IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review On Knowledge Distillation For Efficient Aircraft Detection In Low-Resolution Aerial Images: Addressing Weather And Size

[1]Dhruti Vasava, [2]Dr. Barkha Joshi, [3]Jayna Shah, [4]Brijesh Panchal

[1]Student - GTU, [2]Assistant Professor – SVIT, [3]Assistant Professor – SVIT, [4]Assistant Professor – SVIT
[1]Computer Engineering,
[1]Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

Abstract: Aircraft detection in aerial imagery is critical for civil and defense applications, yet it faces significant challenges, primarily due to the low resolution of images captured from long distances. These images often suffer from poor visibility due to adverse weather (fog, haze) and extreme variations in object scale, where aircraft may appear as just a few pixels. Deep learning models offer high accuracy but are often too computationally demanding for real-time deployment on resource-constrained platforms like Unmanned Aerial Vehicles (UAVs). This review paper comprehensively examines the role of Knowledge Distillation (KD) as an effective solution to this trade-off. KD allows a small, lightweight student model to inherit the high accuracy and generalization capabilities of a large teacher model, thereby enhancing detection efficiency and robustness without compromising inference speed. We classify existing KD techniques, review their application in aerial object detection—particularly for small and rotated objects—and discuss key datasets and evaluation metrics, concluding with promising future directions for developing weather-adaptive and multi-sensor fusion KD frameworks.

Index Terms – Knowledge distillation, aerial imagery, object detection, small objects, UAVs, lightweight models, weather robustness.

I. Introduction

Aircraft detection is an essential task in modern surveillance, border security, and civil aviation, relying heavily on images captured by satellites, high-altitude aircraft, or Unmanned Aerial Vehicles (UAVs). The ability to detect and track aerial targets rapidly and accurately is vital for situational awareness and time-critical decision-making.

Despite the advancements in deep learning, aerial object detection presents unique and difficult challenges. Firstly, the immense distance from which images are captured often results in aircraft targets appearing as small objects with low pixel resolution, making their features ambiguous and hard to discern from background clutter. Secondly, aerial imagery is highly susceptible to adverse environmental factors such as fog, rain, haze, and cloud cover, which severely degrade image quality and detection reliability [8], [10], [13]. Finally, high-performing detection models (e.g., Faster R-CNN) are typically large and computationally expensive, rendering them impractical for deployment on resource-limited edge devices like UAVs, where low latency and energy efficiency are paramount.

To overcome the trade-off between accuracy and efficiency, this paper focuses on Knowledge Distillation (KD) as a potent solution. Knowledge Distillation is a model compression technique where a smaller, more efficient student model is trained to mimic the comprehensive output (or "knowledge") of a larger, pre-trained teacher model. This process allows the student model to achieve performance levels comparable to the teacher while maintaining a lightweight architecture suitable for real-time applications.

The primary objective of this review is to provide a comprehensive overview of KD techniques specifically tailored for robust and efficient aircraft detection in challenging aerial environments. This paper is structured as follows: Section II reviews related literature and existing benchmarks. Section III details the core concepts and types of Knowledge Distillation. Section IV discusses how object detection methods, particularly those based on deep learning, are applied to address specific challenges in aerial imagery. Section V emphasizes the role of lightweight models and KD in facilitating on-platform deployment. Finally, Section VI concludes the paper and outlines future research directions

II. Literature Review and Related Work

This section summarizes prior works on Knowledge Distillation, aerial object detection, and related model compression and efficiency methods. Table I provides a snapshot of relevant research.

