UDC: 528.8:004.932.72]:632.51
COBISS.SR-ID 198434057
doi: https://doi.org/10.61837/mbuir040126127m

ORIGINAL SCIENTIFIC PAPER

RECEIVED: 4. 4. 2026.
ACCEPTED: 24. 4. 2026.

COMPARATIVE BASELINE EVALUATION OF YOLOV5 AND YOLO11 FOR AGRICULTURAL WEED DETECTION

Aleksandar MARKOVIĆ

University of Pristina, Faculty of Sciences and Mathematics, Kosovska Mitrovica, Serbia

aleksandar.markovic@pr.ac.rs
https://orcid.org/0000-0001-5230-010X

Nebojsa DENIĆ

University of Pristina, Faculty of Sciences and Mathematics, Kosovska Mitrovica, Serbia

nebojsa.denic@pr.ac.rs
https://orcid.org/0000-0003-2584-259X

Milovan MINIĆ

University of Pristina, Faculty of Sciences and Mathematics, Kosovska Mitrovica, Serbia

milovan.minic@gmail.com
https://orcid.org/0009-0005-2019-3035

Abstract: *This paper presents a comparative baseline evaluation of YOLOv5 and YOLO11 object detection architectures applied to ambrosia detection in agricultural field imagery. Using a constrained dataset of 96 images (64 training, 32 validation), we systematically analyze the architectural characteristics, scaling mechanisms, and anchor box configurations of both frameworks. Baseline experiments establish performance benchmarks of mAP50=0.624 for YOLOv5 and mAP50=0.710 for YOLO11 under identical training conditions. Our analysis identifies key limitations of general-purpose architectures when applied to specialized botanical detection tasks with limited training data, providing a foundation for targeted domain-specific modifications. Results demonstrate that YOLO11's enhanced backbone design with C2PSA modules achieves superior baseline precision (0.800 vs 0.694) while both architectures fall short of the target mAP50 › 0.85, motivating further architectural and post-processing research.*

Keywords: *YOLO11 architecture; Adaptive Weighted Box Fusion (WBF); C2PSA attention mechanism; Sliced Aided Hyper Inference (SAHI); P2/4 micro-scale detection head; Precision agricultural eradication; Domain-specific anchor optimization*

1. INTRODUCTION

Ambrosia poses significant challenges to agricultural productivity across Central and Eastern Europe. Automated detection systems offer a scalable solution for precision weed management, yet the application of deep learning to specialized agricultural domains faces inherent obstacles: limited annotated datasets, extreme variation in plant scale and density, and the high cost of false positive detections in treatment applications.

The YOLO (You Only Look Once) family of object detectors [1] has established itself as the dominant framework for real-time detection tasks, with subsequent architectural improvements leading to significant advances in speed-accuracy trade-offs [2]. YOLOv5 and YOLO11 representing two widely-adopted variants with substantially different architectural philosophies. YOLOv5 employs C3-based cross-stage partial networks with traditional anchor-based detection, while YOLO11 introduces C3k2 blocks and C2PSA (Cross-Stage Partial with Polarized Self-Attention) modules for enhanced feature discrimination.

This paper contributes a systematic comparative analysis of these two architectures under identical training conditions on a small ambrosia dataset, establishing performance baselines and identifying architectural limitations that motivate subsequent domain-specific optimization.

2. YOLOv5 Architecture Overview

YOLOv5, developed by Ultralytics [3], follows an efficient backbone-neck-head architecture that balances speed and accuracy through compound scaling. The model is defined using a structured YAML configuration specifying every network component.

2.1 Model Scaling and Configuration

The architecture supports multiple model sizes through two scaling parameters:

Table 1. YOLOv5 core scaling parameters and their values for the small variant.

Parameter	Description	YOLOv5s Value
depth_multiple	Controls bottleneck block repetition in C3 modules	0.33
width_multiple	Scales channel count in each layer	0.50

This enables the full YOLOv5 model family:

Table 2. YOLOv5 model family variants with scaling parameters and computational cost.

