International Journal of Advanced Engineering and Technology Research
ISSN: 3106-8677 (Print), ISSN: 3106-8685 (Online) | Vol. 4, No. 1, 2026

YOLO11-Based Object Detection against Same-Color Backgrounds

Jie Xiong [1], Bo Huang [2], Chengguo Xiao [1], Wangjun Zou [1, *]

[1] School of Applied Mathematics, Chengdu University of Information Technology, Chengdu, 610225, China

2 School of Software Engineering, Chengdu University of Information Technology, Chengdu, 610225, China

[*] **Corresponding author**: (Email: Vangen0623@163.com)

Abstract: Objects that closely match background color are difficult to localize because chromatic cues weaken and boundaries can be obscured. This study evaluates a YOLO11 baseline using a dataset containing at least 12 object categories and approximately 630 images per category, trained for 100 epochs. Training and validation curves are summarized alongside confidence-based and precision-recall analyses, and an exploratory grayscale-Canny-YOLO fusion path is documented separately from the evaluated baseline. The evaluation plot reports an overall mAP@0.5 of 0.607; category AP ranges from 0.433 for Table to 0.831 for Bus, while the F1-confidence curve peaks at 0.60 near a confidence threshold of 0.284. The category-count plot indicates substantial imbalance among annotated instances. These results establish a measurable baseline and expose class-dependent performance variation. The findings support dataset balancing, fixed evaluation partitions, and ablation testing as necessary next steps for reliable same-color-background detection.

Keywords: Same-Color Background, Object Detection, YOLO11, Edge Cues, Class Imbalance, mAP.

1. Introduction

Objects that resemble their surroundings in color are difficult to detect because chromatic contrast provides little evidence and object boundaries may be weak or partially obscured. This problem is closely related to camouflaged-object detection, where low contrast, distracting context, and incomplete contours challenge both localization and segmentation [1]. Recent studies address these difficulties through CNN-Transformer feature fusion, dual-branch fusion and self-similarity constraints, multi-information guidance, and edge-aware prediction [2-5]. Their results show that contextual and boundary information can be useful when target and background appearance overlap, while also illustrating that architecture-specific gains must be established on the relevant data.

Single-stage YOLO detectors offer a practical baseline for multi-class bounding-box detection, but small targets and uneven category frequencies can weaken feature learning. Recent YOLO-derived studies investigate small-object detection through efficient multi-scale fusion and dynamic feature aggregation [6-7]. Published YOLO11 evaluations in remote sensing and maritime imagery further show that its performance remains dependent on the target domain and training protocol [8-9]. These results motivate an empirical baseline for the same-color setting rather than an untested claim of universal robustness.

The dataset comprises at least 12 object categories, with approximately 630 images per class; category-count and bounding-box distribution plots are also included. Its current implementation trains a YOLO11 detector for 100 epochs and reports training/validation losses, precision, recall, F1, and mean average precision curves. The overall class-averaged AP at an intersection-over-union threshold of 0.5 is marked as 0.607; category AP values vary from 0.433 for table to 0.831 for bus. The evaluation supports a baseline characterization, but not a claim that a new module improves detection.

This paper organizes these results as a baseline study. It summarizes the data and annotation profile, documents YOLO11 training and threshold-based evaluation, and examines category-wise performance and qualitative detections. A grayscale-Canny-YOLO fusion workflow is also described because it is included in the processing design; however, no comparative scores are reported for that extension, so it is treated as an unverified method design rather than an achieved improvement. Two design pressures follow. Color alone is insufficient, so grayscale and edge cues are considered complementary; meanwhile, performance must be interpreted per class because object counts are unequal. The study does not assume preprocessing improves detection: baseline scores remain separate from untested fusion. This distinction defines needed experiments, including balanced sampling, fixed partitions, confidence calibration, and comparison with an unmodified detector, without treating qualitative examples as evidence of genuine causal improvement. The analysis also notes missing split and hardware details, which limit generalization claims. Such a protocol would also clarify whether edge processing adds useful cues or merely changes the input distribution. Without that comparison, deployment decisions and claims of broad robustness under the stated acquisition conditions would remain premature.

