ORIGINAL RESEARCH Open access

Tracking-by-Detection for Precision Laser Weeding Using Spatial Attention and Adaptive Filtering

Omar Zayed[1*], Leila Ibrahim[1], Ahmed Mansour[2]

Abstract

Within precision agriculture, combining artificial intelligence with high-power laser systems for weed management yields substantial improvements in both efficiency and accuracy. Nonetheless, the detection and tracking of weed targets remain considerable challenges. In object detection, the diverse sizes and traits of weeds can lead to smaller targets being missed. In the domain of tracking prediction, the resemblance among weed forms may degrade the precision of pose estimation. At the same time, the erratic camera movements inherent to laser weeding platforms further heighten the chance of tracking loss. To overcome these hurdles, this work proposes a spatial attention mechanism to improve weed detection precision. It applies refined multi-feature layer extraction and optimal feature-matching methods to generate motion estimates. An adaptive extended Kalman filtering scheme is then incorporated, yielding a weed-tracking method that fuses temporal and spatial cues, enabling fast and accurate weed detection and tracking suited to laser weeding conditions. The detection accuracy of the refined algorithm was examined on both open-source and custom-collected detection datasets, reaching a mean Average Precision (mAP) of 97.29% and 85.83%, respectively. Tracking capability was further assessed using the LettuceMOT and self-collected WeedsMOT datasets, yielding gains in Higher-Order Tracking Accuracy (HOTA) of 12.01% and 8.75% over the ByteTrack and DeepOCSORT algorithms, respectively. The experimental results verify the effectiveness of the introduced weed detection and tracking pipeline, serving as a useful benchmark for advancing laser weeding systems in precision agriculture.

Keywords Object detection, Laser weeding, Weed detection and tracking, Motion estimation, Adaptive extended Kalman filtering

*Correspondence:
Omar Zayed
omar.zayed@cu.edu.eg
[1] Department of AI for Wheat Rust and Pest Prediction, Faculty of Agriculture, Cairo University, Cairo, Egypt
[2] Department of Precision Nutrient Management and Soil Sensors, Faculty of Agriculture, Alexandria University, Alexandria, Egypt

Introduction

With the ongoing rapid expansion of the global population, the strain on food production and security intensifies, placing the improvement of current agricultural productivity and long-term viability at the forefront of agricultural development efforts [1]. Weeds battle with crops for vital nutrients and sunlight during seedling establishment, representing a major factor limiting crop output. Conventional weed suppression techniques rely on manual removal and chemical herbicides, which are both costly and ineffective and pose environmental hazards, thereby failing to meet modern agriculture's goals of efficiency and ecological soundness. Standard mechanical weed control devices are limited to removing weeds between rows, often leading to elevated rates of crop damage and poor weed management within rows [2]. Merging artificial intelligence with high-intensity laser technology in thermal weed control systems enables precise identification

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

and tracking of all weeds across video streams, facilitating swift, effective, and residue-free weed suppression that meets the requirements of precision agriculture [3].

For laser-driven intelligent weed detection and tracking, both precision and speed are essential. Presently, deep learning-oriented object detection approaches have seen widespread adoption in numerous fields, such as identifying leaf ailments in rice plants [4, 5], recognizing leaves of tea seedlings [6], spotting pests in field crops [7], tracking growth in orchards [8], and detecting weeds in cotton and soybean crops [9, 10]. Deep learning detectors generally fall into two primary groups. Compared to two-stage detectors, one-stage detectors bypass the object proposal generation step, enabling faster inference and better suitability for resource-constrained contexts such as laser weeding platforms. Among one-stage detectors, the You Only Look Once (YOLO) family, originally proposed by Redmon [11], is the most broadly utilized. It exhibits superior results relative to prominent methods such as Single Shot multibox Detector (SSD), Faster R-Convolutional Neural Network (Faster R-CNN), EfficientDet, and CenterNet, standing out in precision, speed, and model size [12]. The YOLO lineage has advanced to include significant iterations such as YOLOv3 [13], YOLOv4 [14], YOLOv5 [15], and YOLOv8 [16], with the latest being YOLOv11 [17]. Because of their well-balanced detection capability and computational performance, YOLO-series detectors have been extensively deployed in agricultural surveillance. For instance, Wang *et al.* [18] leveraged a modified YOLOv7 algorithm to categorize and identify weeds and crops in wheat plantations. In contrast, An *et al.* [19] applied an upgraded YOLOX framework to monitor strawberry developmental stages. Furthermore, Liu *et al.* [20] deployed the YOLOv4-Weeds approach for spotting weeds among corn seedlings.

The YOLO algorithm family has achieved notable progress in agricultural sensing and now constitutes a leading methodology for weed identification [9]. Still, difficulties persist in pinpointing small-dimension weed targets, resulting in detection gaps. Certain investigators have sought to boost detection fidelity by embedding attention mechanisms. For example, Chen *et al.* [21] adapted the YOLOv4 architecture with an attention module to detect weeds in sesame plantations, resulting in a marked improvement in overall weed detection rates. Likewise, Wang *et al.* [22] paired the YOLOv5 detection framework with attention mechanisms to recognize invasive weed species, increasing the average precision (AP) from 0.9017 to 0.9272. While attention methods employing dimensionality reduction and downsampling have modestly improved detection performance, they have also increased the model's network complexity, diminishing overall throughput. By leveraging the static spatial arrangements of weeds and crops typical of farming environments, and by adopting a lightweight multi-attention mechanism to produce output channel weightings, this approach can support efficient and precise recognition of target zones.

Once weed detection outputs are available, forming dependable cross-frame weed trajectories becomes imperative to avoid duplicating targeting the same weed. Tracking reliability hinges on the outcomes of inter-frame motion assessment and system motion forecasting. Within motion estimation pipelines, characteristic point data is pulled from successive image frames, and matched point correspondences are filtered to compute the pose transformation linking them. Classic visual feature extraction techniques include oriented FAST and rotated BRIEF (ORB) [23], scale-invariant feature transform (SIFT) [24], and speeded-up robust features (SURF) [25], which align descriptors with keypoint coordinates. In agricultural settings, weed visual characteristics are often highly uniform, yielding near-identical geometric attributes and descriptor values. This uniformity can induce mismatches during alignment, consequently lowering the fidelity of motion estimation.

Deep learning-driven feature detection strategies improve the precision and speed of feature extraction and pairing by incorporating attention-based network architectures [26-29]. The SuperPoint architecture devised by DeTone *et al.* [30] unifies keypoint identification and description within a single model, enabling concurrent learning that accounts for inter-component relationships. This technique uses a self-supervised methodology to produce pixel-resolution feature points and their descriptors. Reinforced SP [31] applies reinforcement learning to select keypoints and match descriptors, optimizing parameters in an end-to-end fashion. SuperGlue [32] enables data interchange across descriptor collections via attention-based graph neural networks and optimal transport formulations, yielding highly accurate correspondences. The MakeGNN [33] model enhances the flow of relevant information by incorporating bilateral perception sampling alongside context-sensitive keypoint pooling. By combining deep-learning-based feature detection and matching

© 2023 The Author(s). | Published by Agro Intelligence Research Press

paradigms, more faithful feature representations and associations can be obtained when managing analogous weed appearances, thereby improving the overall precision of motion estimation.

Building upon inter-frame weed motion estimates, and combining them with filtering techniques for system motion prediction, cross-frame weed association and tracking becomes feasible. Classic tracking pipelines commonly rely on Kalman Filters (KF) for movement forecasting, with the simple online and real-time tracking (SORT) [34] framework connecting bounding boxes across frames to realize multi-object tracking (MOT). Yet, associations grounded solely on bounding boxes are vulnerable to false pairings. To address the drawbacks of SORT, observation-centric SORT [35] was introduced, and the concept was further extended into Deep Observation-Centric Simple Online and Real-time Tracking (DeepOCSORT) [36]. In agricultural applications, the LettuceTrack approach advanced by Hu *et al.* [37] integrates the YOLOv5 detector with Kalman filtering to spot and track lettuce during pesticide spraying, reaching tracking accuracies near 80%. Meanwhile, the AgriSORT pipeline presented by Saraceni *et al.* [38] builds upon SORT, blending a refined Kalman filter for grape cluster tracking and enabling on-the-fly robotic processing. Under laser-weeding operational conditions, haphazard camera movements can cause tracking dropouts when linked to existing predictive tracking frameworks.

Addressing the difficulties inherent to laser weeding technology, this study proposes a meticulous weed-detection and tracking paradigm. By leveraging single-frame detection results and merging motion correspondences from feature matching, we deploy an adaptive extended Kalman Filter (EKF) to achieve reliable weed tracking under laser weeding conditions, particularly for minute and similarly shaped targets. To summarize, this work delivers the following contributions:

(1) The assembly of the WeedMOT dataset tailored for weed tracking, along with the substantiation of its utility on standard algorithms, offers data-driven validation for other algorithmic investigations. The code and dataset referenced in this paper are accessible at: https://github.com/PonciJJ/WeedsSORT.

(2) For weed detection, the YOLOv11 detection engine is paired with the lightweight attention mechanism network Triplet Attention [39] to apply differentiated weightings to salient regions. This method fuses motion cues and template-based matching for data linkage, thereby elevating weed detection precision.

(3) Concerning motion pose estimation, the SuperPoint framework is enhanced via a multidimensional feature extraction decoder, and keypoint correspondences are resolved through the SuperGlue matching algorithm, yielding faithful motion estimation outputs.

(4) For weed tracking, the extended Kalman filter methodology is harnessed to bind temporal sequences with weed geometry, embedding a noise-adaptive weighting coefficient. The infusion of motion data enhances weed-tracking accuracy.

(5) The efficacy of the proposed algorithm is corroborated on both publicly accessible and custom datasets, displaying superior detection and tracking results relative to contemporary mainstream approaches, thereby attesting to the soundness of the suggested methodology.

The assimilation of artificial intelligence underpins efficient weed management, subsequently paving the way for the application of the Internet of Things and robotic solutions. For instance, this concept can be deployed for hydroponic plant oversight through IoT [40, 41] and for refining robotic path planning and obstacle evasion via heuristic algorithms [42]. The overarching algorithm flow for weed detection and tracking is portrayed in **Figure 1**. Within this flow diagram, t-1 and t denote the preceding and succeeding time points, F indicates the image frame data, D refers to the object detection output, H signifies the derived homography matrix, and f designates the extracted feature point data. The projection of Dt−1 onto Ft is obtained by fusing the detection output Dt−1 with Ht−1 from time t-1, and a buffer zone can enlarge this

© 2023 The Author(s). | Published by Agro Intelligence Research Press

projection to define a candidate detection region. Performing weed detection constrained to this candidate region can boost both detection fidelity and processing speed.

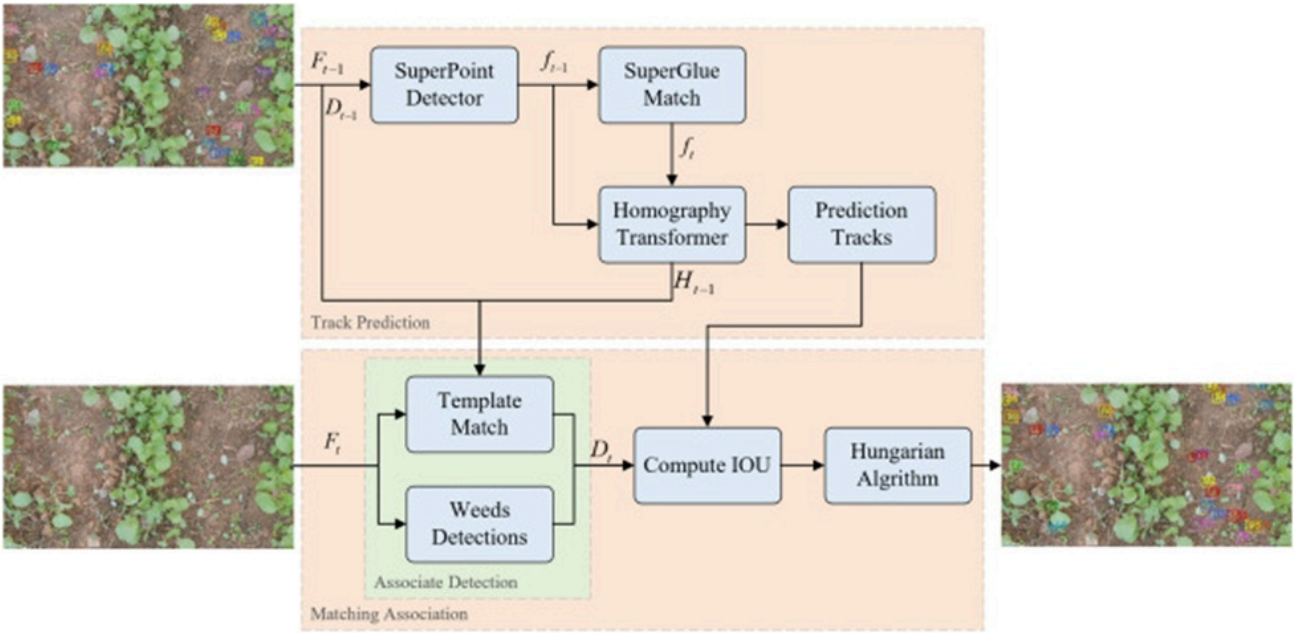


Figure 1. Flowchart of the overall process for weed detection and tracking algorithms.