Variant	Depth Multiple	Width Multiple	Parameters	GFLOPs
YOLOv5n (Nano)	0.33	0.25	~1.9M	4.5
YOLOv5s (Small)	0.33	0.50	~7.2M	15.8 – 16.5
YOLOv5m (Medium)	0.67	0.75	~21.2M	47.9 – 49.0
YOLOv5l (Large)	1.00	1.00	~46.5M	107.7 – 115.5
YOLOv5x (XLarge)	1.33	1.25	~86.7M	204.0 – 215.5

2.2 Backbone Architecture

The YOLOv5 backbone performs progressive feature extraction through a spatial resolution pyramid, starting from the 640×640×3 input and producing multi-scale feature maps:

- [-1, 1, Conv, [64, 6, 2, 2]], # 0-P1/2
- [-1, 1, Conv, [128, 3, 2]], # 1-P2/4
- [-1, 3, C3, [128]],
- [-1, 1, Conv, [256, 3, 2]], # 3-P3/8
- [-1, 6, C3, [256]],
- [-1, 1, Conv, [512, 3, 2]], # 5-P4/16
- [-1, 9, C3, [512]],
- [-1, 1, Conv, [1024, 3, 2]], # 7-P5/32
- [-1, 3, C3, [1024]],
- [-1, 1, SPPF, [1024, 5]], # 9

The C3 blocks implement a split-transform-merge strategy: input is split into two paths, one processed through bottleneck layers (1×1 → 3×3 → 1×1 convolutions) and one bypassed via identity connection, before concatenation and final fusion. This design reduces computation while maintaining gradient flow through residual connections [4]. The SPPF module at the backbone terminus applies sequential 5×5 max pooling to aggregate multi-scale context efficiently [5].

2.3 Neck Architecture (PANet)

The neck implements a bidirectional Feature Pyramid Network with top-down and bottom-up pathways [6]:

- [-1, 1, Conv, [512, 1, 1]],
- [-1, 1, nn.Upsample, [None, 2, “nearest”]],
- [[-1, 6], 1, Concat, [1]], # cat backbone P4
- [-1, 3, C3, [512, False]], # 13

- [-1, 1, Conv, [256, 1, 1]],
- [-1, 1, nn.Upsample, [None, 2, “nearest”]],
- [[-1, 4], 1, Concat, [1]], # cat backbone P3
- [-1, 3, C3, [256, False]], # 17 (P3/8-small)

- [-1, 1, Conv, [256, 3, 2]],
- [[-1, 14], 1, Concat, [1]], # cat head P4
- [-1, 3, C3, [512, False]], # 20 (P4/16-medium)

- [-1, 1, Conv, [512, 3, 2]],
- [[-1, 10], 1, Concat, [1]], # cat head P5
- [-1, 3, C3, [1024, False]], # 23 (P5/32-large)

- [[17, 20, 23], 1, Detect, [nc, anchors]], # Detect(P3, P4, P5)

The top-down pathway propagates semantic information from deep layers to higher resolutions via upsampling and concatenation. The bottom-up pathway then reinforces spatial precision, enabling each detection head to benefit from both semantic depth and spatial accuracy.

2.4 Anchor Box Configuration

YOLOv5 uses predefined anchor boxes optimized for the COCO dataset distribution:

Table 3. Default COCO-optimized anchor box configurations for YOLOv5 detection scales.

Detection Scale	Resolution	Anchor Sizes (w×h pixels)
P3/8 (Small)	80×80	[10,13], [16,30], [33,23]
P4/16 (Medium)	40×40	[30,61], [62,45], [59,119]
P5/32 (Large)	20×20	[116,90], [156,198], [373,326]

These COCO-optimized anchors represent a significant domain mismatch for ambrosia detection, where plant bounding boxes follow a substantially different size and aspect ratio distribution.

3. YOLO11 Architecture Overview

YOLO11 [7] represents the latest evolution in the YOLO family, introducing key architectural innovations while maintaining computational efficiency. The architecture features C3k2 blocks, C2PSA attention modules, and a refined three-dimensional compound scaling strategy.

3.1 Compound Scaling Strategy

YOLO11 introduces a three-dimensional scaling approach with an additional max_channels constraint:

Table 4. YOLO11 model family variants with scaling parameters and computational cost.

