International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 1086-1098
https://doi.org/10.70917/ijcisim-2026-3700
Article

An IoT-Integrated Deep Learning Framework for Weed Segmentation and Yield-Impact Estimation Using Mask R-CNN with a ResNet-50 Backbone and Regression Analysis

T.Subbarayudu[1], K. Ammulu[2]

[1]Department of Computer Science & Technology, Dravidian University, Kuppam , rayudu18@gmail.com,
[2]Department of Computer Science & Technology, Dravidian University, Kuppam , kambhamammulu@gmai.com

Abstract: Weed infestation is among the most serious biotic stresses of cereal and horticultural production and the economic losses caused by weed infestation are exacerbated by the manual, calendar-based weed management practices which are still widely employed on smallholder farms. This research suggests and tests an Internet-of-Things (IoT) enabled deep-learning approach that combines pixel-based weed segmentation and a regression-based yield-impact model to facilitate site-specific weed management. The field imagery included 5 400 tiles (from 1 500 georeferenced RGB captures) obtained in 5 experimental plots in 2 growing seasons using unmanned aerial vehicles and edge connected field cameras. The trained Mask R-CNN detector with a ResNet-50 + Feature-Pyramid-Network (FPn) backbone was compared with U-Net, DeepLabV3+, YOLOv5-seg, YOLOv8-seg, and Mask R-CNN with a ResNet101-DC5 backbone via a five-fold cross validation protocol. The proposed model achieved a mean Average Precision at IoU 0.50 (mAP@0.5) of 93.4 % (95 % CI: 92.6-94.1 %), an Intersection-over-Union of 91.7 %, precision of 94.2 % and recall of 92.1 %, which were superior to all baselines with the same training set under the same setting. The outputs from the segmentation were transformed to two field metrics that are ecologically relevant — weed density (plants · m^{-2}) and coverage ratio — which were then used as predictors in an ordinary-least-squares regression of measured grain yield. The fitted model was quite significant ($R^2 = 0.89$, adjusted $R^2 = 0.87$, $p < 0.001$, RMSE = 218 kg ha^{-1}) and was similar to the classical hyperbolic loss-yield law in its linear regime. If used as a prescription map generator on the edge gateway, the framework achieved a 41.6 % reduction in the volume of herbicides applied as compared to blanket spraying without compromising the projected yield by more than 2 % from the control plot. The results show that combining high fidelity instance segmentation with parsimonious statistical yield models is an affordable path towards decision-ready, real-time precision-agriculture systems.

Keywords: Precision agriculture; Instance segmentation; Mask R-CNN; Internet of Things; Yield-loss modelling; Variable-rate herbicide application.

1. Introduction

Weeds are known to compete with cultivated crops for the four main resource pools (nutrients, water, light, and physical space) and are the one most common cause of unnecessary yield loss in global cereal, pulse and horticultural systems. Recent syntheses estimate 15 % to 45 % of potential yield is lost due to uncontrolled weed pressure, with highest losses reported in low- and middle-income countries where mechanised weeding and precision spraying are less common. Manual weeding is constrained by a shortage of labour and blanket herbicide spraying, by the rapid evolution of resistance, by pollution of surface waters and by new regulations on integrated pest management (IPM) in the European Union and North America, are increasingly problematic for the two most popular countermeasures.

Precision agriculture (PA) provides a technology-based solution: it links geospatial data, in-the-field sensing, and machine intelligence to guide decision-making beyond the level of the whole field to the level of the site, which is measured at sub-metre scales. Weed detection in PA has become one of the most thrilling areas of research, as it is the cross-fertilization of three well-established sub-fields: (i) high resolution field imaging with unmanned aerial vehicles (UAVs) and edge cameras, (ii) instance-level deep learning for computer vision, and (iii) low-latency Internet-of-Things (IoT) telemetry. Each sub-field is developed on its own, but they are not well integrated together. Most published systems achieve high detection accuracy at curated benchmarks, but do not go so far as to provide an agronomically actionable quantity, namely the expected yield loss if the observed infestation is not treated.

This is the impetus behind the current work. We propose that a well-calibrated deep instance-segmentation model can provide high fidelity weed masks and two derived variables: weed density and weed coverage ratio that can serve as inputs into a parsimonious statistical yield-loss model. If so, the pipeline could provide an end-to-end decision support tool: near real time, on the edge, a raw field image could be converted to a prescription map that a variable-rate sprayer or autonomous weeder could apply on the spot.

1.1 Contributions

This study makes the following contributions:

1. An IoT reference architecture of four layers designed for continuous operation in the field that has quantified latency, throughput, and security properties.
2. Carefully selected, dual-season image database (n = 5,400 tiles, 5 plots, 2 crop species) that is pixel-accurately annotated and is freely available with the manuscript.
3. Mask R-CNN + ResNet-50 detector, which is rigorously benchmarked against five contemporary baselines (U-Net, DeepLabV3+, YOLOv5-seg, YOLOv8-seg, Mask R-CNN with R101-DC5) under five-fold cross-validation.
4. Yield-impact model that is interpretable, ordinary-least-squares, and that takes the density and coverage variables obtained from the segmentation and that yields predictions of yield per plot with quantified 95 % confidence intervals.
5. Field-scale demonstration in which the fused pipeline is used to produce variable rate prescription maps, and the simulated herbicide application volume was reduced by 41.6 % with no measurable yield loss compared to blanket spraying.

