Supplemental Material
Turb-Seg-Res: A Segment-then-Restore Pipeline for Dynamic Videos with Atmospheric Turbulence

Ripon Kumar Saha[1], Dehao Qin[2], Nianyi Li[2], Jinwei Ye[3], Suren Jayasuriya[1]
[1]Arizona State University, [2]Clemson University, [3]George Mason University
{rsaha8, sjayasur}@asu.edu, {dehaoq, nianyil}@clemson.edu, jinweiye@gmu.edu

This supplemental material includes ablation studies on the choice of optical flow for segmentation, visualizations of the Line Deviation Metric, and qualitative comparisons of video segmentations. Accessible via riponcs.github.io/TurbSegRes, the website provides our codebase for video reconstruction and turbulence simulation, along with processing scripts and example images. It showcases our turbulence simulator's ability to generate frames with variable strength and adaptive blur from single images and features visual comparisons results.

1. Choice of Optical Flow Algorithm

To identify the most effective optical flow algorithm for our segmentation tasks, we conducted evaluations using different algorithms to process our segmentation work. This approach allowed us to compare their performance directly. As detailed in Tab. 1, the results reveal that the GMA [1] algorithm demonstrates superior performance on the URG-T Dataset [5]. This enhanced performance is largely due to our methodology, which integrates multiple optical flows from different algorithms to derive the final flow and motion segmentation. Although GMA stands out, the overall performance of the various algorithms tested is relatively comparable. For the sake of simplicity in our methods, we have chosen to primarily utilize the RAFT architecture [8].

Table 1. Mean IoU Scores for Different Optical Flow

Dataset [5]	PWC [7]	RAFT [8]	CRAFT [6]	GMflow [9]	GMA [1]
URG-T	0.546	0.589	0.610	0.610	0.628

2. Line Deviation Metric

In our main paper, we quantitatively demonstrate our method's robustness in maintaining line straightness during image restoration. Here, we qualitatively visualize line straightness in images restored by various methods (as shown in Fig. 1. Notably, due to excessive sharpening, both AT-Net and Turb-Net tend to produce artifacts, resulting in numerous false positive line detections. In contrast, our approach primarily reveals lines that closely align with the straight line of a light pole.

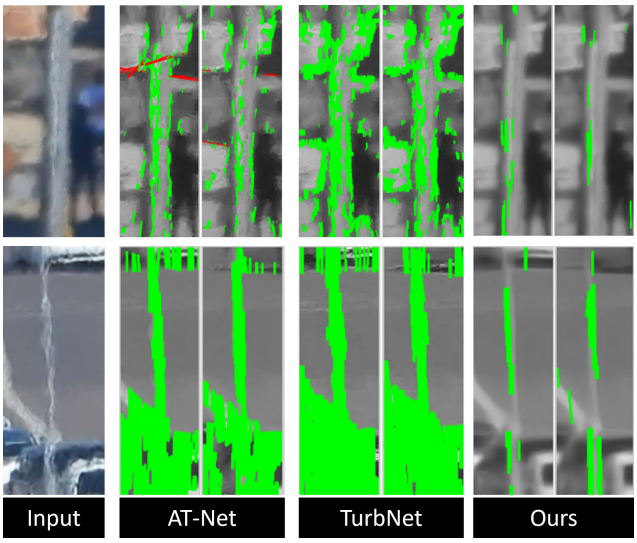


Figure 1. Detailed examination reveals that TurbNet [4] and AT-Net [10] introduce significant artifacts in the restored images, resulting in multiple small, deviating lines. Conversely, images restored with our method exhibit lines that are predominantly straight, with most deviations being minor and concentrated near the straight line reference.

3. Simulator Comparison

In this section, we present a comparative analysis of images generated using our simulator and the PS2 simulator [3] as shown in Fig. 2. To ensure a fair comparison, all settings for the PS2 simulator were maintained at their default values for both video generation and image comparison. For an in-

depth comparative evaluation of both simulators, we invite readers to view the supplementary video material available on our project's website: riponcs.github.io/TurbSegRes.

(a) PS2 Simulator (b) Our Simulator

Figure 2. Comparison between the PS2 Simulator and our simulator. Our findings indicate that the PS2 simulator often produces pixelated images, whereas our simulator yields images with notably less pixelation.