Variant	Depth	Width	Parameters	GFLOPs
YOLO11n (Nano)	0.50	0.25	2.6M	6.6
YOLO11s (Small)	0.50	0.50	9.5M	21.7
YOLO11m (Medium)	0.50	1.00	20.1M	68.5
YOLO11l (Large)	1.00	1.00	25.4M	142.1
YOLO11x (XLarge)	1.00	1.50	57.0M	196.4

3.2 Backbone Architecture

The YOLO11 backbone replaces C3 blocks with C3k2 (Enhanced Cross Stage Partial) modules and adds a C2PSA attention layer at the deepest feature level [8]:

[from, repeats, module, args]
- [-1, 1, Conv, [64, 3, 2]] # 0-P1/2
- [-1, 1, Conv, [128, 3, 2]] # 1-P2/4
- [-1, 2, C3k2, [256, False, 0.25]]
- [-1, 1, Conv, [256, 3, 2]] # 3-P3/8
- [-1, 2, C3k2, [512, False, 0.25]]
- [-1, 1, Conv, [512, 3, 2]] # 5-P4/16

- [-1, 2, C3k2, [512, True]]
- [-1, 1, Conv, [1024, 3, 2]] # 7-P5/32
- [-1, 2, C3k2, [1024, True]]
- [-1, 1, SPPF, [1024, 5]] # 9
- [-1, 2, C2PSA, [1024]] # 10

The C2PSA module integrates polarized self-attention mechanisms operating on both channel and spatial dimensions. Unlike full self-attention, the polarized variant restricts attention computation to enhance efficiency while maintaining feature discrimination capability. This module is particularly beneficial for distinguishing visually similar vegetation classes.

3.3 Neck Architecture (Enhanced PANet)

YOLO11 employs an enhanced Path Aggregation Network with C3k2 blocks replacing the C3 modules used in YOLOv5, and improved concatenation points for better information flow across scales [6]:

- [-1, 1, nn.Upsample, [None, 2, "nearest"]]
- [[-1, 6], 1, Concat, [1]] # cat backbone P4
- [-1, 2, C3k2, [512, False]] # 13

- [-1, 1, nn.Upsample, [None, 2, "nearest"]]
- [[-1, 4], 1, Concat, [1]] # cat backbone P3
- [-1, 2, C3k2, [256, False]] # 16 (P3/8-small)

- [-1, 1, Conv, [256, 3, 2]]
- [[-1, 13], 1, Concat, [1]] # cat head P4
- [-1, 2, C3k2, [512, False]] # 19 (P4/16-medium)

- [-1, 1, Conv, [512, 3, 2]]
- [[-1, 10], 1, Concat, [1]] # cat head P5
- [-1, 2, C3k2, [1024, True]] # 22 (P5/32-large)

- [[16, 19, 22], 1, Detect, [nc]] # Detect(P3, P4, P5)

The neck maintains the same bidirectional top-down/bottom-up structure as YOLOv5's PANet, but with three key differences: C3k2 blocks replace C3 for richer feature fusion, the concatenation at P5 level connects directly to the C2PSA output (layer 10) preserving attention-enhanced features, and channel dimensions are better balanced across scales reducing information bottlenecks during upsampling.

3.4 Key Architectural Differences from YOLOv5

Table 5. Key architectural differences between YOLOv5 and YOLO11.

Component	YOLOv5	YOLO11
Feature blocks	C3 (split-transform-merge)	C3k2 (enhanced bottleneck, 0.25 expansion)
Attention	None	C2PSA at P5 level
Scaling dims	2 (depth, width)	3 (depth, width, max_channels)
Anchor strategy	Predefined (COCO-optimized)	Anchor-free detection
Backbone depth	Shallower at P2/P3	Richer feature extraction throughout

4. EXPERIMENTAL SETUP

4.1 Dataset

All experiments were conducted on a custom ambrosia detection dataset consisting of 96 images: 64 for training and 32 for validation. Images were collected in agricultural field conditions representing varied growth stages, plant densities, and environmental conditions. The dataset was annotated with bounding boxes following standard YOLO format.

4.2 Training Configuration

Both architectures were trained under identical conditions to ensure fair comparison:

Table 6. Shared training configuration applied to both YOLOv5s and YOLO11s experiments.

Parameter	Value
Input resolution	640×640 pixels
Epochs	300
Batch size	8
Optimizer	Adam
Early stopping patience	300 (disabled)
Pretrained weights	COCO (YOLOv5s / YOLO11s)

YOLOv5 training used the hyp.paper.yaml hyperparameter configuration established in Chen's "An improved algorithm based on YOLOv5 for detecting Ambrosia in UAV images" [9], with lr0=0.01, momentum=0.937, and mosaic augmentation enabled. YOLO11 training used default Ultralytics augmentation settings [7].

