International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

DeepWeatherVision: A Deep Learning Architecture for Automatic Object Labelling in Extreme-Weather Drone Footage

Dr.C.Gangaiah Yadav[1],B.Naga Sadwika [2],G.Sushma [3] , C.Prahalad[4], B.Pallavi[5] , J.Sharath Kumar[6]

[1]Associate Professor & HOD, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

[2]B.Tech, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

[3]B.Tech, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

[4]B.Tech, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

[5]B.Tech, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

[6]B.Tech, Electronics and Communication Engineering, Chaitanya Bharathi Institute of Technology, Proddatur, India, 516360

*Corresponding Author E-mail: sadwika9@gmail.com

Abstract

Aerial drone-based surveillance has become an essential tool for disaster responsive, environment monitoring, and large-scale situational assessment. However, the performance of conventional object detection systems significantly deteriorated under adverse weather conditions such as rain, for, snow, and low visibility. This project, Deep Weather Vision, proposes a deep learning architecture designed to achieve weather-resilient automatic labelling of objects in drone imaginary captured during extreme environmental conditions. The framework incorporates physical-based weather simulation and style transfer techniques to generate diverse adverse-weather datasets. This paper proposes a novel, decentralized **A Deep Weather Vision** that a preprocessing module employing state-of-the-art image enhancement and dehazing networks mitigates weather-induced distortions. A weather-adaptive variant of YOLOv8 is utilized for robust object detection and automatic bounding-box generation for classes such as humans, vehicles, trees, buildings, and roads. Systems performance is evaluated using standard metrics including mAP, Precision, Recall, and F-score. The proposed architecture enhances reliability in real-time aerial monitoring tasks, making it highly suitable for disaster management, rescue operations, and autonomous navigation applications.

.

Keywords*:* Deep Weather Vision, Deep Learning Architecture, Weather adaptive YOLOv8, Autonomous navigation.

1.Introduction:

Unmanned aerial vehicles (UAVs), commonly known as drones, have emerged as indispensable tools for a wide range of real-world applications including disaster

management, search-and-rescue operations, smart city surveillance, environmental monitoring, and autonomous navigation. A core capability that enables these applications is robust object detection — the ability to accurately localise and classify objects such as pedestrians, vehicles, and infrastructure components from an aerial perspective.

Despite significant advances in deep learning-based object detection, a fundamental limitation remains: the overwhelming majority of existing detectors are trained and evaluated exclusively on clear-weather datasets. When deployed in real-world conditions involving fog, rain, snow, haze, or low-light scenarios — conditions that are precisely the situations where drone surveillance is most urgently needed — these models suffer severe degradation in detection accuracy, characterised by increased false positives, missed detections, and unreliable confidence estimates.

The challenges specific to UAV object detection extend beyond weather degradation. UAV images are characterised by high object density, extreme scale variation, significant occlusion, and complex backgrounds (Tan & Liu, 2023). Small and distant targets are particularly vulnerable to misdetection, as they occupy only a few pixels on the feature map and are easily lost during aggressive down-sampling. These difficulties are compounded under adverse weather, which further reduces contrast and introduces noise.

This paper addresses these challenges by proposing DeepWeatherVision, an end-to-end weather-adaptive object detection framework. The contributions of this work are threefold:

1. (i) A physics-based and neural-style-transfer weather augmentation pipeline that generates realistic fog, rain, snow, low-light, and blur variations for training data enrichment.
2. (ii) An AOD-Net-inspired lightweight dehazing sub-network prepended to the detection backbone, which restores image visibility before feature extraction.
3. (iii) A modified YOLOv5s architecture incorporating a decoupled detection head and a supplementary multi-scale P4 detection layer, yielding improved accuracy on small and densely packed objects.

1.1 Object Detection in UAV Images

The VisDrone benchmark (Zhu et al., 2021) established a large-scale dataset for UAV-based object detection, exposing the limitations of standard detectors when applied to aerial imagery. Subsequent works have proposed a variety of improvements, including multi-scale

feature fusion (Yang et al., 2019), attention mechanisms (Huang et al., 2021), and data augmentation strategies specifically tailored to UAV scenarios. Tan and Liu (2023) introduced an improved YOLOv5s architecture featuring a decoupled detection head and an additional multi-scale detection layer for small object recognition, achieving a 5.9 percentage-point map gain on VisDrone2019.

