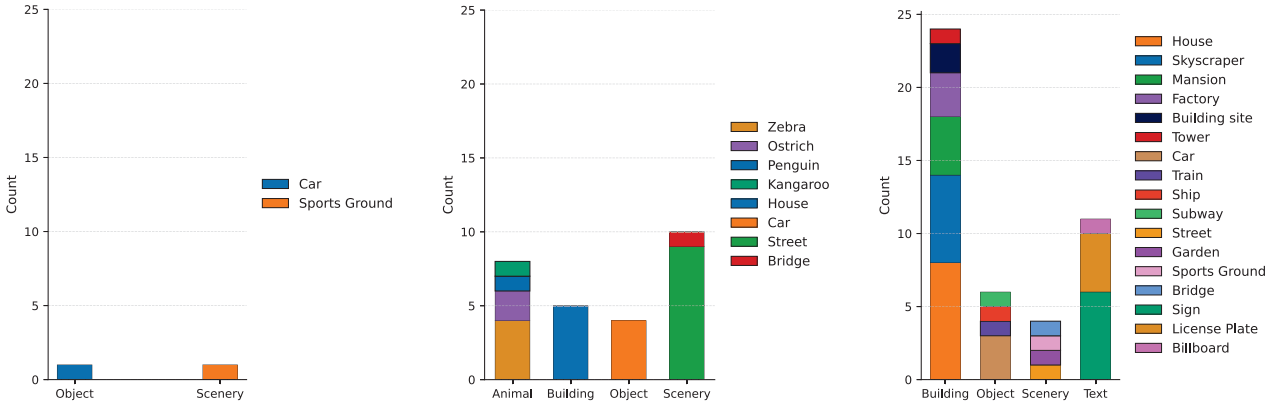
The combination of diverse scene types and imaging conditions enhances the dataset's representativeness of real-world turbulence scenarios, thereby establishing a realistic and diverse evaluation set for developing and evaluating atmospheric turbulence mitigation algorithms.

B. Supplementary Experiments

B.1. Real-World Turbulence Mitigation Performance

To evaluate the practical applicability of our ET-Turb dataset, we evaluate state-of-the-art turbulence mitigation algorithms trained on our synthetic data when applied to real-world turbulent imagery. Fig. 12 presents qualitative results from four representative methods: TSR-WGAN [23], TMT [54], DATUM [53], and MambaTM [55].

The results demonstrate that models trained exclusively on ET-Turb successfully generalize to real atmospheric turbulence conditions. All evaluated algorithms achieve meaningful restoration of details, such as text on road signs and vehicle license plates, which are typically severely degraded by turbulence effects. Notably, while TSR-WGAN [23] produces relatively realistic detail recovery, it exhibits color shift artifacts compared to other methods. TMT [54], DATUM [53], and MambaTM [55] maintain better color fidelity while providing competitive detail restoration.



(a) Captured by Canon EOS-1D Mark IV [2] (b) Captured by Nikon D750 [23] (c) Captured by Nikon Coolpix P1000 [50]

Figure 7. ET-Turb-Real dataset composition: multi-camera systems and diverse scene distribution. Real-world turbulence data is collected across three different imaging devices, providing comprehensive coverage of diverse environmental and anthropogenic scenes such as urban landscapes, entertainment venues, industrial infrastructure, traffic scenarios, and commercial establishments.

Table 4. Quantitative evaluation on the ET-Turb test set. TMT [54] achieves the best performance across all metrics, highlighting the effectiveness of its component-wise training strategy.

Method	PSNR(dB) ↑	SSIM ↑	LPIPS ↓
TSR-WGAN [23]	18.90	0.5927	0.5608
TMT [54]	**24.97**	**0.7507**	**0.3807**
DATUM [53]	22.85	0.6436	0.4700
MambaTM [55]	23.77	0.6658	0.4490

B.2. Qualitative and Quantitative Evaluation on ET-Turb Test Set

Tab. 4 presents a comprehensive quantitative comparison of four algorithms evaluated on the ET-Turb test set. Performance is assessed using three complementary metrics: PSNR for pixel-level accuracy, SSIM for structural similarity, and LPIPS [52] for perceptual quality assessment.