Materials and Methods

Hardware platform and data acquisition

Laser weeding robot platform

The robotic laser weeding base used in this investigation uses stepper motor propulsion. It integrates a high-output CO2 laser, a depth-sensing camera, edge artificial intelligence processors, an electrical supply unit, and additional hardware subsystems. The robot's framework and its tangible embodiment are shown in **Figure 2**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

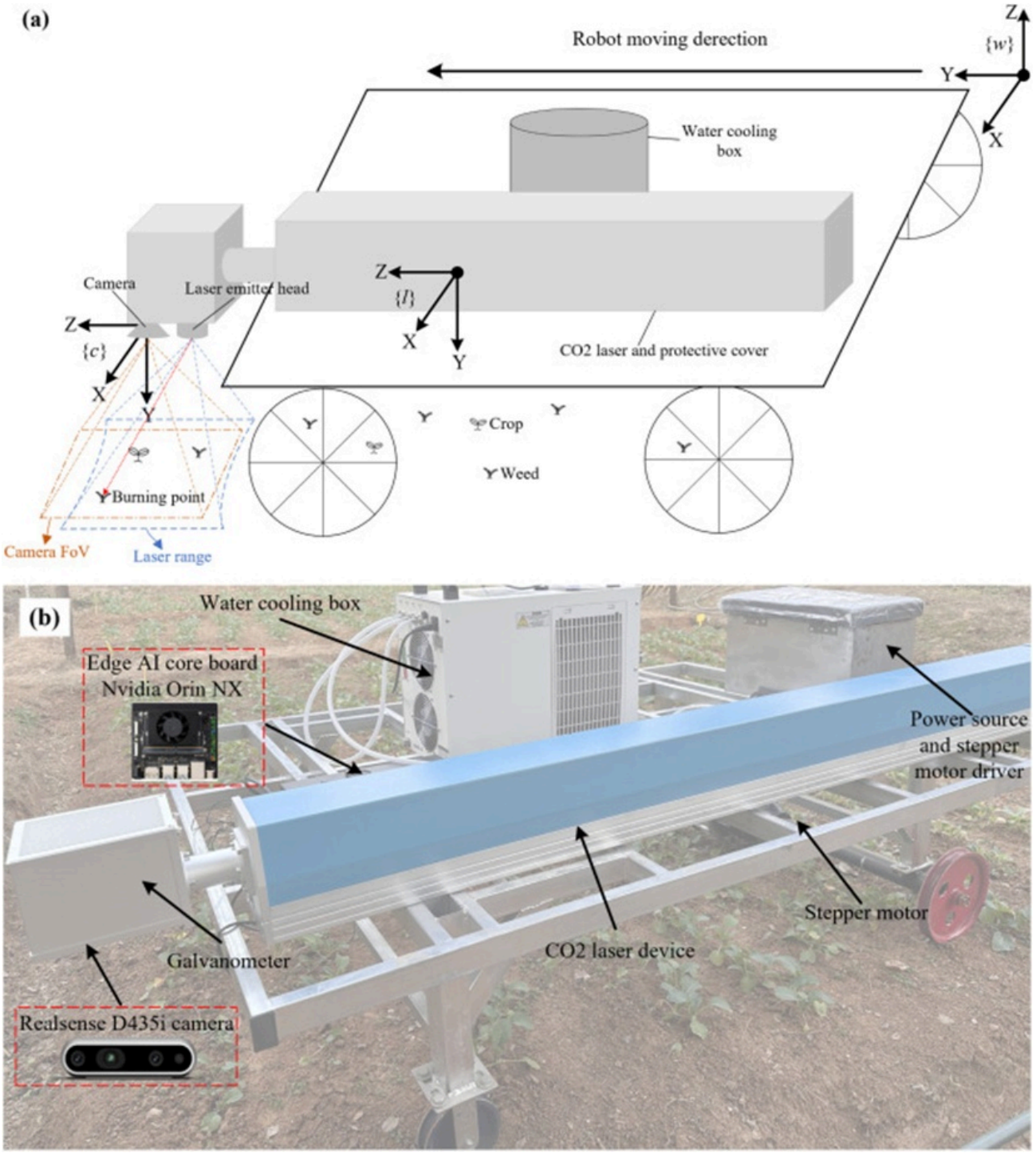


Figure 2. (a) The three-dimensional schematic of the mobile platform encompasses three reference systems: the imaging, the beam, and the global coordinate frames, as well as the perceivable zone of the camera and the projection scope of the laser. (b) This section illustrates the physical configuration of the mobile platform operating in a real cultivated field.

Mounted on the weeding robot are an Nvidia Jetson Orin NX 16GB processing board and an Intel RealSense D435i depth-sensing camera. Image capture is performed in a downward orientation, after which the processing unit computes the geospatial positions of weeds and forwards this information to the galvanometer controller to steer the laser output.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

The optical sensor and beam projector share a coaxial alignment; the CO2 laser device supports a pulse repetition rate reaching 100Hz and a peak emission strength of 180W.

The LettuceMOT dataset

The LettuceMOT [43] collection stems from the VegeBot platform, configured with an RTK-GPS module, a USB RGB imaging unit, and an Intel RealSense D435i depth camera. This dataset was assembled through uninterrupted recording sessions that documented lettuce specimens over two developmental periods. Every recording provides annotated bounding rectangles for each lettuce entity, accompanied by persistent labels spanning adjacent frames. Three distinct motion profiles are represented in the data: “Straight” for purely linear traversal, “B&F” for forward and backward displacement patterns, and “O&I” for situations where the robot blends straight driving with rotational maneuvers. During straight segments, the robot’s translational rate ranges from 0.3 m/s to 0.4 m/s, decreasing to 0.1 m/s to 0.2 m/s through curved paths. Within the scope of this article, B&F recordings drawn from an open-source corpus were used to validate performance in recognizing and following the re-emergence of weeds and crop plants as the platform moved forward and then reversed. The span between consecutive lettuce plants in a row is approximately 0.3-0.35 meters, with inter-row gaps of approximately 0.3 meters. Across the entire compilation, 5,466 picture frames and 707 separately identified object entries yield a cumulative count of 42,735 tagged items.

The WeedMOT dataset

Supplementing existing open-source resources, the present work has assembled the WeedMOT corpus to confirm the dependability of weed identification and trajectory maintenance. The collection splits into two groupings: radish sprout crops and adjacent weed growth; representative photographs and class designations appear in **Figure 3**. Weed gathering concentrated chiefly on the varieties shown in subfigures (b) and (c), both to receive a consolidated “weed” classification label.



Figure 3. (a) portrays a crop seedling from the radish family, (b) depicts a weed belonging to the chickweed group, and (c) exhibits a weed of the thornberry type.

Imagery was obtained via a sensor operating at 1920×1080 pixel resolution and a capture rate of 10 frames per second (FPS), ensuring that the typical overlap between temporally adjacent frames exceeds two-thirds of the total frame area. Collection procedures followed two movement modalities: advancing in a straight line and executing turns, with straight speed set between 0.3 m/s and 0.4 m/s and turning speed between 0.1 m/s and 0.2 m/s. In this independently collected dataset, the spacing between two sequential radish sprouts within a row is near 0.4-0.5 m, whereas the spacing between neighboring rows is approximately 0.35 m. The harvested material comprises 1902 picture frames and 384 individual object instances, totaling 35,773 manually specified objects. Upon completion of recording, crop and weed boundary annotations were applied through DarkLabel, formatting the outputs to match the MOT convention.

$$label = \{frame_{id}, id, x, y, w, h, 1, 1, 1\}, \quad (1)$$

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Within this structure, frameid carries the index value assigned to picture frames inside the footage, id supplies the identifier allocated to each weed instance, whereas x and y deliver the pixel coordinates of the upper-left vertex of the delineated region. Dimensions w and h encode the bounding shape's width and height, respectively; the closing triplet of entries stays unoccupied and is filled using the numeral 1.

Weeds detection

Reliable weed recognition underpins effective trajectory linking and distinguishes weeds from surrounding elements. Weed specimens tend toward diminutive proportions and patchy distribution across the scene, circumstances that may lead to detection omissions. The current study builds on the YOLOv11 recognition architecture, infusing it with the Triplet Attention module—an approach that capitalizes on spatial layout relationships—and coupling it with a similarity-based template alignment routine for measurement association, thereby improving overall weed identification hit rate.

Triplet attention

The Triplet Attention construct establishes cross-channel information dependencies via rotational permutations and residual mapping operations, entirely sidestepping the need for spatial downscaling of the source images. Its parameter-lean composition encapsulates positional intelligence for the sought objects while demanding negligible extra computation, thereby contributing spatial separation and relative importance cues that are indispensable for pinpointing weeds.

As depicted in **Figure 4**, the Triplet Attention component distributes emphasis values across the incoming frames slated for analysis, allocating non-uniform importance to different spatial coordinates within the target in each channel. After this weight-assignment stage in the network pipeline, the result consists of per-channel detection proposals with distinct emphasis factors, visually encoded so that varying color intensities represent zones within a channel map that merit greater scrutiny. Positions drifting toward the red end of the spectrum signal potentially richer feature content, matched with elevated weighting coefficients; in reverse fashion, zones shaded near blue equate to suppressed emphasis, translating to diminished focus throughout the recognition phase. The predictive logic for the initial two sub-paths of the architecture is specified below:

$$X_{H+} = R^{H+}\left(X_H \cdot \sigma\left(ConvBN\left(Z - Pool\left(X_H\right)\right)\right)\right), \quad (2)$$

$$X_{W+} = R^{W+}\left(X_W \cdot \sigma\left(ConvBN\left(Z - Pool\left(X_W\right)\right)\right)\right), \quad (3)$$

Where XH corresponds to the multidimensional array produced by rotating the source image X anticlockwise by 90 degrees around the H-axis, while XW corresponds to the array derived by rotating X clockwise by 90 degrees around the W-axis; the symbol σ denotes the Sigmoid nonlinearity, and ConvBN encapsulates the sequence of convolution followed by batch normalization. Z-pool denotes the compression step performed within the pooling stage. The expressions RH+ and RW+ map to actions that rotate the resultant feature representations clockwise by 90 degrees with respect to the H- and W-axes, respectively. Owing to these geometric manipulations, the representations flowing through the two sub-paths acquire attentional biases and spatial context descriptors. Inside the concluding sub-path, input features remain untransformed; weighting is instead imposed straightforwardly:

$$X_C = X \cdot \sigma\left(ConvBN\left(Z - Pool\left(X\right)\right)\right). \quad (4)$$