2. Materials and Methods

2.1. Dataset and Annotation Profile

The working dataset covers at least 12 object categories, with approximately 630 images per category under different backgrounds. The category plot names Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, and Table, and the annotation summaries include object counts, box positions, and box dimensions. Figure 1 shows the category frequencies and bounding-box position and size distributions. The instance-count plot is visibly imbalanced, with People substantially more frequent than several other categories. Images are associated with YOLO-style object labels. The

available description does not specify the train-validation-test split, image resolution policy, annotation quality checks, or random seed; these omissions constrain interpretation of the reported evaluation.

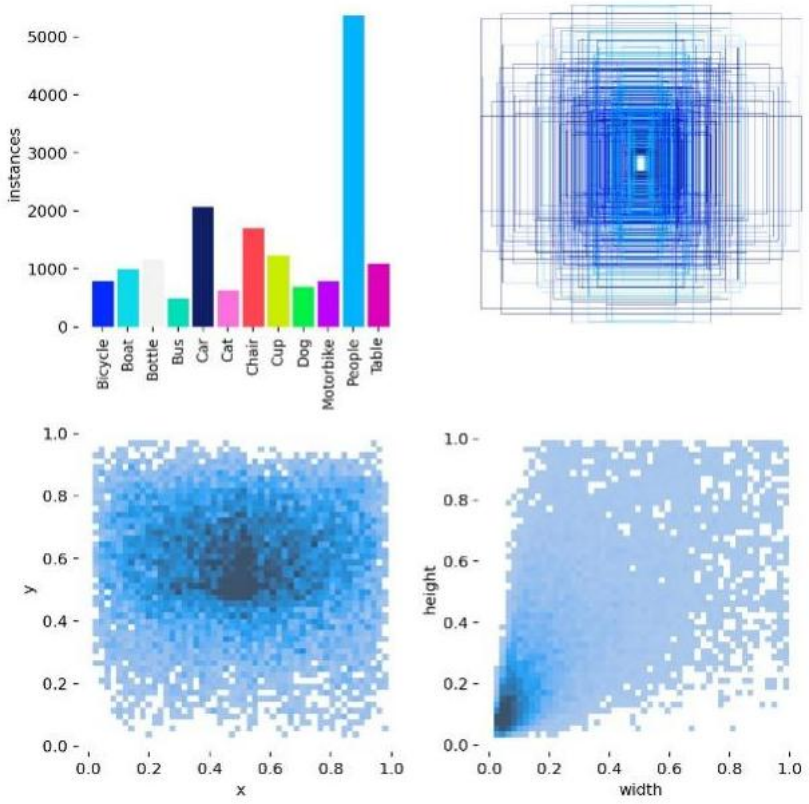


Figure 1. Category frequency and bounding-box position and size distributions.

2.2. YOLO11 Baseline and Training

A YOLO11 detector is used as the baseline for the 12-category task. The training record spans 100 epochs and plots box-regression, classification, and distribution-focal losses for both training and validation, together with precision, recall, mAP@0.5, and mAP@0.5:0.95. The plotted losses decrease over the training interval, while the detection metrics rise. No optimizer, batch size, image size, hardware configuration, checkpoint-selection rule, or independent test protocol is given, so the curves are treated as optimization diagnostics rather than a fully specified reproducibility package. Figure 2 shows the training and validation curves and the corresponding detection metrics.