Among the algorithms tested, TMT [54] achieves the best overall performance across all metrics (PSNR: 24.97 dB, SSIM: 0.7507, LPIPS: 0.3807), due to its approach of separately training the tilt and blur components. MambaTM [55] follows closely with balanced results, then followed by DATUM [53]. TSR-WGAN [23] shows a notable performance gap compared with other methods.

In addition to the quantitative comparison, qualitative assessments, shown in Fig. 13, corroborate the numerical results, with TMT producing the most visually accurate images. However, TSR-WGAN exhibits visible artifacts.

These results validate the effectiveness of the ET-Turb dataset for training turbulence mitigation algorithms and suggest that existing methods can be improved, particularly with more realistic turbulence datasets.

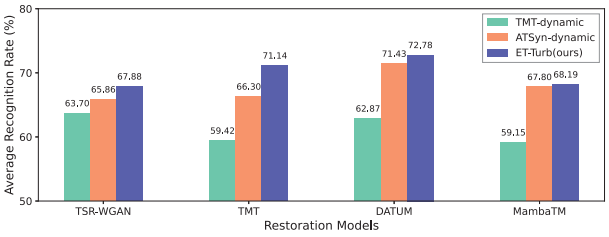


Figure 8. Text recognition accuracy on the turbulence-text dataset [54], evaluated using ASTER [44].

B.3. Text Recognition Evaluation

To evaluate the downstream utility of the synthesized turbulence data, we conduct a text recognition experiment on the Turbulence-Text dataset [54] using ASTER [44] as the recognition backbone and measure accuracy on restored frames.

As shown in Fig. 8, models trained on ET-Turb achieve superior recognition performance compared to those trained on TMT-dynamic [54] and ATSyn-dynamic [53]. In particular, ET-Turb improves recognition accuracy by 7.21% over TMT-dynamic [54] and 1.15% over ATSyn-dynamic [53], indicating better downstream generalization.

B.4. Exposure-Time Decoupling and Brightness Statistics

In practical imaging systems, exposure-time τ, brightness, and sensor noise are physically coupled. However, the primary objective of this work is turbulence synthesis under continuous exposure modeling, rather than full radiometric simulation. In real-world surveillance scenarios, cameras typically operate with auto-exposure mechanisms, where τ does not uniquely determine image brightness due to ISO

Table 5. Robustness evaluation under exposure-dependent Gaussian noise added to the ET-Turb test set. Models trained on ET-Turb and ATSyn-Dynamic [53] are compared.

	Trained on ET-Turb			Trained on ATSyn-Dynamic [53]		
Method	PSNR(dB)↑	SSIM↑	LPIPS↓	PSNR(dB)↑	SSIM↑	LPIPS↓
Noise-input	20.57 (↓1.36%)	0.58 (↓9.38%)	0.57 (↑16.33%)	20.57 (↓1.36%)	0.58 (↓9.38%)	0.57 (↑16.33%)
TMT [54]	21.22 (↓0.38%)	0.66 (↓1.49%)	0.51 (↑2.21%)	21.19 (↓0.42%)	0.66 (↓2.94%)	0.51 (↑2.37%)
DATUM [53]	22.43 (↓0.80%)	0.61 (↓4.69%)	0.51 (↑2.89%)	22.14 (↓1.07%)	0.62 (↓7.46%)	0.53 (↑8.16%)
MambaTM [55]	22.12 (↓1.07%)	0.64 (↓1.54%)	0.52 (↑4.49%)	22.10 (↓1.34%)	0.63 (↓4.55%)	0.52 (↑8.33%)