© 2023 The Author(s). | Published by Agro Intelligence Research Press

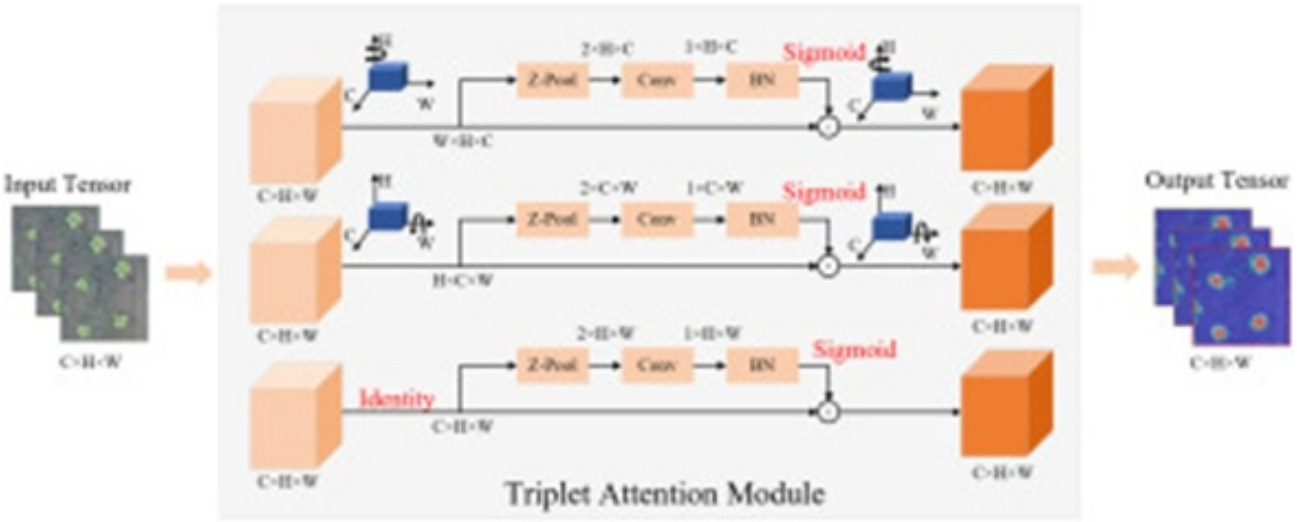


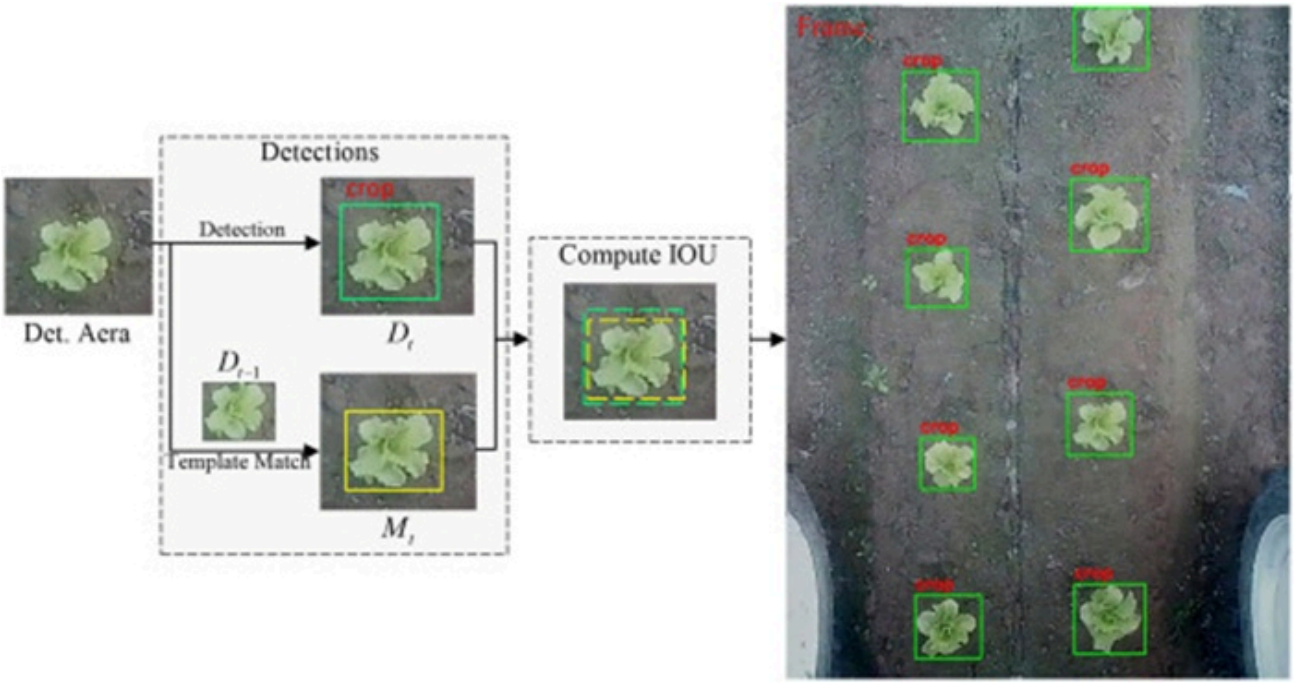
Figure 4. Triplet Attention module in weed detection.

Upon completion, three tensors originating from the separate sub-paths are obtained, and the attention stage proceeds to unify and average these three modules, generating the conclusive tensor:

$$y = \frac{1}{3}(X_{H+} + X_{W+} + X_C). \quad (5)$$

Template matching for data association

When operating in a laser-weeding context, it is reasonable to treat the robot as undergoing rigid-body displacements over brief time increments. Consequently, one can obtain the template alignment mapping Mt linking the earlier frame to the present one, which, when merged with the present frame's detection set Dt, facilitates measurement-to-target linkage and ultimately yields the object set assigned to the present frame, as portrayed in **Figure 5**.



© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 5. Data association based on template matching and target detection.

The template alignment technique delivers high reliability under conditions where localized appearance shifts remain minimal; however, it remains susceptible to false alignments when weeds are concealed or when pronounced illumination variations occur. Meanwhile, recognition paradigms grounded in convolutional neural architectures are better equipped to manage certain forms of occlusion. However, their tendency to overlook targets generally intensifies when the surveyed region grows larger. Through fusion of outputs from both recognition routes and dependence on the pre-existing motion estimate, objects spotted in the prior image frame may be warped forward, enabling calculation of the Intersection over Union (IoU) metric as described:

$$IoU = \frac{S_i}{S_M + S_D - S_i}, \quad (6)$$

Where SM denotes the region produced via template alignment, SD denotes the region encompassed by the detection proposal from the recognition algorithm, and Si denotes the overlapping segment linking the pair. A recognition hypothesis is deemed valid only when the IoU score exceeds the configured acceptance bound. This event thereafter authorizes refinement of the object trajectory record based on that recognition. Embracing a paradigm that merges attention-based weighting with template alignment ensures that the generated focus zones and data linkage can markedly elevate the recognition rate for weed targets of small physical extent.

Weed tracking and prediction

Motion pose estimation

When inferring movement from characteristic patterns observed across successive video frames, classical geometric descriptor extraction and correspondence schemes often yield erroneous pairings because the visible contours of distinct weed specimens are highly similar. To remedy this shortcoming, the present article locates salient image points using an enhanced deep-learning-based framework, SuperPoint, and then characterizes and sifts candidate correspondences using SuperGlue. The structural layout of this algorithmic chain is visualized in **Figure 6**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

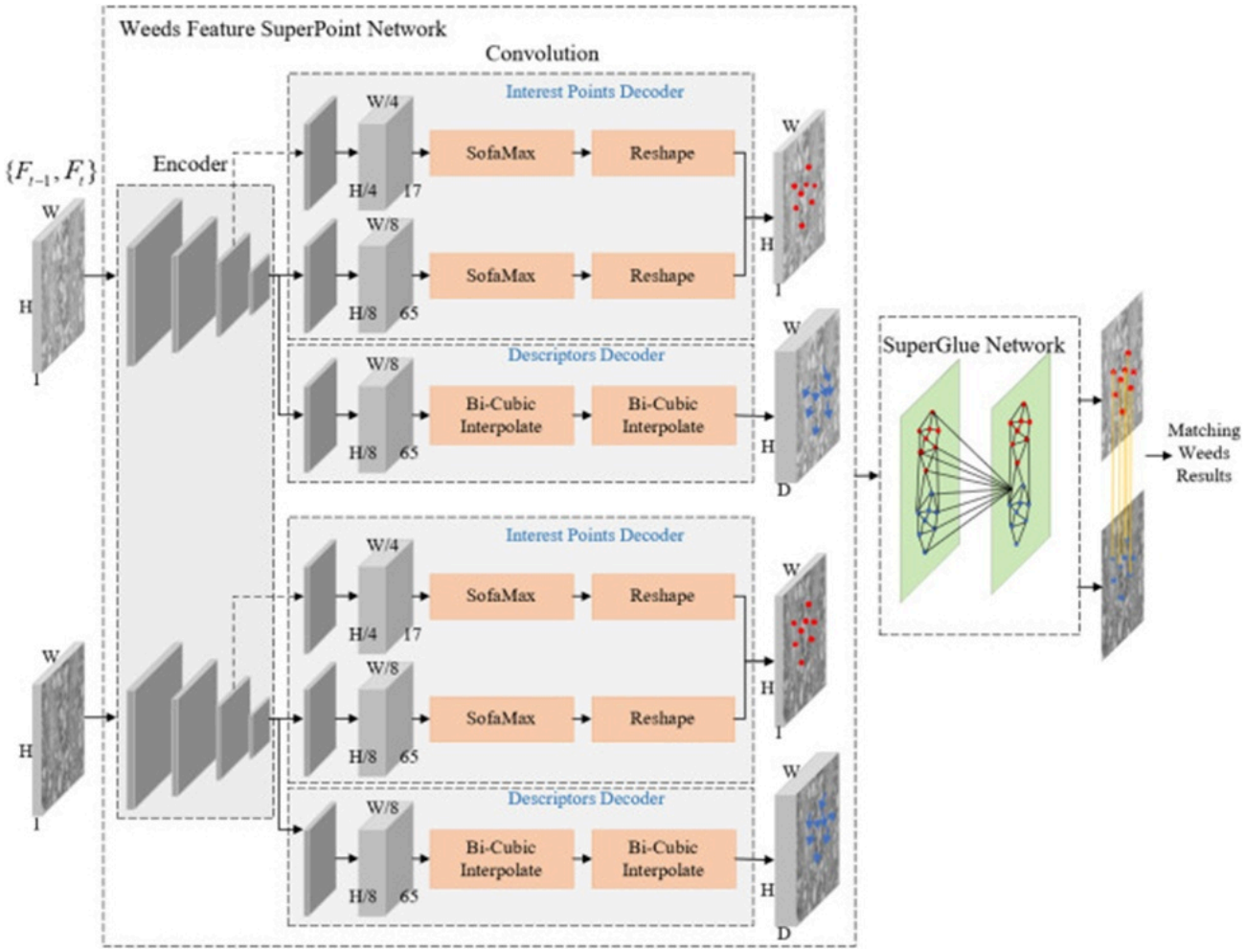


Figure 6. Feature extraction and matching optimized network structure diagram.

For the source batch of weed imagery, parameterized by horizontal span W (width) and vertical extent H (height), the encoding stage performs a spatial reduction. Thereafter, discriminative traits are distilled from the pictures following successive convolutional filtering steps. To incorporate the holistic scene setting, feature representations spanning multiple spatial resolutions are synthesized and examined, while guarding against pitfalls associated with global convergence. Throughout the keypoint localization phase, a quantifiable attrition of both interest-point coordinates and associated descriptors naturally occurs, formalized as:

$$L(F, F', D, D', M, M', S) = L_p(F, M) + L_p(F', M') + \lambda L_d(D, D', S). \quad (7)$$

Within this formulation, F and M denote the spatial maps encoding keypoints and the ground-truth supervisory signals, respectively. In contrast, F ′ and M′ denote the input views subjected to geometric warping. D and D′ pertain to the activation volumes earmarked for descriptor computation and the imagery undergoing homographic perturbation, respectively. Lp and Ld constitute the error terms for keypoint presence and descriptor fidelity, respectively, while λ serves as a fixed scaling factor that governs the trade-off between them.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

For the localized keypoint ensemble, pairing and pruning are performed using an attentional graph network architecture, along with a best-fit feature association stratum. The network pools keypoints from two separate image sources, weaving interconnections within each set of keypoints while also forming bonds with keypoints in the alternate view, collectively designated {self, cross}. The consolidated encoding uniting all keypoint positions and descriptor vectors affiliated with image A is stated as:

$$x_i^{A'} = x_i^A + MLP\left(\left[x_i^A \parallel m_{e-i}\right]\right), \quad (8)$$

Where captures the latent code of the i-th constituent within image A at the intermediate depth of layer L, $x_i^{A'}$ constitutes the refreshed latent code for image A, and [‖] marks the optimal linkage outcome post multi-layer perceptron processing. me−i embodies the alternating operational regime experienced by the network across odd-indexed and even-indexed layers. When L takes an odd value, e adopts ecross; conversely, when L takes an even value, e switches to eself. This oscillating nature of L replicates the iterative reasoning inherent in the algorithmic workflow. With respect to the descriptor attached to image A, the computed output is:

$$f_i^A = W \cdot^{(L)} x_i^A + b, \forall i \in A, \quad (9)$$

Where W and b encapsulate the weight matrix offsets and bias terms relative to the full computational system. Considering both image A and image B, the affinity measure tied to each putative correspondence pair emerges after the feature-matching stage, expressed compactly as:

$$S_{ij} = < f_i^A, f_i^B >, \forall (i,j) \in A \times B, \quad (10)$$

Where Sij quantifies the degree of likeness between the two views, putative matching landmark pairs can subsequently be filtered based on their assigned affinity ratings. Building on the screened point correspondences, an affine transformation model is fitted to resolve the geometric mapping, ultimately delivering the frame-to-frame transformation encoding the coherent motion trajectory of the weed target.

