UAV-to-Camera Distance Estimation Based on a Passive Monocular Locator

Iryna Yurchuk[1,†], Taras Semenchenko[1,†] and Danyil-Mykola Obertan[1,*,†]

[1] *Taras Shevchenko National University of Kyiv, Bohdan Hawrylyshyn str. 24, Kyiv, UA-04116, Ukraine*

Abstract

In modern development of information technologies, object-to-camera distance estimation is an important task that involves extracting additional information from an image. Successfully solving such a task contributes to the automation of many processes, both microscopic (distance from atomic nuclei, etc.) and macrocosmic phenomena (estimation of distances between galaxies, stars, etc.). UAV-to-camera distance estimation has medium complexity, since the measurement accuracy is in the hundreds of meters.

In this work, the authors of the study consider three passive monocular approaches to UAV-to-camera distance estimation: geometry-calibrated models, methods of combining state-of-the-art 2D detectors (YOLO) with classical machine learning regressors and end-to-end multitask networks. On a custom dataset covering distances up to 983.96 m under multiple optical configurations, it is obtained that the geometry-calibrated size-agnostic model achieved the best overall accuracy under ideal localization, with an MAE of 15.33 m and R2 of 0.9672. To remark that the two-stage YOLO-based pipelines provided the best practical balance between accuracy and deployability, with MAE values around 22 m.

Keywords

UAV distance estimation, monocular vision, geometry-based ranging, multitask learning

1. Introduction

The growing use of Unmanned Aerial Vehicles (UAVs) requires advanced tactical defense systems. Traditional active sensors, such as radar and LiDAR, emit signals that can accidentally reveal defensive positions, which creates a strong need for passive optical detection. While vision-based systems provide accurate and stealthy 2D target detection, obtaining 3D spatial information, specifically metric distance, from a single passive 2D sensor remains a major technical limitation.

Estimating depth from a single 2D image is a naturally difficult mathematical problem, because many different 3D scenes can produce the same 2D projection. Although deep learning has improved monocular depth estimation for ground-level tasks [1-5] such as autonomous driving, aerial targets create several additional challenges. One of the popular approaches to determining the distance between the camera that takes the picture and the object that is captured in the frame is to find elements of the object whose dimensions are known in real life [6-8]. In UAV estimation, algorithms must work at very long distances, handle strong scale changes caused by different focal lengths, and operate against plain sky backgrounds that lack common geometric depth cues.

This paper evaluates three fundamentally different architectural approaches for passive monocular ranging at extreme tactical distances:

- Geometry-calibrated models: using perspective projection and bounding-box scaling.
- Two-stage pipelines: combining state-of-the-art 2D detectors (YOLO) with classical machine learning regressors.
- End-to-end multitask networks: jointly learning classification and depth estimation directly from raw pixels.

[1,*] Corresponding author.
[†] These authors contributed equally.
i.a.yurchuk@gmail.com (I. Yurchuk); taras.semenchenko@knu.ua (T. Semenchenko); danyilmykolaobertan@gmail.com (D.-M. Obertan)
0000-0001-8206-3395 (I. Yurchuk); 0009-0007-3259-7007 (T. Semenchenko); 0009-0003-8667-6967 (D.-M. Obertan)

The purpose of this work is to study the peculiarity of UAV-to-camera distance estimation and methods for obtaining them based on passive monocular approaches with high efficiency within distances up to 1 kilometer.

To address an important gap in the literature, we benchmark these methods on the dataset with distances of up to 1000 meters across different optical setups. This comparative analysis gives practical insights into the trade-offs between geometric limits and lack of data in deep learning.

2. Related works

2.1. Vision-based UAV detection

Recent studies have explored passive optical UAV detection as an alternative to traditional radio-frequency sensing, especially for compact aerial targets with low radar visibility [9, 10]. Vision-based systems have shown strong performance in UAV recognition and localization while also supporting intuitive scene interpretation and low-cost deployment [10]. In particular, recent works have evaluated real-time detectors such as YOLOv10 and RT-DETR for tactical UAV detection on resource-constrained platforms, showing that fine-tuned optical models can provide a practical solution for field deployment [11, 12]. However, most of these studies focus mainly on detection and classification, while accurate distance estimation remains less developed, and it is still unclear how well such detectors support metric ranging when the target shrinks to a few pixels.

2.2. Monocular ranging and depth estimation for Aerial Targets

Monocular depth estimation has been widely studied in areas such as autonomous driving and indoor robotics, where deep learning models can use rich scene structure and contextual depth cues [13]. In contrast, aerial UAV scenarios are much less represented in the literature. Small target size, long observation distances, scale variation across optical setups, and low-texture sky backgrounds make direct transfer of standard monocular depth methods difficult [10, 14]. Existing UAV ranging studies therefore often rely on geometric formulations based on camera parameters and target size [14, 15, 16] or on hybrid pipelines that combine detection outputs with separate regression models [12, 17]. End-to-end depth prediction has also been explored, but performance usually drops at long distances and under limited-data conditions [18, 19, 20]. Overall, large-scale comparisons of different monocular ranging strategies for UAVs at long operational ranges remain limited, and each family has its own weak point: geometric methods rely on accurate localization and known target size, hybrid pipelines inherit detector errors, and end-to-end networks need more data than is typically available for aerial targets.

