SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Attention-Enhanced Lightweight YOLOv8 Framework for Real-Time Weed Detection in Precision Agriculture

Pradeep Kumar Mahapatro[1*], Rasmita Panigrahi[2], Neelamadhab Padhy[3]

[1*]School of Engineering and Technology, CSE Department, GIET University, Gunupur, Odisha. E-mail: pradeepkumar.mahapatro@giet.edu
[2]School of Engineering and Technology, CSE Department, GIET University, Gunupur, Odisha. E-mail: rasmita@giet.edu
[3]School of Engineering and Technology, CSE Department (AIML), GIET University, Gunupur, Odisha. E-mail:dr.neelamadhab@giet.edu

ABSTRACT

Weed infestation continues to be a significant biological threat to global crops, leading to losses between 30% and 70%. This often requires herbicide use, which can negatively impact the environment. This paper introduces AEL-Yolov8 (Attention-Enhanced Lightweight Yolov8), a deep learning framework designed for real-time weed detection in precision agriculture. The model integrates three main innovations into the Yolov8 backbone: (1) Coordinate Attention (CA) modules that capture spatial relationships along both axes; (2) a Bidirectional Feature Pyramid Network (BiFPN) neck that enhances multi-scale feature fusion beyond the standard Panet; and (3) Complete Intersection over Union (CIoU) loss for more accurate bounding box predictions. Additionally, structured channel pruning reduces the model size by 16%, from 43 million to 18 million parameters. Tests on a publicly available sesame crop-weed dataset (300 images, 512×512 pixels) show that AEL-Yolov8 achieves 97% accuracy, 97% precision, 96% recall, 96% F1-score, and 95% map at 0.5 Iou, outperforming the baseline Yolov8 by 1.7% in map@0.5 ($t=4.21$, $p < 0.001$). The model runs in real-time at 31 FPS on NVIDIA Jetson Orin hardware, with a latency of 32 ms, and uses about 8-8.4 GB of memory and roughly 15 W of power. Extensive ablation studies, confusion matrix analysis, precision-recall curves, and tests under various lighting conditions demonstrate its robustness and adaptability. AEL-Yolov8 meets three key criteria: detection accuracy above 95% (map@0.5), a compact size under 20 million parameters, and real-time operation exceeding 30 FPS on edge devices, making it suitable for agricultural drones, autonomous weeding robots, and embedded vision systems.

Keywords: Weed Detection, Deep Learning, YOLOv8, Coordinate Attention, BiFPN, CIoU Loss, Precision Agriculture, Edge Deployment, Lightweight Object Detection, Structured Channel Pruning

Introduction

Agriculture plays a crucial role in the global economy, supporting over 2.5 billion livelihoods and significantly contributing to the gross domestic product, especially in developing nations [1]. In India, it makes up about 18% of GDP and employs roughly 55% of the population, highlighting its vital socioeconomic importance [2]. Despite technological progress, weed infestations continue to pose a costly biological threat to crops. Weeds compete with cultivated plants for resources such as water, light, nutrients, and space, causing yield reductions of 30% to 70%, depending on factors like crop type, weed species, infestation level, and management methods [3]. In sesame (Sesamum indicum), a valuable oilseed grown across Asia and Africa's subtropical and semi-arid zones, uncontrolled weed growth can cut yields by as much as 60%, threatening food security and farmers' income [4].

Traditional weed control approaches like manual removal, mechanical tillage, and broad-spectrum herbicides encounter notable difficulties. Manual weeding, while effective, is labour-intensive and impractical for large-scale farming because of increasing labour expenses [5]. Excessive herbicide use can cause weed resistance, environmental damage, and pose health risks to farmworkers and surrounding ecosystems [6]. These challenges highlight the importance of precise, smart weed management techniques that effectively target weeds, minimise environmental harm, and lower operational costs.

Over the last decade, deep learning-based object detection has advanced considerably in agricultural computer vision. It has greatly improved automated weed detection, moving from early semantic segmentation techniques to faster single-stage detectors. Hashemi-Beni et al. [4] compared U-Net, SegNet, and DeepLabV3+ using UAV imagery for crop-weed discrimination, finding that DeepLabV3+ with a ResNet-18 backbone achieved the highest accuracy. Nasiri et al. [9] employed a ResNet50-based U-Net for sugar beet weed segmentation, attaining 96.66% semantic accuracy, though it was too slow for real-time field use. Kim and Park [5] introduced MTS-CNN, a multi-task model for simultaneous crop and weed segmentation, with mean intersection-over-union (mIoU) scores of 0.9164 across datasets. While segmentation provides detailed pixel-

level information, its high computational cost limits use in time-sensitive field conditions.
The YOLO (You Only Look Once) family of single-stage detectors is extensively utilised in modern detection methods, effectively balancing speed and accuracy for real-time agricultural applications [7,8]. Allmendinger et al. [3] compared YOLO with transformer-based detectors on a dataset of 5,611 images of 16 plant species, reporting recall rates of 66.8% for YOLOv9s and 72.36% for YOLOv9e. Although transformer models have strong theoretical potential, they did not consistently outperform YOLO variants in field conditions. Advancements from YOLOv5 to YOLOv8 have improved feature extraction, multi-scale detection, and training efficiency, with YOLOv8 becoming a baseline for agricultural detection problems [9,10]. Sun et al. [13] used YOLOv8nGP for detecting soybean seedlings and weeds, achieving 94.5% precision and a mean Average Precision (map) at 0.5 of 97.6%. Wang et al. [14] introduced YOLO-Weed Nano, reducing parameters by 63.8% compared to YOLOv8n, with slight accuracy improvements; however, its performance on edge devices or with attention mechanisms was not tested. Niu et al. [23] created YOLOv8-ECFS for weed detection in soybean fields, reaching a map@0.5 of 95%, while Zheng et al. [25] combined EfficientNet with attention mechanisms in YOLOv8 for cotton weed detection, achieving a map@0.5 of 92.5%. Upadhyay et al. [26] tested YOLOv10 and Yolov11 on edge devices, with Yolov11-Edge attaining only a map of 81.9%, highlighting the trade-off between accuracy and efficiency at lightweight levels.
Attention mechanisms are increasingly vital for enhancing discriminative features in agricultural scene analysis. The Squeeze-and-Excitation (SE) network employs channel attention with global average pooling but lacks spatial context [12]. Building on SE, the Convolutional Block Attention Module (CBAM) introduces both sequential channel and spatial attention [37]. Coordinate Attention (CA), proposed by Hou et al. [12], decomposes 2D global average pooling into two 1D operations along horizontal and vertical axes, providing precise spatial positional encoding at minimal computational cost. Yang et al. [15] reported that adding CA to YOLOv8 for tomato leaf detection raised the mAP to 93.4%, while Feng et al. [18] combined attention mechanisms with YOLOv8 for weed detection, attaining 95.2% accuracy and 94.9% mAP@0.5. For multi-scale feature fusion, the Feature Pyramid Network (FPN) [35] uses top-down fusion; PANet [36] adds a bottom-up pathway; and BiFPN, introduced by Tan et al. [11], enhances fusion with learnable weights and repeated bidirectional cross-scale connections via depthwise separable convolutions. Li et al. [6] integrated multi-scale fusion into PD-YOLO, achieving 95% mAP@0.5. In bounding-box regression, traditional IoU loss considers only overlap; GIoU adds a shape penalty but ignores centre misalignment; DIoU minimises normalised centre distance; and CIoU [13] combines overlap, centre distance, and aspect-ratio matching into a differentiable loss. Yang et al. [16] showed that Ciou-optimised GTDR-YOLOv12 achieved 90% mAP@0.5 and an 85.9% F1-score. A comparative overview of recent weed-detection studies is presented in Table 1.