Tracking prediction based on adaptive EKF

When processing video footage, the reappearance of an identical weed specimen in frame after frame creates complications owing to the extreme likeness and tight clustering of such targets. These circumstances can easily lead to erroneous linkages when relying solely on shape-based matching. The Kalman Filter framework unites sequential cues with morphological descriptors to extrapolate motion paths; however, its underlying premise of strictly linear robot displacement constitutes a limitation. In open-field operating conditions where weeding assignments become complicated, imaging sensor motion often includes unpredictable components. The Extended Kalman Filter circumvents this shortcoming by harnessing nonlinear representations to model both the state transition dynamics and the measurement process, consequently yielding more reliable forecasts. Throughout the position correction cycle, it incorporates motion correspondence to build adaptive multipliers that tune the observation noise covariance matrix. The system state encapsulating each target takes the form:

$$x_k = \left[x_c(k), y_c(k), w(k), h(k), v_{x_c}(k), v_{y_c}(k), v_w(k), v_h(k)\right]^T. \quad (11)$$

© 2023 The Author(s). | Published by Agro Intelligence Research Press

In this notation, k denotes the discrete time index, (xc,yc) identifies the centroid coordinates of the enclosing box, and (w,h) captures its width and height. The velocity components along each axis—vxc, vyc, vw, and vh—quantify the rate of dimensional change. The observation vector follows the specification:

$$z_k = Hx_k + v_k, \quad (12)$$

Where H serves as the observation operator, and vk denotes the stochastic measurement disturbance inherent to the sensing apparatus.

Within this formulation, the nonlinear expressions responsible for trajectory projection and refinement read as follows:

$$x_k = f(x_{k-1}) + w_k, \quad (13)$$

$$z_k = h(x_k) + v_k, \quad (14)$$

Within which f(·) and h(·) designate the nonlinear transformation functions, and performing a first-order Taylor series decomposition with EKF-based linearization delivers:

$$f(x_{k-1}) \approx f(\widehat{x}_{k-1}) + F_k(x_{k-1} - \widehat{x}_{k-1}), \quad (15)$$

$$h(x_k) \approx h(\widehat{x}^{-}_{k}) + H_k(x_k - \widehat{x}^{-}_{k}), \quad (16)$$

Where Fk corresponds to the Jacobian derivative of f(·) at the point $\hat{x}_{k-1}$ and Hk to the Jacobian derivative of $h(\cdot)$ at $\hat{x}_k$, the differentiated functions are expressed as:

$$F_k = \frac{\partial f}{\partial x}\Big|_{x=\widehat{x}_{k-1}}, \quad (17)$$

$$H_k = \frac{\partial h}{\partial x}\Big|_{x=\widehat{x}^{-}_{k}}. \quad (18)$$

Exploiting the state quantity inferred at the preceding time step along with the linearized pseudo-truth obtained through EKF approximation, the a priori state estimate and error covariance advanced to the current time step are forecast as:

$$\widehat{x}^{-}_{k} = f(\widehat{x}_{k-1}), \quad (19)$$

$$\widehat{P}^{-}_{k} = F_k \widehat{P}_{k-1} F_k^T + Q, \quad (20)$$

$$F_k = \frac{\partial f}{\partial x}\Big|_{x=\widehat{x}_{k-1}}. \quad (21)$$

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Given the unavoidable presence of observation-related system perturbations, an adaptively tuned coefficient is introduced during the measurement update phase:

$$\tilde{R} = (1 - m_k)\, R + b, \quad (22)$$

In which R˜ represents the recalculated noise covariance term, mk captures the instantaneous measurement confidence level, which is adjustable on the fly in response to speed and illumination variations, and b designates a fixed offset reflecting the hardware setup's immutable noise floor. Synthesizing the foregoing observed and predicted quantities, the maximum a posteriori state correction resulting from the nonlinear refinement cycle is obtained as:

$$\tilde{y}_k = z_k - h\left(\widehat{x}_k^{-}\right), \quad (23)$$

$$S_k = H_k \widehat{P}_k^{-} H_k^T + \tilde{R}, \quad (24)$$

$$K_k = \widehat{P}_k^{-} H_k^T S_k^{-1}, \quad (25)$$

$$\widehat{x}_k = \widehat{x}_k^{-} + K_k \tilde{y}_k, \quad (26)$$

$$H_k = \frac{\partial h}{\partial x}\Big|_{x=\widehat{x}^{-}{}_k}, \quad (27)$$

$$\widehat{P}_k = (I - K_k H_k)\, \widehat{P}_k^{-}. \quad (28)$$

Here, Sk denotes the innovation discrepancy, Kk is the Kalman amplification gain, and the triplet y˜k, xˆk, and Pˆk collectively constitute the refreshed best-guess observation.

Results and Discussion

Experimental setup and evaluation index

Experimental setup

The object detection architecture was trained and performance assessed on a workstation equipped with an NVIDIA GeForce 4090 graphics card and an i7 14900KF central processor. The detection algorithms were benchmarked against both the internally amassed Weeds Dataset and the Cotton4—v3i public compilation. Altogether, four algorithmic pipelines received comparative scrutiny—YOLOv5, YOLOv8, YOLOv11, and the methodology put forward in this investigation. The class distribution and sample counts allocated to training and testing across the two datasets are enumerated in **Table 1**.

Table 1. Number of targets detected in public and homemade data sets

Dataset	Class	total	Test set	Validation set	Train set

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Cotton4.v3i	Crop	2088	214	457	1417
	Weed	3576	357	741	2478
Weeds Dataset	Crop	6394	757	1223	4414
	Weed	29379	2992	9060	17327

Regarding the data partitioning scheme, the publicly sourced collection consisted of 4,203 images, with 2,941 for training, 841 for validation, and 410 for final testing. The custom-built dataset encompasses 1,902 images, subdivided into 1,331 for training, 380 for validation, and 191 for testing. The initial training hyperparameters are itemized in **Table 2**.

Table 2. Training initialization parameters.

Input size	Batch size	Momentum	Learning rate	Epoch
416×416	120	0.937	0.001-0.00001	300

During model fitting, the input image resolution was maintained at 416×416 pixels, employing a batch count of 120 per iteration. The momentum parameter was fixed at 0.937, while the learning rate followed a decay trajectory from an initial 0.001 down to a terminal 0.00001 over the course of 300 complete epochs. Once the detection network's training and evaluation stages concluded, the weight configuration that yielded peak performance was transferred to the tracking pipeline for maintaining weed trajectories. The weed target tracking experiments were conducted on a hardware platform centered on the NVIDIA Jetson Orin NX 16GB processing module, using both the open-access LettuceMOT corpus and the proprietary WeedsMOT compilation as evaluation benchmarks.

Evaluation index

The assessment criteria used in this study are organized into two categories: those related to object detection and those relevant to weed target tracking. The detection model is measured through mAP, recall, precision, and the F1 metric. Their formal definitions are provided below:

$$\text{Precision} = \frac{TP}{TP+FP}, \quad (29)$$

$$\text{Recall} = \frac{TP}{TP+FN}, \quad (30)$$

$$\text{F1Score} = 2 * \frac{\text{Precision}*Recall}{\text{Precision}+Recall} = \frac{2TP}{2TP+FP+FN}, \quad (31)$$

Where true positive (TP) denotes instances where the model successfully labels Positive samples as positive, false positive (FP) captures cases where Negative samples are incorrectly branded as positive, and false negative (FN) counts situations where positive samples are erroneously classified as negative. mAP holistically balances Precision and Recall, using the IoU overlap metric, with the customary acceptance threshold set to 0.5.

Performance quantification for the target tracking model rests on Higher-Order Tracking Accuracy (HOTA), multi-target tracking accuracy (MOTA), DetA, Identity F1 Score (IDF1), mostly tracked (MT), and mostly lost (ML). HOTA merges

© 2023 The Author(s). | Published by Agro Intelligence Research Press

several dimensional criteria—including detection quality, association correctness, and spatial alignment—and derives an aggregated IoU-grounded score according to:

$$HOTA = \sqrt{DetA \cdot AssA}, \quad (32)$$

Within which DetA captures the detection sub-score, and AssA encapsulates the association and localization sub-score. All evaluations are carried out via the TrackEval toolkit, the standard software suite for HOTA-oriented metric computation.

Weed target detection experiment

Public data set testing

Through a side-by-side comparison of mean average precision, recall, accuracy, the quartet of precision-oriented indices including F1 score, and runtime throughput in FPS across multiple detection frameworks, the comparative outcomes on the Cotton4.v3i public benchmark is catalogued in **Table 3**. The processing latency of each algorithm, expressed in frames per second (FPS), was likewise logged.

Table 3. The performance of four model types was tested on a public data set.

Model	FPS	F1Score (%)	Recall (%)	Precision (%)	mAP@0.5 (%)
YOLOv5	**49.7**	96.02	95.93	96.12	94.49
YOLOv8	43.8	96.12	95.99	96.26	95.25
YOLOv11	46.1	96.64	96.45	96.84	95.96
YOLOv11-TA(Ours)	39.7	**98.31**	**97.71**	**98.92**	**97.29**

From **Table 3**, it is apparent that among the three off-the-shelf algorithms, YOLOv11 offers the best trade-off between detection quality and computational speed. Building on this foundation and fusing the Triplet Attention (TA) module yields results attributable to the proposed approach. When set against the unmodified baseline, a measurable uptick in precision is evident, whilst the incremental time penalty proves so slight as to be practically inconsequential.

Figure 7 displays the detection outputs generated by the refined algorithm on the public data resource. A batch of 410 images was evaluated against the public corpus, encompassing 214 crop and 357 weed annotated instances; a real-time statistical breakdown of the detection outcomes for the various algorithmic classes is presented in **Table 4**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

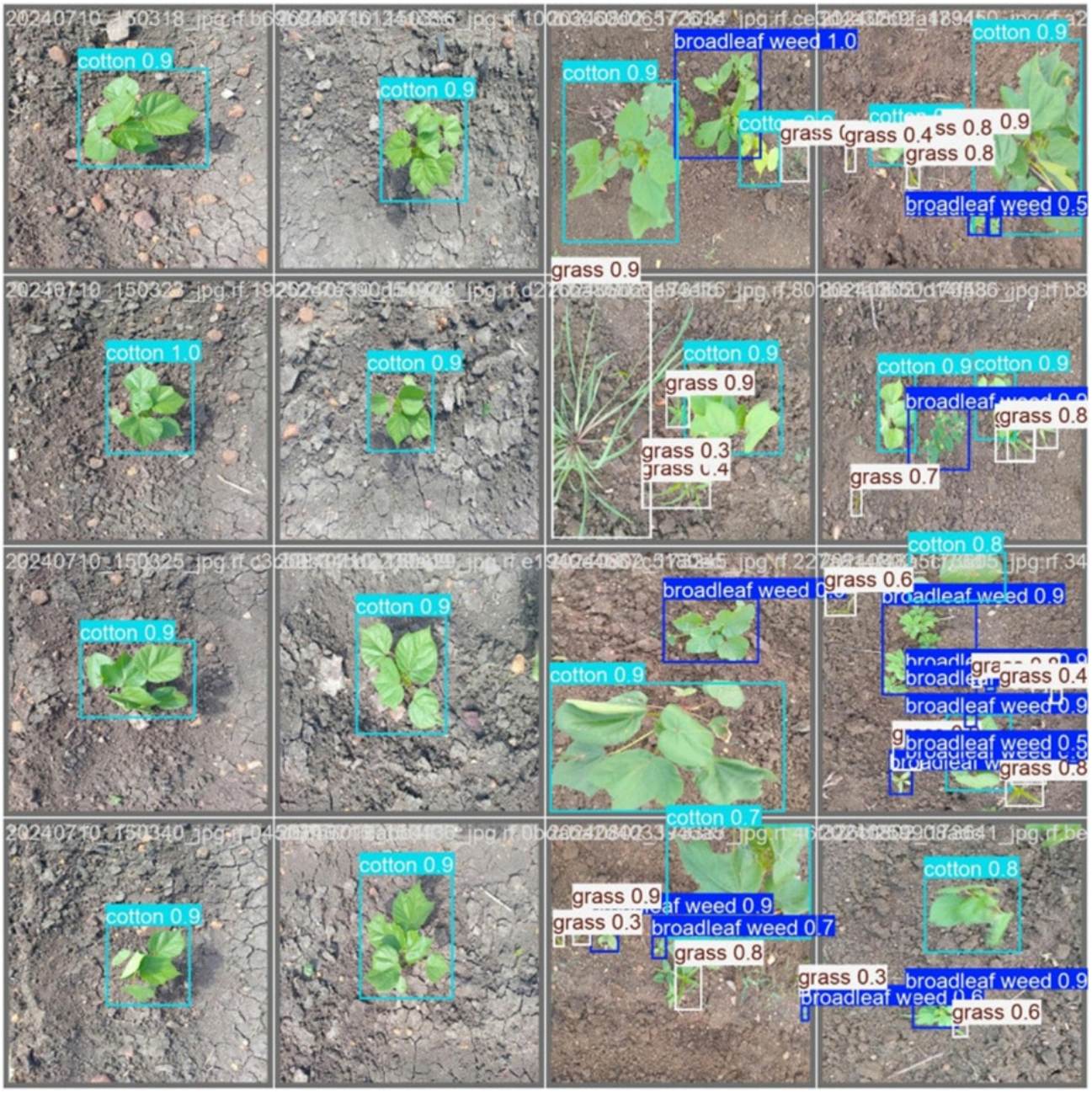


Figure 7. The algorithm in this paper detects the results on the public dataset.