Table 6. Sensitivity analyses with respect to r_0 and wind speed. The Fried parameter r_0 is perturbed by $\pm 10\%$, while wind speed is increased from the training range (1–10 m/s) to 10–20 m/s.

	r_0 ±10%			wind speed within 10-20m/s		
Method	PSNR(dB)↑	SSIM↑	LPIPS↓	PSNR(dB)↑	SSIM↑	LPIPS↓
Input	20.71 (↓0.07%)	0.78 (↓0.14%)	0.34 (↑0.25%)	20.43 (↓0.51%)	0.76 (↓2.07%)	0.36 (↑1.88%)
TMT [54]	21.20 (↓0.06%)	0.83 (↓0.14%)	0.33 (↑0.12%)	20.83 (↓0.82%)	0.80 (↓1.19%)	0.36 (↑2.22%)
DATUM [53]	22.27 (↓0.11%)	0.79 (↓0.17%)	0.33 (↑0.18%)	21.75 (↓0.87%)	0.76 (↓1.48%)	0.36 (↑2.37%)
MambaTM [55]	21.57 (↓0.06%)	0.78 (↓0.13%)	0.36 (↑0.00%)	21.19 (↓0.60%)	0.75 (↓1.23%)	0.38 (↑1.99%)

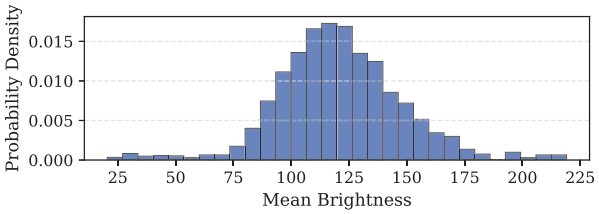


Figure 9. Brightness distribution statistics of ET-Turb.

and aperture adjustments that vary across camera designs. Explicitly modeling these factors would introduce additional parameters beyond the scope of this work.

Although brightness is not explicitly coupled with τ in our formulation, ET-Turb spans a broad brightness range. As shown in Fig. 9, the brightness distribution follows a smooth normal-like distribution, ensuring sufficient radiometric diversity for model training.

B.5. Robustness to Exposure-Dependent Noise

We introduce exposure-dependent Gaussian noise defined as $\sigma(\tau) = K/\sqrt{\tau}$ with $K = 0.001$ and $\tau \in [0.5, 40]$ ms to evaluate robustness under radiometric perturbations. Noise is added to a randomly selected subset (10%) of the ET-Turb test set (100 videos), then models trained on ET-Turb and ATSyn-Dynamic [53] are evaluated.

As summarized in Tab. 5, all models exhibit performance degradation under noise. However, models trained on ET-Turb demonstrate consistently smaller performance drops compared to those trained on ATSyn-Dynamic [53], indicating improved robustness to radiometric perturbations compared to ATSyn-Dynamic [53].

Table 7. Processing time for a 256^2 frame.

