Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38807

Comprehensive Evaluation of UAV Object Detection Models Using the Analytic Hierarchy Process and Entropy Weight Method

Ruoyu Zhang

School of Integrated Circuits, Beijing University of Technology, Beijing, China
z.mc_4079@qq.com

Abstract. A single accuracy metric cannot adequately describe how object detection models perform in unmanned aerial vehicle (UAV) applications. This study develops a multidimensional evaluation method by combining the analytic hierarchy process (AHP) with the entropy weight method. The index system is built around three dimensions, usability, efficiency, and robustness. AHP assigns subjective weights according to task requirements, while the entropy weight method calculates objective weights from variation in the experimental data. These two sets of weights are then combined to evaluate each model as a whole. Experiments were carried out on the VisDrone2019 dataset using five representative object detection models: YOLOv8s, YOLOv7-tiny, DINO-DETR, FCOS, and RetinaNet. The results indicate that YOLOv8s obtains the highest overall score of 86.48. YOLOv7-tiny and DINO-DETR follow with scores of 80.43 and 79.09.This method distinguishes differences among models across detection performance, computational efficiency, and robustness under complex imaging conditions. It therefore offers a quantitative basis for comparing and selecting UAV object detection models..

Keywords: UAV object detection, comprehensive evaluation, analytic hierarchy process, entropy weight method, combined weighting

1. Introduction

With the development of unmanned aerial vehicle (UAV) platforms and computer vision technology, UAV object detection has been widely applied to traffic monitoring, public safety, and environmental perception. Compared with ordinary natural images, UAV aerial images typically contain small and densely distributed targets, complex backgrounds, frequent occlusion, and substantial viewpoint variations, which impose higher requirements on detection accuracy, computational efficiency, and adaptability to complex environments [1-3].

YOLO, FCOS, RetinaNet, and DETR-family models based on Transformers have all developed quickly in recent years. Their network structures, feature modeling methods, and computational complexity differ quite a lot, so each model places different weight on detection accuracy, inference speed, resource consumption, and robustness [4, 5]. Most object detection studies still rely on metrics such as mAP and AP to judge model performance. That is not enough. In practical

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38807

applications for UAV platforms, real-time capability, hardware limits, and stability under complex imaging conditions also affect whether a model is actually suitable.

Multi-criteria evaluation methods make it possible to assess the same object from several dimensions at once. Among them, the analytic hierarchy process (AHP) determines the relative importance of indicators by building a hierarchical structure and pairwise judgment matrices, which makes it suitable for multi-level and multi-criteria decision problems [6, 7]. Existing research indicates that AHP can be used effectively in the comprehensive evaluation of complex alternatives. Some studies have also combined it with objective weighting methods to blueuce the limitations that come with relying only on subjective weights [7-9].

Based on this need, this study builds a multi-level evaluation index system for UAV object detection models from three dimensions: usability, efficiency, and robustness. It then combines AHP with the entropy weight method to evaluate five representative object detection models: YOLOv8s, YOLOv7-tiny, DINO-DETR, FCOS, and RetinaNet. After mapping heterogeneous indicators onto a unified scale and calculating comprehensive scores across all indices; we compare how these models perform across multiple dimensions so as to provide a reference for selecting UAV object detection models.

2. Related work

2.1. Object detection models

As deep learning has advanced, object detection has split into several main approaches, including anchor-based one-stage detection, anchor-free detection, and Transformer-based end-to-end detection [4, 5]. RetinaNet follows an anchor-based dense detection framework and introduces Focal Loss to blueuce the imbalance between foreground and background samples [5]. FCOS removes pblueefined anchors and predicts classes and bounding boxes directly on feature maps [5]. The YOLO family places more emphasis on balancing detection accuracy with real-time performance, while its lightweight variants further lower computational cost and make deployment easier [5]. More recently, Transformer-based detectors represented by DETR have made end-to-end pblueiction possible through object queries and global feature interaction, which has further broadened the range of object detection methods [5].

These frameworks differ clearly in how they model features, how much computation they require, and how they generate target pblueictions. Their performance priorities are therefore not the same. Recent studies have also kept their attention on dense small-object detection from drone viewpoints and remote-sensing object detection [10, 11]. Based on this background, YOLOv8s, YOLOv7-tiny, DINO-DETR, FCOS, and RetinaNet are selected as the evaluation objects so that multiple representative object detection paradigms are included.