Table 4. The performance of four detection models tested on self-collected data sets

Model	Grass			Broadleaf			Cotton		
	FN	FP	TP	FN	FP	TP	FN	FP	TP
YOLOv5	8	8	196	7	6	142	1	6	211
YOLOv8	8	8	196	6	5	143	2	6	211
YOLOv11	6	6	198	6	5	143	2	5	212
YOLOv11-TA(Ours)	5	2	200	3	2	145	1	3	213

© 2023 The Author(s). | Published by Agro Intelligence Research Press

In this dataset, weed specimens are labeled as “Broadleaf” and “Grass”, yet the detection tallies for both weed subtypes are combined during statistical aggregation. As **Table 4** shows, the algorithm enhanced with the attention mechanism achieves superior detection performance, yielding the highest counts of correctly identified weeds and the fewest occurrences of non-weed objects incorrectly flagged as weeds.

Self-collected weed dataset testing

To confirm the algorithm’s capacity to accurately localize weed targets, the current study tests multiple algorithmic formulations on independently collected imagery; the corresponding detection metrics are presented in **Table 5**.

Table 5. The performance of four detection models tested on self-collected data sets

Model	FPS	F1Score (%)	Recall (%)	Precision (%)	mAP@0.5 (%)
YOLOv5	**32.5**	78.72	78.04	79.42	77.72
YOLOv8	29.3	82.07	81.32	82.83	80.23
YOLOv11	28.7	83.67	82.72	84.64	82.03
YOLOv11-TA(Ours)	28.3	**87.16**	**86.19**	**87.54**	**85.83**

In authentic field imagery, individual frames contain more weed and crop entities, while the targets themselves occupy fewer pixels, conditions that collectively erode overall detection accuracy. Augmenting the pipeline with an attention mechanism module strengthens detection performance by accounting for spatial interdependencies among weeds. Subsequently, using the optimized detection model in conjunction with template matching, the weed-detection outputs across consecutive frames are shown in **Figure 8**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

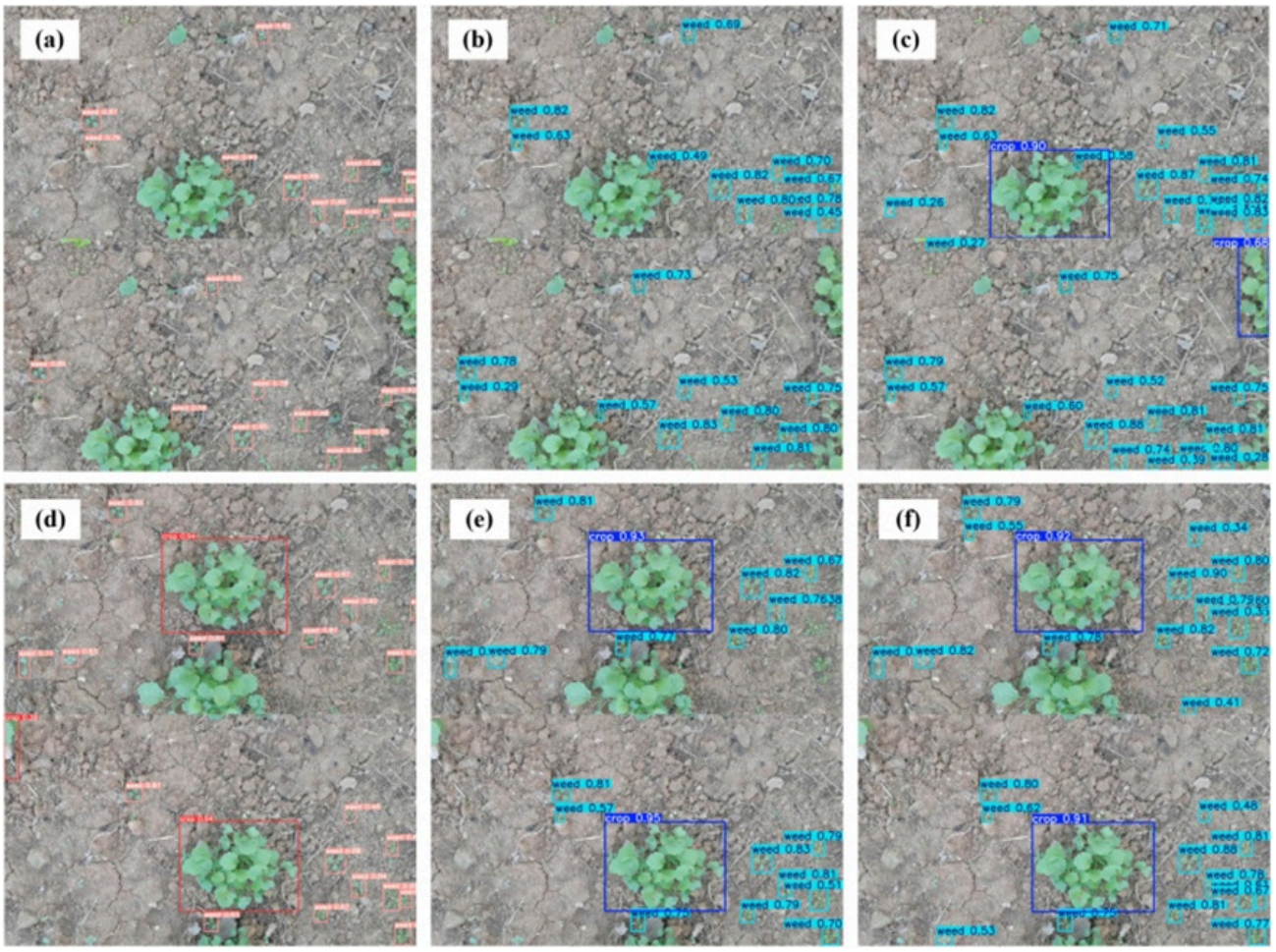


Figure 8. Results of continuous detection of image frames in a realistic scene dataset: (a) and (d) are detection results of the YOLOv5 model; (b) and (e) are detection results of the YOLOv11 model; (c) and (f) are detection results of the optimized model proposed in this paper.

As shown in **Figure 8**, when the objects of interest occupy a small footprint within the frame, globally operating detection strategies tend to miss a subset of identifiable targets. A total of 191 images were evaluated in the custom dataset, which contains 757 crop and 2992 weed ground-truth annotations. It should be noted that species are no longer split into weed categories but are instead treated uniformly as a single generic class. The real-time statistical summary of detection outcomes across multiple algorithmic variants is reported in **Table 6**.

Table 6. The performance of four detection models was evaluated on self-collected datasets.

Model	Weed			Crop		
	FN	FP	TP	FN	FP	TP
YOLOv5	657	605	2335	175	172	582
YOLOv8	559	503	2433	140	128	617
YOLOv11	518	449	2475	131	114	626
YOLOv11-TA(Ours)	413	367	2578	104	93	653

© 2023 The Author(s). | Published by Agro Intelligence Research Press

Across the self-collected datasets, the overall recognition rates proved somewhat underwhelming, predominantly because the targets are physically quite small. The tabulated numbers indicate that the missed-detection rates for all algorithmic variants are substantial, as weeds are frequently occluded as the platform advances and may consequently shed considerable discriminative feature information. By incorporating the attention mechanism module, the frequency of overlooked weed instances declines; furthermore, when viewed alongside the algorithm's processing throughput, the temporal overhead experiences essentially no meaningful escalation. Within the context of consecutive frame analysis, data linkage is established through template matching. This robust strategy performs well in scenarios with weed targets and objects of constrained spatial dimensions.

Weed tracking experiment

Ablation experiment

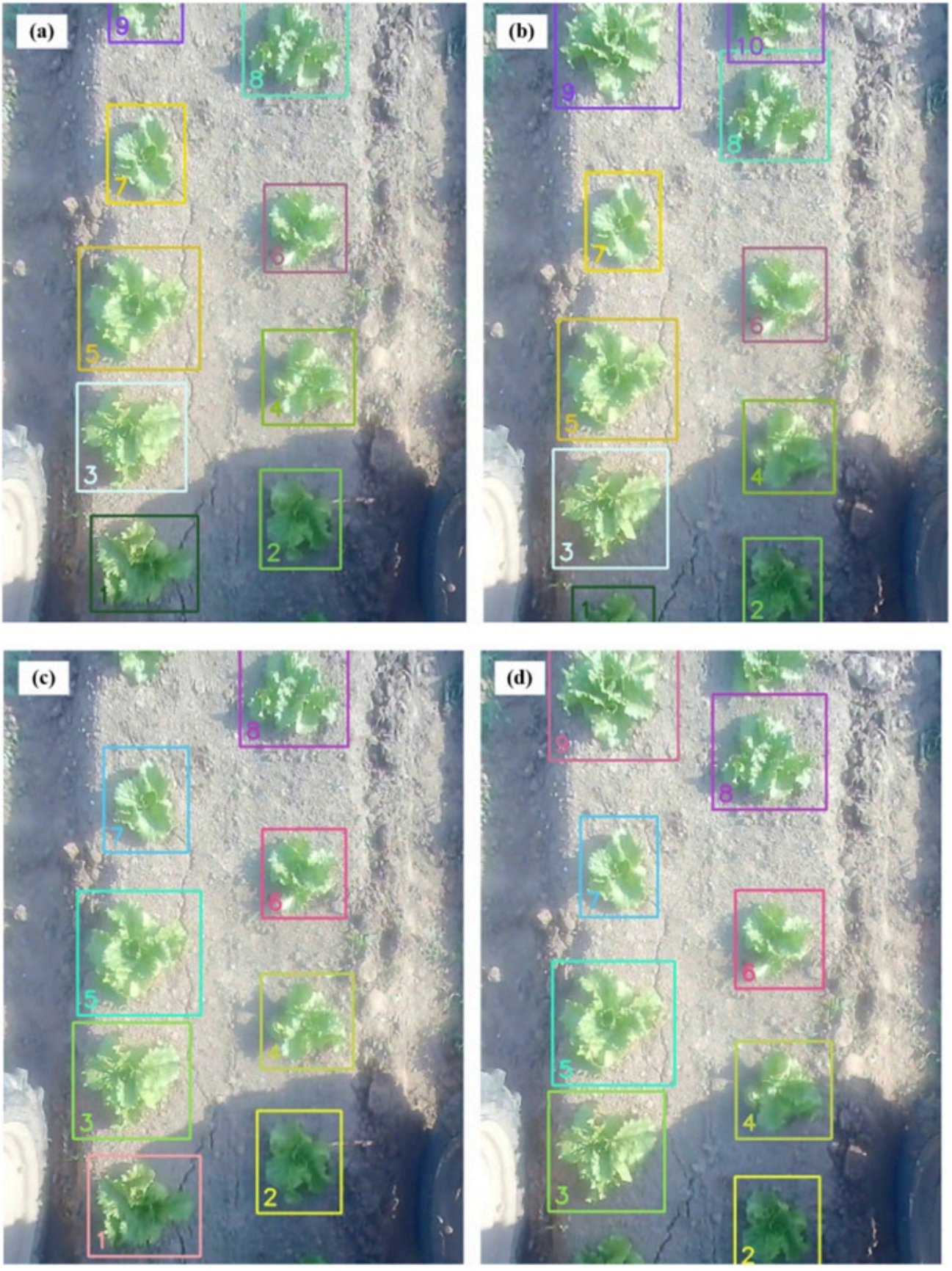
To substantiate the individual contribution of each optimised module, the gains from motion pose estimation refinement and weed target tracking filter enhancement were separately quantified on the publicly available LettuceMOT corpus, using the mainstream ByteTrack [44] algorithm as a comparative reference. The evaluation centred on the B&F1 and B&F2 sequence subsets, given that both encompass forward traversal and return-path motion, thereby constituting a more complex operational setting, as presented in **Table 7**.

Table 7. Tracking algorithms' performance on LettuceMOT.

	B&F2			B&F1		
	WS-2	**WS-1**	**ByteTrack**	**WS-2**	**WS-1**	**ByteTrack**
HOTA (%)	45.51	**48.64**	43.65	62.55	**63.92**	60.46
MOTA (%)	52.75	**53.15**	46.76	65.73	**66.18**	63.53
DetA (%)	57.84	**58.42**	39.63	84.13	**85.32**	77.58
AssA (%)	41.64	**44.02**	30.53	55.54	**58.25**	46.57
IDF1(%)	54.87	**57.24**	39.41	71.48	**74.52**	55.37

Within **Table 7**, the label WS-1 corresponds to tracking outputs produced by the optimised motion estimation pipeline, designated as WeedsSORT-1 throughout the remainder of this discussion. The label WS-2 captures the results from the refined object-tracking prediction component, subsequently referred to as WeedsSORT-2. The collected evidence shows that WeedsSORT-1 achieves the highest score on the HOTA metric, outperforming ByteTrack by approximately 8%. Likewise, the HOTA figure registered by WeedsSORT-2 exceeds that of ByteTrack by around 3.8%. The DetA scores of WeedsSORT-1 and WeedsSORT-2 remain largely comparable, a consequence of their mutual reliance on the same enhanced detector; nevertheless, WeedsSORT-1 benefits from the matching optimisation outcomes used during cross-validation, resulting in a marginally superior DetA value. Each of the three algorithm families yields better results on the B&F1 sequence than on B&F2, a discrepancy attributable to the poorer, spatially uneven illumination characterising the latter. Such lighting inconsistency can cause substantial shifts in the pixel-level representation of targets across both detection and matching stages, potentially leading to tracking dropouts. A visual juxtaposition of the tracking behaviours exhibited by WeedsSORT-1 and ByteTrack on the B&F2 sequence is provided in **Figure 9**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press



© 2023 The Author(s). | Published by Agro Intelligence Research Press

Figure 9. Comparison of the algorithm tracking effect on the B&F2 sequence: (a) to (b) are the effects of Weedssor-1 tracking, and (c) to (d) are the effects of ByteTrack tracking.