The training phase was executed using a highly standardized configuration to ensure that the resulting performance metrics reflect inherent architectural capabilities rather than discrepancies in hyperparameter tuning. By employing a uniform 300-epoch schedule with the Adam optimizer and a 640×640 input resolution across both pipelines, the experiment provides a controlled benchmark for specialized botanical detection. As illustrated in the following Python implementation snippets, both models utilized their respective CLI frameworks to initialize from COCO-pretrained weights, allowing the networks to leverage general feature hierarchies while adapting to the unique textures and geometries of the *Ambrosia trifida* dataset.

While the training statistics (Figures 1 and 2) track the convergence of loss functions and mean average precision over time, the real-world efficacy of these models is most apparent in the qualitative detection results provided in Figures 3 and 4. A direct comparison of these field imagery samples reveals that YOLOv11 achieves significantly better results than the YOLOv5 baseline. Where the YOLOv5s output often exhibits conservative detection patterns and less precise bounding box alignment, the YOLOv11s framework-driven by its anchor-free paradigm and C2PSA attention modules-produces markedly cleaner and more geometrically accurate detections. This visual superiority is substantiated by the quantitative jump in baseline precision, which rises from 0.694 in YOLOv5 to 0.800 in YOLO11, underscoring the latter's ability to successfully discriminate between target weeds and complex background vegetation.

Training commands:

```
# YOLOv5s
python train.py --img 640 --batch 8 --epochs 300 --data ambrosia.yaml \
    --cfg models/yolov5s.yaml --optimizer Adam --patience 300


# YOLO11s
yolo detect train data=ambrosia.yaml model=yolo11s.yaml \
epochs=300 imgsz=640 batch=8
```