2.2. Multi-criteria comprehensive evaluation

Multi-criteria comprehensive evaluation methods bring indicators with different units and different levels of importance into one evaluation framework. AHP turns qualitative judgment into quantitative weight through a hierarchical structure, judgment matrices, and consistency tests, and it has been used to evaluate complex multi-level alternatives [6, 7]. Earlier studies have also combined AHP with methods such as TOPSIS. After normalizing indicators and carrying out a comprehensive calculation, these studies evaluated multiple alternatives in the same framework [6, 7, 12]. This

DOI: 10.54254/2755-2721/2026.38807

combination suggests that subjective decision-making methods can work together with data-driven approaches and be used in a wider range of evaluation tasks.

Most existing comparisons of object detection models center on single metrics such as mAP and FPS. That creates a problem. These metrics do not show detection performance, computational efficiency, and adaptability to complex environments at the same time [1, 3, 4]. For this reason, this study combines AHP with the entropy weight method to carry out a multidimensional comprehensive evaluation of different object detection models while taking both task requirements and differences in indicator data into account.

3. Method

3.1. Construction of a multidimensional evaluation index system

Object detection models differ substantially in detection accuracy, computational efficiency, and adaptability to complex environments; a single indicator cannot adequately characterize their overall performance. Therefore, model comprehensive performance is defined as the target level, and a hierarchical evaluation system is constructed from three dimensions: usability, efficiency, and robustness. Let the sets of second-level indicators corresponding to the three first-level indicators be $\mathscr{C}_U$, $\mathscr{C}_E$, and $\mathscr{C}_R$, respectively. The evaluation index system can then be formally expressed as:

$$\mathscr{C} = \{\mathscr{C}_U, \mathscr{C}_E, \mathscr{C}_R\}, \qquad |\mathscr{C}_U| = 3,\ |\mathscr{C}_E| = 3,\ |\mathscr{C}_R| = 6. \tag{1}$$

Here, usability refers to a model's detection and localization performance, measured by mAP, AP@IoU=0.5, and AP@IoU=0.75. Efficiency refers to inference speed and resource overhead, measured by FPS, FLOPs, and GPU memory usage. Robustness describes how stable the model remains under complex imaging conditions and is evaluated by mAP in six perturbation scenarios: overexposure, underexposure, partial occlusion, Gaussian noise, Gaussian blur, and contrast variation. Together these form a multidimensional evaluation framework with 3 first-level indicators and 12 second-level indicators. This framework provides a common basis for later normalization, weighting, and comprehensive evaluation.

This hierarchical structure also defines how the second-level indicators enter the evaluation process. They are first compablue and aggregated within their own first-level category before the three dimensions are consideblue together at the overall level. That distinction matters. It prevents accuracy, computational cost, and perturbation stability from being treated as though they were the same kind of quantity. When one second-level indicator changes, it first affects its own dimension score and then enters the comprehensive score through its assigned weight. In this way, the link between raw measurements and final ranking is retained.

The three first-level dimensions correspond to different practical concerns in UAV detector selection: usability focuses on detection and localization capability; efficiency concerns inference speed and resource consumption; robustness addresses performance stability under common image degradations. The second-level indicators make these dimensions measurable in quantitative terms. Organizing them hierarchically also avoids combining semantically different metrics too early.

Within usability (mAP), AP@0., while AP@0.. The efficiency dimension combines FPS with FLOPs so runtime speed is judged alongside computational cost rather than on its own. Robustness uses after six image perturbations covering illumination change visibility noise blur contrast all of which can affect aerial images during acquisition conditions Taken together these 12 describe

DOI: 10.54254/2755-2721/2026.38807

complementary aspects of model behavior blueucing risk that final comparison depends on one narrow characteristic

3.2. Indicator quantification and normalization

Because the indicators differ in units, scales, and preference directions, they must be quantified on a common scale before comprehensive evaluation. Suppose there are m candidate models and n second-level indicators. Let the raw observation of model i on indicator j be x_{ij}. The original evaluation matrix is defined as:

$$\mathbf{X} = (x_{ij})_{m\times n}, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, n. \tag{2}$$