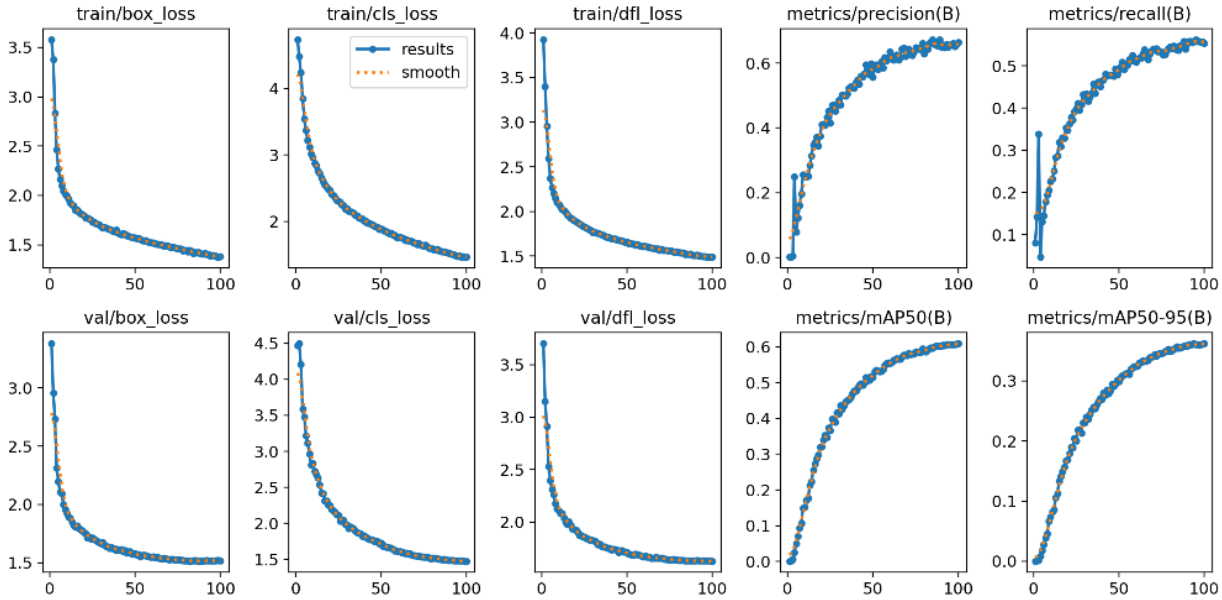


Figure 2. Training and validation losses and detection metrics over 100 epochs.

2.3. Candidate Grayscale and Edge-Fusion Workflow

The processing design contains a grayscale branch, Canny edge extraction, YOLO detection, preservation of the grayscale image, fusion with the original image, and a second edge-detection stage before output. Recent work continues to use Canny edge detection for boundary extraction [10]. As shown in Figure 3, the original image also feeds the fusion stage directly, while the first Canny edge map is passed to

YOLO. The workflow does not specify the fusion operator, edge thresholds, or detector input encoding in sufficient detail to reproduce the extension. It is therefore documented as a candidate pipeline; the numerical results below belong to the YOLO11 baseline, not to a controlled evaluation of this fusion design.

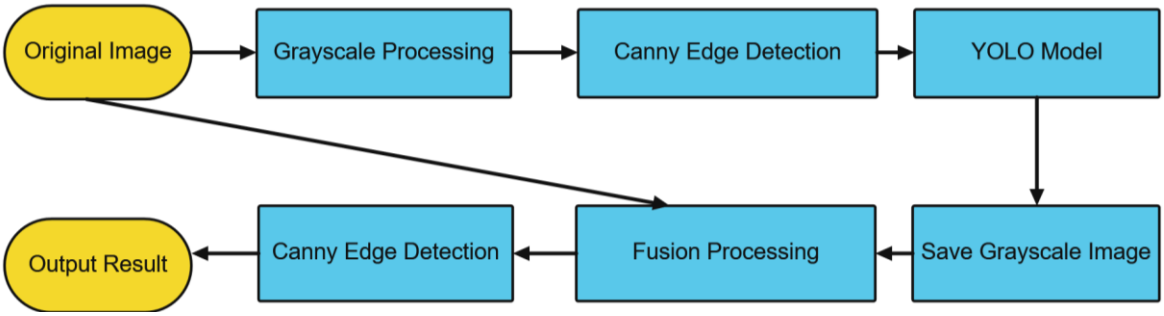


Figure 3. Candidate grayscale, Canny, YOLO, and image-fusion workflow.