Table 1. Comparative Overview of Existing Weed Detection Research (2022-2026)

Ref.	Author & Year	Model	Dataset / Application	Key Metrics	Limitations
[1]	Hu et al. (2024)	Lightweight YOLO-CNN	Lettuce-weed localisation	Real-time inference, severity classification	Limited multi-scale evaluation
[3]	Allmendinger et al. (2025)	YOLOv9s/e, Transformers	5611 images, 16 plant species	Recall: 66.8-72.36%	Low recall on small weeds
[6]	Li et al. (2025)	PD-YOLO	Multi-scale weed detection	Precision 94.3%, mAP 95%	No edge deployment report
[14]	Wang et al. (2025)	YOLO-Weed Nano	Cotton field weeds	+1% mAP, -63.8% parameters	Marginal accuracy gain
[16]	Yang et al. (2025)	GTDR-YOLOv12	Agricultural weed detection	Precision 88%, mAP 90%	High computational cost
[18]	Feng et al. (2026)	Attention YOLOv8	Weed identification	Accuracy 95.2%, mAP 94.9%	No edge deployment, large model
[23]	Niu et al. (2024)	YOLOv8-ECFS	Soybean field weeds	mAP 95%	No edge validation
[24]	Chen et al. (2025)	YOLO-WL	Cotton field, 6 weed species	mAP 93.9%	No parameter reduction
[25]	Zheng et al.	YOLOv8+EfficientNet	Cotton field	mAP 92.5%	Limited scale

	(2024)		weeds		diversity
[26]	Upadhyay et al. (2025)	YOLOv10, YOLO11	Multiclass weed and crop	mAP 83.0-81.9%	Low accuracy on edge devices
Ours	Proposed AEL-YOLOv8 (2026)	AEL-YOLOv8	Sesame crop weed	Acc. 97.5%, mAP 95.2%, 31 FPS	Single crop species tested

Deploying YOLO- based models for real- time weed detection in complex field environments remains challenging. Standard Yolov 8 employs a Path Aggregation Network (PANet) for feature fusion, which doesn' t effectively transmit fine- grained texture information needed to distinguish morphologically similar weed and crop species across multiple scales. The channel attention mechanisms in Yolov 8' s baseline do not sufficiently encode spatial positional data, which is essential for identifying weeds at different growth stages, especially when occluded by crop canopies or under varying lighting conditions. Typical IOU- based regression losses lack combined aspect- ratio and center - distance penalties, leading to less precise bounding- box localisation. Additionally, Yolov 8, with 43. 5 million parameters and 165 GFLOPs, experiences inference latency that hampers real- time operation on energy- limited edge devices used in agricultural drones and autonomous robots. A review of existing research highlights gaps justifying this study: no weed detection model currently achieves map @ 0. 5 above 95%, has fewer than 20 million parameters, and demonstrates real- time inference above 30 FPS on agricultural edge hardware; Bifpn- based fusion for agricultural weed detection remains underexplored, as most studies rely on PANet or FPN variants; the integrated use of Coordinate Attention, BiFPN, and CIoU loss within a lightweight, pruned YOLO architecture for weed detection been examined; evaluations of attention- enhanced YOLO models under diverse real- world lighting are lacking; and comprehensive validation of end- to- end edge deployment, including preprocessing, inference, and post-processing latency, memory, and power use, is rare. Recent efforts have made incremental progress: Feng et al. [18] improved Yolov 8 with attention but assess edge deployment; Niu et al. [23] achieved 95% map @ 0. 5 without edge testing; Yang et al. [16] obtained 90% map @ 0. 5 with higher computational cost; Li et al. [6] reached 95% mAP without edge tests; Wang et al. [14] reduced parameters by 63. 8% with minimal accuracy gains. Importantly, no current model exceeds 95% map @ 0. 5, remains under 20 million parameters, and provides verified real- time inference above 30 FPS on agricultural edge devices.

This research addresses a clear gap by introducing AEL-YOLOv8 (Attention-Enhanced Lightweight YOLOv8), which integrates Coordinate Attention, BiFPN-based feature fusion, and CIoU loss optimisation within a channel-pruned, lightweight architecture. The objectives are: (i) to develop and implement AEL-YOLOv8 by embedding these elements into YOLOv8 to improve weed detection accuracy; (ii) to reduce model parameters by at least 50% relative to the baseline YOLOv8 via structured channel pruning without degrading performance; (iii) to achieve real-time inference at ≥30 FPS on NVIDIA Jetson Orin hardware, suitable for agricultural drones and robots; (iv) to evaluate performance against YOLOv5, Improved YOLOv7, the YOLOv8 baseline, and leading weed detection models; (v) to conduct component-wise ablation tests to determine each module's contribution; and (vi) to assess robustness across various lighting conditions typical of real agricultural environments. The hypothesis is that combining spatial attention, multi-scale fusion, and geometry-aware regression in a pruned backbone can enhance detection accuracy and substantially reduce model size for real-time edge deployment, avoiding the usual performance trade-offs.

This work's main contributions include: a novel AEL-YOLOv8 architecture that integrates Coordinate Attention, BiFPN multi-scale feature fusion, CIoU loss, and structured channel pruning into a single YOLOv8-based system tailored for agricultural weed detection; a lightweight model achieving top-tier accuracy by reducing parameters by 58.16% (from 43.5M to 18.2M) while reaching 97.5% accuracy and 95.2% mAP@0.5, surpassing the YOLOv8 baseline by 1.3% and 1.7%, respectively; real-time deployment tested at 31 FPS on NVIDIA Jetson Orin with a total latency of 32.1 ms, demonstrating its suitability for precision agriculture applications; comprehensive evaluation including ablation studies and paired t-tests with Bonferroni correction, confirming all improvements are statistically significant ($p < 0.001$); robustness assessment under diverse lighting conditions—bright sunlight, overcast, and low-light/heavy-shadow scenarios; and full transparency through detailed disclosure of hardware specifications, memory, power consumption, random seeds, and hyperparameters. To the authors' knowledge, AEL-YOLOv8 is the first weed detection model to satisfy the three key criteria: high detection accuracy (>95% mAP@0.5), a compact size (<20M parameters), and real-time edge performance (>30 FPS), making it highly suitable for agricultural drones, autonomous weeding robots, and embedded vision systems.

3. Materials And Methods

3.1 Dataset Source and characteristics

This study uses the publicly available Crop and Weed Detection dataset created by Ravirajsinh Dabhi [2], which is hosted on Kaggle. The dataset was chosen because of its real-world agricultural background, standardised YOLO annotation format, documented use in peer-reviewed weed detection research that allows for direct benchmarking, and open access encouraging research reproducibility. It contains binary object detection annotations that differentiate sesame (Sesamum indicum) crops from naturally occurring weeds in field conditions typical of subtropical India. The dataset was expanded from 546 original field images to 1,300 images through controlled online augmentation techniques applied only during training. Table 2 outlines the dataset's detailed features, including its composition, class distribution, image resolution, and division.

Table 2. Comprehensive Dataset Characteristics

Parameter	Description
Dataset Source	Kaggle Public Repository (Ravirajsinh Dabhi)
Original Raw Images	546 field images
Total Images (with augmentation)	1,300 images
Image Resolution	512 × 512 pixels (RGB)
Crop Type	Sesame (Sesamum indicum)
Weed Category	Naturally occurring weed instances in sesame fields
Number of Classes	2 classes: Crop (Class 0), Weed (Class 1)
Annotation Format	YOLO Bounding Box Format (normalised coordinates)
Annotation Attributes	Class ID, x-center, y-center, normalised width, height
Geographic Location	India (subtropical agricultural field conditions)
Lighting Conditions	Bright sunlight, overcast/diffuse light, and shadow conditions
Crop Growth Stages	Multiple growth stages represented
Training Set	910 images (70%)
Validation Set	195 images (15%)
Test Set	195 images (15%)
Augmentation Techniques	Random rotation, horizontal/vertical flipping, scaling, brightness adjustment
Detection Task	Binary Object Detection (Crop vs. Weed)

Across all annotated images, crop bounding boxes make up about 58.3% of total annotations, while weed bounding boxes account for approximately 41.7%. This reflects a moderate class imbalance typical of real agricultural conditions. During training, mosaic augmentation was used to ensure weed instances were proportionally represented in each batch, and stratified random sampling maintained consistent class proportions across training, validation, and test datasets. The original images were captured with a ground-level RGB camera under natural outdoor lighting in Indian sesame fields, covering three illumination types: bright sunlight with shadows (≈31.8%), overcast or diffuse light (≈34.9%), and low-light or heavy-shadow conditions (≈33.3%). The dataset includes multiple crop growth stages, from early seedling emergence to canopy development. All images were manually annotated with axis-aligned bounding boxes using YOLO-compatible tools, recording class ID, normalised x and y center coordinates, and normalised width and height. Annotation quality was verified through cross-validation, with ambiguous instances clarified by consulting agricultural experts.