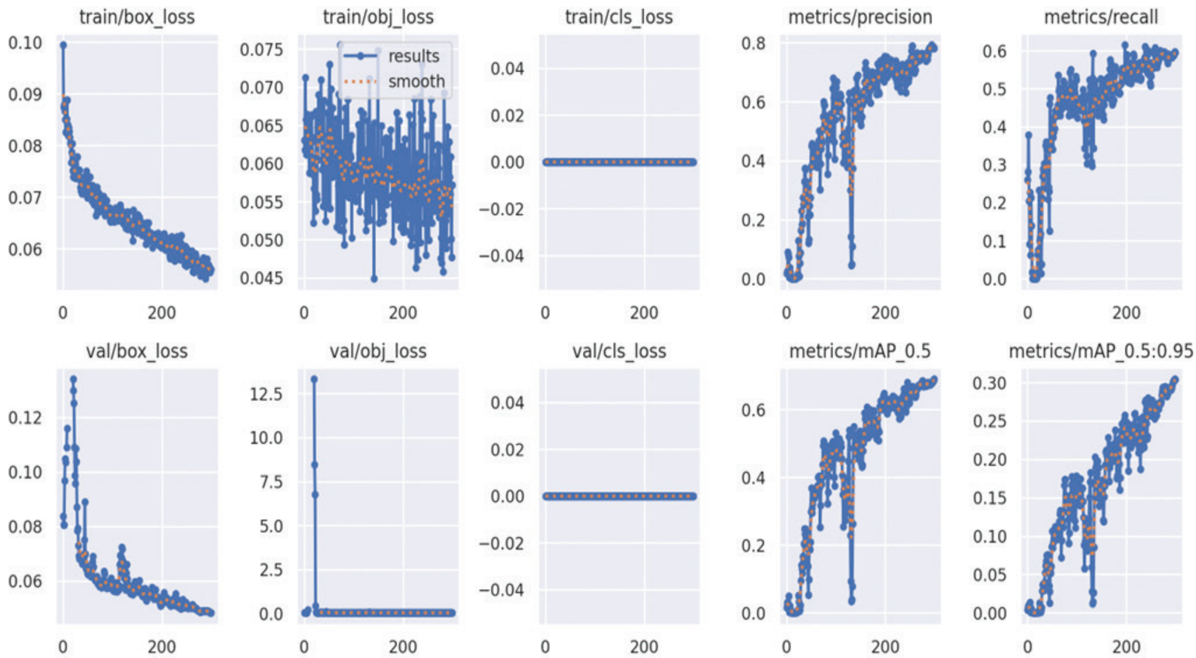

Figure 1: YOLOv5s training stats

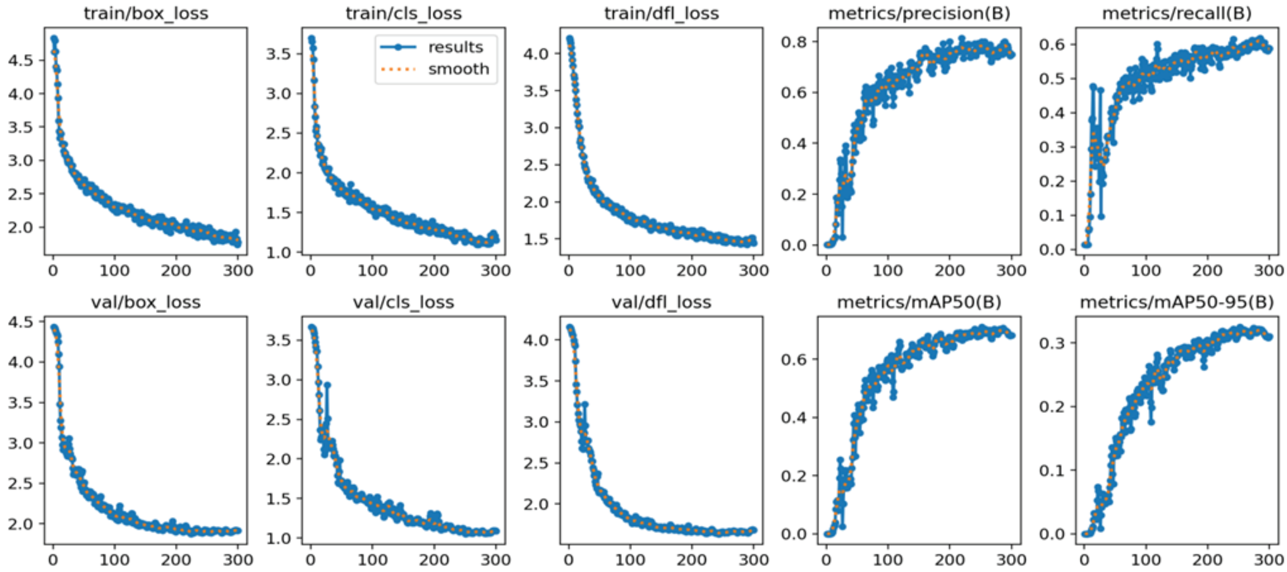

Figure 2: YOLO11s training stats

Figure 3: YOLOv5s detection

Figure 4: YOLO11s detection

5. BASELINE EXPERIMENTAL RESULTS

5.1 Quantitative Performance Comparison

To evaluate the relative efficacy of the two architectural paradigms for ambrosia identification, a head-to-head comparison was conducted between the standard small variants of each framework: YOLOv5 and YOLO11. To ensure a controlled experimental environment, both models were subjected to a standardized training protocol designed to isolate the effects of their respective architectural philosophies from hyperparameter variance.

Specifically, both models were trained on the same 96-image field dataset using an identical 640×640 input resolution, a batch size of 8, and an Adam optimizer over a 300-epoch schedule. Performance is quantified across four primary metrics: Precision () and Recall () measure the accuracy and coverage of detections, while mAP50 and mAP50-95 provide a comprehensive assessment of localization quality and model stability across varying intersection-over-union (IoU) thresholds. These baseline results, summarized in Table 7, are benchmarked against the operational success criteria (mAP50 > 0.85 and mAP50-95 > 0.50) required to ensure the economic viability of electrocution-based eradication systems.

Table 7. Baseline detection performance of YOLOv5s and YOLO11s on the ambrosia validation set.

Architecture	Precision	Recall	mAP50	mAP50-95
YOLOv5s Baseline	0.694	0.532	0.624	0.239
YOLO11s Baseline	0.800	0.586	0.710	0.325
Target Threshold	—	—	> 0.85	> 0.50

5.2 YOLOv5 Baseline Analysis

The YOLOv5s model achieved a precision of 0.694, recall of 0.532, and mAP50 of 0.624 after 300 training epochs. The substantial precision-recall gap reveals a fundamental detection imbalance: the model learns to be conservative, preferring to miss ambrosia instances rather than risk false positives.