In circumstances where fresh crop specimens enter the frame or previously monitored ones gradually exit, WeedsSORT-1 maintains durable, interruption-free tracking. To afford a more immediate visual grasp of the performance differentials, several metric classes are depicted as bar charts in **Figure 10**, specifically for the B&F2 sequence.

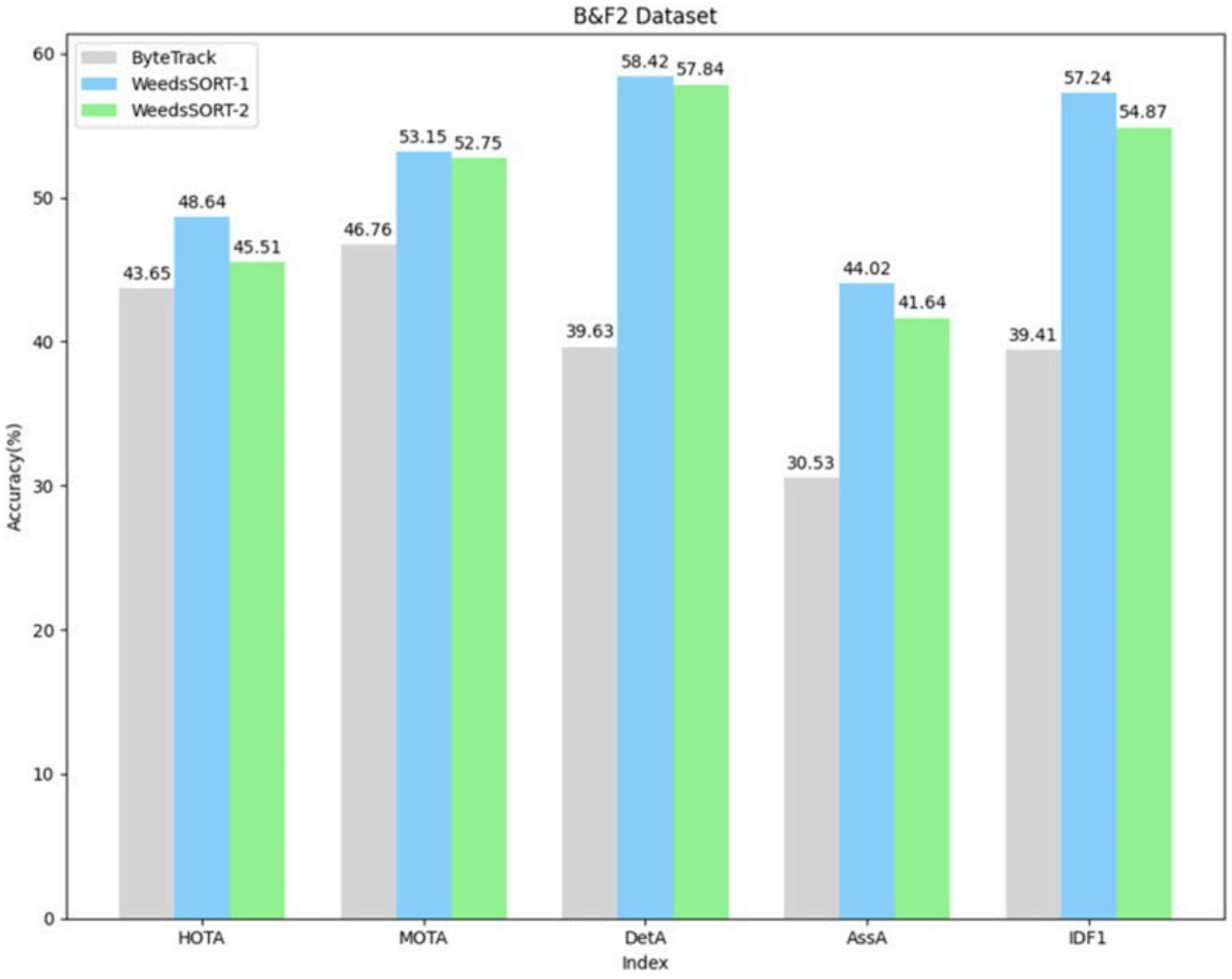


Figure 10. Comparative effect of several types of tracking indicators.

The graphical evidence confirms that both WeedsSORT-1 and WeedsSORT-2 outperform ByteTrack across the full spectrum of metrics, with WeedsSORT-1 posting the strongest overall performance. This observation suggests that motion correspondences and constraints bridging successive frames exert a notable impact on aggregate tracking precision. All optimisation components were subsequently consolidated into a single tracking framework and collectively termed WeedsSORT. Its object tracking fidelity was then evaluated on the public LettuceMOT dataset and benchmarked against ByteTrack and DeepOCSORT, with the full comparative breakdown provided in **Table 8**.

Table 8. The performance of three tracking algorithms was evaluated on LettuceMOT.

Dataset	Method	ML	MT	IDF1(%)	DetA (%)	MOTA (%)	HOTA (%)
Straight1	ByteTrack	4	5	94.12	83.97	85.12	84.01

© 2023 The Author(s). | Published by Agro Intelligence Research Press

	DeepOCSORT	5	6	95.32	84.43	85.43	85.19
	WeedsSORT	**3**	**8**	**96.32**	**90.35**	**88.13**	**87.44**
Straight2	ByteTrack	4	7	89.45	83.41	81.23	79.23
	DeepOCSORT	5	6	90.62	82.19	83.85	80.12
	WeedsSORT	**3**	**7**	**91.56**	**88.31**	**83.98**	**83.23**
B&F1	ByteTrack	5	4	71.43	77.58	63.53	60.46
	DeepOCSORT	6	4	72.57	76.93	61.45	59.62
	WeedsSORT	**5**	**6**	**77.47**	85.54	**68.67**	**65.76**
B&F2	ByteTrack	5	3	52.51	39.63	46.76	43.65
	DeepOCSORT	3	4	57.63	44.62	49.43	46.23
	WeedsSORT	**2**	**6**	**60.48**	**59.14**	**55.93**	**52.48**

In **Table 8**, the sequences labelled Straight1 and Straight2 correspond to operating modes in which the mobile robot advances along a purely linear path. Under these tracking conditions, there is no need to contend with turning manoeuvres or reverse-direction movement, which generally allows the algorithms to achieve higher accuracy levels. When set against the other two algorithm classes, WeedsSORT delivers markedly more consistent track continuity from one frame to the next. By exploiting detection linkages across preceding and subsequent frames, combined with constraints imposed by the EKF, it achieves highly dependable target following. All three algorithms ingest target detections supplied by the YOLOv11 detector. With respect to the DetA dimension, embedding an attention mechanism provides WeedsSORT with antecedent spatial awareness during the detection stage. Moreover, by merging motion signals with template-based alignment, it establishes data correspondences with the detection outputs, thereby improving overall detection precision.

Real scene data experiment

In addition to testing against the public benchmark, this investigation evaluated the trio of algorithms on the custom-collected dataset, which was split into two separate sequences; the resulting performance figures are collated in **Table 9**.

Table 9. The performance of three tracking algorithms was evaluated on the WeedMOT dataset.

Dataset	Method	ML	MT	IDF1(%)	DetA (%)	MOTA (%)	HOTA (%)
Weeds1	ByteTrack	7	4	30.72	41.84	36.52	35.65
	DeepOCSORT	7	5	34.82	37.27	38.13	37.19
	WeedsSORT(Ours)	**4**	**7**	**37.32**	**47.72**	**41.03**	**40.25**
Weeds2	ByteTrack	4	5	37.31	39.18	35.26	33.55
	DeepOCSORT	6	5	38.38	38.52	36.23	34.13
	WeedsSORT(Ours)	**3**	**8**	**40.18**	**43.83**	**38.23**	**37.28**

The data summarised in **Table 9** indicate that the sheer abundance of weeds in the self-collected dataset, combined with the extremely limited pixel area occupied by individual specimens, degrades the overall tracking capability of all algorithms examined. Focusing on HOTA accuracy, WeedsSORT surpasses ByteTrack by a margin spanning roughly

© 2023 The Author(s). | Published by Agro Intelligence Research Press

11.12%–12.9%, and exceeds DeepOCSORT by an estimated 8.2%–9.3%. A pictorial illustration capturing how the three algorithms handle weed tracking is furnished in **Figure 11**.

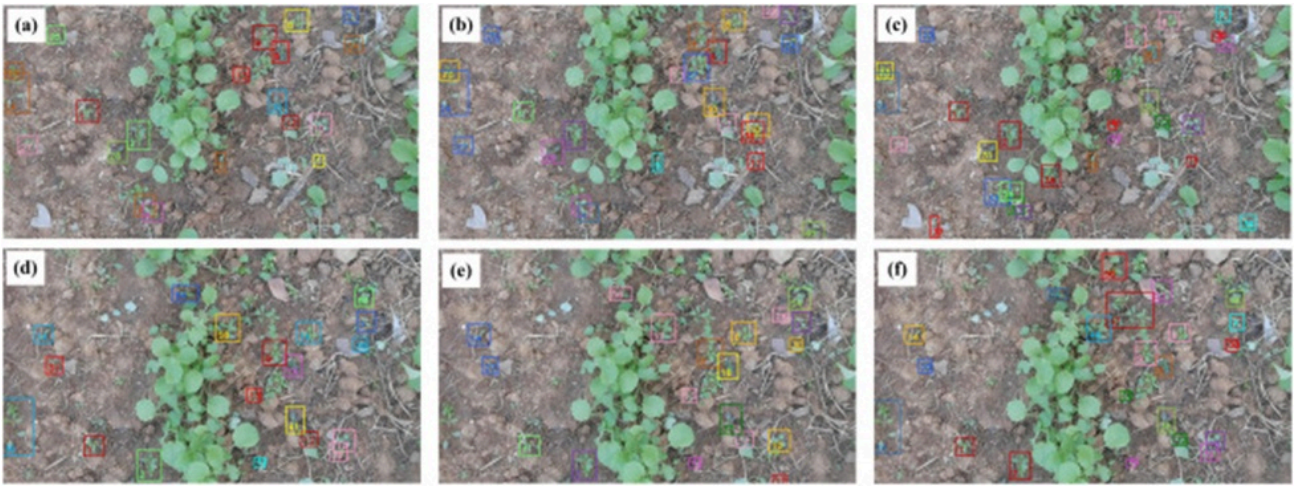


Figure 11. (a) to (d) illustrates the tracking results obtained using ByteTrack, while (b) to (e) presents the outcomes from DeepOCSORT, and (c) to (f) shows the results from WeedsSORT.

The qualitative observations indicate that WeedsSORT produces the most reliable weed-tracking behaviour, distinguishing itself particularly by achieving higher recognition rates for targets of very limited spatial extent. In contrast, the other two algorithms exhibit deficiencies, such as incorrect temporal associations and track fragmentation, when confronted with small entities, such as weeds. This underperformance arises primarily because these competing algorithms neither incorporated detection refinements specifically designed for minute weed targets nor embedded suitable constraints in the motion estimation stage, leading to tracking failures. Through the tight coupling of accurate weed detection with motion pose estimation outputs, together with the application of extended Kalman filtering and noise-adaptive scaling factors, the empirical evidence and quantitative data demonstrate that this work successfully realises precise tracking of small-scale, highly similar weed specimens, thereby satisfying the demands of exacting weed management in laser weeding contexts.

Discussion

Treatment results of occluding weeds

Beyond standard experimental protocols and analyses, the masking of weeds by crop canopy as the robot traverses the field poses a considerable hurdle for weeding tasks and warrants closer scrutiny. This study presents a comparative evaluation of several algorithms in scenarios where weeds are partially hidden, with the resulting performance measures graphed.

