DMAT: An End-to-End Framework for Joint Atmospheric Turbulence Mitigation and Object Detection

Paul Hill, Zhiming Liu, Alin Achim, David Bull, and Nantheera Anantrasirichai
Visual Information Laboratory, University of Bristol, Bristol, UK

WACV 2026 TUCSON, AZ 3/6 - 3/10

Abstract

- Atmospheric Turbulence (AT) degrades the clarity and accuracy of surveillance imagery, posing challenges not only for visualization quality but also for object classification and scene tracking. Deep learning-based methods have been proposed to improve visual quality, but spatio-temporal distortions remain a significant issue.
- In this paper, we propose a novel framework that learns to compensate for distorted features while simultaneously improving visualization and object detection.
- This end-to-end training strategy leverages and exchanges knowledge of low-level distorted features in the AT mitigator with semantic features extracted in the object detector.
- Specifically, in the AT mitigator a 3D Mamba-based structure is used to handle the spatio-temporal displacements and blurring caused by turbulence.
- Optimization is achieved through back-propagation in both the AT mitigator and object detector.
- Our proposed DMAT outperforms state-of-the-art AT mitigation and object detection systems up to a 15% improvement on datasets corrupted by generated turbulence.

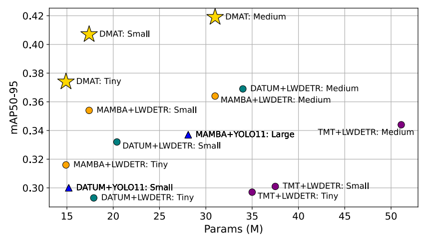


Figure 1. Object detection results for all object sizes and classes of our proposed method (DMAT) compared to individual AT-mitigation and object detection methods.

- Code: https://github.com/pui-nantheera/DMAT
- Dataset: https://zenodo.org/records/17673509

Methods

- Registration module: Incorporating a UNet-like architecture with deformable 3D convolutions. The registration module has a depth scale of 4. Larger kernels are used in the initial layers to provide a wider field of view.
- Enhancement module: Extracting features using 3D convolutions, then being processed in the 3D Mamba-based UNet-like network and simultaneously fed to the object detector.
- Object detection module: Extracting the enhanced features using a ResNet50 backbone, then being flattened and supplemented with positional encoding before being fed into a transformer encoder. We employ cross-stage feature fusion to improve gradient flow between the encoder and decoder.
- Loss functions:
  (1) Reconstruction loss: Charbonnier loss
  (2) Detection loss: L1 for box size and location + the generalized IoU loss + Binary Cross-Entropy loss

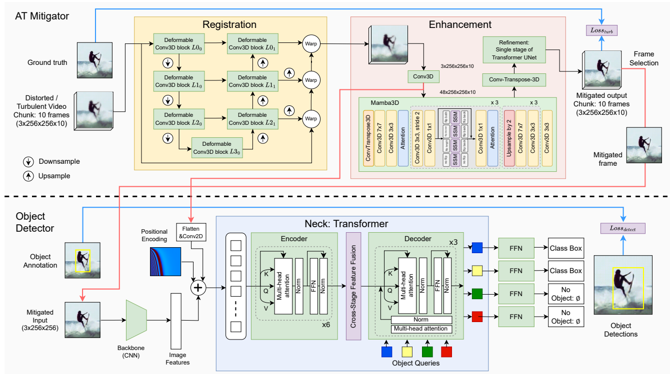


Figure 2. Architecture of the DMAT framework for atmospheric distortion mitigation and object detection.

Synthetic datasets

- AT distortions using the Phase-to-Space (P2S) Transformation
- Static scenes: COCO2017 (1) all 80 classes, (2) Top10, (3) CarPerson
- Dynamic scenes: GOT-10k dataset, Person or Cars, 50-frame clips, In total, 1,761 videos are used for training and 441 for testing

Results

Table 1. Object detection performance.

Mitigator	Detector	Params (M)	Static scenes: All classes, All sizes	Static scenes: All classes, Small	Static scenes: Top 10 classes, All sizes	Static scenes: Top 10 classes, Small	Static scenes: Car and person, All sizes	Static scenes: Car and person, Small	Dynamic scenes: Car and person, All sizes	Dynamic scenes: Car and person, Small
DATUM	YOLO11x	5.8+56.9	0.317	0.060	0.166	0.032	0.291	0.132	0.133	0.013
	DETR-Large	5.8+60.0	0.243	0.100	0.120	0.051	0.216	0.087	0.102	0.007
	LWDETR-Med	5.8+28.2	0.369	0.200	0.231	0.098	0.306	0.140	0.154	0.015
TMT	YOLO11x	22.9+56.9	0.307	0.038	0.168	0.026	0.256	0.092	0.112	0.008
	DETR-Large	22.9+60.0	0.224	0.087	0.129	0.040	0.178	0.052	0.093	0.003
	LWDETR-Med	22.9+28.2	0.344	0.165	0.226	0.088	0.271	0.119	0.126	0.007
MAMBA	YOLO11x	2.8+56.9	0.350	0.037	0.191	0.031	0.292	0.128	0.147	0.011
	DETR-Large	2.8+60.0	0.275	0.097	0.135	0.041	0.217	0.076	0.128	0.008
	LWDETR-Med	2.8+28.2	0.364	0.186	*0.249*	0.105	0.310	0.138	0.167	0.017
DMAT-Tiny (ours)		2.8+12.1	*0.374*	*0.188*	0.214	*0.108*	*0.348*	*0.178*	*0.214*	*0.127*
DMAT-Small (ours)		2.8+14.6	0.407	**0.234**	0.276	0.134	0.373	0.194	0.251	0.143
DMAT-Med (ours)		2.8+28.2	**0.419**	**0.234**	**0.282**	**0.137**	**0.385**	**0.199**	**0.269**	**0.152**