3. Results

3.1. Training Behavior

Across the 100 plotted epochs, the training and validation box, classification, and distribution-focal losses decline, and the displayed precision, recall, and mean-average-precision curves increase. The corresponding evaluation plots provide class-level precision-recall curves and confidence-dependent F1, precision, and recall. These trends indicate that the model learns from the training data, but without a documented held-out split or repeated runs they do not establish robustness to new backgrounds or acquisition conditions.

3.2. Confidence and Category-Wise Performance

The performance plot marks the all-class mAP@0.5 as 0.607. Its F1-confidence curve reaches a maximum of 0.60 at a confidence threshold of 0.284. The same plot reports an all-class precision of 1.00 at confidence 0.940 and recall of 0.88 at confidence 0.000, illustrating the expected operating-point trade-off rather than a single threshold that is optimal for every application. The class-wise AP values reproduced in Table 1 range from 0.433 for Table to 0.831 for Bus. This spread, together with the category-count imbalance in Figure 1, indicates that the overall score conceals meaningful differences among categories. Figure 4 presents the confidence and precision-recall curves used for this summary.

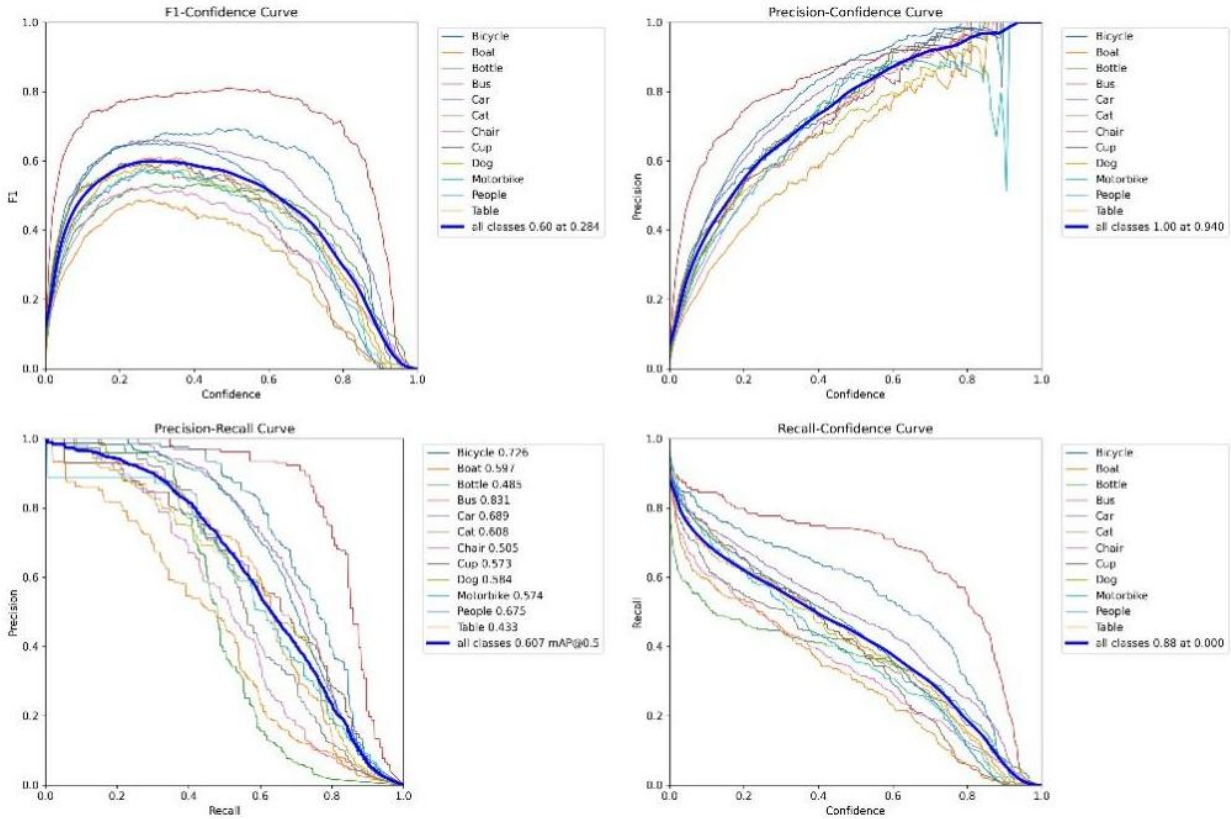


Figure 4. F1, precision-recall, precision-confidence, and recall-confidence curves.