According to the original interval-based scoring rules, an indicator mapping function $\phi_j(\cdot)$ is introduced to map observations with different units to the interval $[0, 100]$:

$$s_{ij} = \phi_j(x_{ij}) \in [0, 100]. \tag{3}$$

To ensure that a higher score consistently represents better performance, the mapping function must satisfy the corresponding monotonicity constraints:

$$\frac{\partial \phi_j(x)}{\partial x} > 0, \quad j \in \mathscr{J}_P; \qquad \frac{\partial \phi_j(x)}{\partial x} < 0, \quad j \in \mathscr{J}_N. \tag{4}$$

Here, $\mathscr{J}_P$ denotes the set of benefit-type indicators, including mAP, $\mathrm{AP@IoU} = 0.5$, $\mathrm{AP@IoU} = 0.75$, FPS, and mAP under each perturbation scenario; $\mathscr{J}_N$ denotes the set of cost-type indicators, including FLOPs and GPU memory usage. After normalization, the unified score matrix $\mathbf{S} = (s_{ij})_{m\times n}$ is obtained.

The scoring transformation is applied before either weighting method is used, so both AHP and the entropy weight method operate on the same standardized matrix. For benefit-type indicators, the mapping must preserve the positive relationship between the original observation and the score; for cost-type indicators, the mapping reverses that relationship so that a lower resource cost corresponds to a higher evaluation score. This common orientation is necessary for additive aggregation. It ensures that an increase in any standardized indicator always has the same meaning in the later calculation and prevents the final score from combining mutually contradictory optimization directions.

A further purpose of this normalization step is to prevent the numerical scale of an indicator from influencing the evaluation independently of its actual importance. Without a common score range, an indicator expressed with large numerical values could contribute more strongly to an additive calculation simply because of its unit. Mapping all indicators through the same scoring framework separates scale effects from weighting effects, so that the later AHP and entropy weights determine importance rather than the original magnitude of the measurements.

3.3. Subjective weight modeling using the analytic hierarchy process

To represent the relative importance of evaluation indicators in UAV object detection tasks, AHP is used to model indicator importance hierarchically. Pairwise comparisons are performed among

DOI: 10.54254/2755-2721/2026.38807

indicators at the same level using Saaty's 1-9 scale to construct a judgment matrix [6]. For a level containing indicators, the judgment matrix is expressed as:

$$\mathbf{A} = (a_{pq})_{q\times q}, \qquad a_{pq} > 0, \quad a_{pq}a_{qp} = 1, \quad a_{pp} = 1. \tag{5}$$

The subjective weight vector $\boldsymbol{\omega}^{A}$ is obtained from the maximum eigenvalue of the judgment matrix and its corresponding eigenvector:

$$\mathbf{A}\boldsymbol{\omega}^{A} = \lambda_{\max}\boldsymbol{\omega}^{A}, \qquad \sum_{j=1}^{q}\omega_{j}^{A} = 1. \tag{6}$$

To test the consistency of the judgment matrix, the consistency index CI and consistency ratio CR are used:

$$\mathrm{CI} = \frac{\lambda_{\max}-q}{q-1}, \qquad \mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}. \tag{7}$$

When $\mathrm{CR} < 0.1$, the judgment matrix is considered to have acceptable consistency. Consistency tests are performed for the first-level indicators and their corresponding second-level indicators, and the eigenvectors of matrices that pass the tests are used as the AHP subjective weights.

In the AHP procedure, the Saaty 1-9 scale is used to compare the relative importance of two indicators at the same hierarchical level. The judgment matrix formed in this step is reciprocal. If one indicator is given a certain importance value relative to another, the opposite position in the matrix takes its reciprocal. The priority vector is calculated from the principal eigenvector and then normalized to obtain the subjective weight vector. Since these comparisons are carried out separately for the first-level indicators and for the second-level indicators within each dimension, AHP can capture both overall task priorities and the internal importance relationships among specific criteria.

A consistency test is also required in AHP because pairwise judgments can contain logical inconsistencies if they are not checked. For instance, if one criterion is considered more important than a second, and that second criterion more important than a third, then their implied relationship should remain broadly consistent across all three criteria. The consistency ratio gives a quantitative way to test this point. Only judgment matrices that meet the accepted threshold are used to calculate subjective weights. This matters because it reduces arbitrary comparisons and makes the resulting priority vector more reliable within the proposed evaluation structure.