2.3. Geospatial integration and the data gap

Accurate passive ranging is important not only as a standalone computer vision task, but also as a component of larger defense-oriented monitoring systems. For example, integrating vision-based detections into geospatial movement analysis frameworks requires reliable distance estimates to recover approximate target position in 3D space [21].

Despite this need, there is still a clear gap in the literature. Most existing UAV datasets and ranging benchmarks are limited to short operational distances, usually below 100 m [14, 18]. Systematic comparisons of monocular ranging methods under extreme tactical distances on real-world UAV data are still largely missing. Existing benchmarks also tend to fix a single optical setup and few UAV types, which makes it hard to separate scale effects from target-type effects. This paper addresses these gaps by evaluating multiple ranging strategies on a custom dataset covering distances from 2.02 m to 983.96 m across 14 focal lengths and 5 UAV classes.

3. Dataset and exploratory data analysis (EDA)

This study uses a custom monocular UAV dataset [22] of 4074 annotated images covering 5 UAV classes: fpv-quadcopter, shahed-131, zala-421-16e2, lancet, and zala-lancet-z-52. Each image contains one target annotated in COCO format with a class label, bounding box, focal length metadata (lens_mm), and ground-truth target distance (distance_m). All images have a fixed resolution of 1920 × 1080, providing a consistent geometric reference. Representative samples from the dataset are shown in Figure 1.

Figure 1: Sample images from the custom monocular UAV dataset.

To increase variation in imaging conditions, the dataset was collected with controlled changes in viewpoint, target pose, lighting, and scene appearance. The optical setup includes 14 focal lengths from 14 mm to 400 mm, allowing the same target class to appear at very different image scales. Target distance ranges from 2.02 m to 983.96 m, with a median of 61.85 m, a mean of 138.81 m, and a 90th percentile of 388.84 m. This makes the task more difficult than standard monocular ranging benchmarks because the target size varies from large close-range objects to extremely small long-range detections.

Although class counts are relatively balanced overall, the dataset is strongly imbalanced across distance ranges. Across all classes, 1796 samples fall into the 0–50 m interval (44.1%), 1140 into 50–150 m (28.0%), 741 into 150–400 m (18.2%), and only 397 beyond 400 m (9.7%). The hardest long-range cases are therefore the least represented. This imbalance also varies by class: for example, fpv-quadcopter appears only at shorter distances, while shahed-131 is better represented at long range.

Overall, the dataset combines a wide focal-length range with target distances spanning nearly three orders of magnitude, making it suitable for evaluating both geometric and learning-based ranging methods.

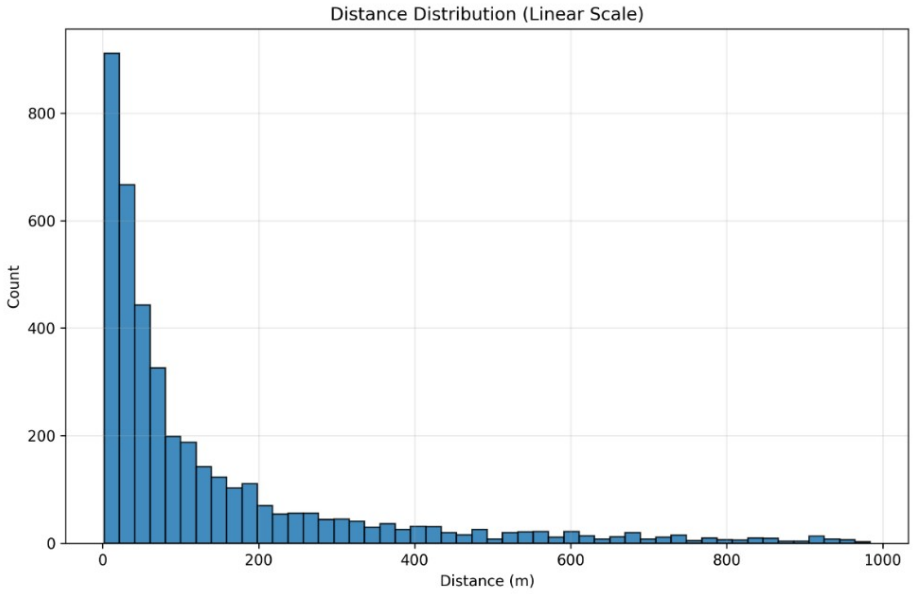

Figure 2: Distance distribution of the dataset.

Figure 2 shows a long-tail distance distribution: most samples are concentrated at short ranges, while sample density drops substantially beyond 400 m. This confirms that the most difficult long-range regime is also the least represented in the data.