Table 1. Class-wise AP@0.5 values annotated in the evaluation plot.

Category	AP@0.5	Category	AP@0.5
Bicycle	0.726	Boat	0.591
Bottle	0.485	Bus	0.831
Car	0.689	Cat	0.608
Chair	0.505	Cup	0.573
Dog	0.584	Motorbike	0.574
People	0.675	Table	0.433
All classes (mAP@0.5)	0.607		

3.3. Qualitative Detection Examples

The composite detections show bounding boxes and predicted labels for the dataset categories. Figure 5 presents these examples. They demonstrate that image-level inference outputs were generated for visual inspection. Because the examples are not accompanied by a ground-truth comparison, error counts, or a matched baseline, they are interpreted qualitatively and are not used to claim a causal performance gain.

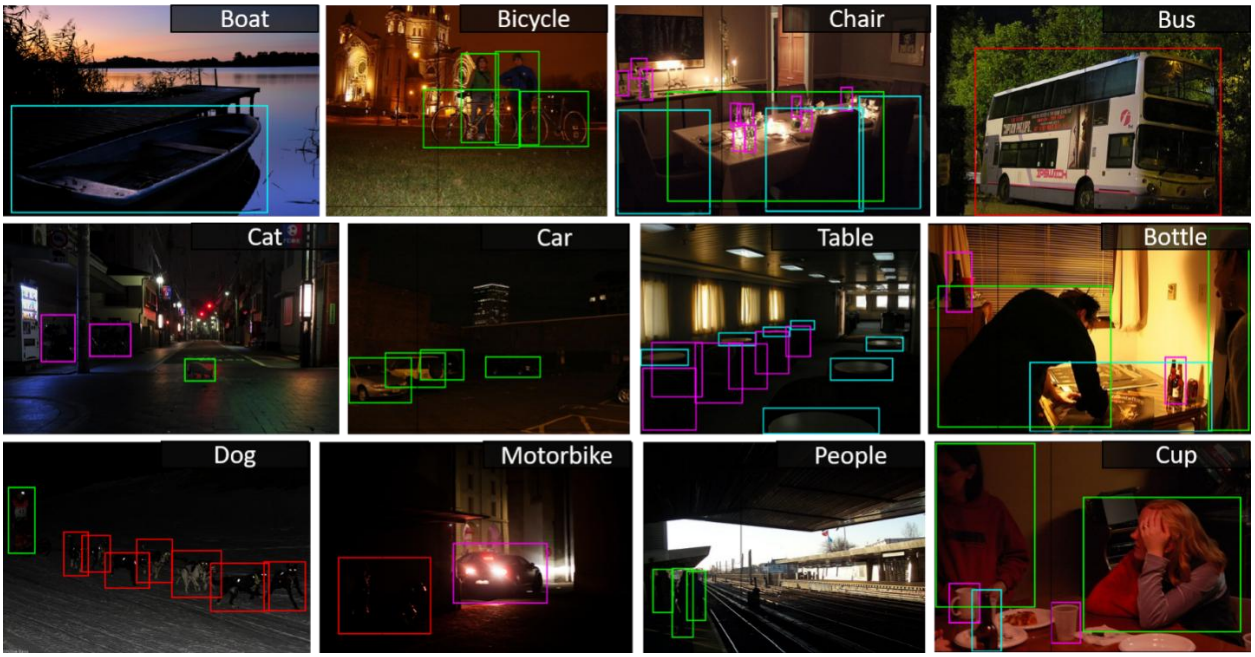


Figure 5. Representative YOLO11 detections across the object categories.