3.4. Objective weight modeling using the entropy weight method

AHP reflects task requirements but still incorporates subjectivity from expert judgment. To introduce information on variation within the data, the entropy weight method is further used to calculate objective weights [8, 9, 12]. Based on the normalized score matrix , the proportion of model under indicator is first calculated as:

$$p_{ij} = \frac{s_{ij}}{\sum_{i=1}^{m} s_{ij}}. \tag{8}$$

The information entropy of indicator j is then calculated as:

$$e_{j} = -\frac{1}{\ln m}\sum_{i=1}^{m} p_{ij}\ln p_{ij}, \qquad 0 \le e_{j} \le 1. \tag{9}$$

DOI: 10.54254/2755-2721/2026.38807

The difference coefficient and entropy weight of the indicator are respectively expressed as:

$$d_j = 1 - e_j, \qquad \omega_j^E = \frac{d_j}{\sum_{r=1}^{n} d_r}. \tag{10}$$

A smaller entropy value indicates greater dispersion of the indicator across models and thus stronger discriminative ability, resulting in a larger objective weight. In this way, experimental data can complement the subjective weights obtained by AHP.

For the entropy weight method, the standardized score matrix is first converted into proportions that describe the relative contribution of each model under each indicator. Information entropy is then calculated from these proportions, after which the difference coefficient is obtained and normalized to generate the objective weights. The logic of this calculation is tied to dispersion across the five candidate models: an indicator with similar values for all models contributes less information for discrimination, whereas an indicator with clearer differences contributes more. Consequently, the objective weight is determined by the observed comparative information in the experiment rather than by an externally assigned preference.

The entropy weight should be interpreted as a measure of discriminative information rather than as a direct measure of practical importance. A large objective weight means that an indicator differentiates the candidate models more clearly in the observed data, but it does not necessarily mean that the indicator is more important for every UAV application. This distinction is one reason for combining entropy weighting with AHP: the former reflects data variation, while the latter represents the task-oriented importance assigned within the evaluation hierarchy.

3.5. Combined weighting and comprehensive evaluation

To account for both task prior knowledge and data variation, a linear combination is used to integrate the AHP subjective weights and entropy-based objective weights [8, 9, 12]. Let the two weights for indicator be and , respectively. The combined weight is:

$$\omega_j = \alpha\omega_j^A + (1-\alpha)\omega_j^E, \qquad 0 \le \alpha \le 1. \tag{11}$$

This study follows the equal-weight fusion setting of the original scheme, i.e., $\alpha = 0.5$, while the specific weights of individual indicators are not listed separately. After the combined weights are obtained, the scores of all second-level indicators are aggregated using a weighted sum. The comprehensive evaluation score of model i is:

$$F_i = \sum_{j=1}^{n} \omega_j s_{ij}, \qquad i = 1, \ldots, m. \tag{12}$$

Local weighting can also be carried out separately within the three first-level dimensions, usability, efficiency, and robustness, to obtain the score for each dimension. In this way, detection performance, computational cost, and adaptability to complex environments are brought into one quantitative framework, so different detection models can be compared and ranked on the same basis.

In the present scheme, equal-proportion fusion gives subjective and objective components the same overall influence. This avoids a final ranking driven too heavily by either expert preference or data dispersion. After the combined weights are set, the standardized second-level scores are aggregated into first-level scores and then into a comprehensive score. The interpretation is direct.

DOI: 10.54254/2755-2721/2026.38807

Because all stages use scores with the same positive orientation, a larger comprehensive score means that a model performs better overall across the selected criteria. The framework therefore provides both an overall ranking and dimension-specific results that can support deployment decisions.

This weighting structure also allows interpretation at the dimension level. The second-level indicators under usability, efficiency, and robustness are first aggregated within each dimension; these three first-level results are then consideblue together in the overall evaluation. As a result, it is possible to trace a model's final ranking back to the dimensions and indicators that shape it most clearly. The comprehensive score is therefore best understood as a summary of the evaluation process rather than as a substitute for the more detailed dimension-level results used to explain model behavior.