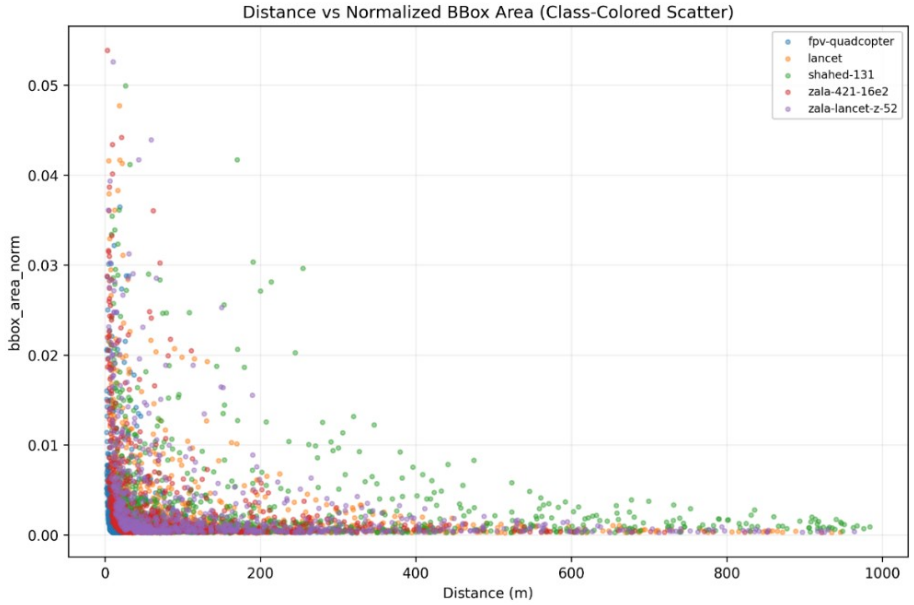

Figure 3: Relationship between target distance and normalized bounding-box area.

Figure 3 shows a clear inverse relationship between target distance and normalized bounding-box area. More distant UAVs occupy a much smaller image region, consistent with perspective geometry. At the same time, noticeable spread remains at similar distances because of focal length, target shape, and viewing pose.

Overall, the dataset is challenging because of the combined effects of wide distance range, scale variation, class-dependent visibility, and limited long-range coverage. These properties make it suitable for evaluating how different monocular ranging methods behave under realistic UAV observation conditions.

4. Methodology

All methods were trained and evaluated on the same dataset split generated with a fixed random seed, using a 70/15/15 split for training, validation, and testing. The evaluated approaches fall into three groups: geometry-calibrated models, a two-stage detection-regression pipeline, and an end-to-end multitask network.

4.1. Geometry-calibrated models

The first group of methods is based on the physical pinhole camera model and related geometry-based monocular ranging formulations [14, 15]. Under perspective projection, the visible size of an object decreases as distance increases, while the projected size becomes larger with a longer focal length. Using the normalized bounding-box area $bbox_{\text{area\_norm}}$, the raw distance estimate can be written as proportional to:

$$distance_{\text{raw}} \propto \frac{lens_{\text{mm}}}{\sqrt{bbox_{\text{area\_norm}}}} \tag{1}$$

To reduce the effect of dataset-specific bias, imperfect lens metadata, annotation noise, and deviations from the ideal projection model, a global affine calibration was fitted on the training split and then applied to all raw predictions:

$$distance_{\text{pred}} = max\left(0, a \cdot distance_{\text{raw}} + b\right) \tag{2}$$

During the experiments, it was found that the long or flat aerodynamic shapes of different UAVs can strongly affect the projected bounding-box area depending on the viewing angle and pose. Because of this, accounting for shape-related geometric distortion is important for accurate distance prediction.

Although the theoretical evaluation of these geometry-calibrated models uses ground-truth bounding boxes to isolate the quality of the ranging formulation itself, practical deployment depends on an object detector. In a fully autonomous pipeline, YOLO provides the bounding-box coordinates and class labels required for the geometric equations.

4.1.1. Ground truth UAV size

This variant uses the annotated physical target size $target_{\text{size\_m}}$ and estimates one global scale coefficient from the training set:

$$distance_{\text{raw}} = \frac{k_{\text{global}} \cdot lens_{\text{mm}} \cdot target_{\text{size\_m}}}{\sqrt{bbox_{\text{area\_norm}}}} \tag{3}$$

The coefficient k_{global} is estimated robustly as the median scale implied by the training samples. This formulation assumes that the true physical UAV size is known during inference.

4.1.2. Size agnostic

The size-agnostic variant removes the explicit physical-size term and includes all scale and geometric information in class-specific coefficients:

$$distance_{\text{raw}} = \frac{k_{\text{class}} \cdot lens_{\text{mm}}}{\sqrt{bbox_{\text{area\_norm}}}} \tag{4}$$

This design is more practical when the exact UAV dimensions are not available. Instead of using known physical size, the model relies on the predicted object class and the observed image geometry to estimate the missing scale information. By learning a separate k_{class} for each UAV category from the training split, this formulation also adapts to class-specific shape and appearance differences. For example, it can partly account for the flatter profile of fixed-wing drones compared with the more volumetric shape of quadcopters without requiring explicit size input.

4.2. Two-stage pipeline

The second group of methods follows a modular two-stage design, combining recent advances in vision-based UAV detection with separate distance regression [11, 12]. In Stage 1, a YOLO detector is trained for autonomous UAV localization and classification. The detector was initialized from yolov10n.pt and trained for 100 epochs with image size 960 and batch size 16. For each image, it outputs a bounding box, predicted class, and detection confidence.