Methods	Hardie	Chimitt	Lau	P2S	Ours
Sec./frame	119.63	5.88	3.13	0.35	3.24

B.6. Sensitivity Analyses

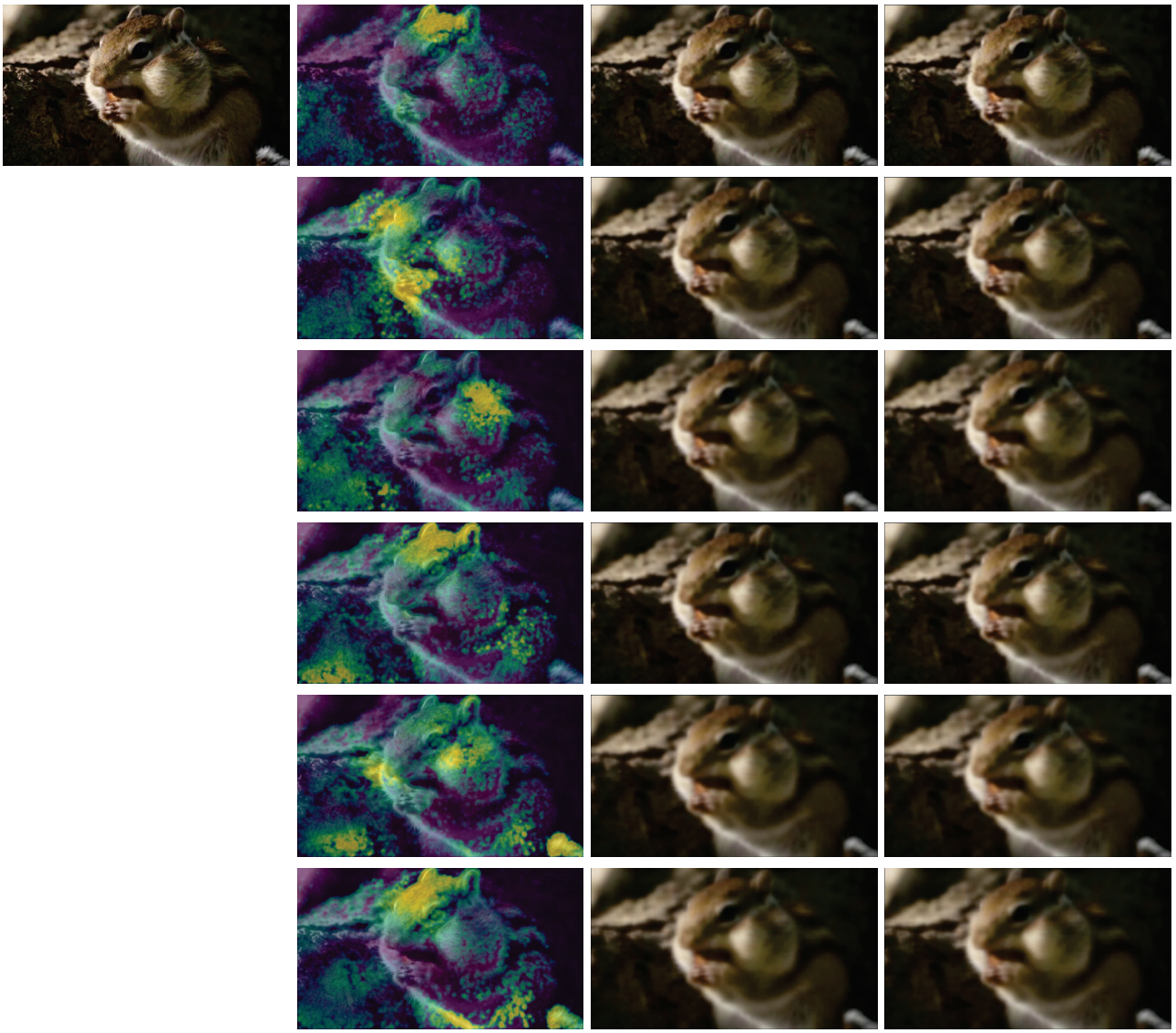
To further evaluate the robustness of models trained on ET-Turb, we conduct additional sensitivity analyses with respect to key atmospheric parameters, including the Fried parameter r_0 and wind speed. Specifically, r_0 is perturbed by $\pm 10\%$, and wind speed is varied from the training range (1-10 m/s) to an unseen higher range (10-20 m/s). The results in Tab. 6 demonstrate stable performance under both parameter variations, with only marginal degradation across all evaluation metrics. This indicates the robustness of models trained on ET-Turb to atmospheric parameter shifts.

B.7. Computational Cost

We compare the run time of the proposed method with several representative approaches [12, 19, 26, 29]. The processing time for a 256^2 frame is summarized in Tab. 7, where the runtimes of competing methods are taken from [29]. Our method requires 3.24 s/frame, which is comparable to Lau [26] (3.13 s) but slower than P2S [29] (0.35 s). When accelerated with GPU, our method consumes 448 MB.