As shown in **Figure 12**, after embedding the attention mechanism optimisation, the algorithm achieves a higher detection hit rate for weeds with patchy spatial arrangements and those partly concealed, ultimately yielding superior recognition of occluded weeds. The recognition tallies achieved by the three classes of algorithms for obstructed weeds and crops are assembled in **Table 10**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

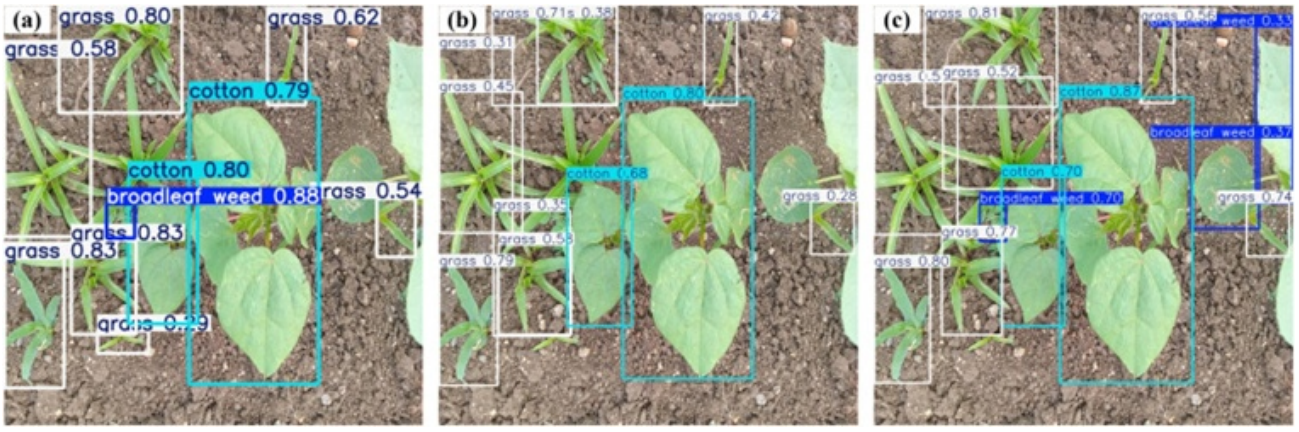


Figure 12. In the case of weed occlusion, (a)YOLOv5 detection result, (b) YOLOv11 detection result, and (c) YOLOV11-TA detection result.

Table 10. Identify the situation when weeds are obscured.

Method	Grass	Broadleaf	Cotton
YOLOv5	7	1	2
YOLOv11	8	0	2
YOLOv11-TA(Ours)	**9**	**3**	2

Operating beneath a cotton plant canopy, the detection algorithm was refined throughout this investigation and posted higher weed-detection percentages for both broadleaf and grassy weed varieties. In particular, the detection efficacy for broadleaf weeds outperformed that of YOLOv11. Equally, the detection performance targeting grassy weeds also outpaced that of the competing pair of algorithms, indicating that attention-mechanism-based optimisation yields meaningful benefits in conditions with occlusion.

The results of weed treatment in high-density distribution

Beyond circumstances in which weeds are occluded, laser weeding environments routinely harbor elevated weed populations, rendering analysis and discussion of detection performance under high-density conditions necessary. This paper selects two illustrative scenarios and pits the three algorithms against one another. The side-by-side comparisons are shown in **Figure 13**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

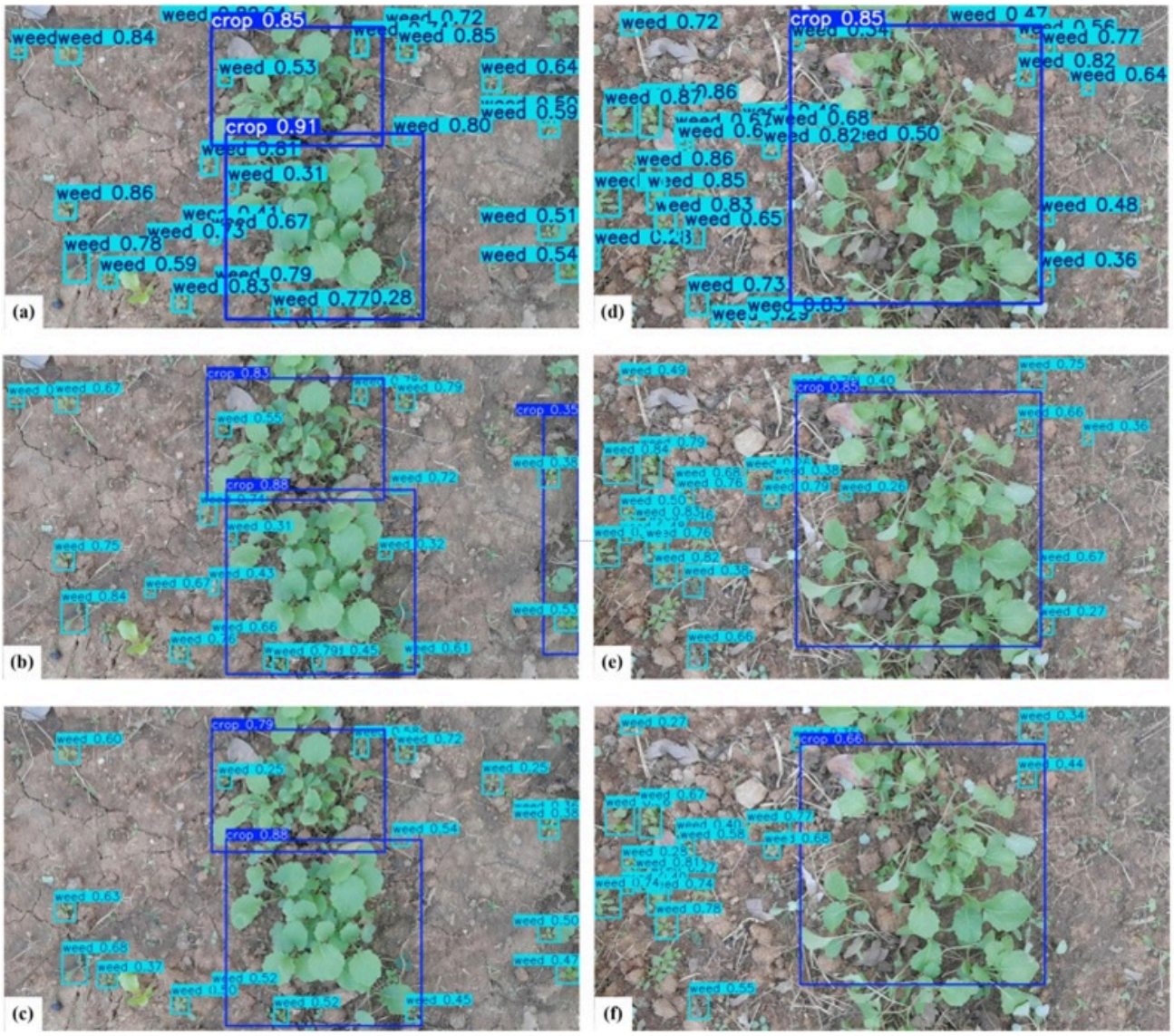


Figure 13. (a), (b), and (c) are the detection results of YOLOv5, YOLOv11, and Yolov11-TA in Scenario 1, respectively; (d), (e), and (f) are the detection results of YOLOv5, YOLOv11, and Yolov11-TA in Scenario 2, respectively.

As illustrated in **Figure 13**, when the scene is saturated with excessive weed specimens, the detection outputs from YOLOv5 and YOLOv11 are inadequate, with numerous missed detections. Following the integration of the attention mechanism, the detection algorithm proposed in this paper delivers markedly stronger detection performance for small targets and densely clustered weeds compared with the other two algorithms; the numerical results are tabulated in **Table 11**.

Table 11. Weed identification in high-density scenarios.

Method	Scenario 2		Scenario 1	
	Weed	Crop	Weed	Crop
YOLOv5	18	1	17	2
YOLOv11	25	1	21	3

© 2023 The Author(s). | Published by Agro Intelligence Research Press

YOLOv11-TA(Ours)	**31**	**1**	**33**	3

In line with the evidence presented in **Table 11**, the weed detection rate accomplished by the YOLOv11-TA algorithm substantially exceeded that attained by both the unoptimised variant and the YOLOv5 algorithm. In situations with uneven spatial distribution, integrating the detection pipeline with the attention mechanism yielded substantial gains in weed-detection precision. This achievement furnishes a high-accuracy underpinning for the execution of laser weeding tasks.

The results under different weather conditions

Sustaining the weeding robot's uninterrupted operation requires that detection precision remain robust across a wide range of environmental conditions. For this reason, the current study offers a comparative examination of detection strategies across differing settings, including rainy periods and scenes with minimal ambient light. The resultant detection examples are depicted in **Figure 14**.

As the evidence in **Figure 14** makes plain, the detection effectiveness of both YOLOv5 and YOLOv11 degrades severely when illumination levels drop, for instance, during nocturnal hours and in rainfall. This shortfall most probably originates from the algorithms' inability to interpret the deteriorated image signal meaningfully. By contrast, the variant incorporating the attention mechanism, advocated in this work, demonstrates markedly stronger detection prowess, enabling dependable weed spotting even under these adverse settings. A numerical breakdown contrasting the identification and detection tallies produced by the trio of algorithms is supplied in **Table 12**.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

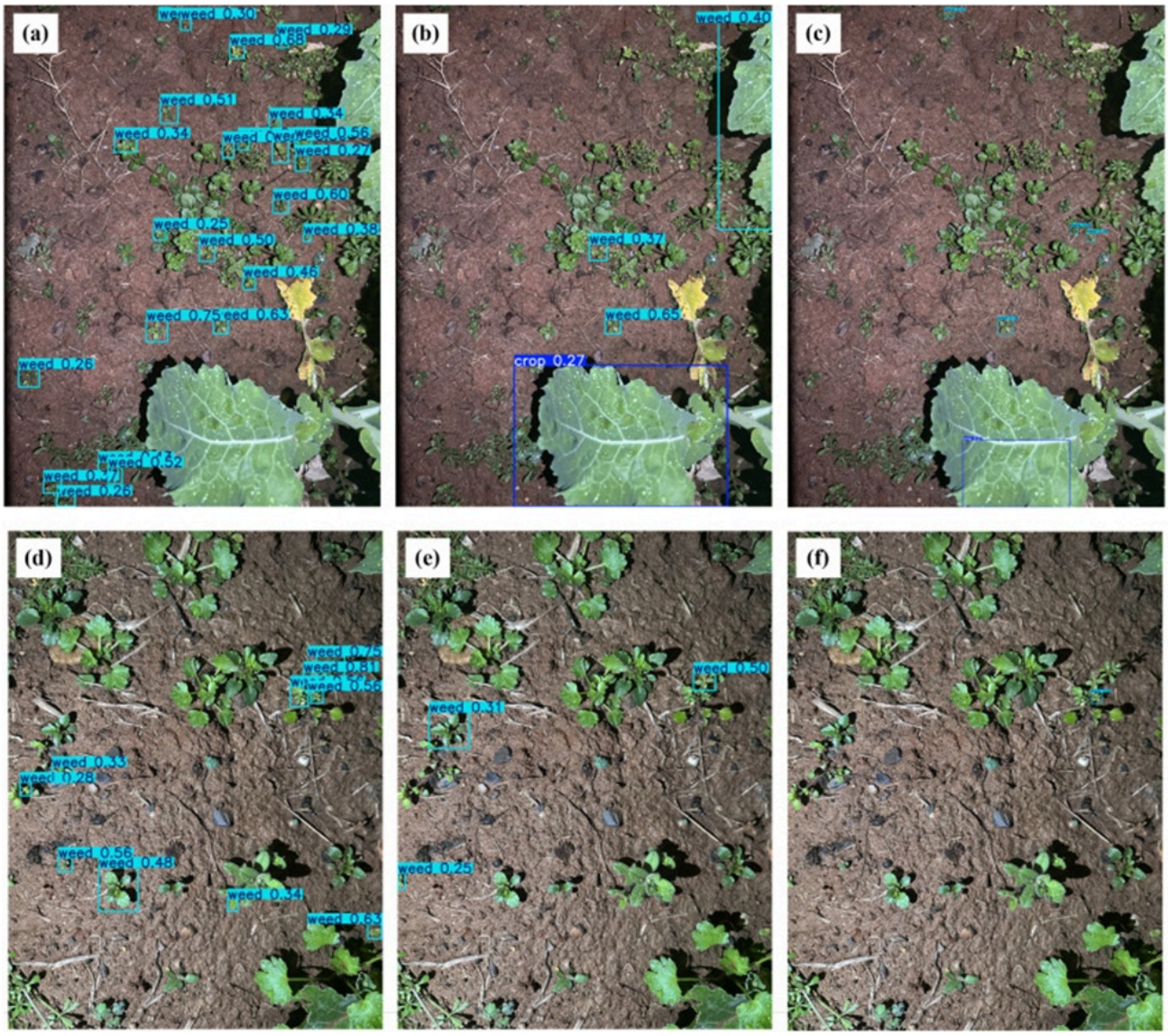


Figure 14. (a), (b) and (c) are the detection results of YOLOV11-TA, YOLOv11 and YOLOv5 algorithms under low light at night, respectively; (d), (e) and (f) are the detection results of YOLOv11-TA, YOLOv11 and YOLOv5 algorithms on rainy nights, respectively.