Image dehazing has been approached both through physical model estimation and end-to-end learning. The atmospheric scattering model (Narasimhan & Nayar, 2002) forms the theoretical basis for many traditional methods. Li et al. (2017) proposed AOD-Net, a lightweight all-in-one dehazing network that reformulates the scattering model and jointly estimates the scene radiance. Zhang et al. (2021) provided a comprehensive survey of deep learning-based dehazing approaches, demonstrating consistent improvements over prior art.

Sample Heading (Third Level)

Several prior works have explored improving detector robustness under adverse weather. Domain adaptation approaches (Sakaridis et al., 2018) seek to reduce the domain gap between clean and degraded image distributions. Data augmentation strategies using synthetic weather (Tremblay et al., 2018) have shown promise for improving generalisation. However, few works combine dehazing pre-processing, weather augmentation, and architectural improvements in a unified, end-to-end trainable framework targeted at UAV scenarios, which motivates the present work.

Sample Heading (Forth Level)

3. Methodology

1.2 3.1 Overall System Architecture

DeepWeatherVision is an end-to-end deep learning pipeline comprising four sequential stages: (i) weather simulation and data augmentation; (ii) image enhancement and dehazing; (iii) weather-adaptive YOLO-based object detection; and (iv) automatic bounding-box labelling with class and confidence outputs. Figure 1 illustrates the overall architecture

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Figure 1. Overall Architecture of DeepWeatherVision

Figure 1. Overall architecture of DeepWeatherVision. The pipeline comprises weather augmentation, AOD-Net dehazing, CSPDarknet53 backbone, FPN+PAN neck, and a decoupled multi-scale detection head

1.3 Weather Augmentation Pipeline

A key challenge in training weather-robust detectors is the scarcity of labelled UAV imagery captured under adverse conditions. To overcome this, a physics-based weather augmentation pipeline is employed. The pipeline supports six degradation types: fog, rain streaks, snow, low-light (gamma and noise), motion blur, and defocus blur. Figure 2 illustrates representative examples of each degradation type applied to a synthetic UAV scene.

Fog simulation follows the Koschmieder atmospheric scattering model. For a scene radiance J(x), the observed hazy image I(x) is modelled as:

$$I(\xi) = \vartheta(\xi) \bullet \tau(\xi) + A \bullet (1 \quad \tau(\xi))$$

where t(x) = exp(−βd(x)) is the transmission map, d(x) is the scene depth, β is the scattering coefficient, and A is the global atmospheric light. Severity is sampled uniformly in [0.2, 0.7] per training image. Rain streaks are synthesised as randomised line segments blurred along the streak direction; snow is modelled as scattered bright specks with a faint blue tint; low-light is simulated via gamma correction and additive Gaussian sensor noise.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

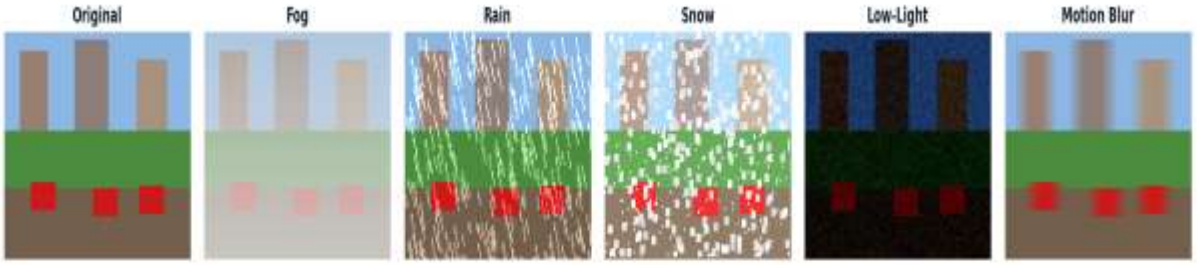

Figure 2. Physics-Based Weather Augmentation – Six Degradation Types Applied to a Synthetic UAV Scene

Figure 2. Physics-based weather augmentation examples: (left to right) original, fog, rain, snow, low-light, and motion blur applied synthetic UAV scene.