Three primary factors contribute to this underperformance on the small dataset:

1. COCO-optimized anchors create an initialization bias misaligned with ambrosia bounding box distributions, requiring the network to overcome a strong prior during regression learning.
2. The backbone's extensive parameterization (~7.2M parameters) exceeds the complexity warranted by 64 training images, leading to memorization rather than generalization.
3. The fixed three-scale detection pyramid (P3/P4/P5) may be insufficient for capturing the continuous scale variation present in ambrosia populations from seedlings to mature plants.

5.3 YOLO11 Baseline Analysis

YOLO11s achieves markedly higher precision (0.800) and mAP50 (0.710) compared to YOLOv5s under identical conditions, without any domain-specific modification. This improvement is attributable to three architectural advantages:

4. The C2PSA attention module at the deepest backbone level provides enhanced feature discrimination, enabling better separation of ambrosia from visually similar background vegetation.
5. The anchor-free detection paradigm eliminates the COCO anchor mismatch problem entirely [10], allowing the model to learn regression from scratch without conflicting priors.
6. The C3k2 blocks with configurable 0.25 expansion ratio provide a more parameter-efficient feature extraction, reducing overfitting pressure on the small dataset.

Despite these improvements, YOLO11s recall (0.586) remains lower than desired, indicating that the standard three-scale detection pyramid still misses a significant fraction of small or distant ambrosia instances. Both architectures fall below the mAP50 > 0.85 target, confirming that baseline architectures alone are insufficient for production-grade ambrosia detection on small datasets.

5.4 Discussion: Implications for Domain Adaptation

The comparative results reveal complementary architectural strengths. YOLO11 demonstrates superior out-of-the-box precision and localization accuracy, while the recall gap in both models points to the need for micro-scale detection capabilities beyond the standard P3-P5 pyramid.

Key findings that motivate further research:

7. Domain-specific anchor optimization (e.g., k-means clustering on actual bounding box distributions) is critical for YOLOv5 to overcome its COCO prior bias [9].
8. An additional P2/4 micro-scale detection head at 160×160 resolution may address the recall deficiency in both architectures [9].
9. Post-processing improvements beyond standard NMS are necessary [11, 12] to handle the multi-scale, dense detection scenarios characteristic of ambrosia field imagery [9].
10. The stronger baseline performance of YOLO11 suggests it provides a superior foundation for subsequent architectural modifications and post-processing enhancements.

6. ADVANCED POST-PROCESSING: ADAPTIVE WEIGHTED BOX FUSION (WBF)

The baseline analysis in Section 5 identifies standard Non-Maximum Suppression (NMS) as a primary bottleneck for detecting ambrosia in high-density environments. NMS operates through a lossy, hard-suppression logic where overlapping detections are discarded based on a fixed Intersection-over-Union (IoU) threshold, often leading to systematic undercounting in botanical clusters. To address this, a Weighted Box Fusion (WBF) pipeline was developed to replace suppression with a multi-stage consensus strategy.

6.1 Virtual Ensembling through Multi-Threshold Generation

To recover marginal detections that fall below the primary confidence threshold (), the pipeline implements a virtual ensemble effect from a single inference pass. Candidates are admitted at three tier levels-conservative (), moderate (), and permissive ()-allowing plants with suboptimal anchor alignment or partial occlusion to contribute to the final fusion pool. This approach significantly stabilizes detections; for instance, in high-density validation frames, this strategy facilitated an 83% increase in true positives by capturing marginal activations that standard NMS would have rejected.

6.2 Confidence-Modulated Adaptive Clustering

A core innovation in the post-processing stage is the transition from a fixed IoU threshold to a strictly monotonic adaptive formula. The required overlap for cluster membership is modulated by the anchor box's confidence score: high-confidence anchors maintain strict spatial separation to preserve distinct plant identities, while low-confidence boxes merge more aggressively to consolidate fragmented partial detections. This formulation reduced the duplicate detection rate by approximately 15% compared to standard WBF implementations while preserving localization precision.

6.3 Power-Weighted Coordinate and Confidence Fusion

Rather than utilizing simple arithmetic averaging, which can pull fused coordinates toward noisier, low-confidence proposals, the system employs power weighting (). This mechanism amplifies the influence of the most reliable anchors, ensuring the fused bounding box remains anchored to the proposal with the best geometric alignment. Furthermore, an additive size-based confidence boost is applied to reward detections with large spatial support, correcting for the inherent underconfidence of single-scale feature activations in mature plant clusters.