Table 12. Identification of weeds at night and on rainy days.

Method	Rain scene		Night scene	
	Weed	Crop	Weed	Crop
YOLOv5	1	0	4	1
YOLOv11	3	0	3	1
YOLOv11-TA(Ours)	**10**	**0**	**24**	0

Consistent with the data shown in **Table 12**, the shortage of adequate lighting cripples the detection capabilities of YOLOv5 and YOLOv11 to a considerable degree. In contrast, the YOLOV11-TA variant described in this article produces

© 2023 The Author(s). | Published by Agro Intelligence Research Press

far more favourable detection results. Anchored by a spatial attention distribution mechanism combined with a small-target detection framework, the system delivers highly accurate detection across a range of weather phenomena.

Weed tracking computing requirements

Owing to the edge-based deployment architecture of the laser weeding system, a thorough investigation of the system's runtime velocity is called for. The hardware assembly under scrutiny is the NVIDIA Jetson Orin NX 16GB module, which furnishes a processing capacity of 100 TOPS. The frame-processing throughput for each weed-detection and tracking algorithm was recorded, along with the corresponding live memory footprint. This section also recapitulates the statistical data for WS-1 and WS-2, previously discussed in Section 3.3, and consolidates the measurements in **Table 13**.

Table 13. The running speed and memory usage on the core board.

Dataset	Method	Memory (%)	FPS	IDF1(%)	DetA (%)	HOTA (%)
Straight1	ByteTrack	**47.4**	**25.83**	94.12	83.97	84.01
	DeepOCSORT	55.1	21.31	95.32	84.43	85.19
	WS-1	57.9	21.04	95.49	88.43	85.85
	WS-2	52.4	23.65	95.21	87.96	84.94
	WeedsSORT	63.2	20.42	**96.32**	**90.35**	**87.44**
Straight2	ByteTrack	**43.8**	**28.91**	89.45	83.41	79.23
	DeepOCSORT	51.3	25.72	90.62	82.19	80.12
	WS-1	55.8	25.43	90.89	86.92	81.49
	WS-2	50.1	26.75	89.98	85.13	80.76
	WeedsSORT	58.3	24.97	**91.56**	**88.31**	**83.23**

Reviewing the contents of **Table 13**, ByteTrack posts the most advantageous FPS rate and memory consumption profile, a trait explained by the comparatively fewer restrictions the algorithm places on motion pose estimation and detection processing during tracking. Considering the aggregate HOTA metric, WeedsSORT delivers the best performance, whilst its processing FPS remains above 20, largely meeting real-time demands. Throughout the algorithm's runtime, the core board's occupied memory fraction never exceeds 65%, indicating no memory saturation risk when deploying the weed detection and tracking pipeline on this hardware configuration.

Conclusion

This manuscript introduces WeedsSORT, an original algorithm crafted for weed detection and tracking in the realm of precision agriculture, expressly designed for laser weeding apparatus. To improve recognition of small weed specimens in single-frame captures, we combine YOLOv11 with a Triplet Attention module, boosting detection precision while incurring only a trivial extra computational cost. In the tracking stage, we mitigate potential false correspondences by applying a refined SuperPoint feature detection workflow with a multi-feature-layer image feature decoder. In the subsequent step, the SuperGlue algorithm prunes the set of candidate matching keypoints, constructing connections both within individual feature groups and across different feature sets. Acknowledging the inherent unpredictability of farm settings, we deploy an EKF formulation endowed with a noise-adaptive factor to adjust nonlinear variables on the fly, thereby unifying temporal and spatial data. Empirical tests conducted on publicly and independently collected datasets show that the detection algorithm achieves mAP scores of 97.29% and 85.83%, outpacing YOLOv11, YOLOv8, and YOLOv5. Tracking capability benchmarks set against ByteTrack and DeepOCSORT using the LettuceMOT and WeedsMOT collections

© 2023 The Author(s). | Published by Agro Intelligence Research Press

record respective HOTA accuracy lifts of 12.01% and 8.75%. In addition, the algorithm proposed in this study has been tested under conditions of weed occlusion, densely packed weed clusters, pitch-black nights, and rainy nights, consistently achieving test results that surpass those of competing state-of-the-art approaches. At the same time, the running speed and memory utilisation on the core board remain broadly sufficient for real-time operation. These results highlight the effectiveness of WeedsSORT as a reliable solution for weed detection and tracking in the agricultural robotics domain.

That said, the proposed model, despite being capable of exact weed detection and tracking and fit for deployment on edge-class hardware, still encounters certain limitations. (1) The spectrum of input modalities is constrained; the present algorithm relies exclusively on visual characteristics for both detection and tracking, without harnessing additional sensor streams, such as data from an IMU, which could enrich motion estimation. (2) The terrain structure remains comparatively unvaried; the operational surface for the laser vehicle is chiefly level ground. Future deployments in more taxing landscapes, such as irregular or rugged terrain, are likely to pose additional obstacles for the current algorithm. In upcoming research endeavours, we plan to fuse measurements from IMUs and wheel speed encoders to supply extra constraints and refinements for the tracking prediction mechanism, and we will explore the design of a more resilient and faithful weed detection and tracking algorithm suited to other intricate and demanding operational contexts.

Acknowledgements

None

Conflict of interest

None

Financial support

This work was supported by Hubei Provincial Technology Innovation Program (2024BBB053).

Ethics statement

None

Received: 24 Jan 2022 Revised: 30 Mar 2022 Accepted: 14 Apr 2022
Published online: 25 May 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Hernandez D, Pasha L, Yusuf DA, Nurfaizi R, Julianingsih D. The role of artificial intelligence in sustainable agriculture and waste management: towards a green future. Int Trans Artif Intell. 2024;2(2):150-7.

2. Woyessa D. Weed control methods used in agriculture. Am J Life Sci Innov. 2022;1(1):19-26.

3. Yang Q, Du X, Wang Z, Meng Z, Ma Z, Zhang Q. A review of core agricultural robot technologies for crop productions. Comput Electron Agric. 2023;206:107701.

4. Pal A, Kumar V. AgriDet: Plant leaf disease severity classification using agriculture detection framework. Eng Appl Artif Intell. 2023;119:105754.

5. Dogra R, Rani S, Singh A, Albahar MA, Barrera AE, Alkhayyat A. Deep learning model for detection of brown spot rice leaf disease with smart agriculture. Comput Electr Eng. 2023;109(PA):108659.

6. Li J, Li J, Zhao X, Su X, Wu W. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4. Comput Electron Agric. 2023;211:107955.

7. Ahmad I, Yang Y, Yue Y, Ye C, Hassan M, Cheng X, et al. Deep learning based detector YOLOv5 for identifying insect pests. Appl Sci. 2022;12(19):10167.

8. Roy AM, Bhaduri J. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. Comput Electron Agric. 2022;193:106694.

9. Dang F, Chen D, Lu Y, Li Z. YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Comput Electron Agric. 2023;205:107655.

10. Hasan AM, Sohel F, Diepeveen D, Laga H, Jones MG. A survey of deep learning techniques for weed detection from images. Comput Electron Agric. 2021;184:106067.

11. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proc IEEE Conf Comput Vis Pattern Recogn. 2016:779-88.

12. Badgujar CM, Poulose A, Gan H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. Comput Electron Agric. 2024;223:109090.

13. Redmon J, Farhadi A.YOLOv3: An incremental improvement, 2018.

14. Bochkovskiy A, Wang CY, Liao HYM. Yolov4: Optimal speed and accuracy of object detection, 2020.

15. Jocher G, Stoken A, Borovec J, Chaurasia A, Changyu L, Hogan A, et al. ultralytics/yolov5: v5.0-YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations. Zenodo. 2021.

16. Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach Learn Knowl Extr. 2023;5(4):1680-716.

17. Khanam R, Hussain M.Yolov11: An overview of the key architectural enhancements , 2024.

18. Wang K, Hu X, Zheng H, Lan M, Liu C, Liu Y, et al. Weed detection and recognition in complex wheat fields based on an improved YOLOv7. Front Plant Sci. 2024;15:1410096.

19. An Q, Wang K, Li Z, Song C, Tang X, Song J. Real-time monitoring method of strawberry fruit growth state based on YOLO improved model. IEEE Access. 2022;10:124363-72.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

20. Liu S, Jin Y, Ruan Z, Ma Z, Gao R, Su Z. Real-time detection of seedling maize weeds in sustainable agriculture. Sustainability. 2022;14(22):15088.

21. Chen J, Wang H, Zhang H, Luo T, Wei D, Long T, et al. Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. Comput Electron Agric. 2022;202:107412.

22. Wang Q, Cheng M, Huang S, Cai Z, Zhang J, Yuan H. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings. Comput Electron Agric. 2022;200:107244.

23. Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF, 2011 International Conference on Computer Vision, 2011.

24. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-4.

25. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). Comput Vis Image Underst. 2008;110(3):346-59.

26. Liu Y, Shen Z, Lin Z, Peng S, Bao H, Zhou X. GIFT: Learning transformation-invariant dense visual descriptors via group CNNs. Adv Neural Inf Process Syst. 2019;32:15527-37.

27. Lee J, Jeong Y, Kim S, Min J, Cho M. Learning to distill convolutional features into compact local descriptors, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision , 2021.

28. Barroso-Laguna A, Riba E, Ponsa D, Mikolajczyk K. Key.Net: Keypoint detection by handcrafted and learned CNN filters. Proc IEEE/CVF Int Conf Comput Vis. 2019.

29. Shi Y, Cai JX, Shavit Y, Mu TJ, Feng W, Zhang K. Clustergnn: Cluster-based coarse-to-fine graph neural network for efficient feature matching, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition , 2022.

30. DeTone D, Malisiewicz T, Rabinovich A.Self-supervised interest point detection and description, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops . 2018.

31. Bhowmik A, Gumhold S, Rother C, Brachmann E.Reinforced feature points: Optimizing feature detection and description for a high-level task, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition . 2020.

32. Sarlin PE, DeTone D, Malisiewicz T, Rabinovich . Superglue: Learning feature matching with graph neural networks, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

33. Li Z, Ma J. Learning feature matching via matchable keypoint-assisted graph neural network. IEEE Trans Image Process. 2024

34. Bewley A, Ge Z, Ott L, Ramos F, Upcroft B. Simple online and realtime tracking, In: 2016 IEEE International Conference on Image Processing. 2016.

35. Cao J, Pang J, Weng X, Khirodkar R, Kitani K.Observation-centric sort: Rethinking sort for robust multi-object tracking. Proc IEEE/CVF Conf Comput Vis Pattern Recognit. 2023.

36. Maggiolino G, Ahmad A, Cao J, Kitani K. Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification. IEEE Int Conf Image Process. 2023.

37. Hu N, Su D, Wang S, Nyamsuren P, Qiao Y, Jiang Y, et al. LettuceTrack: Detection and tracking of lettuce for robotic precision spray in agriculture. Front Plant Sci. 2022;13:1003243.

38. Saraceni L, Motoi IM, Nardi D, Ciarfuglia TA. AgriSORT: A Simple Online Real-time Tracking-by-Detection framework for robotics in precision agriculture. IEEE Int Conf Robot Autom. 2024:2675-82.

© 2023 The Author(s). | Published by Agro Intelligence Research Press

39. Zhou H, Li J, Peng J, Zhang S, Zhang S. Triplet attention: Re-thinking the similarity in transformers. Proc ACM SIGKDD Conf Knowl Discov Data Min. 2021:2378-88.

40. Duangpakdee K, Thananta G, Sukpancharoen S. IoT enhanced deep water culture hydroponic system for optimizing Chinese celery yield and economic evaluation. Smart Agric Technol. 2024;9:100545.

41. Duangpakdee K, Sukpancharoen S. Vertical Smart Farm System for off-Season Crop Production using Hydroponics and IoT-based Environmental Control. IEEE Int Conf Adv Robot Mechatron. 2024:284-9.

42. Promkaew N, Thammawiset S, Srisan P, Sanitchon P, Tummawai T, Sukpancharoen S. Development of metaheuristic algorithms for efficient path planning of autonomous mobile robots in indoor environments. Results Eng. 2024;22:102280.

43. Hu N, Wang S, Wang X, Cai Y, Su D, Nyamsuren P, et al. LettuceMOT: A dataset of lettuce detection and tracking with re-identification of re-occurred plants for agricultural robots. Front Plant Sci. 20222022;13:1047356.

44. Zhang Y, Sun P, Jiang Y, Yu D, Weng F, Yuan Z, et al. ByteTrack: Multi-object tracking by associating every detection box. Lect Notes Comput Sci. 2022;13682:1-21.

© 2023 The Author(s). | Published by Agro Intelligence Research Press