Title	Source	Year	Datasets	Summary
A Review of KD in Object Detection [1]	IEEE Access	2024	COCO, VOC, KITTI	Reviews KD methods; teacher–student misalignment issues
Novel KD for UAV small-object detection [2]	Complex Intell. Syst.	2025	UAV aerial datasets	Blur-robust KD for tiny objects; UAV-specific
Model compression using KD [3]	IEEE Access	2022	CIFAR-10/100, ImageNet	Comparative KD study; focused on classification
High-frequency feature fusion [4]	Remote Sens.	2025	Aerial imagery	Improves small-object detection (not KD-specific)
Instance-Conditional KD [5]	NeurIPS	2021	COCO	Instance-aware KD; boosts detection accuracy with overhead
EMD-based KD [6]	J. Phys. Conf. Ser.	2021	Detection datasets	Feature alignment using Earth Mover's Distance
Multiscale keypoint detection [7]	IEEE JSTARS	2021	Aerial datasets	Multiscale handcrafted detection (not KD)
DOTA dataset [8]	CVPR	2018	DOTA	Large-scale aerial dataset benchmark
PENet [9]	arXiv	2020	Aerial datasets	Point-based detection (no KD)
Dot Distance for Tiny Objects [10]	CVPRW	2021	Aerial datasets	Anchor-free detection for tiny objects (no KD)
KD Survey [11]	IJCV	2022	ImageNet, COCO, CIFAR	Comprehensive taxonomy of KD methods
Lightweight UAV detection [12]	Sensors	2023	UAV datasets	Lightweight detector; no KD
Rotated object detection survey [13]	IEEE Access	2024	Aerial datasets	Rotation-aware survey (no KD)
Recent KD survey [14]	ML Appl.	2024	Multiple benchmarks	Reviews advanced KD: attention, progressive, multitask

Lightweight aerial detection [15]	Sci. Rep	2024	Aerial datasets	Lightweight UAV detector; application focus
Intra-class Patch Swap for Self-Distillation [16]	ArXiv	2025	Image datasets	Self-distillation method; uses intra-class patch swapping
Tiansuan Constellation... [17]	SAGC 2022	2022	Microsatellite data	Software-defined microsatellite; orbital deep neural networks

Table 1- Summary Of Related Work on Knowledge Distillation and Aerial Object Detection

A. Knowledge Distillation Surveys and Model Compression

Numerous surveys have established the foundational role of KD in model compression, focusing on how a compact student model can inherit the generalization capabilities of a complex teacher model [3], [11], [14]. These works provide a taxonomy of KD methods, including response-based and feature-based distillation, essential for understanding the underlying mechanisms that enable high accuracy in resource-efficient architectures.

B.KD in Aerial and Remote Sensing Applications

The application of KD in aerial and remote sensing has been rapidly expanding. Specific strategies developed for this domain include:

- **Attention-Based Distillation [2], [5]:** Guiding student models to focus on critical image regions by mimicking the teacher's attention maps, which is crucial for tiny object detection.
- **Multi-Scale Feature Distillation [6]:** Transferring knowledge across multiple feature scales to ensure that fine-grained spatial details, necessary for recognizing distant or small targets, are preserved in the student.
- **Robust Distillation [2]:** Techniques specifically designed to improve the student's robustness against common aerial image degradation factors like blur and haze.

C.Rotated and Small Object Detection Challenges

The unique characteristics of aerial imagery—such as arbitrary object rotations and severe scale variations—have driven the development of specialized detection models [10], [13]. KD plays an integral role in this area by compressing high-performance, rotation-aware detectors (often computationally heavy) into faster, lightweight variants suitable for real-world deployment.

III. Knowledge Distillation (KD) Techniques

Knowledge Distillation is an effective model compression technique where a high-capacity teacher network T is used to train a compact student network S. The primary objective is to transfer the generalization capability of T to S, thereby maintaining high performance while significantly reducing computational and memory requirements [1], [11].

The teacher model is first trained to capture detailed feature representations and produce "soft targets" (e.g., probability distributions or feature maps). These soft targets act as an enriched supervisory signal for the student model, complementing the standard ground-truth labels.