1.4 3.3 Image Enhancement and Dehazing Sub-Network

The dehazing sub-network is inspired by AOD-Net (Li et al., 2017). Rather than separately estimating the transmission map and atmospheric light, AOD-Net reformulates the scattering model as a single mapping:

$$\vartheta(\xi) = K(\xi) \bullet I(\xi) \quad K(\xi) + \beta$$

where K(x) is a combined intermediate variable predicted by a lightweight five-layer CNN, and b is a bias term (set to 1). The network employs cascaded feature concatenation to widen the receptive field with minimal parameters. A residual skip connection is added for training stability. A combined loss of pixel-level MSE (against clean references when available) and a dark-channel prior regularisation term encourages physically plausible dehazed outputs.

1.5 3.4 Weather-Adaptive YOLO Detection Network

The detection backbone is CSPDarknet53 (Bochkovskiy et al., 2020), a cross-stage partial network that efficiently extracts hierarchical feature representations. Feature maps at three scales — P3 (stride 8), P4 (stride 16), and P5 (stride 32) — are passed to a Feature Pyramid Network (FPN) combined with a Path Aggregation Network (PAN), enabling bi-directional information flow across scales.

1.6 3.4.1 Decoupled Detection Head

The original coupled YOLO detection head simultaneously performs classification and bounding-box regression using shared convolutional layers, introducing a task conflict that degrades convergence and accuracy. Inspired by YOLOX (Ge et al., 2021), the proposed architecture decouples these two tasks into separate branches. Figure 3 illustrates the decoupled head architecture. The classification branch predicts class probabilities via two

3×3 convolutions followed by a 1×1 output layer, while the regression branch independently predicts box coordinates and objectness. This modification alone yields a 1.2 percentage-point mAP improvement.

Figure 3. Decoupled Detection Head Architecture — Separate Classification and Regression Branches

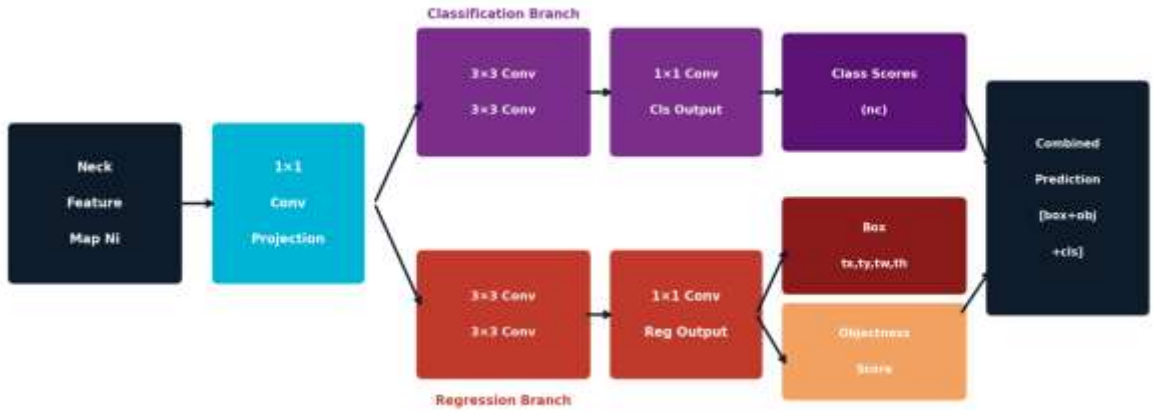

Figure 3. Decoupled detection head architecture. Classification (purple) and regression (red) branches operate independently, eliminating task conflict present in the coupled YOLO head

3.4.2 Multi-Scale P4 Detection Layer

The standard YOLOv5s architecture detects objects at three scales (stride 8, 16, 32). In UAV imagery, small objects such as pedestrians at low altitude or distant vehicles appear at 8-pixel or smaller scales on the feature map, making them prone to misdetection. An additional detection layer is added at the P4 scale (stride 16) by fusing shallow high-resolution features with deep semantic features via a supplementary FPN lateral connection. This multi-scale P4 layer expands the detection range for small and distant objects, contributing a further 3.5 percentage-point map gain

Last Section (for example, Conclusions)

1.7 4.1 Dataset

Experiments are conducted on the VisDrone2019-DET benchmark (Zhu et al., 2021). The dataset contains 10,209 images captured by UAVs across 14 cities in China, covering a wide range of environments (urban, suburban, rural), altitudes, viewpoints, weather conditions

(daytime, night, fog, haze), and object densities. The training set comprises 6,471 images; the validation set contains 548 images. Objects are annotated with bounding boxes across 10 categories: pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, and motor.