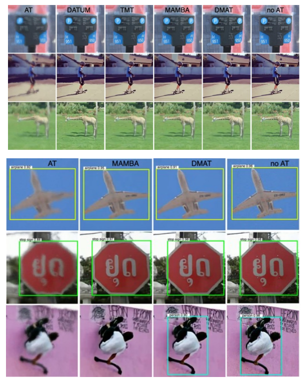


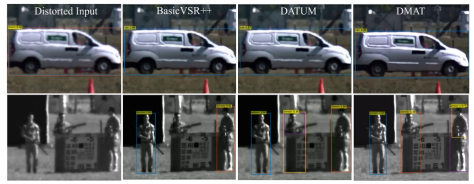
Figure 3. Performance on synthetic AT dataset.

Results

Table 2. Atmospheric turbulence mitigation performance.

AT Mitigation	Static, All: PSNR ↑	Static, All: SSIM ↑	Static, All: LPIPS ↓	Static, Top10: PSNR ↑	Static, Top10: SSIM ↑	Static, Top10: LPIPS↓	Static, CarPerson: PSNR ↑	Static, CarPerson: SSIM ↑	Static, CarPerson: LPIPS↓	Dynamic, CarPerson: PSNR ↑	Dynamic, CarPerson: SSIM ↑	Dynamic, CarPerson: LPIPS↓
AT	20.946	0.512	0.546	20.668	0.513	0.550	20.210	0.488	0.564	22.252	0.591	0.466
DATUM	23.092	0.644	0.356	22.910	0.649	0.343	22.390	0.632	0.360	22.564	0.693	*0.373*
TMT	*23.332*	*0.654*	**0.348**	*23.167*	*0.661*	**0.331**	*22.592*	*0.641*	**0.350**	*22.747*	*0.712*	**0.367**
MAMBA	23.763	0.667	0.378	23.693	0.677	0.360	23.084	0.657	0.380	23.297	0.783	0.379
DMAT (ours)	**23.841**	**0.671**	*0.373*	**23.861**	**0.683**	*0.357*	**23.220**	**0.663**	*0.376*	**23.312**	**0.831**	0.371

Figure 4. Performance on real AT dataset **CLEAR**.



Ablation Study

- Experiment 1: When the AT mitigator is removed, the performance of object detection drops by approximately 50% on average, with an even larger drop for smaller objects.
- Experiment 2: When the object detection module is removed, the performance of AT mitigation drops by up to approximately 1.5% in term of PSNR.
- Experiment 3: When the feature sharing between the two modules is removed, the performance degradation is reduced by up to 1% in terms of PSNR and 2% in terms of mAP.

Conclusions

The main contributions of this paper are summarized as follows.

- A first end-to-end architecture for joint AT mitigation and object detection, **DMAT**.
- Integrating a 3D Mamba-based restoration module with a transformer-based detector.
- New optimized sets of synthetic turbulent videos based on subsets of the COCO and GOT-10k datasets for static and dynamic scenes.

Contact

Pui Anantrasirichai
University of Bristol
Email: N.Anantrasirichai@Bristol.ac.uk
Website: https://pui-nantheera.github.io/

References

1. CLEAR: N. Anantrasirichai, A. Achim, N.G. Kingsbury, and D.R. Bull. Atmospheric turbulence mitigation using complex wavelet-based fusion. IEEE Transactions on Image Processing, 22(6):2398–2408, 2013
2. TMT: Xingguang Zhang, Zhiyuan Mao, Nicholas Chimitt, and Stanley H. Chan. Imaging through the atmosphere using tur- bulence mitigation transformer. IEEE Transactions on Computational Imaging, 10:115–128, 2024
3. DATUM: Zhang, Xingguang and Chimitt, Nicholas and Chi, Yiheng and Mao, Zhiyuan and Chan, Stanley H, Spatio-Temporal Turbulence Mitigation: A Translational Perspective, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024
4. MAMBA: Haifan Gong, Luoyao Kang, Yitao Wang, Yihan Wang, Xi- ang Wan, Xusheng Wu, and Haofeng Li. Nnmamba: 3D biomedical image segmentation, classification and landmark detection with state space model. IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pages 1–5, 2025
5. DETR: icolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to- end object detection with transformers. In European Conference on Computer Vision (ECCV), pages 213–229, 2024
6. LW-DETR: Qiang Chen, Xiangbo Su, Xinyu Zhang, Jian Wang, Jiahui Chen, Yunpeng Shen, Chuchu Han, Ziliang Chen, Weixi- ang Xu, Fanrong Li, et al. LW-DETR: A transformer re- placement to yolo for real-time detection. arXiv preprint arXiv:2406.03459, 2024
7. P2S: Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Accelerating atmospheric turbulence simulation via learned phase- to-space transform. IEEE/CVF International Conference on Computer Vision (ICCV), 2021