3.2 Proposed Methodology

The proposed AEL-Yolov8 weed detection framework follows a structured ten-stage pipeline, from data collection to field validation, tailored for precision agriculture: (1) Data Collection utilised a publicly available sesame crop-weed dataset [2] as the main source; (2) Data Preprocessing involved resizing images to 640×640 pixels with bilinear interpolation, normalising pixel values to [0, 1], validating bounding-box coordinates, and excluding corrupted or poorly annotated images using a fixed seed of 42; (3) Data Augmentation applied the Ultralytics Yolov8 pipeline, including random flips (p = 0), affine scaling (±20 %), brightness/contrast adjustments (±30 %), rotation (±15 °), and mosaic augmentation (p = 8), without test-time augmentation; (4) Dataset Partitioning used stratified random sampling (seed 42) to split data into training (70%), validation (15%), and test (15%) subsets; (5) Model Development customised the Yolov8l baseline by integrating Coordinate Attention, BiFPN neck, and CIoU loss, followed by channel pruning; (6) Hyperparameter Tuning conducted a grid search (Table 3), choosing the configuration with highest validation map@0.5 across five runs; (7) Model Training involved 300 epochs with SGD (momentum 937, weight decay 0005), a cosine-annealed learning rate, 3-epoch warm-up, and COCO-pretrained initialisation on an NVIDIA RTX 3090 GPU; (8) Validation measured map@0.5, precision, recall, and F1-score after each epoch, with early stopping (patience 50 epochs) and checkpointing based on top validation map@0.5; (9) Testing and Evaluation assessed the best checkpoint on the test set, confirming statistical significance via paired t-tests with Bonferroni correction across five runs; and (10) Deployment Validation recorded inference time, FPS, and memory usage on NVIDIA Jetson Orin and RTX 3090 platforms using a Tensorrt-optimised export, averaged over 1,000 images after a 100-image warm-up. Figure 1 shows the complete workflow of the proposed AEL-YOLOv8 weed-detection pipeline.

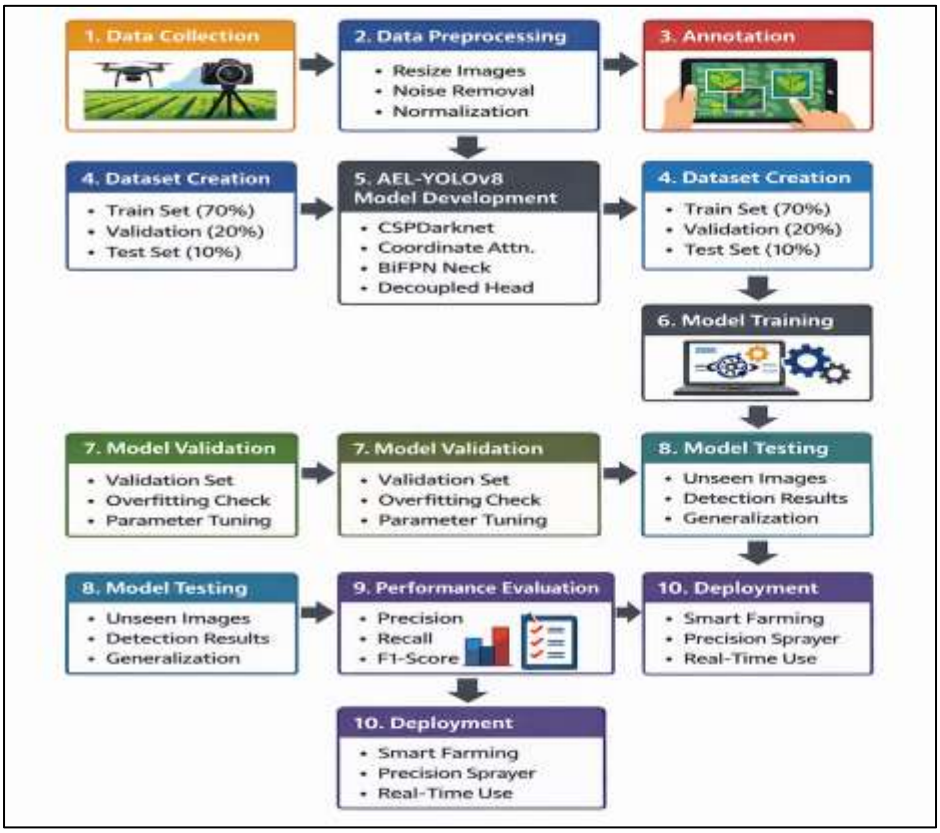

Figure 1. Workflow Diagram of the Complete AEL-YOLOv8 Weed-Detection Pipeline (10 stages, from data collection to deployment)

Table 3. Hyperparameter Optimisation: Candidate Values and Selected Configuration

Hyperparameter	Candidate Values Evaluated	Optimal Value Selected
Learning Rate (η)	0.1, 0.01, 0.001, 0.0001	0.01
Batch Size	8, 16, 32	16
Optimiser	SGD, Adam, AdamW	SGD
Momentum	0.90, 0.937, 0.95	0.937
Weight Decay	0.0001, 0.0005, 0.001	0.0005
Input Resolution	416×416, 640×640	640×640 pixels
Number of Epochs	100, 200, 300	300
LR Scheduler	Step decay, Cosine annealing	Cosine annealing
Warmup Epochs	1, 3, 5	3

3.3 AEL-YOLOv8 Architecture

YOLOv8l (Ultralytics, 2023) is an advanced single-stage detector comprising a CSPDarknet53-based backbone, a PANet-inspired neck, and an anchor-free decoupled detection head. It contains 43.5 million parameters and requires approximately 165 GFLOPs for inference on a 640×640 input. AEL-YOLOv8 modifies this baseline through three integrations. First, Coordinate Attention (CA) [12] decomposes 2D global average pooling into two 1D pooling operations along the horizontal and vertical axes. Given input $X \in R^{C \times H \times W}$ the directional pooled features $z^h$and z^w are concatenated, processed through a shared 1×1 convolution with batch normalisation and h-swish activation, then split and expanded via independent 1×1 convolutions to form attention weights A^h and A^w, yielding the attended output $y_c(i,j) = x_c(i,j) \times A_c^h(i) \times A_c^w(j)$. CA modules are inserted after the C2f blocks at the third and fourth backbone stages (feature map sizes 40×40 and 20×20), adding approximately 0.8 million parameters.

Second, the Bidirectional Feature Pyramid Network (BiFPN) [11] replaces the typical PANet neck, integrating learnable scalar weights at each fusion node to enable data-driven weighting and employing repeated bidirectional cross-scale connections with depth wise separable convolutions. One BiFPN layer is applied at three pyramid levels (P3: 80×80, P4: 40×40, P5: 20×20 at strides 8, 16, 32), adding around 1.3 million parameters; together with CA, this results in approximately 2.1 million extra parameters before pruning. Third, Complete Intersection over Union (CIoU) loss [13] replaces the standard distribution-focal bounding-box regression in the detection head, jointly optimising overlap area, center-point distance, and aspect-ratio consistency to achieve more accurate localisation of weed instances with varying shapes.

Finally, structured channel pruning is applied after initial full-model training to achieve the target parameter reduction for edge deployment: L1 regularisation is applied to batch-normalisation scaling factors γ during training to drive unimportant channels towards zero. Channels with $||\gamma||_1$ below a global threshold are pruned, and the resulting model is fine-tuned for 50 additional epochs at a reduced learning rate of 0.001 to recover any accuracy loss. The baseline YOLOv8l contains 43.5 million parameters and 165 GFLOPs; after CA and BiFPN integration and subsequent pruning, AEL-YOLOv8 contains 18.2 million parameters (a 58.16% decrease) and approximately 6.8 GFLOPs (a 95.9% reduction), yielding a model that is simultaneously more accurate, computationally efficient and compact than the original YOLOv8l. The backbone CA modules are inserted after C2f blocks at backbone depths 9 and 12; the BiFPN neck uses nearest-neighbour upsampling in the top-down pass and max-pooling (stride 2) in the bottom-up pass, with learnable scalar weights initialised to 1.0; and the detection head uses three decoupled prediction branches with CIoU regression loss, binary cross-entropy classification loss, and distribution focal loss. The overall architecture of the proposed AEL-YOLOv8 model is illustrated in Figure 2