In Stage 2, these outputs are converted into a structured feature vector and passed to a traditional machine learning regressor for distance estimation. The feature set includes:

- predicted UAV class;
- one-hot encoded predicted class indicators;
- detection confidence;
- normalized bounding-box geometry;
- focal length;
- binary indicator showing whether a detection was produced.

The bounding-box representation is normalized by image width and height so that the downstream regressors receive scale-consistent inputs. Distance targets are learned in $\log\left(1+distance_{\text{m}}\right)$ space and converted back to metric distance during inference.

Three regression heads were evaluated in this work:

- Support Vector Regression with an RBF kernel ($C=10.0$, $\epsilon=0.05$);
- LightGBM regression (2000 boosting rounds, learning rate 0.03, num leaves 63, with early stopping on the validation set);
- Random Forest regression (400 trees).

Model selection in this group was based on validation MAE. This pipeline is fully autonomous at inference time because both localization and class prediction are produced by YOLO before regression is applied.

4.3. End-to-End multitask model

The third approach is an end-to-end multitask convolutional neural network that jointly learns UAV classification and distance estimation directly from the input image, following the general direction of learning-based monocular distance prediction [18]. The model uses a ResNet18 backbone initialized with ImageNet weights to extract high-level visual features from resized RGB images. We did not apply additional pretraining on monocular depth datasets such as KITTI or NYU Depth v2, as these cover ground-level and indoor scenes and provide limited transfer to aerial targets observed against low-texture sky backgrounds. Images were resized to 384×384, and the focal-length metadata ($lens_{mm}$) was normalized by the maximum lens value in the dataset.

Along with the image features, the scalar focal-length metadata is also given directly to the network. After feature extraction, the normalized focal length value is concatenated with the flattened backbone representation and passed through a fusion layer consisting of a linear projection, ReLU activation, and dropout. This fused representation is then shared by task-specific heads for UAV classification and distance regression, where distance is predicted in $\log\left(1+distance_{\mathrm{m}}\right)$ space.

The network was trained for 40 epochs with batch size 16 using AdamW with a learning rate of 1×10^{-3} and weight decay 1×10^{-4}. The multitask objective combines cross-entropy loss for classification and L1 loss for distance prediction. During training, model selection is based on the combined validation loss, and the best checkpoint is kept for final evaluation.

Compared with the geometry-calibrated models, this approach does not require explicitly defined bounding boxes to compute geometric priors. Compared with the two-stage pipeline, it replaces hand-designed feature extraction and separate sequential regressors with a single jointly optimized network.

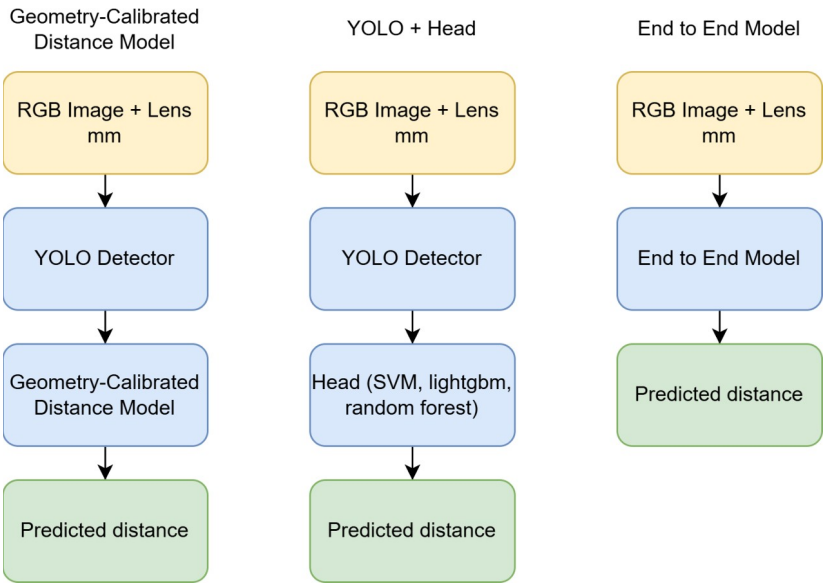

Figure 4: Comparison of the evaluated monocular UAV ranging pipelines.

5. Experimental setup and evaluation

5.1. Evaluation metrics

Distance estimation performance was evaluated using the following regression metrics: Mean Absolute Error (MAE); Root Mean Square Error (RMSE); Median Absolute Error (MedAE); Mean Absolute Percentage Error (MAPE); Coefficient of Determination (R^2). In addition, two practical accuracy measures were reported: accuracy within 10 meters and accuracy within 10%.

To examine robustness across distance ranges, the test samples were divided into four intervals: 0–50 m, 50–150 m, 150–400 m, and above 400 m. Error statistics were computed for each interval to show how performance changed with increasing distance.

For the autonomous pipeline and the multitask model, evaluation also included detector-related information such as detection success, classification correctness, and bounding-box overlap. These values were not the primary focus of the study, but they helped explain downstream distance estimation performance.

5.2. Fairness of comparison and evaluation scenarios

The evaluation considered two scenarios: theoretical ranging performance and fully autonomous ranging performance.

The geometry-calibrated methods were evaluated in an oracle-box setting, where ground-truth bounding boxes were provided to the distance model. This measured the upper-bound performance of the ranging formulation under ideal localization.