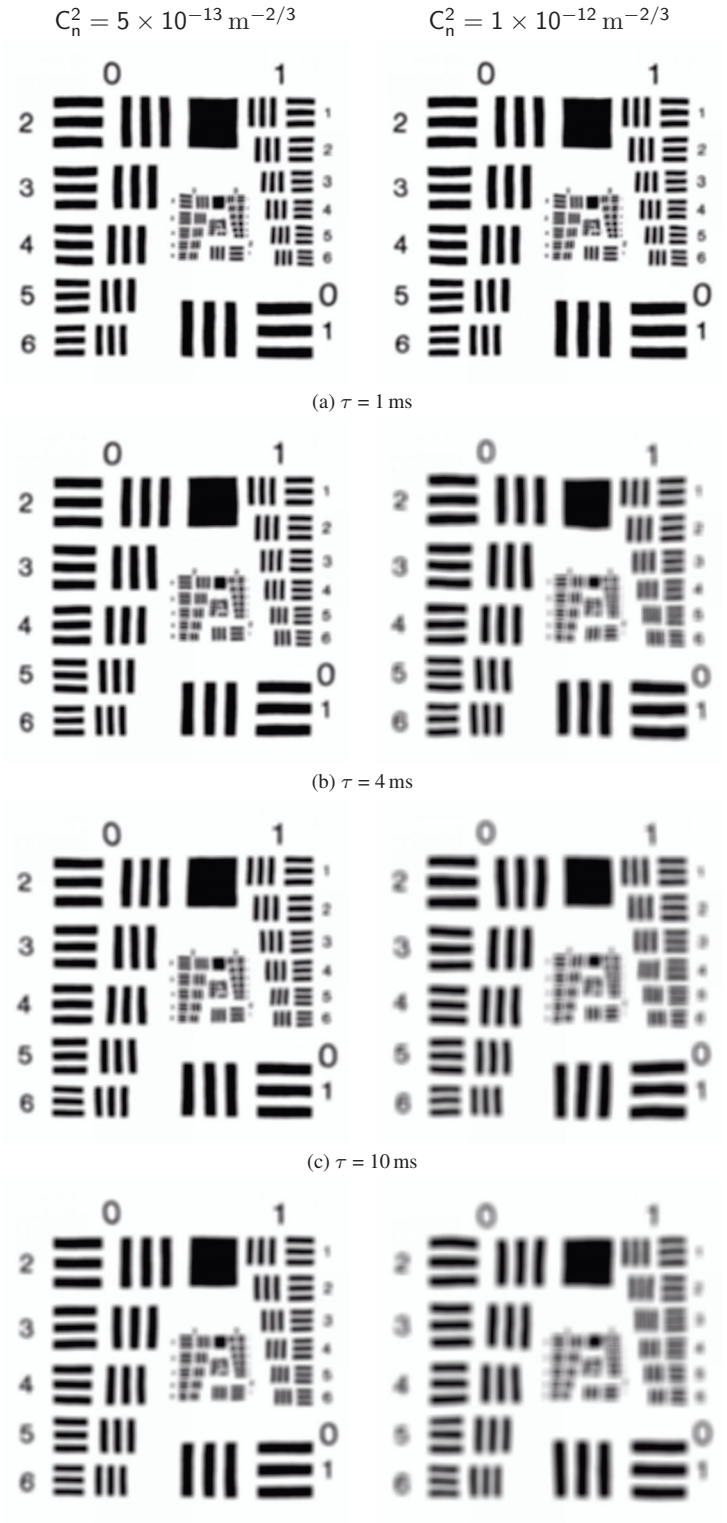
Table 8. Physical and optical parameters for the six turbulence synthesis configurations. These parameters span realistic ranges encountered in practical atmospheric imaging scenarios.