4. Discussion

The reported mAP@0.5 of 0.607 establishes a measurable starting point for same-color and visually complex backgrounds, while the per-class AP range shows that performance is not uniform. The class-count plot suggests that frequent categories may contribute disproportionately to aggregate metrics. A balanced sampling or reweighting study is a reasonable next experiment, but its effect cannot be inferred from the present results. Similarly, grayscale and edge cues are plausible complements when color contrast is weak, yet the documented fusion path has no ablation against the YOLO11 baseline. The training curves, confidence curves, and qualitative examples should therefore be read as a baseline profile, not as validation of the proposed preprocessing or network modifications.

Several experimental details remain necessary for a stronger claim: a fixed and disclosed data partition, class-wise test metrics, repeated runs, a baseline comparison, an ablation of grayscale and Canny stages, and inference-time measurements on stated hardware. The dataset description also needs a precise account of how same-color backgrounds are defined and how images are selected across classes. These checks would determine whether the observed score reflects the target visual challenge or the broader object-recognition distribution represented in the images.

5. Conclusion

A 100-epoch YOLO11 baseline was evaluated for a multi-category dataset containing same-color or complex-background examples. The performance plot reports mAP@0.5 of 0.607, with class AP from 0.433 to 0.831 and a peak F1 of 0.60 at confidence 0.284. Training losses decline, but category imbalance and incomplete evaluation details limit generalization claims. Grayscale-Canny fusion remains a candidate extension until controlled comparisons quantify its contribution. The results provide a concrete baseline and identify the data and ablation tests needed for a defensible assessment of same-color-background detection.

References

[1] Xiao, F., Hu, S., Shen, Y., et al. (2024). A survey of camouflaged object detection and beyond. CAAI Artificial Intelligence Research, 3, 9150044. https://doi.org/10.26599/AIR.2024.9150044

[2] Zhang, D. D., Wang, C. P., Wang, H. Y., et al. (2025). An effective CNN and Transformer fusion network for camouflaged object detection. Computer Vision and Image Understanding, 259, 104431. https://doi.org/10.1016/j.cviu.2025.104431

[3] Yang, H. Z., Zhu, Y., Sun, K., et al. (2025). Camouflaged object detection via dual-branch fusion and dual self-similarity constraints. Pattern Recognition, 157, 110895. https://doi.org/10.1016/j.patcog.2024.110895

[4] Shi, C. J., Zhao, L., Wang, R., et al. (2025). Multi-information guided camouflaged object detection. Image and Vision Computing, 156, 105470. https://doi.org/10.1016/j.imavis.2025.105470

[5] Ma, J. Y., Chen, T. Y., Xiao, J., et al. (2025). An edge-aware high-resolution framework for camouflaged object detection. Image and Vision Computing, 157, 105487. https://doi.org/10.1016/j.imavis.2025.105487

[6] Peng, S. F., Fan, X., Tian, S. W., et al. (2024). PS-YOLO: A small object detector based on efficient convolution and multi-scale feature fusion. Multimedia Systems, 30, 241. https://doi.org/10.1007/s00530-024-01447-0

[7] Chen, J., & Er, M. J. (2024). Dynamic YOLO for small underwater object detection. Artificial Intelligence Review, 57, 165. https://doi.org/10.1007/s10462-024-10788-1

[8] He, L. H., Zhou, Y. Z., Liu, L., et al. (2025). Research on object detection and recognition in remote sensing images based on YOLOv11. Scientific Reports, 15, 14032. https://doi.org/10.1038/s41598-025-96314-x

[9] Zhao, B. G., Zhao, J. W., Song, R., et al. (2025). Enhanced YOLO11 for lightweight and accurate drone-based maritime search and rescue object detection. PLOS ONE, 20(7), e0321920. https://doi.org/10.1371/journal.pone.0321920

[10] Song, Y., Li, C., Xiao, S., et al. (2024). A parallel Canny edge detection algorithm based on OpenCL acceleration. PLOS ONE, 19(1), e0292345. https://doi.org/10.1371/journal.pone.0292345