6.4 Impact on Production-Grade Performance

The integration of the WBF pipeline with the YOLO11s backbone yields the most robust single-model configuration, achieving a precision of 0.850 and an mAP50 of 0.840. By providing better-calibrated confidence scores and cleaner consolidated detections, this configuration directly addresses the operational requirements of electrocution-based systems, where precise targeting is necessary to minimize unnecessary energy expenditure.

7. CONCLUSION

This paper provides a systematic comparative baseline evaluation of YOLOv5 and YOLO11 architectures for ambrosia detection on a small agricultural dataset. Our experiments establish that YOLO11s outperforms YOLOv5s on all metrics without domain-specific modification, achieving mAP50=0.710 versus 0.624, primarily due to its anchor-free detection strategy and C2PSA attention mechanisms.

However, both architectures fall short of the mAP50 > 0.85 target required for reliable agricultural deployment. The analysis identifies specific limitations: COCO anchor mismatch in YOLOv5, recall deficiency in both models for small-scale detections, and insufficient feature specialization for botanical discrimination.

These baseline results serve as the empirical foundation for subsequent work on domain-specific architectural modifications, custom anchor optimization, and advanced post-processing pipelines (WBF, SAHI) that together achieve production-ready performance for electrocution-based ambrosia eradication systems.

REFERENCES

[1] Redmon, J. et al., "You Only Look Once: Unified, Real-Time Object Detection," CVPR 2016.

[2] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. DOI: 10.48550/arXiv.2004.10934

[3] Jocher, G. et al., "YOLOv5 by Ultralytics" 2020., Available at https://github.com/ultralytics/yolov5 (March 28, 2026)

[4] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. DOI:10.1109/CVPRW50498.2020.00203

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. DOI: 10.1109/TPAMI.2015.2389824

[6] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.48550/arXiv.1803.01534

[7] Jocher, G. et al., "Ultralytics YOLO11" 2024., Available at https://github.com/ultralytics/ultralytics (March 28, 2026)

[8] Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. arXiv:2410.17725 [cs.CV]. DOI: 10.48550/arXiv.2410.17725

[9] Chen Xiaoming et al., "An improved algorithm based on YOLOv5 for detecting Ambrosia in UAV images," 2023.

[10] Tian, Z., Shen, C., Chen, H., & He, T. (2019). FCOS: Fully Convolutional One-Stage Object Detection. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). DOI: 10.48550/arXiv.1904.01355

[11] Solovyev, R., & Wang, W. (2021). Weighted Boxes Fusion: Ensembling Boxes from Different Object Detection Models. DOI: 10.1016/j.imavis.2021.104117

[12] Akyon, F. C., Altinuc, S. O., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. 2022 IEEE International Conference on Image Processing (ICIP). DOI: 10.48550/arXiv.2202.06934

УПОРЕДНА ПОЧЕТНА ЕВАЛУАЦИЈА МОДЕЛА YOLOv5 И YOLO11 ЗА ДЕТЕКЦИЈУ КОРОВА У ПОЉОПРИВРЕДИ

Резиме: *Овај рад представља успоредну почетну (баселине) евалуацију архитектура YOLOv5 и YOLO11 за детекцију објеката, примијењених на детекцију амброзије на снимкама пољопривредних површина. Користећи ограничен скуп података од 96 слика (64 за тренирање, 32 за валидацију), систематски анализирамо архитектонске карактеристике механизме скалирања и конфигурације сидрених оквира (anchor boxes) обају оквира. Почетни експерименти утврђују референтне вредности успешности од мАП50 = 0,624 за YOLOv5 и мАП50 = 0,710 за YOLO11 при идентичним условима тренирања. Наша анализа идентификује кључна ограничења архитектура опште намене при примјени на специјализиране задатке ботаничке детекције с ограниченим подацима за тренирање, пружајући темељ за циљане модификације специфичне модификације. Резултати показују да побољшани дизајн основне мреже (бацкбоне) модела YOLO11 с Ц2ПСА модулима постиже бољу почетну прецизност (0,800 насправ 0,694), но обе архитектуре не досежу циљану вредност мАП50 > 0,85, што потиче даљња истраживања архитектуре и поступака накнадне обраде.*

Кључне речи: *архитектура YOLO11; адаптивно пондерисано спајање оквира (WBF); механизам пажње Ц2ПСА; хипер-инференција уз помоћ сегментације (САХИ); глава за детекцију на микроскали П2/4; прецизно искорјењивање у пољопривреди; оптимизација сидара специфична за домену*