The two-stage pipeline and the end-to-end multitask model were evaluated in a detected-box setting, where localization and classification were produced automatically. Their reported distance errors therefore included both regression error and upstream detection error.

For this reason, direct numerical comparison between geometry-based methods and autonomous pipelines should be interpreted carefully. The oracle-box setting reflects the best-case potential of geometric ranging, while the detected-box setting reflects the practical performance of deployable end-to-end systems.

6. Results and discussion

6.1. Quantitative results

The size-agnostic geometry-calibrated model achieved the best overall results, showing that geometric ranging remains highly effective when target localization is ideal. The class-aware geometric variant performed slightly worse, which suggests that adding class-specific size assumptions did not improve robustness on this dataset.

Among the fully autonomous methods, the two-stage YOLO-based pipelines delivered the strongest practical performance. Their results were broadly similar, indicating that the main benefit comes from separating detection and ranging. Within this group, the LightGBM variant showed the most balanced behavior.

In contrast, the end-to-end multitask model performed substantially worse than both the geometry-based and two-stage approaches. Overall, these results show that geometry-based methods provide the best upper-bound accuracy, while two-stage pipelines offer the best balance between accuracy and deployability in realistic settings.

Table 1
Quantitative comparison of the evaluated monocular UAV ranging methods

Method	MAE (m)	RMSE (m)	Median AE (m)	MAPE (%)	R^2	Within 10m (%)	Within 10% (%)
Geometry-Calibrated Distance Model (Size Agnostic)	15.33	32.89	5.66	21.24	0.9672	70.26	45.1
Geometry-Calibrated Distance Model	19.92	35.88	8.62	28.49	0.961	55.56	27.94
Two-stage (YOLO + svm_global)	22.11	47.3	6.68	21.92	0.9322	59.64	40.36
Two-stage (YOLO + lightgbm)	22.6	54.96	6.13	18.8	0.9085	60.46	45.42
Two-stage (YOLO + random_forest_global)	24.22	53.12	6.81	20.92	0.9145	58.99	41.83
End-to-end	59.79	123.15	21.1	58.09	0.5406	31.05	13.73

6.2. Distance degradation analysis

To better understand how target distance affects reliability, we analyzed both the error distributions and the performance within separate distance ranges.

The error cumulative distribution function (CDF) shows that the geometry-based and two-stage models reach low errors much faster than the end-to-end model. The Size Agnostic model reaches its 90th percentile error (P90) at only 35.6 m. The two-stage LightGBM and SVM models show very similar and stable behavior, with P90 values of 55.9 m and 57.8 m. In comparison, the end-to-end model has a much heavier error tail, with a P90 of 154.5 m.

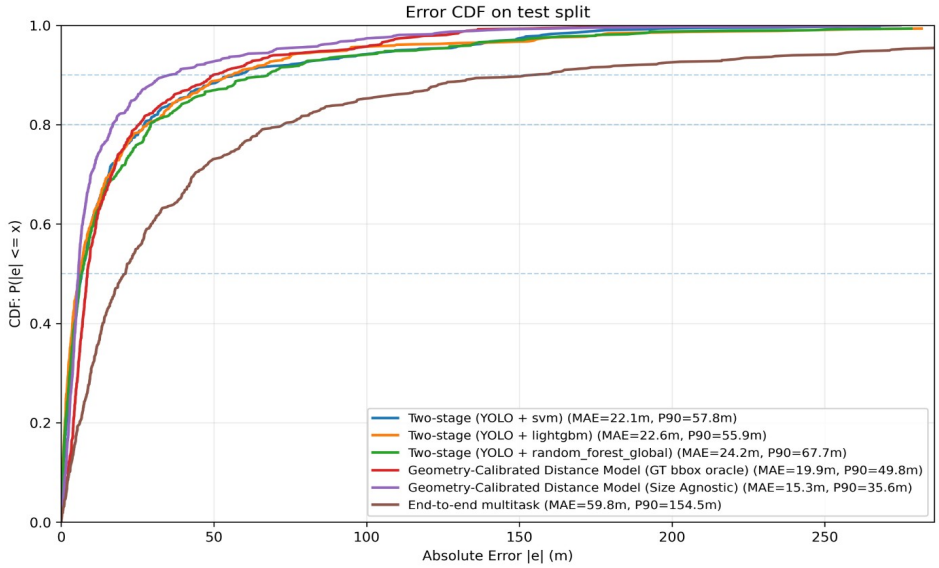

Figure 5: Error cumulative distribution functions of the evaluated ranging methods on the test split.

This drop in reliability becomes even clearer when looking at MAE across different distance bins. In the short-range interval of 0-50 m, all models perform very well, and the geometry-based and YOLO-based methods keep MAE below 7 m. However, as the target moves farther away, especially beyond 300-400 m, the UAV becomes very small in the image. At this stage, even a one-pixel error in the bounding box can lead to a large error in estimated distance. For example, the YOLO + SVM model reaches an MAE of 85.0 m in the 400 m+ bin. The end-to-end model performs especially poorly at this range, with an MAE rising to 285.0 m, which shows that it cannot reliably recover distance information from such small visual targets.