4. Experiments

4.1. Experimental setup

To evaluate the applicability of the proposed multidimensional comprehensive evaluation method, five representative object detection models—YOLOv8s, YOLOv7-tiny, DINO-DETR, FCOS, and RetinaNet—are compared. Experiments use the VisDrone2019 object detection dataset, which is characterized by small and densely distributed targets, complex backgrounds, frequent occlusion, and substantial viewpoint variations [1, 2]. All models are trained and tested using the same data split, evaluation metric definitions, and data processing pipeline. The basic evaluation metrics include mAP, AP@IoU=0.5, AP@IoU=0.75, FPS, FLOPs, and GPU memory usage. In addition, six perturbation scenarios—overexposure, underexposure, partial occlusion, Gaussian noise, Gaussian blur, and contrast variation—are constructed from the original test images to evaluate model robustness. All indicators are normalized and combined according to the method described in Section 3; only the final evaluation results and analysis are presented in this section.

4.2. Experimental results and analysis

4.2.1. First-level indicator evaluation

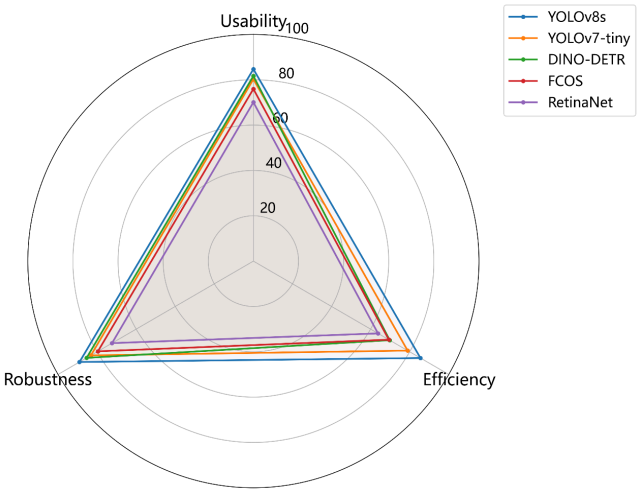


Figure 1. Radar chart of first-level indicators: usability, efficiency, and robustness

DOI: 10.54254/2755-2721/2026.38807

Figure 1 shows the evaluation results of the five models for the three first-level indicators of usability, efficiency, and robustness. YOLOv8s achieves relatively high scores in all three dimensions and exhibits the most balanced overall performance; YOLOv7-tiny shows relatively small differences across dimensions, indicating good overall applicability; DINO-DETR performs strongly in usability and robustness but has relatively low efficiency; FCOS and RetinaNet score below the top three models overall, with RetinaNet showing comparatively weaker performance across all three dimensions.

The first-level radar chart further reveals performance trade-offs that cannot be captured by the scalar comprehensive score alone. YOLOv8s exhibits the largest and most balanced profile across the three axes, indicating that its overall advantage is distributed across multiple dimensions rather than being driven by a single exceptionally high score. YOLOv7-tiny shows a relatively compact profile with stronger efficiency, whereas DINO-DETR retains competitive usability and robustness but exhibits relatively lower efficiency. FCOS and RetinaNet occupy smaller overall profiles. These patterns further justify multidimensional evaluation, because models with similar overall capability may still differ substantially in dimensions that are critical to particular deployment scenarios.