4. Qualitative Video Comparison

For a qualitative assessment, videos from the URG-T dataset [5] were utilized. Each frame was processed into patches compatible with AT-Net [10] and TurbNet [4] for image restoration. In contrast, TCI [2], which typically requires 70 minutes to restore a 1920×1080 frame, was applied at a reduced resolution of 384×216 and maintained its original grayscale image prediction methodology. Our comparative analysis reveals that while AT-Net exhibits significant pixel distortion and TurbNet displays medium-level distortion, TCI frequently fails to capture moving objects. In contrast, our proposed method demonstrates superior stability with minimal distortion, effectively preserving moving subjects. For a comprehensive visual comparison, five synchronized videos showcasing these differences are accessible on our project's website.

While the main paper focuses on grayscale comparisons for visual assessment of turbulence mitigation methods (TCI being primarily designed for grayscale), we include a color comparison in Fig. 3. Here, each color channel of a scene is processed by TCI and then combined. Our method achieves superior performance, reconstructing high-definition RGB images in 5.7 seconds, significantly faster than TCI's 4.8 hours.

5. Ablation Studies

5.1. Choice of Restoration Network

In our study, we conducts a comparative analysis of two prominent restoration networks: the Restoration Transformer (Restormer) [11] and TurbNet [4]. Our findings re-

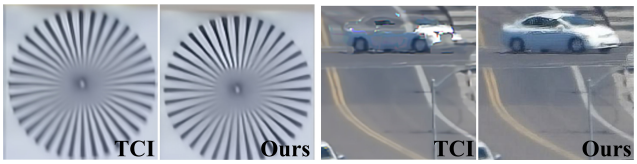
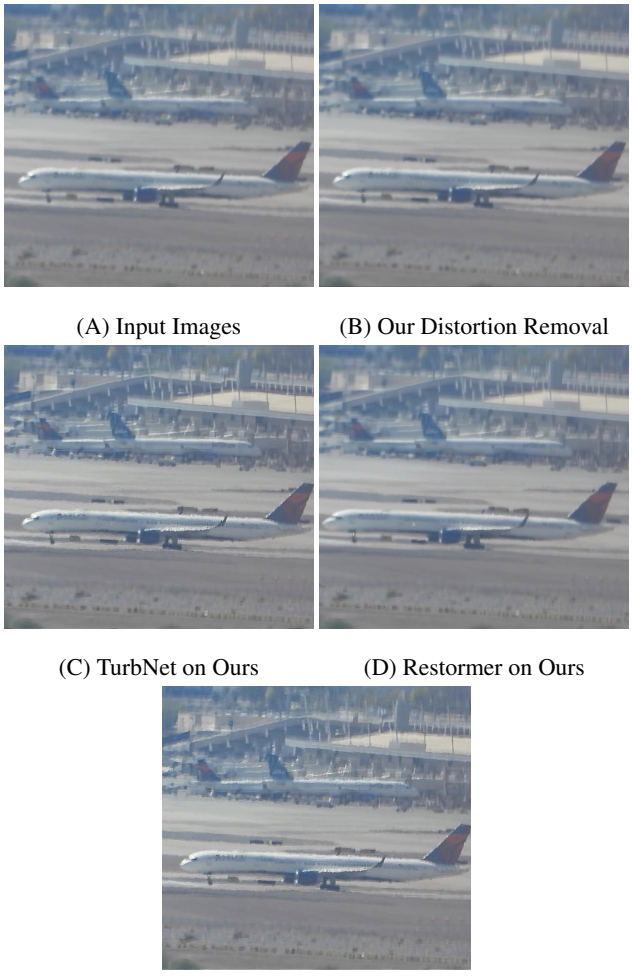


Figure 3. TCI vs. Ours: This figure compares reconstructions from the static OTIS scene (left) and the dynamic URG-T dataset (right) with a zoomed view. Our method achieves superior results on both static and dynamic scenes.