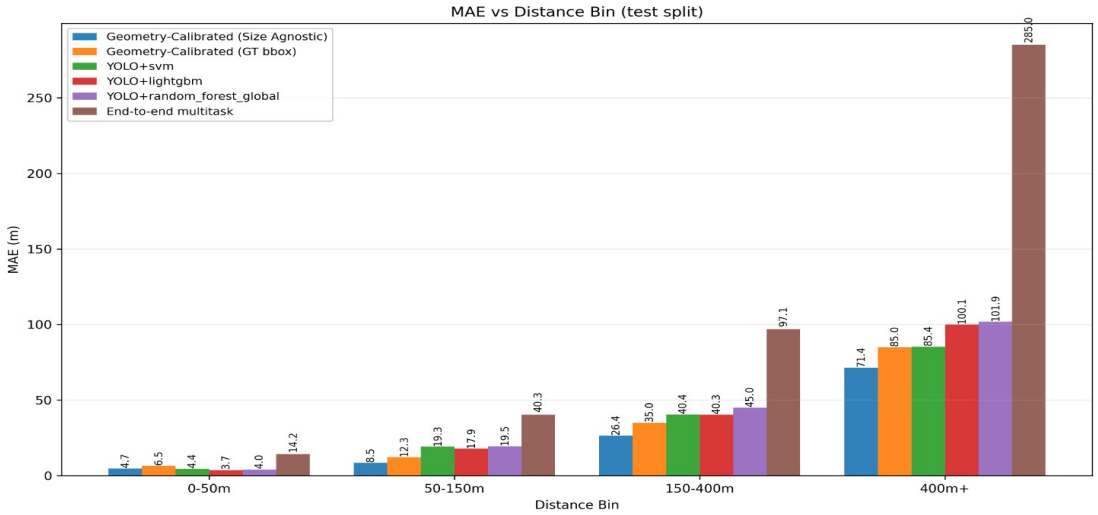

Figure 6: Mean absolute error across distance bins for the evaluated ranging methods on the test split.

6.3. Ablation and error analysis

To better explain these results, we carried out an error analysis focused on both the geometric models and the weak points of the deep learning approach.

6.3.1. Calibration evolution and model equivalence

An important result is that the Size Agnostic geometry model (MAE 15.33 m) performs better than the standard Geometry-Calibrated model based on fixed physical object sizes (MAE 19.92 m). A model that assumes one fixed real-world size is too rigid for real flight conditions. The apparent size of a UAV in a bounding box depends not only on distance, but also on its type, viewing angle, pose, and shape. For example, a fixed-wing UAV and a quadcopter can produce very different box areas even at the same distance.

The Size Agnostic model avoids this limitation. Instead of relying on strict class-specific physical dimensions, it learns effective spatial relationships directly from the data and then uses affine calibration to correct systematic errors. In this way, it can better handle perspective effects, partial visibility, annotation noise, and other real-image problems. This makes it more flexible and more accurate than the stricter geometry-based version.

6.3.2. End-to-end limitations and data starvation

The poor performance of the ResNet18 end-to-end multitask model (MAE 59.79 m) is most likely caused by data starvation. The ImageNet features used for initialization are tuned for generic object recognition rather than metric depth reasoning, and around 4000 images is not enough to adapt them to joint classification and long-range distance regression. We did not attempt pretraining on large-scale monocular depth datasets (KITTI, NYU Depth v2, DIODE), since their ground-level and indoor scenes differ too much from aerial views at distances up to 1 km to expect a useful transfer. Depth pretraining on aerial imagery remains an open direction.

This becomes even more clear when looking at the confusion matrices. The two-stage pipeline uses YOLO, which already benefits from large-scale pre-training. Because of this, the YOLO + SVM model achieves high classification accuracy, correctly identifying classes such as lancet with 94.4% accuracy and fpv-quadcopter with 98.4% accuracy. In contrast, the end-to-end model has to learn all visual features only from our limited dataset, which leads to strong class confusion. For example, it misclassifies shahed-131 almost 40% of the time and also struggles to separate zala-lancet-z-52 from lancet, with class accuracies dropping to 63.6% and 54.2%. Without stable

classification features, the model also fails to learn the class-dependent visual cues needed for accurate distance estimation.