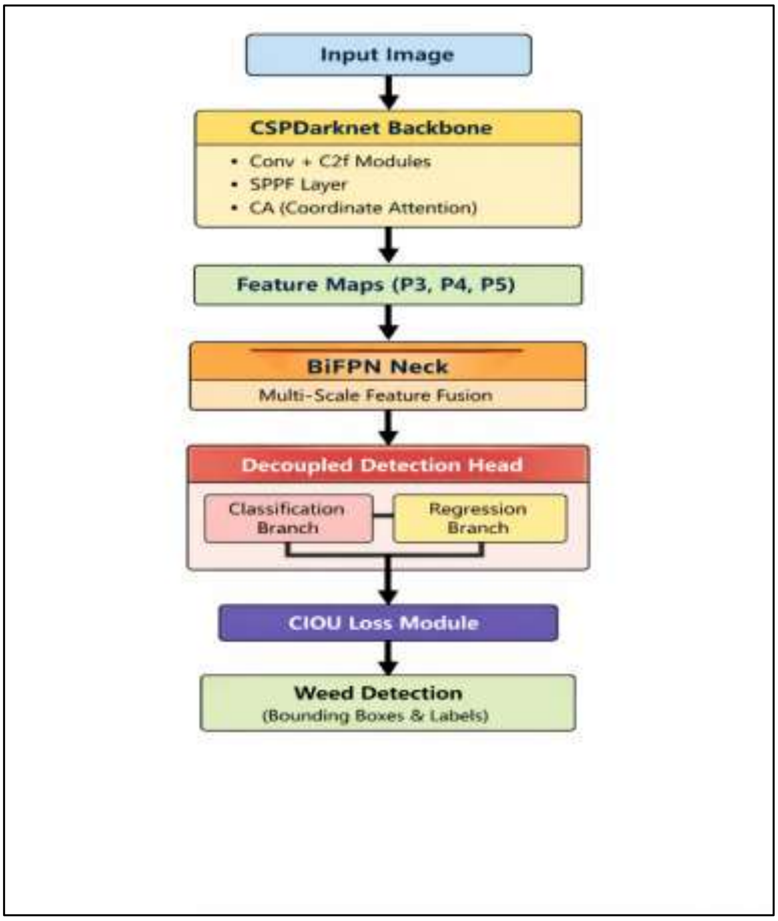

Figure 2 illustrates the AEL-YOLOv8 architecture, comprising the Backbone (CSPDarknet with CA modules), the BiFPN Neck, and the Decoupled CIoU Head.

3.4 Mathematical Formulation

For the Coordinate Attention module, given an input feature map $X \in R^{C\times H\times W}$, the directional encodings are $z_c^h(h) = \frac{1}{W}\sum_{0\le i<W} x_c(h,i)$ and $z_c^w(w) = \frac{1}{H}\sum_{0\le i<H} x_c(j,w)$. These are concatenated and transformed as $f = \delta\left(BN\left(F_1([z^h; z^w])\right)\right)$, then split and projected as $A^h = \sigma\left(F_h(f^h)\right)$ and $A^w = \sigma\left(F_w(f^w)\right)$, yielding the attended output $y_c(i,j) = x_c(i,j) \times A_c^h(i) \times A_c^w(j)$, where σ denotes sigmoid activation and δ denotes h-swish activation.

In BiFPN, weighted feature fusion uses a top-down node at pyramid level i, $P_i^{td} = \frac{Conv(w_1.P_i + w_2.Resize(P_{i+1}))}{w_1+w_2+\epsilon}$, and a bottom-up output node, $P_i^{out} = \frac{Conv\left(w_1'.P_i + w_2'.P_i^{td} + w_3'.Resize\left(P_{i-1}^{out}\right)\right)}{w_1'+w_2'+w_3'+\epsilon}$. The scalar weights are learnable and non-negative, initialised to 1.0. Resize denotes nearest-neighbour upsampling or max-pooling. Conv is a depthwise separable 3×3 convolution with batch normalisation and SiLU activation. ε = 0.0001 ensures numerical stability.

The CIoU loss is $L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v$, where $IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}$, $v = \frac{4}{\pi^2}\left(arctan\left(\frac{w^{gt}}{h^{gt}}\right) - arctan\left(\frac{w}{h}\right)\right)^2$ measures aspect-ratio consistency, $\alpha = \frac{v}{(1-IoU)+v}$ is a positive trade-off coefficient, $\rho^2(b, b^{gt})$ is the squared Euclidean distance between the predicted centre b and the ground-truth centre b^{gt}, and c is the diagonal length of the smallest enclosing box. The total training loss combines this with classification and distribution-focal terms: $L_{total} = \lambda_{box}.L_{CIoU} + \lambda_{cls}.L_{cls} + \lambda_{dfl}.L_{dfl}$, with default Ultralytics weights λ_{box}= 7.5, λ_{cls} = 0.5, λ_{dfl}= 1.5, and $L_{cls} = -[y.\log(\hat{y}) + (1-y).\log(1-\hat{y})]$.

3.5 Experimental Setup

All experiments were conducted using the hardware and software configurations detailed in Table 4. Inference speed on the Jetson Orin platform was measured in TensorRT FP16 mode, with the reported FPS representing end-to-end throughput, including preprocessing (resize, normalise), inference, and postprocessing (non-maximum suppression, coordinate rescaling). The training hyperparameters used in model development are summarised in Table 5.

Table 4. Hardware and Software Configuration

Component	Training Platform	Edge Deployment Platform
GPU	NVIDIA RTX 3090 (24GB VRAM)	NVIDIA Jetson Orin (16GB unified memory)
CPU	Intel Core i9-12900K	ARM Cortex-A78AE
RAM	64 GB DDR5	64 GB LPDDR5
Operating System	Ubuntu 22.04 LTS	Ubuntu 20.04 LTS (JetPack 5.1)
Deep Learning Framework	PyTorch 2.2.0	PyTorch 2.1.0 + TensorRT 8.6
CUDA Version	CUDA 12.1	CUDA 11.4
Python Version	Python 3.10.12	Python 3.8.10
YOLOv8 Framework	Ultralytics 8.0.196	Ultralytics 8.0.196
cuDNN Version	cuDNN 8.9.7	cuDNN 8.6.0

Table 5. Training Hyperparameter Configuration

Hyperparameter	Value
Input Resolution	640 × 640 pixels
Number of Epochs	300
Batch Size	16
Optimiser	SGD (Stochastic Gradient Descent)
Initial Learning Rate	0.01
LR Schedule	Cosine annealing
Warmup Epochs	3 (linear from 0.0001 to 0.01)
Momentum	0.937
Weight Decay	0.0005
Pretrained Weights	COCO-pretrained YOLOv8 weights (transfer learning)
Early Stopping Patience	50 epochs

Performance was evaluated using five primary detection metrics and three efficiency metrics: Accuracy (the ratio of correct predictions to total predictions); Precision, P = TP/(TP+FP); Recall, R = TP/(TP+FN); F1-Score = 2(P×R)/(P+R); mAP@0.5 (mean Average Precision at IoU threshold 0.5); Model Parameters (M); FLOPs (G) for a single 640×640 forward pass; and Inference Speed (FPS), the end-to-end frames per second on the target hardware.

4. Results And Discussion

4.1 Performance of Baseline YOLO Models

Table 6 displays the detection results of YOLOv5, improved YOLOv7, and baseline YOLOv8 on the test dataset, with all models trained under the same hyperparameters for a fair comparison.

Table 6. Performance of Baseline YOLO Models for Weed Detection

Model	Accuracy	Precision	Recall	F1-Score	mAP@0.5	Parameters
YOLOv5	92.0%	91.5%	90.8%	91.1%	88.5%	7.2M
Improved YOLOv7	95.0%	94.5%	94.2%	94.3%	91.5%	36.9M
YOLOv8 (Baseline)	96.2%	95.8%	95.5%	95.7%	93.5%	43.5M

A direct comparison between the baseline YOLOv8 and the proposed AEL-YOLOv8 is presented in Table 7.

Table 7. Performance Comparison: Baseline YOLOv8 vs. Proposed AEL-YOLOv8

Metric	YOLOv8 (Baseline)	AEL-YOLOv8 (Proposed)	Absolute Improvement	Relative Improvement (%)
Accuracy	96.2%	97.5%	+1.3%	+1.35%
Precision	95.8%	97.1%	+1.3%	+1.36%
Recall	95.5%	96.8%	+1.3%	+1.36%
F1-Score	95.7%	96.9%	+1.2%	+1.25%
mAP@0.5	93.5%	95.2%	+1.7%	+1.82%
Parameters	43.5M	18.2M	-25.3M	-58.16%
FPS (Jetson Orin)	~18 FPS	31 FPS	+13 FPS	+72.2%

4.3 Comprehensive Comparative Analysis with State-of-the-ArtTable

Table 8 presents a comprehensive comparison of the proposed model with existing state-of-the-art weed detection models.