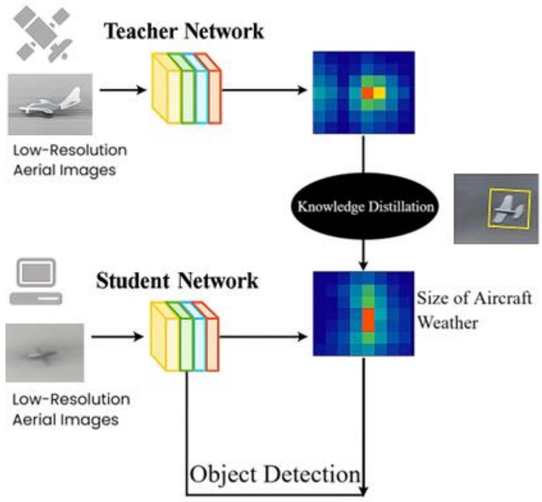

fig. 1. General illustration of the knowledge distillation framework

A. Types of Knowledge Distillation

- Response-Based Distillation (Logit Distillation) [3] - This is the original KD method, training the student model using the softened probability distributions (logits) produced by the teacher. The softening is typically achieved using a temperature hyperparameter (T) in the SoftMax function. These soft outputs reveal crucial class relationships that aid the student in better generalization, particularly useful in multi-class detection scenarios.
- Feature-Based Distillation (Feature Map Distillation) [5], [6] – Instead of using only the final outputs, this method transfers knowledge by aligning intermediate feature maps between the teacher and student. By learning from these deeper representations, the student model can absorb more spatial and semantic context, which is highly advantageous in aerial image detection where targets are often small or lack clear boundaries.
- Attention-Based Distillation [5] – Attention mechanisms are utilized to highlight the most relevant regions of an image. The student model is trained to replicate the teacher's attention maps, thereby learning to focus on critical areas, such as tiny aircraft, while effectively suppressing background noise. This targeted focus significantly improves detection accuracy in complex aerial scenes.
- Self-Distillation [2] – In this setup, a single model acts as both the teacher and the student. Knowledge is transferred internally, often from deeper layers to shallower layers, or from an earlier training stage to a later one. This enhances feature learning and boosts overall performance without the need for a separate, large teacher network.

IV. Object Detection In Aerial Images

Object detection in aerial images is fundamental for applications ranging from defence surveillance to disaster monitoring. Aerial views, acquired from high altitudes, present difficulties such as severe scale variations, arbitrary object rotations, and significant image degradation due to atmospheric effects.

A. Deep Learning-Based Detectors and Challenges:
Modern aerial detection heavily relies on deep learning architectures, typically divided into two categories:

1) Two-stage detectors (e.g., Faster R-CNN): These models first propose regions of interest (RoIs) and then classify and refine them. They offer high accuracy, especially for small objects [7], [13], but are slow.

2) Single-stage detectors (e.g., YOLO, SSD): These models perform localization and classification in one forward pass, emphasizing inference speed for real-time applications [12], [15]. However, they often trade off some precision, particularly with very small targets.

B.KD for Addressing Scale and Weather Challenges:
The limitations of conventional detectors in aerial environments are mitigated through KD-based strategies:

1) Scale Robustness: KD improves detection of small aircraft by transmitting features across various resolution levels (Multi-Scale Feature Distillation), helping the student preserve the subtle details of tiny targets. Anchor-free detection methods, when guided by KD, further improve robustness to varying object sizes [2], [5].

2) Weather Robustness: The teacher model, trained on diverse and augmented data, learns to generalize across various weather conditions. KD transfers this robust generalization ability to the lightweight student, enabling reliable detection even in fuzzy, hazy, or low-contrast aerial photographs.

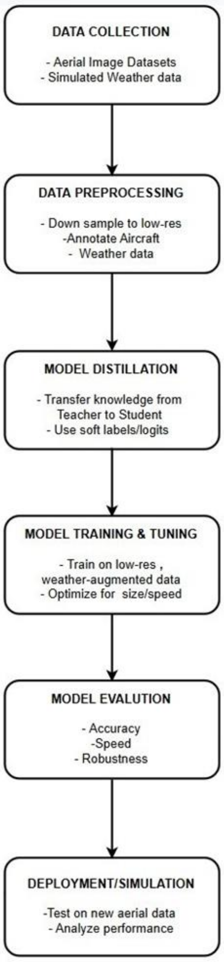

fig. 2. Typical flow of an object detection pipeline in aerial image analysis

V. Lightweight Models and Deployment Efficiency

A. Importance for UAVs and Embedded Systems
For applications involving UAVs, drones, or other embedded devices, the models must operate under strict constraints on computational power, memory, and energy consumption. Deploying massive models is often impossible, making lightweight and efficient object detectors a necessity for real-time performance in aerial surveillance [12], [15].

B.Role of Knowledge Distillation
KD is the most crucial technique for achieving this deployment efficiency. By compressing the knowledge from high-capacity networks (like VGG, ResNet) into compact student architectures (like MobileNet, ShuffleNet, or smaller YOLO variants), KD enables these lightweight models to maintain competitive detection accuracy and generalization capabilities while achieving significantly faster

Frames Per Second (FPS). This balance is critical for real-time aerial monitoring where low latency is essential.