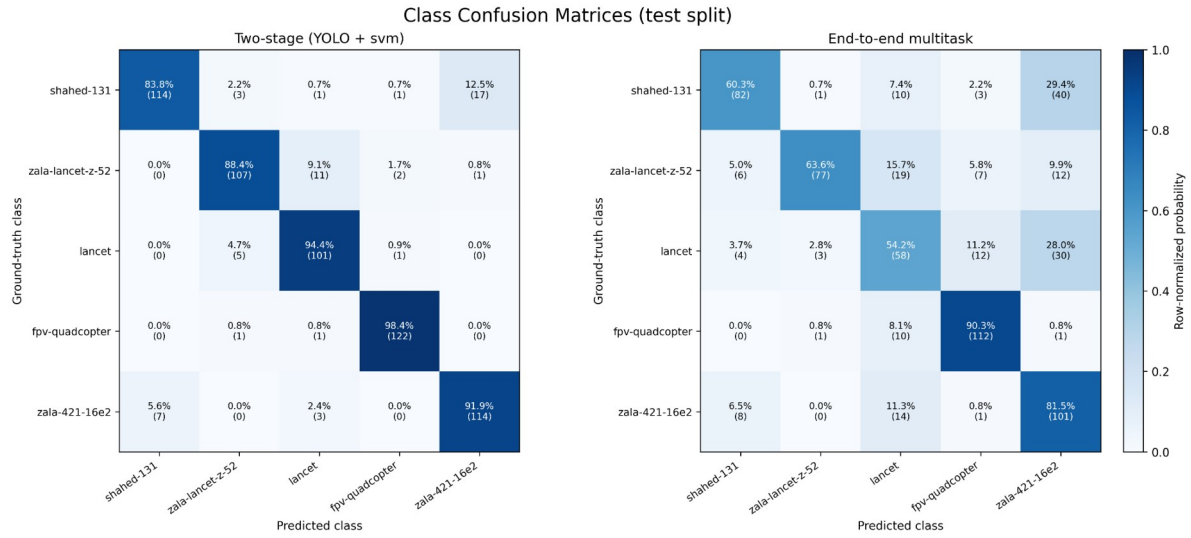

Figure 7: Test-split confusion matrices for the two-stage YOLO-based pipeline with SVM regression and the end-to-end multitask model.

6.3.3. Visual inference context

A visual review of individual predictions supports these findings. In fully automatic mode, the YOLO detector can still localize small targets successfully, but its predicted bounding boxes naturally introduce some extra error compared to manually annotated ground-truth boxes. This explains why the same target can produce a distance estimate of 72.76 m under oracle geometry-based conditions and 55.25 m under the two-stage pipeline. Even with this expected variation, the two-stage approach remains the most practical and reliable solution for autonomous passive UAV distance estimation without using active radar or LiDAR.

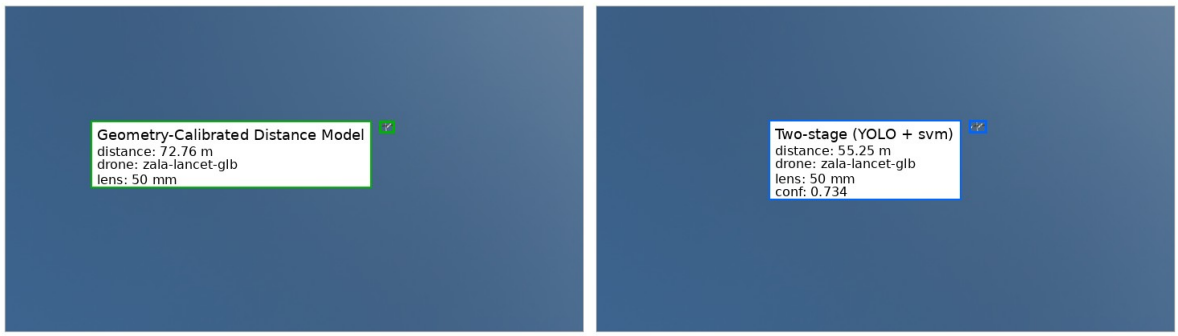

Figure 8: Example predictions for the same UAV instance produced by the geometry-calibrated model and the two-stage YOLO-based pipeline.

7. Conclusions

This paper compared three passive monocular UAV ranging approaches on a custom dataset covering distances up to 983.96 m under multiple optical configurations. The results showed that the geometry-calibrated size-agnostic model achieved the best overall accuracy under ideal localization, with an MAE of 15.33 m and $R^2 = 0.9672$. Among the fully autonomous methods, the two-stage YOLO-based pipelines provided the best practical balance between accuracy and deployability, with MAE values around 22 m. In contrast, the end-to-end multitask model

performed substantially worse, indicating that the available dataset is not large enough to support reliable joint learning of classification and ranging at extreme distances.

The present evaluation was carried out on a single custom dataset, which, although broad in terms of distances, focal lengths, and UAV classes, reflects one specific set of imaging conditions. Cross-validation on independent data would help further confirm how the relative behavior of the three approaches transfers to other sensors, environments, and UAV types.

Overall, the study shows that passive monocular UAV ranging is feasible at distances close to 1000 m, but performance depends strongly on localization quality and data availability. Future work will focus on expanding long-range data, validating the proposed methods on more diverse datasets, improving robustness to detection noise, using temporal information from multi-frame sequences to stabilize distance prediction, and exploring depth pretraining on aerial imagery to address the data starvation observed in the end-to-end model.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] T.-H. Wu, C. Gong, D. Kong, S. Xu, Q. Liu, A Novel Visual Object Detection and Distance Estimation Method for HDR Scenes based on Event Camera, in: Proceedings of the 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 2021, pp. 636–640. doi:10.1109/ICCC54389.2021.9674426.

[2] H. Saleh, S. Saleh, N. T. Toure, W. Hardt, Robust Collision Warning System based on Multi Objects Distance Estimation, in: Proceedings of the 2021 IEEE Concurrent Processes Architectures and Embedded Systems Virtual Conference (COPA), San Diego, CA, USA, 2021, pp. 1–6. doi:10.1109/COPA51043.2021.9541452.

[3] H. Yang, Y. Tang, W. Yu, A Human-Machine Safety Distance Detection Method Based on Computer Vision, in: Proceedings of the 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 2022, pp. 5784–5789. doi:10.1109/CCDC55256.2022.10033461.