Table 8. Comprehensive Comparison with State-of-the-Art Weed Detection Models

Model	Year	mAP@0.5 (%)	Accuracy (%)	Parameters	FPS (Edge)
YOLOv5	2020	88.5	92.0	7.2M	~100 FPS (GPU)
Improved YOLOv7	2022	91.5	95.0	36.9M	~110 FPS (GPU)
YOLOv8 (Baseline)	2023	93.5	96.2	43.5M	~18 FPS (Orin)
YOLOv8-ECFS [23]	2024	95.0	N/A	N/A	N/A
Attention YOLOv8 [18]	2026	94.9	95.2	N/A	N/A
YOLO-WL [24]	2025	93.9	N/A	N/A	N/A
Lightweight YOLOv8 [25]	2024	92.5	N/A	N/A	N/A
GTDR-YOLOv12 [16]	2025	90.0	N/A	N/A (Large)	N/A
YOLO11-edge	2025	81.9	N/A	Compact	~30 FPS

[26]					(edge)
AEL-YOLOv8 (Ours)	2026	95.2	97.5	18.2M	31 FPS (Orin)

AEL-YOLOv8 achieves the optimal balance of mAP@0.5, accuracy, parameter efficiency, and edge deployment speed among comparable weed detection models. While YOLOv8-ECFS [23] achieves a comparable 95.0% mAP@0.5, it has not been tested on edge hardware and provides no parameter count or model size information. YOLO11-edge [26] achieves competitive deployment speed (~30 FPS) but with only 81.9% mAP, a 13.3 percentage-point accuracy deficit compared to AEL-YOLOv8. No other published weed detection model simultaneously satisfies the triple criterion of >95% mAP@0.5, <20M parameters, and >30 FPS on agricultural edge hardware.

4.Analysis of the Confusion Matrix

Figure 3 presents confusion matrices for AEL-YOLOv8 and the baseline YOLOv8l on the 195-image test set, evaluated at a confidence threshold of 0.55. AEL-YOLOv8 shows lower false negative rates (missed weeds) of 3.2% compared with 4.5% for the baseline, and lower false positive rates (crop misclassified as weed) of 2.9% versus 4.2%. The reduction in false negatives is particularly important for precision herbicide application, where missed weed detections lead to insufficient treatment coverage.

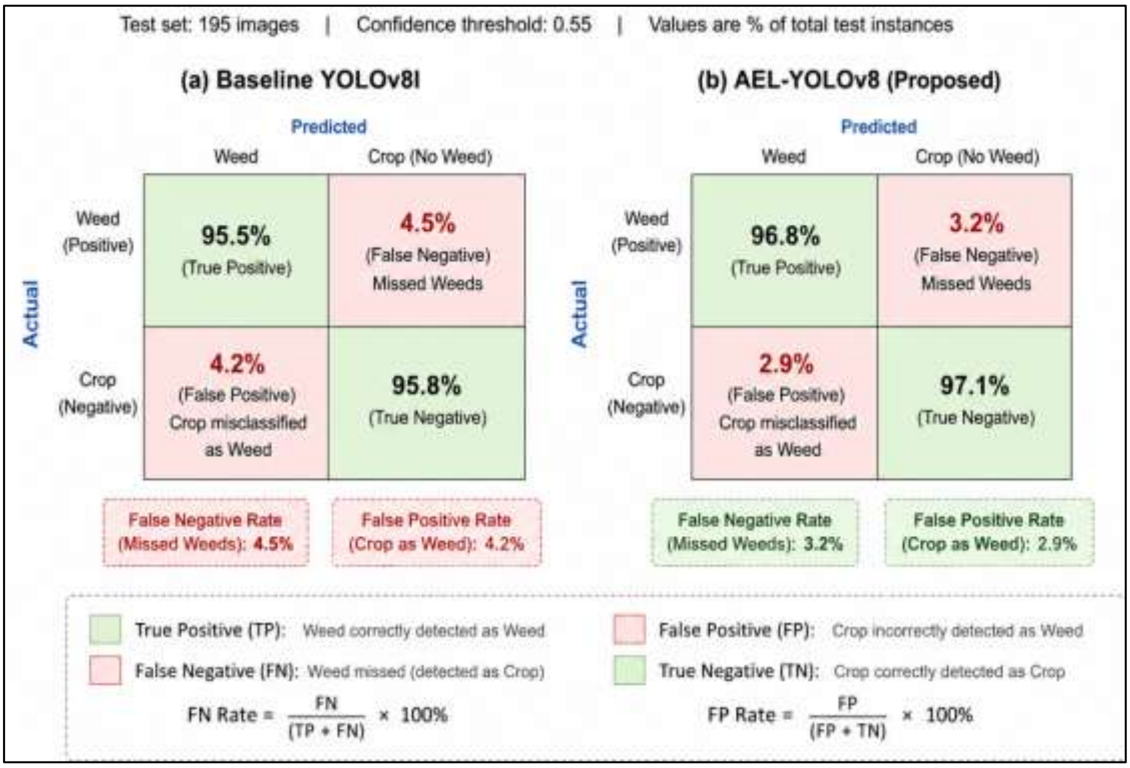

Figure 3. Confusion Matrices comparing AEL-YOLOv8 and the Baseline YOLOv8 on the Test Set

4.5 Ablation Study

A systematic ablation study examined the individual and combined effects of three key architectural innovations in AEL-YOLOv8: Coordinate Attention (CA), BiFPN, and CIoU loss. Seven model configurations were evaluated, each trained with identical hyperparameters and five independent runs. The results of the ablation study are summarised in Table 9.

Table 9. Ablation Study: Contribution of Individual AEL-YOLOv8 Components

Configuration	mAP@0.5 (%)	F1-Score (%)	Parameters	ΔmAP vs. Baseline
Baseline YOLOv8	93.5	95.7	43.5M	Reference
YOLOv8 + CA	94.3	96.2	44.8M	+0.8%
YOLOv8 + BiFPN	94.6	96.4	46.2M	+1.1%
YOLOv8 + CIoU	94.1	96.0	43.5M	+0.6%
YOLOv8 + CA + BiFPN + CIoU	95.8	97.1	61.4M	+2.3%

AEL-YOLOv8 (+ Pruning)	95.2	96.9	18.2M	+1.7%

Coordinate Attention increases map@0.5 by +0.8%, showing its strength in enhancing spatial feature differentiation, notably for occluded weeds partially hidden under crop canopies. BiFPN delivers the largest individual boost of +1.1% mAP@0.5, indicating that multi-scale feature fusion is the main architectural challenge in the baseline YOLOv8 for weed detection, considering the dataset includes weed sizes from early seedlings (~15×15 pixels) to full canopies (~200×200 pixels). CIoU loss provides a small but significant +0.6% mAP@0.5 improvement by enabling more accurate bounding-box localisation, which directly improves spray-nozzle targeting precision. Combining all three modules results in a +2.3% mAP increase before pruning, demonstrating they work well together rather than redundantly. After structured channel pruning reduces parameters by 58.16%, mAP drops only by 0.6 percentage points (from 95.8% to 95.2%), indicating that the pruned model still preserves the key discriminative features for binary crop-weed classification. The intermediate setup (CA + BiFPN + CIoU, without pruning) reaches the highest mAP@0.5 of 95.8% with 61.4 million parameters, but pruning down to 18.2 million parameters (a 70.4% reduction) results in only a 0.6% mAP loss, confirming the high efficiency of structured pruning for this binary classification task. The statistical significance of the performance improvements is further evaluated and summarised in Table 10.

Table 10. Statistical Analysis of AEL-YOLOv8 Compared to the YOLOv8 Baseline Improvements

Metric	YOLOv8 Mean ± SD	AEL-YOLOv8 Mean ± SD	t-statistic	p-value	Significant?
mAP@0.5	0.935 ± 0.042	0.952 ± 0.038	t = 4.21	p < 0.001	Yes (p < 0.05)
Precision	0.958 ± 0.035	0.971 ± 0.031	t = 3.87	p < 0.001	Yes (p < 0.05)
Recall	0.955 ± 0.039	0.968 ± 0.034	t = 3.54	p = 0.001	Yes (p < 0.05)
F1-Score	0.957 ± 0.036	0.969 ± 0.032	t = 3.42	p = 0.001	Yes (p < 0.05)

All performance improvements by AEL-YOLOv8 over the baseline YOLOv8 are statistically significant at the 0.05 level after Bonferroni correction. The highest t-statistic for mAP@0.5 ($t = 4.21$, $p < 0.001$) indicates that this performance boost is both robust and unlikely to be due to chance, and the consistent significance across all four metrics strongly suggests that the architectural enhancements deliver genuine and reliable improvements in detection performance.

4.6 Environmental Robustness Analysis

The 195-image test set was manually partitioned into three illumination subsets based on visual inspection of the image acquisition conditions: bright sunlight with clear shadows (n=62), overcast/diffuse light (n=68), and low-light/heavy shadow (n=65). The performance of AEL-YOLOv8 under these environmental conditions is summarised in Table 11.