4.2.2. Second-level indicator evaluation

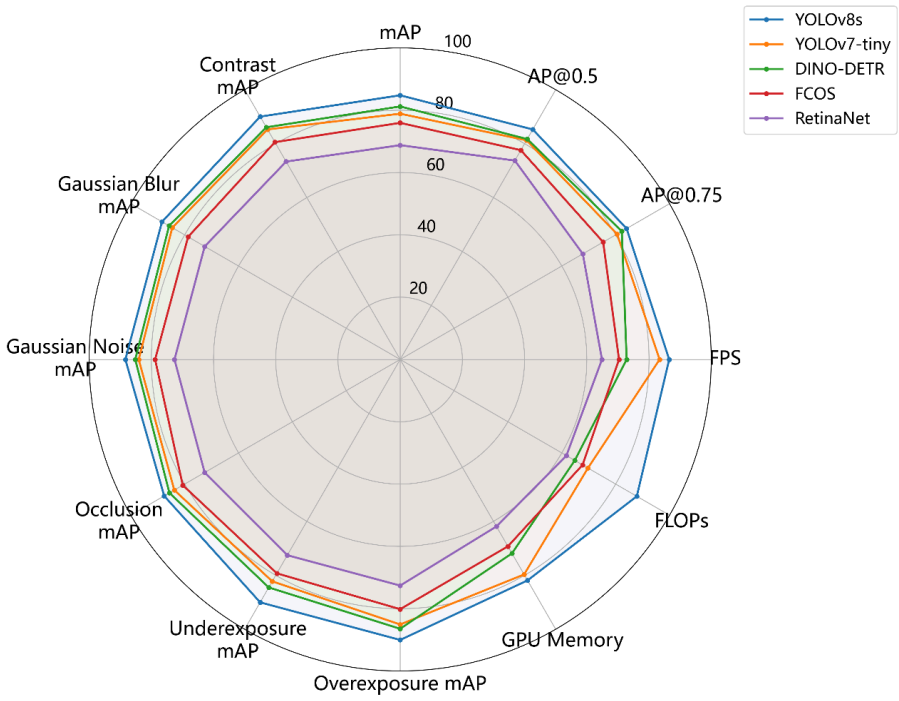


Figure 2. Radar chart of the twelve second-level indicators

Figure 2 presents the evaluation results of the five models across the 12 second-level indicators. YOLOv8s maintains strong performance in detection accuracy, inference efficiency, and multiple complex perturbation scenarios, showing a relatively balanced overall advantage; YOLOv7-tiny is highly competitive on efficiency-related indicators, although some accuracy and robustness indicators are slightly lower than those of YOLOv8s; DINO-DETR performs relatively stably in localization accuracy, occlusion, and complex degradation scenarios, but is comparatively weaker in computational efficiency and resource usage; FCOS and RetinaNet show relatively limited overall performance across most second-level indicators.

DOI: 10.54254/2755-2721/2026.38807

At the second level, the radar chart provides a finer explanation of the first-level differences. YOLOv8s remains close to the outer region on most accuracy, efficiency, and perturbation-related axes, which explains its balanced first-level profile. YOLOv7-tiny is especially competitive on efficiency-related indicators but is slightly weaker on several accuracy and robustness measures. DINO-DETR performs well on localization and several degraded-image conditions, while its computational indicators reduce its efficiency score. The lower and less uniform profiles of FCOS and RetinaNet indicate that their final rankings are caused by accumulated disadvantages across several criteria rather than by a single isolated metric.

4.2.3. Comprehensive evaluation results

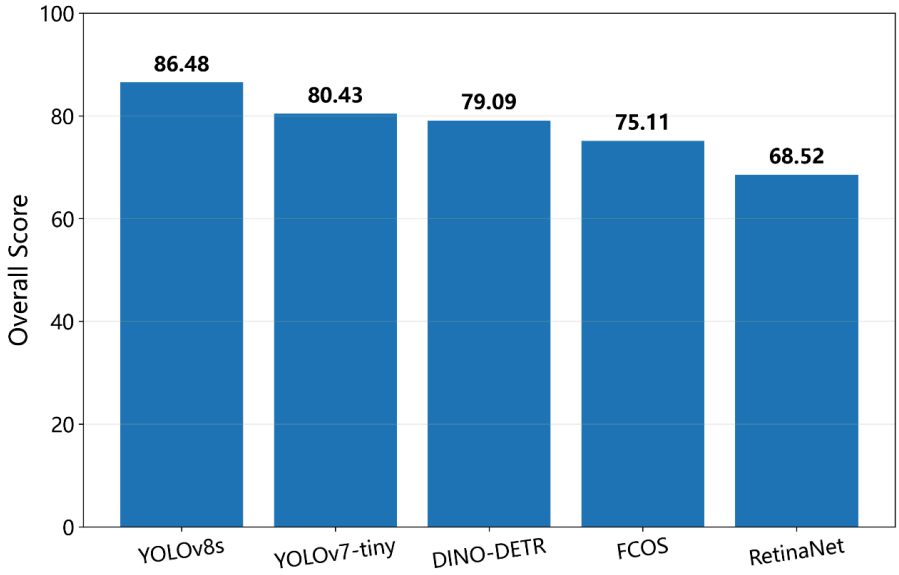


Figure 3. Bar chart of comprehensive evaluation scores

Figure 3 shows the comprehensive evaluation scores of the five models. YOLOv8s obtains the highest score of 86.48; YOLOv7-tiny and DINO-DETR rank second and third with scores of 80.43 and 79.09, respectively; FCOS and RetinaNet score 75.11 and 68.52, ranking fourth and fifth.