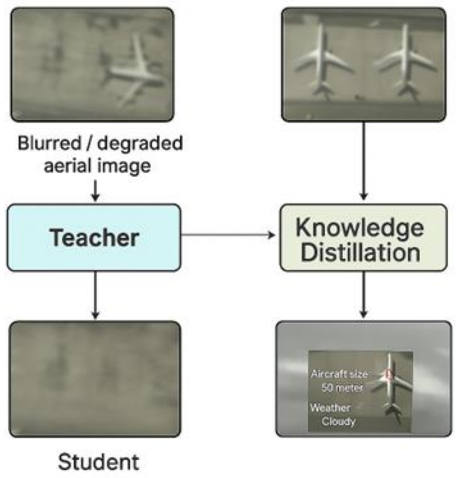

fig. 3. Conceptual diagram illustrating teacher-student knowledge transfer

C.Example of KD in Lightweight Architecture

- YOLOv5s with Knowledge Distillation [15]: Distilling the compact YOLOv5s from a larger YOLO model achieves superior performance and fast inference speeds suitable for high-speed UAV applications.
- MobileNet/ShuffleNet Adaptations through KD [12]: These architectures, specifically designed for mobile and embedded applications, are often enhanced with KD to boost their feature representation capabilities, ensuring robustness in complex aerial scenes while remaining highly efficient.

VI. Datasets and Benchmarks

Evaluating aerial detection performance requires specialized datasets and consistent metrics.

A. Common Aerial Datasets

Benchmark datasets are essential for rigorous evaluation:

1) DOTA [8]: A large-scale dataset featuring objects with extreme scale variance and arbitrary orientations, making it a standard benchmark for complex aerial imaging conditions.

2) VisDrone [10]: Focuses on drone-captured images, featuring small, densely packed objects across various environmental conditions, simulating real-world UAV surveillance.

3) UCAS-AOD [9]: Specifically designed for the detection of aircraft and vehicles in aerial images, providing detailed annotations for these target classes.

4) xView [16]: One of the largest public aerial datasets, covering diverse environments and conditions with over a million annotated instances.

B.Evaluation Metrics

The primary metrics for evaluating aerial object detectors are:

1) Mean Average Precision (mAP): The standard metric for detection accuracy, averaging precision across all categories and Intersection over Union (IoU) thresholds.

2) Frames Per Second (FPS): A critical metric for real-time applications, quantifying the model's inference speed.

3) Model Size/FLOPs: Measures the memory and computational complexity, which directly informs deployability on resource-constrained platforms.

VII. Future Directions and Open Challenges

Despite the success of KD, several avenues remain open for enhancing its efficacy in aerial object detection:

- Weather-Adaptive KD: Future research should focus on developing dynamic KD frameworks that can adjust their distillation loss or strategy based on real-time environmental factors (fog, rain, haze) to maximize detection reliability in rapidly changing aerial conditions.
- KD with Super-Resolution (SR): Integrating KD with image SR methods can address the issue of low-resolution aerial imagery before detection. The distilled knowledge can help student models retain crucial fine details reconstructed by the SR module, enhancing small-object identification [2], [4], [7].
- Multi-Sensor Fusion with KD: Incorporating data from complementary sensors (infrared, optical, radar) and fusing this knowledge via KD could allow student models to achieve enhanced detection precision under difficult conditions like nighttime or heavy cloud cover [17].
- Explainable Knowledge Distillation (XKD): To boost confidence and interpretability in critical areas like defence and UAV navigation, XKD aims to analyse the knowledge elements preserved in the student model, ensuring that the model is making decisions based on reliable and relevant visual cues.

VIII. CONCLUSION

Knowledge Distillation (KD) provides a vital bridge between the high accuracy of complex deep learning models and the stringent efficiency demands of real-time aerial surveillance systems. This technique is particularly effective in addressing the critical challenges of detecting small, low-resolution aircraft and maintaining robustness in adverse weather conditions. While KD successfully transfers the generalization power of large teacher networks to lightweight student models, open challenges remain in dynamic weather adaptation and integrating multi-sensor data. Future work in areas like Weather-Adaptive KD, combining KD with Super-Resolution, and developing Explainable KD frameworks will be essential for realizing highly reliable, universally deployable, and computationally efficient aerial detection systems.