Distance (m)	Focal length (m)	F-number	C_n^2 ($10^{-14}\,\mathrm{m}^{-2/3}$)	Height (m)	Wind speed (m/s)	Exposure time (ms)
50	0.1	4	260	4	2	4
328	0.4	5.6	16	50	2	10
339	0.3	11	29	10	5	20
518	0.5	8	10	50	5	8
517	0.5	11	17	10	5	10
724	0.8	8	11	100	4	20



(a) GT (b) Tilt (c) Blur (d) Turb

Figure 10. Visualization of the turbulence synthesis pipeline across six parameter configurations. Each row corresponds to a distinct set of physical parameters from Tab. 8. The columns show: (a) original undistorted image, (b) tilt with color-coded displacement field, (c) exposure-dependent blur, and (d) final synthesized turbulence combining both degradation components.



(a) τ = 1 ms

(b) τ = 4 ms

(c) τ = 10 ms

(d) τ = 40 ms

Figure 11. Generation of turbulent images under varying intensities and exposure-time using our data synthesis method. As the exposure-time and turbulence intensity increase, the degree of blur becomes more pronounced. The associated parameter set is distance = 500 m, focal length = 300 mm, F-number = 8, height = 50 m, and wind speed = 5 m/s.

(a) Input

(b) TSR-WGAN [23]

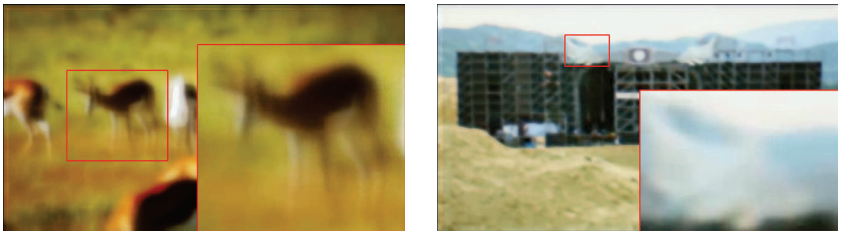
(c) TMT [54]

(d) DATUM [53]

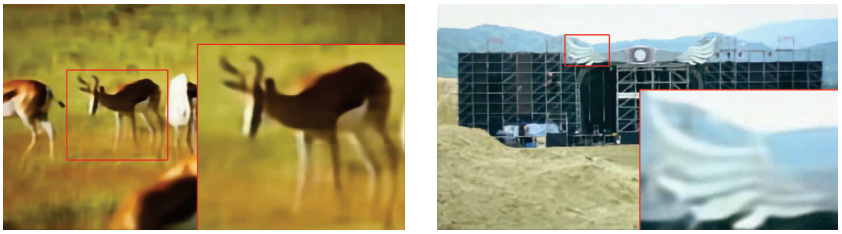
(e) MambaTM [55]

Figure 12. Real-world turbulence mitigation results. Models trained on ET-Turb dataset demonstrate strong generalization to real atmospheric turbulence, successfully restoring fine details such as text and license plates while maintaining varying degrees of color fidelity.

(a) Input



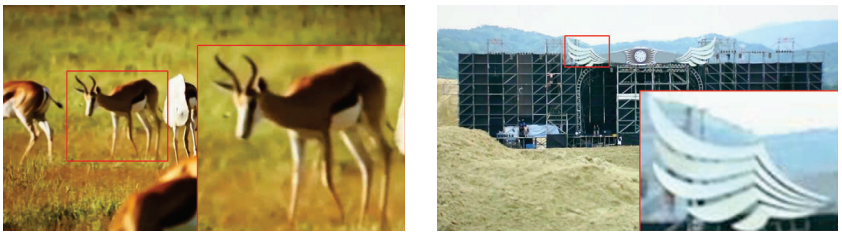
(b) TSR-WGAN [23]



(c) TMT [54]

(d) DATUM [53]

(e) MambaTM [55]



(f) GT

Figure 13. Qualitative comparison of state-of-the-art turbulence mitigation methods trained on ET-Turb dataset. Each row corresponds to a specific method, showing results from two representative test scenes (left and right).