The results reveal clear performance differences among the five models. YOLOv8s achieves the best overall balance in detection performance, efficiency, and robustness, outperforming YOLOv7-tiny and DINO-DETR by 6.05 and 7.39 points, respectively. YOLOv7-tiny and DINO-DETR differ by only 1.34 points, indicating comparable overall performance despite their different strengths. FCOS and RetinaNet rank lower, with relatively weaker performance across the evaluated criteria.

As shown in Figures 1 and 2, YOLOv7-tiny demonstrates greater computational efficiency, whereas DINO-DETR performs better in usability and robustness but is limited by higher computational demands. The first-level and second-level indicators explain these differences, showing that similar comprehensive scores can result from different performance characteristics.

Overall, the combined AHP–entropy weighting framework provides a balanced assessment of detection accuracy, computational efficiency, and robustness. Under the proposed evaluation criteria, YOLOv8s offers the strongest overall performance, while YOLOv7-tiny is more suitable for efficiency-constrained applications and DINO-DETR remains competitive in complex environments. Therefore, comprehensive scores should be interpreted alongside the radar charts to support model selection under different UAV deployment requirements.

DOI: 10.54254/2755-2721/2026.38807

5. Conclusion

To address the limitation that a single performance metric cannot fully characterize the overall performance of object detection models, this study develops a multidimensional comprehensive evaluation method for UAV object detection based on AHP and the entropy weight method, providing a unified quantitative evaluation from the three dimensions of usability, efficiency, and robustness. Validation on the VisDrone2019 dataset with YOLOv8s, YOLOv7-tiny, DINO-DETR, FCOS, and RetinaNet shows that YOLOv8s achieves the best overall performance, while YOLOv7-tiny and DINO-DETR respectively demonstrate strong lightweight comprehensive performance and adaptability to complex scenes. The experimental results indicate that multidimensional comprehensive evaluation can reflect performance differences among models more comprehensively than a single accuracy metric and can provide a reference for comparing and selecting UAV object detection models. The proposed evaluation system is still affected by indicator selection and the specification of subjective weights; future work may further verify its applicability and stability using additional datasets, models, and deployment environments.

References

[1] Hua, W., & Chen, Q. (2025). A survey of small object detection based on deep learning in aerial images. *Artificial Intelligence Review, 58*, 162.
[2] Chen, Z., Wang, H., Wu, X., et al. (2024). Object detection in aerial images using DOTA dataset: A survey. *International Journal of Applied Earth Observation and Geoinformation, 134*, 104208.
[3] Zhao, C., Liu, R. W., Qu, J., et al. (2024). Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons. *Engineering Applications of Artificial Intelligence, 128*, 107513.
[4] Dogra, A. K., Sharma, V., & Sohal, H. (2024). A survey of deep learning techniques for detecting and recognizing objects in complex environments. *Computer Science Review, 54*, 100686.
[5] Sun, Y., Sun, Z., & Chen, W. (2024). The evolution of object detection methods. *Engineering Applications of Artificial Intelligence, 133*, 108458.
[6] Cui, Z., Taiwo, O. L., & Aaron, P. M. (2024). An application of AHP and fuzzy entropy-TOPSIS methods to optimize upstream petroleum investment in representative African basins. *Scientific Reports, 14*, 6956.
[7] Li, S., Zhou, R., & Huang, H. (2025). Multidimensional evaluation of autonomous driving test scenarios based on AHP-EWN-TOPSIS models. *Automotive Innovation, 8*, 237-251.
[8] Liu, J., Zhang, Q., Xie, M., et al. (2024). A blockchain platform selection method with heterogeneous multi-criteria decision-making based on hybrid distance measures and an AHP-EWM weight method. *Expert Systems with Applications, 256*, 124910.
[9] Kalyan, S., Bansal, P., & Kumar, S. (2026). A novel hybrid AHP-entropy weighted dynamic network DEA framework for comprehensive efficiency assessment. *Expert Systems with Applications, 296*, 129137.
[10] Li, W., Xi, Y., Jia, W., et al. (2025). PointSR: Self-regularized point supervision for drone-view object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11707-11716).
[11] Tang, D., Cao, X., Wu, X., et al. (2025). AeroGen: Enhancing remote sensing object detection with diffusion-driven data generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3614-3624).
[12] Xu, C., Xu, C., Zhan, L., et al. (2025). Establishment and validation of a novel composite index for energy efficiency evaluation of data center chilled water systems based on AHP and entropy weight method. *Energy and Buildings, 337*, 115677.