REFERENCES

[1] S. Cheng, Q. Zhao, X. Zhang, N. Yadikar, and K. Ubul, “A Review of Knowledge Distillation in Object Detection,” IEEE Access, Jan. 2024, doi: 10.1109/ACCESS.2023.3288692.
[2] S. Jobaer, X.-s. Tang, Y. Zhang, G. Li, and F. Ahmed, “A novel knowledge distillation framework for enhancing small object detection in blurry environments with unmanned aerial vehicle-assisted images,” Complex Intell. Syst., vol. 11, art. 63, 2025, doi: 10.1007/s40747-024-01676-w.
[3] Y.-W. Hong, J.-S. Leu, M. Faisal, and S. W. Prakosa, “Analysis of model compression using knowledge distillation,” IEEE Access, vol. 10, pp. 85095–85105, Aug. 2022, doi: 10.1109/ACCESS.2022.3197608.
[4] X. Ge, L. Qi, Q. Yan, J. Sun, Y. Zhu, and Y. Zhang, “Enhancing real-time aerial image object detection with high-frequency feature learning and context-aware fusion,” Remote Sens., vol. 17, no. 12, art. 1994, 2025, doi: 10.3390/rs17121994.
[5] Z. Kang, P. Zhang, X. Zhang, J. Sun, and N. Zheng, “Instance-Conditional Knowledge Distillation for Object Detection,” in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 34, 2021, pp. 16468–16480. [Online]. Available: https://arxiv.org/abs/2110.12724
[6] Z. Liang, “A new knowledge distillation method for object detection based on EMD,” J. Phys.: Conf. Ser., vol. 2083, art. 042028, 2021, doi: 10.1088/1742-6596/2083/4/042028.
[7] J. Su, J. J. Liao, D. Gu, Z. Wang, and G. Cai, “Object detection in aerial images using a multiscale keypoint detection network,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. (JSTARS), vol. 14, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9293352
[8] G. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, “DOTA: A large-scale dataset for object detection in aerial images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, 2018, pp. 3974–3983, doi: 10.1109/CVPR.2018.00418.

[9] Z. Tang, X. Liu, G. Shen, and B. Yang, "PENet: Object detection using points estimation in aerial images," arXiv preprint arXiv:2001.08247, Jan. 2020. [Online]. Available: https://arxiv.org/abs/2001.08247
[10] C. Xu, J. Wang, W. Yang, and L. Yu, "Dot Distance for Tiny Object Detection in Aerial Images," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Nashville, TN, USA, June 2021, pp. 1192–1201, doi: 10.1109/CVPRW53098.2021.00150. [Online]. Available: https://ieeexplore.ieee.org/document/9522822
[11] G. Hinton, O. Vinyals, J. Dean, T. Gou, L. Yu, S. Bai, and T. Zhao, "Knowledge distillation: A survey," Int. J. Comput. Vis., vol. 130, no. 12, pp. 3367–3385, Dec. 2022, doi: 10.1007/s11263-022-01623-7.
[12] J. Wang, F. Zhang, Y. Zhang, Y. Liu, and T. Cheng, "Lightweight object detection algorithm for UAV aerial imagery," Sensors, vol. 23, no. 13, art. 5786, 2023, doi: 10.3390/s23135786.
[13] J. He and K. L. E. Law, "Deep learning models for rotated object detection in aerial images: Survey and performance comparisons," IEEE Access, vol. 11, pp. 87912–87934, 2024, doi: 10.1109/ACCESS.2024.3509745.
[14] A. Moslemi, A. Briskina, Z. Dang, and J. Li, "A survey on knowledge distillation: Recent advancements," Mach. Learn. Appl., vol. 18, art. 100605, Dec. 2024, doi: 10.1016/j.mlwa.2024.100605.
[15] L. Deng, Y. Zhang, X. Wang, Z. Li, and H. Chen, "Lightweight aerial image object detection algorithm based on improved YOLOv5s," Sci. Rep., vol. 13, no. 7817, 2023, doi: 10.1038/s41598-023-34892-4.
[16] H. Choi, E. S. Jeon, A. Shukla, and P. Turaga, "Intra-class Patch Swap for Self-Distillation," arXiv, art. 2505.14124, May 2025. [Online]. Available: https://arxiv.org/pdf/2505.14124
[17] F. He, C. Liu, M. Wang, E. Yang, and X. Liu, "Tiansuan Constellation: Intelligent software-defined microsatellite with orbital attention for Sustainable Development Goals," in The 2022 International Conference on Space-air-ground Computing (SAGC 2022), vol. 13864, Fiji, 2022.