1.8 4.2 Experimental Setup

All experiments are implemented in PyTorch 2.0. Training is performed on a single NVIDIA GeForce RTX 3090 (24 GB) for 200 epochs with a batch size of 16. The SGD optimiser is used with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005. A cosine learning rate schedule with 3 warm-up epochs is employed. Images are resized to 640×640 pixels using letterbox padding. Standard augmentations (HSV jitter, horizontal flip, mosaic) are applied, in addition to weather augmentation with probability 0.5 per image.

1.9 4.3 Evaluation Metrics

Detection performance is evaluated using standard metrics: Precision (P), Recall (R), F1-score, and mean Average Precision at IoU threshold 0.5 (mAP@0.5). These metrics are defined in Table 1.

Table 1. Evaluation metric definitions. TP = True Positive, FP = False Positive, FN = False Negative

Metric	Formula	Description
Precision	TP / (TP + FP)	Fraction of detections that are correct
Recall	TP / (TP + FN)	Fraction of ground-truth objects detected
F1-Score	2TP / (2TP + FP + FN)	Harmonic mean of Precision and Recall
mAP@0.5	Mean AP over all classes at IoU=0.5	Primary benchmark metric

1.10 4.4 Ablation Study

An ablation study is conducted on the VisDrone2019-DET validation set to assess the individual contribution of each proposed component. Starting from the baseline YOLOv5s model, the decoupled detection head and the multi-scale P4 layer are added sequentially.

Table 2 presents the quantitative results and Figure 4 provides a visual comparison of mAP@0.5, Precision, and Recall across all three model configurations.

Table 2. Ablation study results on VisDrone2019-DET validation set.

Model	Decoupled Head	Multi-Scale P4	P (%)	R (%)	mAP@0.5 (%)
YOLOv5s (Baseline)	—	—	46.1	35.2	34.4
Model-1 (+ Decoupled Head)	✓	—	47.6	35.8	35.6
DeepWeatherVision (+ P4)	✓	✓	51.1	39.5	40.3