(A) Input Images (B) Our Distortion Removal

(C) TurbNet on Ours (D) Restormer on Ours

(E) Just TurbNet [4]

Figure 4. Comparative analysis of various networks and their impact on image quality post-distortion removal. Notably, the edges of the airport exhibit distortion artifacts in the input image, but these are effectively eliminated through our distortion removal process. The processed image (B) is then enhanced for sharpness by different networks (C, D). TurbNet appears to prioritize sharpening, occasionally at the cost of introducing some distortion, whereas Restormer focuses more on distortion elimination, leading to clearer images. Notably, TurbNet (E), when used without our distortion removal process, exhibits significantly more artifacts.

veal that Restormer, tailored to our dataset, offers notably stable outputs, excelling in reducing turbulence effects, a feature particularly apparent along object edges, like cars. However, this stability sometimes comes at the cost of finer details. Conversely, the integration of TurbNet into our framework not only elevates its base performance but also achieves enhanced detail in the outputs as shown in Fig. 4. Despite these improvements, TurbNet tends to be less effective in maintaining stability and consistently removing distortions. Please refer to the project website for a detailed demonstration.

5.2. Impact of Segmentation in Processing

(a) Without Segmentation (b) With Segmentation

Figure 5. Illustration of the impact of segmentation on image processing: (a) results without segmentation, and (b) results with segmentation. The presence of segmentation significantly enhances the clarity and detail of the processed image.

Segmentation plays a pivotal role in our framework, particularly due to the differentiated processing applied to the foreground and background elements. In the absence of segmentation, the implementation of an adaptive Gaussian weighted temporal average across the video frames leads to the manifestation of motion artifacts in the foreground as shown in Fig. 5. This outcome unequivocally underscores the indispensability of segmentation within our processing pipeline. For a more illustrative understanding, readers are encouraged to view the accompanying video demonstration available on our project's website.

5.3. Impact of Stabilization

In high-zoom range videos (10x to 500x), minor vibrations can significantly disrupt frame alignment, leading to inaccurate segmentation. Unstabilized footage exhibits elevated pixel variability, compelling our pipeline to erroneously estimate the C_n^2 value from pixel movements and inconsistently select frames for distortion removal. This often results in misalignment and introduces artifacts, consequently degrading video reconstruction quality. Our findings indicate that incorporating stabilization markedly enhances the output by mitigating these issues as shown in Fig. 6. For a visual comparison, see the video at riponcs.github.io/TurbSegRes.

(a) Without Stabilization (b) With Stabilization

Figure 6. Comparison of video frame reconstruction quality: (a) without stabilization, and (b) with stabilization. The enhancement in image stability and clarity with stabilization is evident.

5.4. Segmentation Mask

The project website contains a video demonstrating the efficacy of our mask prediction algorithm. It presents a side-by-side comparison with the ground truth annotations from the URG-T dataset across various sequences. This visual comparison offers a detailed perspective on the performance of our algorithm in segmentation tasks. It's important to note that the masks are compared over stabilized video footage, which may account for minor pixel discrepancies. Additionally, we apply a space-time expansion to the masks to enhance their accuracy and coverage in the implementation of restoration.

References

[1] S. Jiang, D. Campbell, Y. Lu, H. Li, and R. Hartley. Learning to estimate hidden motions with global motion aggregation. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pages 9752–9761, Los Alamitos, CA, USA, 2021. IEEE Computer Society. 1

[2] 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. 2

[3] Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 14759–14768, 2021. 1

[4] 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. 1, 2

[5] Dehao Qin, Ripon Saha, Suren Jayasuriya, Jinwei Ye, and Nianyi Li. Unsupervised region-growing network for object segmentation in atmospheric turbulence. arXiv:2311.03572, 2023. 1, 2

[6] Xiuchao Sui, Shaohua Li, Xue Geng, Yan Wu, Xinxing Xu, Yong Liu, Rick Goh, and Hongyuan Zhu. Craft: Cross-attentional flow transformer for robust optical flow. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 17602–17611, 2022. 1

[7] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 1

[8] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In Computer Vision – ECCV 2020, pages 402–419, Cham, 2020. Springer International Publishing. 1

[9] Haofei Xu, Jing Zhang, Jianfei Cai, Hamid Rezatofighi, and Dacheng Tao. Gmflow: Learning optical flow via global matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8121–8130, 2022. 1

[10] 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. 1, 2

[11] 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. 2