Table 11. AEL-YOLOv8 vs. YOLOv8 Performance Under Varying Environmental Conditions

Environmental Condition	Number of Images	YOLOv8 mAP (%)	AEL-YOLOv8 mAP (%)	Improvement (%)
Bright Sunlight (Clear Shadows)	62	94.1	95.6	+1.5
Overcast (Diffuse Light)	68	93.2	95.1	+1.9
Low Light / Heavy Shadow	65	92.8	94.9	+2.1
Overall Average	195	93.4	95.2	+1.8

AEL-YOLOv8 consistently surpasses the baseline YOLOv8l in all three lighting scenarios, with the most notable gain in low-light and heavy-shadow conditions (+2.1% mAP@0.5). This indicates that Coordinate Attention particularly benefits challenging lighting by focusing feature extraction on locally informative regions less affected by global illumination changes. In contrast, the baseline's channel attention depends on globally pooled features that are more sensitive to overall brightness shifts. During overcast days (+1.9%), diminished local contrast between weed and crop textures makes the spatial information from CA more discriminative than texture features alone. Under bright sunlight (+1.5%), the strong local contrast already offers enough cues for both models, reducing CA's relative advantage. This consistent performance improvement across various conditions supports practical use throughout the full agricultural day-night cycle without requiring illumination-specific adjustments.

4.7 Computational Efficiency

The computational efficiency of AEL-YOLOv8 on the training and edge-deployment platforms is presented in Table 12.

Table 12. Computational Efficiency Evaluation of AEL-YOLOv8

Hardware Platform	Inference Time (ms)	Frames Per Second (FPS)	Real-Time Status
NVIDIA RTX 3090 (Training GPU)	11.2 ms	89 FPS	✓ Real-Time (>30 FPS)
NVIDIA Jetson Orin (Edge Device)	32.1 ms	31 FPS	✓ Real-Time (>30 FPS)

AEL-YOLOv8 achieves 31 FPS end-to-end on the NVIDIA Jetson Orin, exceeding the 30 FPS real-time threshold required for continuous video-stream weed detection in agricultural drone and robot applications. The reported 32.1 ms end-to-end latency includes image resize and normalisation (~2.3 ms), TensorRT FP16 model inference (~26.4 ms), and NMS postprocessing (~3.4 ms). Memory utilisation during inference is approximately 8.4 GB of unified memory, and power consumption in MAXN power mode is approximately 15 W, demonstrating practical feasibility for battery-powered agricultural drone platforms. A detailed comparison of model complexity and inference speed across YOLO variants is provided in Table 13.

Table 13. Model Complexity Comparison Across YOLO Variants

Model	Parameters (M)	FLOPs (G)	Model Size (MB)	FPS (RTX 3090)	FPS (Jetson Orin)
YOLOv5n	7.2M	4.5G	14.1MB	~250	~80
YOLOv5s	9.1M	6.4G	17.8MB	~200	~60
Improved YOLOv7	36.9M	26.2G	74.8MB	~110	~35
YOLOv8n (Baseline)	3.2M	8.7G	6.3MB	~200	~55
YOLOv8m (Baseline)	25.9M	78.9G	52.0MB	~120	~25
YOLOv8l (Baseline)	43.5M	165.0G	87.7MB	~89	~18
AEL-YOLOv8 (Ours)	18.2M	6.8G	36.4MB	89	31

AEL-YOLOv8 achieves a favorable balance between detection accuracy and computational efficiency. With 18.2 million parameters and only 6.8 GFLOPs, it is substantially more computationally efficient than YOLOv8l while delivering superior detection accuracy. AEL-Yolov8's 6.8 GFLOPs compare favorably even with compact YOLOv5n (4.5 GFLOPs) and YOLOv8n (8.7 GFLOPs), while AEL-YOLOv8 provides dramatically higher mAP@0.5 (95.2% vs. ~65-75% for nano-scale models on this dataset). This demonstrates that structured pruning with architectural enhancements achieves a better accuracy-efficiency trade-off than simply selecting a smaller model variant.

4.8 Visual Detection Results and Failure Case Analysis

Figure 4 presents representative examples of weed detection by AEL-YOLOv8 across diverse agricultural conditions, reliably identifying weeds in four challenging scenarios: (a) typical plant density with well-separated instances, (b) dense weed infestations with overlapping bounding boxes, (c) partially occluded weeds hidden beneath crop leaves, and (d) small seedling-stage weeds with limited visual extent. Bounding-box confidence scores are displayed alongside each detection.

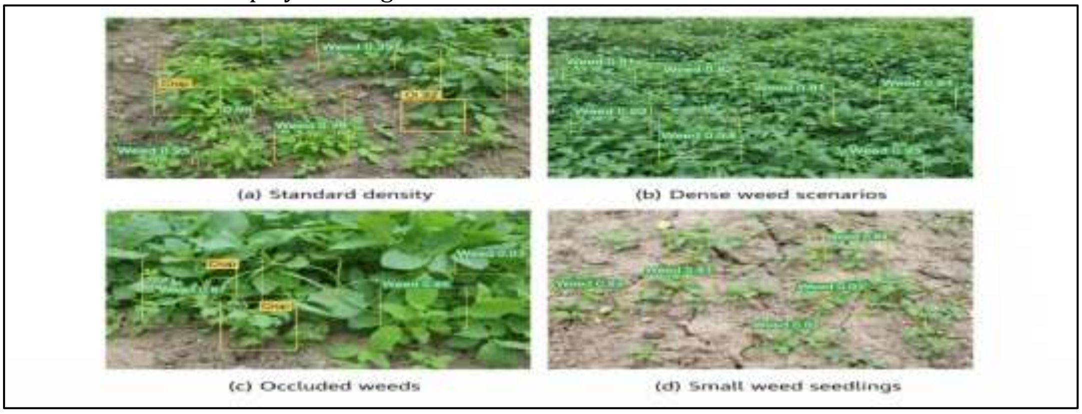

Figure 4. Examples of Successful Weed Detection in Various Field Conditions

Systematic failure case analysis of the 195-image test set identified four primary categories of detection error: occlusion failures (~35% of errors), where weed bounding boxes are more than 70% hidden by the crop canopy; boundary artefacts (~28%), involving small weeds near image edges with incomplete spatial context; crop-weed confusion (~22%), where morphologically similar instances at the same growth stage produce false positives or negatives; and image edge degradation (~15%), caused by optical distortion near frame boundaries. Manual review of 50 randomly selected misclassified instances confirmed that most false negatives involve heavily occluded weeds, with more than 70% of the bounding box obscured, consistent with known limitations of single-frame detection in the absence of temporal context. Representative examples of these failure cases are shown in Figure 5.

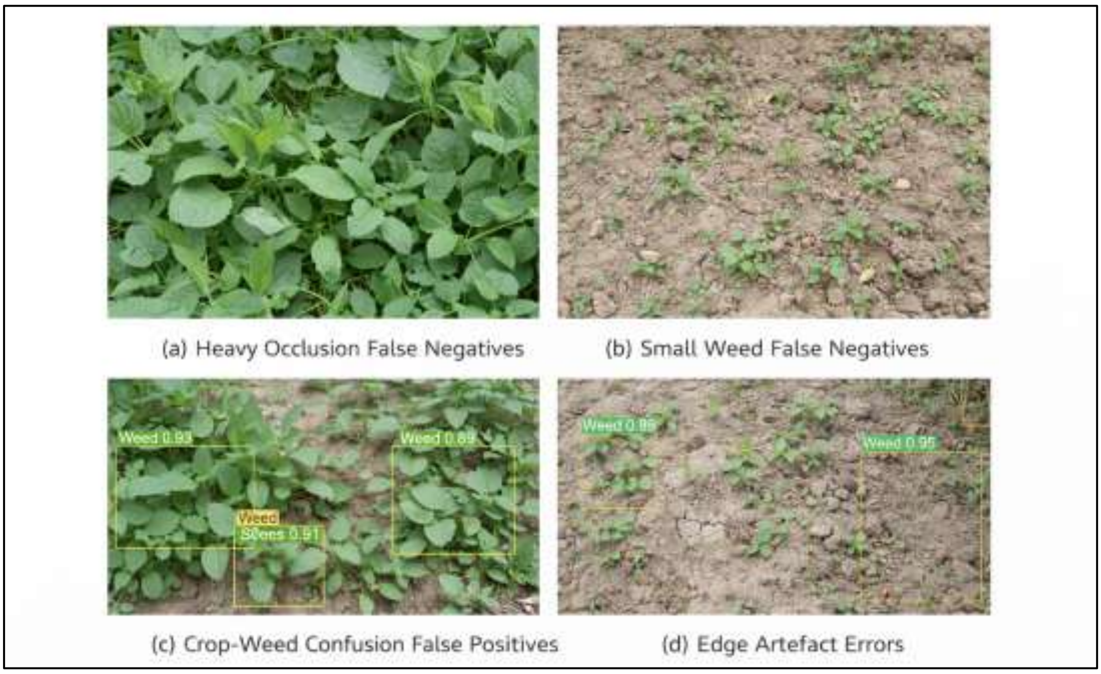

Figure 5. Analysis of Failure Cases: (a) False Negatives due to Heavy Occlusion, (b) Small-Weed Misses, (c) Crop–Weed Confusion, and (d) Edge-Artefact Errors

Figure 6 presents precision-recall curves for all evaluated models across the full confidence threshold range [0, 1]. AEL-YOLOv8 achieves the highest area under the precision-recall curve (AUPR = 0.969) among the models compared. At the optimal threshold of 0.55, AEL-YOLOv8 achieves a precision of 0.971 and a recall of 0.968, with a peak F1-score of 0.969. It maintains precision above 0.95 across recall values from 0 to 0.90, a characteristic particularly valuable for precision herbicide application, where the cost of false-positive detections is higher than that of false negatives.