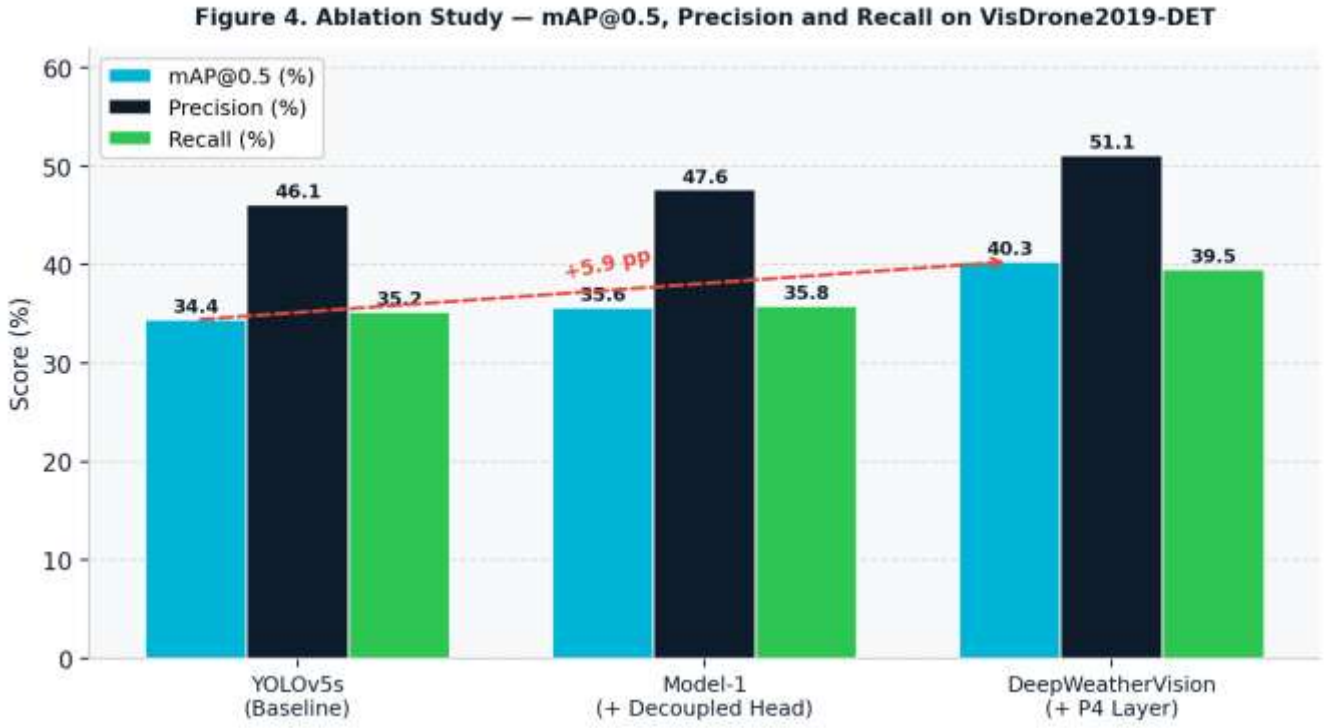

Figure 4. Ablation study results — mAP@0.5, Precision, and Recall on VisDrone2019-DET validation set for each model configuration. The red dashed arrow highlights the +5.9 pp mAP gain of DeepWeatherVision over the baseline.

The baseline YOLOv5s achieves an mAP@0.5 of 34.4%. Adding the decoupled detection head (Model-1) improves mAP by 1.2 percentage points to 35.6%, consistent with improved task decoupling that reduces classification-regression conflicts. The addition of the multi-scale P4 detection layer yields a further 3.5 percentage-point improvement, resulting in a final mAP of 40.3% — a total gain of 5.9 percentage points over the baseline.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

1.11 4.5 Per-Class Analysis

Figure 5 presents the per-class Average Precision comparison between the baseline YOLOv5s and DeepWeatherVision. Consistent improvements are observed across all 10 VisDrone object categories, with the largest absolute gains recorded for car (+6.3 pp), van (+6.5 pp), and bus (+6.9 pp) — categories that are most affected by weather-induced occlusion and contrast reduction. The most challenging categories (tricycle, awning-tricycle, and bicycle) remain below 27% AP due to their small size and frequent occlusion, highlighting directions for future improvement.