[4] R. Waranusast, P. Riyamongkol, P. Pattanathaburt, Distance Estimation Between Camera and Vehicles from an Image using YOLO and Machine Learning, in: Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, pp. 482–488. doi:10.23919/APSIPAASC55919.2022.9980156.

[5] U. Rajapaksha, F. Sohel, H. Laga, D. Diepeveen, M. Bennamoun, Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey, ACM Computing Surveys 56 (2024) 1–51. doi:10.1145/3677327.

[6] S. Ausaf Hussain, Waseemullah, N. Ahmed Khan, Face-to-camera distance estimation using machine learning, in: Proceedings of the 3rd International Conference on Innovations in Computer Science & Software Engineering (ICONICS), Karachi, Pakistan, 2022, pp. 1–8. doi:10.1109/ICONICS56716.2022.10100618.

[7] Y. Zha, F. Li, Y. Li, L. Hong, Q. Ling, MDE-Net: Monocular Distance measurement based on detected license plates, in: Proceedings of the 41st Chinese Control Conference (CCC), Hefei, China, 2022, pp. 5453–5457. doi:10.23919/CCC55666.2022.9902166.

[8] S. Qiao, H. Zhang, F. Xie, Z. Jiang, Deep-Learning-Based Direct Attitude Estimation for Uncooperative Known Space Objects, IEEE Transactions on Aerospace and Electronic Systems 60 (2024) 2526–2541. doi:10.1109/TAES.2023.3325801.

[9] B. Wang, Q. Li, Q. Mao, J. Wang, C. L. P. Chen, A. Shangguan, H. Zhang, A survey on vision-based anti unmanned aerial vehicles methods, Drones 8 (2024) 518. doi:10.3390/drones8090518.

[10] T. Luan, S. Zhou, Y. Zhang, W. Pan, Fast identification and detection algorithm for maneuverable unmanned aircraft based on multimodal data fusion, Mathematics 13 (2025) 1825. doi:10.3390/math13111825.

[11] I. Yurchuk, T. Semenchenko, Vision-based UAV Detection Models for Small-Edge Devices, in: CEUR Workshop Proceedings, volume 4035, 2025, pp. 81–90. URL: https://ceur-ws.org/Vol-4035/Paper7.pdf.

[12] Z. Ni, J. Shi, L. Li, C. Ni, Real-time vehicle detection and distance estimation: Soft-sensor approach using optimized you only look once version 5 and perspective geometry, Sensors and Materials 38 (2026) 329. doi:10.18494/SAM5781.

[13] H. Wang, Y. Qu, Z. Dang, D. Wu, M. Cui, H. Shi, J. Zhao, A ground-based visual system for uav detection and altitude measurement deployment and evaluation of ghost-yolov11n on edge devices, Sensors 26 (2025) 205. doi:10.3390/s26010205.

[14] C. Zhang, X. Weng, Y. Cao, M. Ding, Monocular absolute depth estimation from motion for small unmanned aerial vehicles by geometry-based scale recovery, Sensors 24 (2024) 4541. doi:10.3390/s24144541.

[15] J. Gaigalas, L. Perkauskas, H. Gricius, T. Kanapickas, A. Kriščiūnas, A framework for autonomous uav navigation based on monocular depth estimation, Drones 9 (2025) 236. doi:10.3390/drones9040236.

[16] F. Gökçe, G. Üçoluk, E. Şahin, S. Kalkan, Vision-Based Detection and Distance Estimation of Micro Unmanned Aerial Vehicles, Sensors 15 (2015) 23805–23846. doi:10.3390/s150923805.

[17] Y.-C. Lai, Z.-Y. Huang, Detection of a Moving UAV Based on Deep Learning-Based Distance Estimation, Remote Sensing 12 (2020) 3035. doi:10.3390/rs12183035.

[18] B. Kiefer, Y. Quan, A. Zell, Approximate supervised object distance estimation on unmanned surface vehicles, arXiv preprint (2025). doi:10.48550/arXiv.2501.05567.

[19] D. Silva, N. Jourdan, N. Gählert, Long Range Object-Level Monocular Depth Estimation for UAVs, in: R. Gade, M. Felsberg, J.-K. Kämäräinen (Eds.), Image Analysis, volume 13885 of Lecture Notes in Computer Science, Springer, Cham, 2023, pp. 325–340. doi:10.1007/978-3-031-31435-3_22.

[20] H. Azad, V. Mehta, I. Mantegh, M. Bolic, DroneRanger: Vision-Driven Deep Learning for Drone Distance Estimation, in: Proceedings of the 2024 International Conference on Unmanned Aircraft Systems (ICUAS), 2024, pp. 442–449. doi:10.1109/ICUAS60882.2024.10557013.

[21] I. Yurchuk, D.-M. Obertan, Geospatial detection and movement analysis system for unmanned aerial vehicles based on computer vision methods, International Journal of Information Technology and Computer Science 17 (2025) 16–27. doi:10.5815/ijitcs.2025.04.02.

[22] Fuzzy-Org, Vision-based UAV Ranging Dataset, GitHub repository, 2026. URL: https://github.com/Fuzzy-Org/vision-based-uav-ranging-dataset.