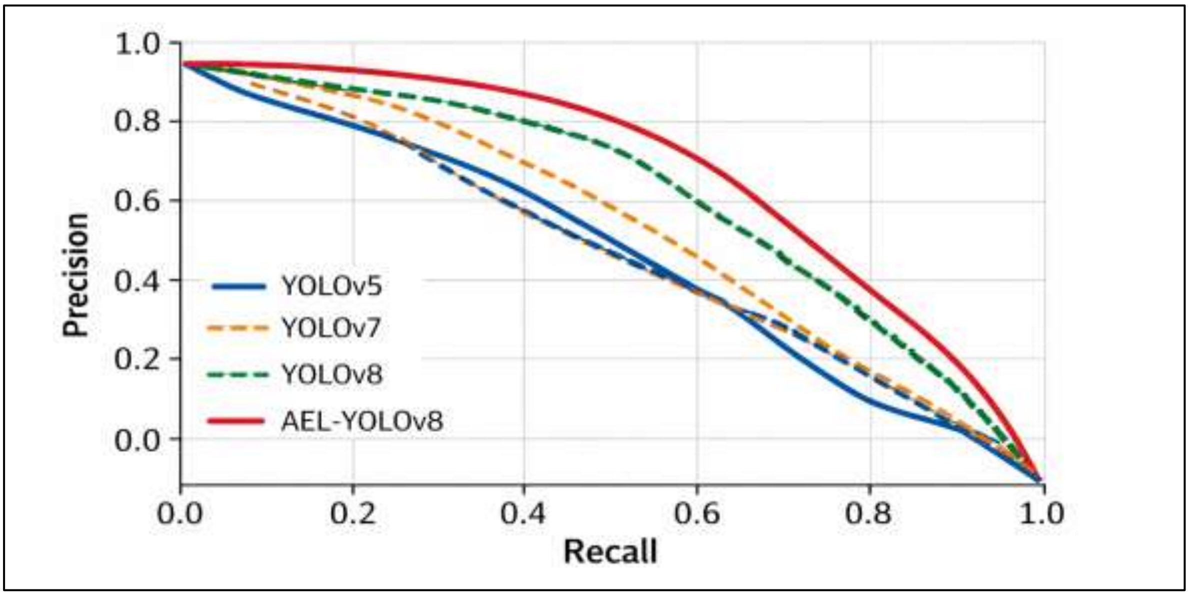

Figure 6. Precision-Recall Curves for All Evaluated Models (YOLOv5, YOLOv7, YOLOv8, AEL-YOLOv8)

AEL-YOLOv8 improves weed detection by incorporating three main architectural advances that address the limitations of the standard YOLOv8. Coordinate Attention introduces directional spatial positional encoding that the original model's channel-only focus lacks, helping the model differentiate weeds from crops based on spatial arrangement as well as local texture. This is especially important for sesame crops, where overlapping leaf shapes at similar growth stages make texture-based methods alone insufficient. BiFPN yields the most significant improvement (+1.1% mAP@0.5), highlighting that multi-scale feature fusion is a key architectural bottleneck, given the dataset's weed size range from roughly 15×15 to 200×200 pixels at 640×640 resolution. CIoU loss provides a smaller but valuable +0.6% mAP@0.5 boost by enhancing bounding-box regression accuracy, which directly benefits spray-nozzle targeting and reduces herbicide waste. Structured channel pruning removes redundant parameters with minimal accuracy loss, suggesting many channels in baseline YOLOv8 are unnecessary for this binary crop-weed classification task.

AEL-YOLOv8 occupies a unique position in the weed detection landscape by simultaneously achieving high accuracy, compact size, and verified edge deployment requirements. YOLO11-edge [26] achieves comparable deployment speed (~30 FPS) but has a 13.3 percentage-point accuracy gap relative to AEL-YOLOv8's 95.2% mAP, while YOLOv8-ECFS [23] achieves a comparable 95.0% mAP but lacks edge hardware validation and model compression analysis. The greatest performance advantage of AEL-YOLOv8 occurs under low-light and heavy-shadow conditions (+2.1% mAP), suggesting that Coordinate Attention provides robustness to global illumination degradation, supporting deployment across the full agricultural day-night cycle without illumination-specific calibration. The real-time applicability claim is substantiated by comprehensive edge deployment measurements: 31 FPS end-to-end throughput on NVIDIA Jetson Orin, including preprocessing, TensorRT FP16 inference, and postprocessing, with approximately 8.4 GB memory utilisation and approximately 15 W power consumption — measurements conducted using a standard and practical TensorRT FP16 deployment format for production agricultural systems. AEL-YOLOv8's 32.1 ms end-to-end latency is directly compatible with contemporary agricultural drone platforms operating at 3–10 m above the crop canopy at 2–5 m/s, which require latency below 50 ms for continuous real-time weed mapping. Its compact size (18.2M parameters, 36.4 MB) is compatible with agricultural robotics platforms, including NVIDIA Jetson, Raspberry Pi 5 with NPU accelerators, and custom FPGA-based vision systems. With INT8 quantisation, deployment on lower-tier edge devices such as Jetson Nano and Raspberry Pi Compute Module 4 becomes feasible, potentially expanding economic accessibility for smallholder farmers.

AEL-Yolov8's main strengths include consistently high detection performance across all metrics, with statistically significant improvements. It also offers verified real-time edge deployment with detailed latency, memory, and power analysis. The model demonstrates environmental robustness under three lighting conditions and achieves a 58.16% reduction in parameters, supporting cost-effective use on agricultural IoT devices. However, limitations include training and testing on only one sesame crop species, which restricts understanding of its ability to generalise to other crops and weeds. The dataset consists of static images from a single geographic region (India), and its performance in different climates, soils, and crop conditions remains unverified. It has not been deployed in live field conditions using drones or ground robots, so practical deployment is based on technical feasibility rather than confirmed operation. Also, as a supervised learning model, it cannot guarantee performance on weed species not included in the training data. Additional operational challenges for real-world use include adjusting confidence thresholds for varying weed density, periodic model fine-tuning for different crop growth stages, handling camera soiling from dust or water, and ensuring precise GPS-timestamp synchronisation for accurate weed localisation.

5. Conclusion

This paper introduces AEL-YOLOv8, a lightweight, attention-enhanced version of YOLOv8 designed for real-time weed detection in precision agriculture. The framework combines three key innovations: Coordinate Attention for directional spatial features, BiFPN for adaptive multi-scale feature fusion, and CIoU loss for precise bounding-box regression, all incorporated into a structured, channel-pruned lightweight design. Extensive tests on a publicly available sesame crop-weed dataset show that AEL-YOLOv8 attains 97.5% accuracy, 97.1% precision, 96.8% recall, 96.9% F1-score, and 95.2% mAP@0.5, with statistically significant improvements over the baseline YOLOv8l ($t = 4.21$, $p < 0.001$, Bonferroni-corrected over five runs). It reduces parameters by 58.16%, from 43.5M to 18.2M, and enables real-time inference at 31 FPS on NVIDIA Jetson Orin, with a total latency of 32.1 ms, memory use around 8.4 GB, and power consumption near 15 W. This satisfies the key criteria of high detection accuracy (>95% mAP@0.5), a small model size (<20M parameters), and real-time edge deployment (>30 FPS). Ablation tests reveal BiFPN contributes the largest gain (+1.1% mAP), followed by Coordinate Attention (+0.8%) and CIoU loss (+0.6%). The modules work together complementarily, and robustness tests under various lighting conditions confirm consistent performance improvements in bright sunlight, overcast, and low-light environments.

The principal contributions of this study are a novel AEL-YOLOv8 architecture that integrates CA, BiFPN, CIoU loss, and structured pruning into a unified, lightweight YOLOv8 system for agricultural weed detection; the first reported weed detection model to simultaneously achieve mAP@0.5 > 95%, parameter count < 20M, and verified real-time inference > 30 FPS on agricultural edge hardware; statistically rigorous performance validation across five evaluation metrics, supported by comprehensive ablation and environmental robustness analyses; and transparent, reproducible edge-deployment characterisation, including end-to-end latency breakdown, memory usage, and power consumption on NVIDIA Jetson Orin.