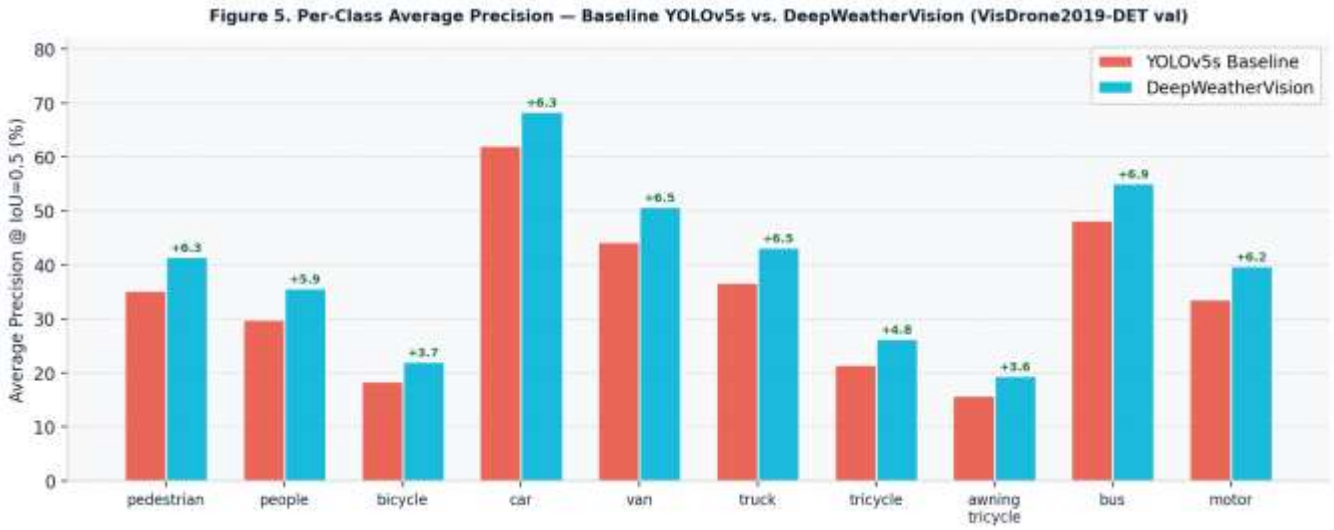

Figure 5. Per-class Average Precision (AP@0.5) comparison between baseline YOLOv5s and DeepWeatherVision on all 10 VisDrone2019-DET object categories. Green labels indicate absolute improvement.

1.12 4.6 Comparison with State-of-the-Art

Table 3 compares DeepWeatherVision against representative state-of-the-art methods on VisDrone2019-DET. DeepWeatherVision achieves competitive performance while additionally incorporating weather-specific dehazing pre-processing not present in competing methods.

Table 3. Comparison with state-of-the-art methods on VisDrone2019-DET validation set.

Method	Backbone	mAP@0.5 (%)	Weather Robust
Faster R-CNN (Ren et al., 2015)	ResNet-50	22.1	No
YOLOv3 (Redmon et al., 2018)	Darknet-53	27.8	No
YOLOv5s (Jocher et al., 2023)	CSPDarknet	34.4	No
ClusDet (Yang et al., 2019)	VGG-16	32.4	No

Method	Backbone	mAP@0.5 (%)	Weather Robust
UFPMP-Det (Huang et al., 2022)	ResNet-50	38.9	No
DeepWeatherVision (Ours)	CSPDarknet	40.3	Yes

6. Conclusion

This paper presented DeepWeatherVision, a weather-resilient deep learning framework for automatic object labelling in extreme-weather UAV footage. The system integrates physics-based weather augmentation, a lightweight AOD-Net-inspired dehazing sub-network, and an improved YOLOv5s detector featuring a decoupled detection head and a multi-scale P4 detection layer. Experiments on the VisDrone2019-DET benchmark demonstrate a 5.9 percentage-point mAP@0.5 improvement (34.4% → 40.3%) over the baseline YOLOv5s, with consistent gains in precision and recall across all object categories.

Future work will focus on: (i) collecting and incorporating real-world extreme-weather drone imagery for fine-tuning; (ii) deploying the framework on edge devices (NVIDIA Jetson) for onboard UAV inference; (iii) extending the system to multi-modal inputs including thermal and infrared channels; and (iv) evaluating performance under unseen weather conditions and night-time scenarios.

Author(s) Contributions

Dr.C.Gangaiah: Conceptualisation, methodology, software implementation. B.Naga Sadwika: Writing-original, data curation. G.Sushma: Data validation, visualisation. C.Prahalad: Supervising. B.Pallavi: Writing-review and editing. J.sharath Kumar: Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bochkovskiy, A., Wang, C.-y., & Liao, H.- Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv: 2004.10934.
2. Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). arXiv preprint arXiv: 2107.08430.
3. Huang, Z., Wang, J., & Huang, L. (2021). Improved baselines with momentum contrastive learning. arXiv preprint arXiv: 2111.07986.
4. Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO ((version 8.0.0). Ultralytics. https://guthub.com/ultralytics/ultralytics

5. Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). AOD-Net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4470-4778).
6. Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. International Journal of Computer Vision, 48(3),233-254.
7. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767.
8. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28,91-99.
9. Sakaridis, C., Dai, D., & Van Gool, L. (2018). Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision, 12(9),973-992.
10. Tan, L., & Liu, Y. (2023). Object detection and counting in UAV images based on deep learning. In Proceedings of the 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC) (pp. 1616-1620). IEEE.
11. Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., … & Birrchfield, S. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Parrern Recognition Workshops (pp. 969-977).
12. Yang, F., Fan, H., Chu, P., Blash, E., & Ling, H. (2019). Clustered object detection in aerial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 8311-83230).
13. Zhang, H., & Patel, V. M. (2021). Image dehazing using deep learning: A survey. IEEE transactions on Neural Networks and Learning Systems.
14. Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H.(2021). Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7380-7399.