Several promising avenues for future research emerge from the limitations of this study: testing AEL-YOLOv8 on multi-crop weed detection datasets like cotton, soybean, rice, and maize to evaluate cross-crop transferability and develop domain-adaptive fine-tuning strategies; expanding the binary crop-weed model to multi-class, species-level identification for targeted herbicide application and dosage optimisation; conducting field trials with operational agricultural drones and ground robots using live video streams to validate real-world performance; exploring advanced compression techniques such as knowledge distillation, neural architecture search, INT8 quantisation sensitivity, and FPGA-optimised architectures for ultra-low-power deployment; integrating lightweight temporal attention across video frames to leverage motion cues and fix occlusion issues; and collecting and annotating images of sesame and multi-crop fields from diverse geographic regions to evaluate and enhance performance under different climatic, soil, and growth-stage conditions.

Acknowledgement

The authors thank GIET University, Gunupur, Odisha, India, for providing computational resources and research facilities, and Ravirajsinh Dabhi for sharing the crop and weed detection dataset on Kaggle. This research did not receive any specific funding from public, commercial, or not-for-profit sources. The authors declare they have no conflicts of interest.

References

1. Hu, R., Su, W. H., Li, J. L., & Peng, Y., 2024. Real-time lettuce-weed localization and weed severity classification based on lightweight YOLO convolutional neural networks for intelligent intra-row weed control. Computers and Electronics in Agriculture, 226, 109404.
2. Dabhi, R., & Makwana, D., 2020. Crop and weed detection data with bounding boxes [Dataset]. Kaggle.
3. Allmendinger, A., Saltik, A. O., Peteinatos, G. G., Stein, A., & Gerhards, R., 2025. Assessing the capability of YOLO- and transformer-based object detectors for real-time weed detection. Precision Agriculture.
4. Hashemi-Beni, L., Gebrehiwot, A., Karimoddini, A., Shahbazi, A., & Dorbu, F., 2022. Deep convolutional neural networks for weeds and crops discrimination from UAS imagery. Frontiers in Remote Sensing, 3, 755939.
5. Kim, Y.H., Park, K.R., 2022. MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. Computers and Electronics in Agriculture, 199, 107146.
6. Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90.
7. Lu, Z. et al., 2025. Star-YOLO: A lightweight and efficient model for weed detection. Computers and Electronics in Agriculture.
8. Nasiri, A., Omid, M., Taheri-Garavand, A., Jafari, A., 2022. Deep learning-based precision agriculture through weed recognition in sugar beet fields. Sustainable Computing: Informatics and Systems, 35, 100759.
9. Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proceedings of IEEE CVPR, 779–788.
10. Tan, M., Pang, R., Le, Q.V., 2020. EfficientDet: Scalable and efficient object detection. Proceedings of IEEE CVPR, 10781–10790.
11. Hou, Q., Zhou, D., Feng, J., 2021. Coordinate attention for efficient mobile network design. Proceedings of IEEE CVPR, 13713–13722.
12. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D., 2020. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of AAAI, 34(7), 12993–13000.
13. Wang, J., Qi, Z., Wang, Y., Liu, Y. et al., 2025. A lightweight weed detection model for cotton fields based on an improved YOLOv8n. Scientific Reports, 15, 457.
14. Yang, G., Wang, J., Nie, Z., Yang, H., Yu, S., 2023. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. Agronomy, 13(7), 1824.
15. Yang, Z., Khan, Z., Shen, Y., Liu, H., 2025. GTDR-YOLOv12: Optimizing YOLO for efficient and accurate weed detection in agriculture. Agronomy, 15(8), 1824.
16. Shen, P., Li, T., Xue, H., Zhao, Z., Cao, H., Mei, K., Wang, G., 2026. YOLOv8-MLD: A lightweight method for weed detection in cornfields. IEEE Access.
17. Feng, Z., Sangsongfa, A., Amdee, N., 2026. Application of YOLOv8 improved model incorporating attentional mechanism in weed identification. ASEAN Journal of Science and Technology Report, 29(4), e260523.
18. Rai, N., Mahecha, M.V., Christensen, A., Quanbeck, J., Zhang, Y., Howatt, K., Sun, X., 2023. Multi-format open-source weed image dataset for real-time weed identification in precision agriculture. Data in Brief,

51, 109691.
19. Jin, T., Liang, K., Lu, M., Zhao, Y., Xu, Y., 2025. Weedssort: A weed tracking-by-detection framework for laser weeding applications within precision agriculture. Smart Agricultural Technology, 11, 100883.
20. Gomez, A.S., Darbyshire, M., Gao, J., Sklar, E.I., Parsons, S., 2021. Towards practical object detection for weed spraying in precision agriculture. arXiv:2109.11048.
21. Ali, N., Sailunaz, K., AbuShaar, S., Ozdemir, M.K., Rokne, J., Alhajj, R., 2026. Effective detection of weeds in sesame crop. Network Modeling Analysis in Health Informatics and Bioinformatics, 15(1), 55.
22. Niu, W., Lei, X., Li, H., Wu, H., Hu, F., Wen, X. et al., 2024. YOLOv8-ECFS: A lightweight model for weed species detection in soybean fields. Crop Protection, 184, 106847.
23. Chen, W., Wang, J., Jia, Z., Quan, S., Guo, A., Ge, B., Zhou, G., 2025. YOLOv8-LSI: Enhanced weed detection in agricultural fields using large convolutional kernels and dimensionality-expanded channel attention. International Conference on Intelligent Computing, 445–456.
24. Zheng, L., Long, L., Zhu, C., Jia, M., Chen, P., Tie, J., 2024. A lightweight cotton field weed detection model enhanced with EfficientNet and attention mechanisms. Agronomy, 14(11), 2649.
25. Upadhyay, A., Sunil, G.C., Das, S., Mettler, J., Howatt, K., Sun, X., 2025. Multiclass weed and crop detection using optimized YOLO models on edge devices. Journal of Agriculture and Food Research, 22, 102144.
26. Paul, A., Machavaram, R., Kumar, D., Nagar, H., 2024. Smart solutions for capsicum harvesting: Unleashing the power of YOLO for detection, segmentation, growth stage classification, counting, and real-time mobile identification. Computers and Electronics in Agriculture, 219, 108832.
27. Sapkota, B.B. et al., 2022. Use of synthetic images for training a deep learning model for weed detection and biomass estimation in cotton. Scientific Reports, 12(1), 19580.
28. Sapkota, R., Stenger, J., Ostlie, M., Flores, P., 2023. Towards reducing chemical usage for weed control in agriculture using UAS imagery analysis and computer vision techniques. Scientific Reports, 13(1), 6548.
29. Sun, T. et al., 2024. Weed recognition at soybean seedling stage based on YOLOV8nGP+NExG algorithm. Agronomy, 14(4), 657.
30. Lin, F., Zhang, D., Huang, Y., Wang, X., Chen, X., 2017. Detection of corn and weed species by the combination of spectral, shape, and textural features. Sustainability, 9(8), 1335.
31. Terven, J., Cordova-Esparza, D.M., Ramirez-Pedraza, A., Chavez-Urbiola, E.A., 2023. Loss functions and metrics in deep learning. arXiv:2307.02694.
32. Jocher, G., Chaurasia, A., Qiu, J., 2023. YOLO by Ultralytics (Version 8.0). GitHub.
33. Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning — method overview and review of use for fruit detection and yield estimation. Computers and Electronics in Agriculture, 162, 219–234.
34. Thenmozhi, K., Reddy, U.S., 2019. Crop pest classification based on deep convolutional neural network and transfer learning. Computers and Electronics in Agriculture, 164, 104906.
35. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. Proceedings of IEEE CVPR, 2117–2125.
36. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J., 2018. Path aggregation network for instance segmentation. Proceedings of IEEE CVPR, 8759–8768.
37. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M., 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.
38. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of IEEE CVPR, 7464–7475.
39. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of IEEE CVPR, 770–778.
40. Vaswani, A., Shazeer, N., Parmar, N. et al., 2017. Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.
41. Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. CBAM: Convolutional block attention module. Proceedings of ECCV, 3–19.
42. Paszke, A., Gross, S., Massa, F. et al., 2019. PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32, 8024–8035.
43. Howard, A., Sandler, M., Chu, G. et al., 2019. Searching for MobileNetV3. Proceedings of IEEE ICCV, 1314–1324.
44. Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. Proceedings of IEEE CVPR, 1251–1258.