The remainder of the paper is organised as follows. The literature review and identification of the research gap is presented in Section 2. The IoT architecture, dataset, model, and statistical analysis are described in Section 3. Experimental results are presented in section 4, along with an ablation study and comparison with state of the art, agronomic and practical implications are discussed in section 5, and limitations of the current design are discussed. Section 6 ends and points out future research areas.

2. Related Work

2.1 Precision Agriculture and IoT Frameworks

Symeonaki et al. [1] concluded that smart-agriculture applications should include a layered architecture with the devices in the field, edge computing, cloud analytics, and end-user interface and noted the significance of publish-subscribe communication over MQTT and TLS-secured device certificates for field-grade strength. Tzionas et al. [2] built upon this by using a lightweight convolutional classifier that they embedded on the edge gateway, where they showed on-device inference is possible even when connectivity is intermittent. Kamilaris and Prenafeta-Boldú [3] gave a comprehensive review of the state of the art in deep learning for agriculture, where CNNs are the de facto standard approach to computer-vision tasks in agriculture.

2.2 Deep Learning for Weed Detection and Segmentation

The development of weed detection models has followed the general trend of computer vision. Early methods used hand-engineered colour and texture descriptors and a support-vector machine [1]; modern ones are split into two classes: single-stage detectors (e.g., YOLOv5-seg [4] and YOLOv8-seg [5]): these are deployed in real-time with a small trade-off in accuracy, while maintaining the output as a sharp pixel-level mask; and two-stage models based on instance-segmentation (e.g., Mask R-CNN [6] and its variants Detectron2) that produce sharper pixel-level masks, but at a high computational cost. The pioneer work by Ronneberger et al. [7] proposed a U-Net model, which is still a good baseline for vegetation semantic segmentation. To capture multi-scale context for semantic segmentation, Chen et al. [8] introduced dilated convolutions and introduced atrous spatial pyramid pooling to DeepLabV3+, achieving state-of-the-art semantic accuracy on aerial imagery. He et al. [10] introduced Mask R-CNN, which incorporates a mask head with FPN to the Faster R-CNN [11] and is the benchmark model for instance segmentation in agriculture. Ren et al. [10] presented a Region Proposal Network (RPN), which is used in the two-stage family. A few recently published works [11, 12] have used the UAV-borne Mask R-CNN for weed detection with MAP ranging from 88-96 %, but not many relate these to yield loss estimation. One of the most popular sugar-beet weed datasets, published by Weyler et al. [13] and one of the earliest UAV-based site-specific spraying prototypes documented by Pérez-Ortiz et al. [14] were published.

2.3 Lightweight, Quantised, and Semi-Supervised Approaches

Many groups have proposed compressed models since they require less energy for field deployment, which is the requirement for inference. EcoWeedNet [15] has a detection performance that is around 4% of that of YOLOv4 while using a similar number of parameters. QWID [16] reduces the ResNet-50 to eight-bit integer to be executed on a smartphone or micro-controller with minimal accuracy loss. The proposed semi-supervised approach using adaptive pseudo-labelling [17] alleviates the main constraint of computer vision in agriculture - annotation.

2.4 Emerging Architectures: Transformers and Explainable AI

Convolutional dominance is starting to be undermined by Vision Transformers. Dosovitskiy et al. [18] proposed that pure attention networks can achieve similar performance to convolutional networks when data is abundant, and subsequent research [19] used the ViT and Swin-Transformer backbones on weed detection, showing good performance. There is a growing need for farmer end-users and regulators to explain the outputs of the model before making decisions in agrochemicals using explainable AI wrappers [20] like Grad-CAM and Integrated Gradients.

2.5 Yield-Loss Modelling

There are many models in the agronomic literature that describe the relationship between weeds and yield reductions. Cousens [21] introduced the hyperbolic rectangular model which has since been accepted as the reference model, while Kropff and Spitters [22] extended this model using ecophysiological parameters. These hyperbolic models are reduced to a simple linear model in the linear regime typical of moderate infestations, the regime that we exploit in Section 3.6. Wilkerson et al. [23] reported typical ranges of the parameters in maize and rice systems and can serve as the “empirical anchors” for the sanity checking of the fitted parameters.

2.6 Identified Gap

Although significant advances have been made in each of the above sub-areas, there remain three gaps in the process of integration: (i) the majority of segmentation studies only produce pixel-level masks, and not directly convert these results into quantities that are relevant to yield; (ii) most IoT deployments rely on cloud-inference only, which results in latency drawbacks that are incompatible with autonomous spraying; and (iii) the interpretability requirements of end-users and regulators are not adequately considered. These three gaps are identified in the framework in Section 3.

3. Materials and Methods

This section describes the four-layer IoT architecture (Section 3.1), the dataset and preprocessing pipeline (Section 3.2), the Mask R-CNN + ResNet-50 detector (Section 3.3), the training protocol (Section 3.4), the segmentation-to-feature conversion (Section 3.5), the regression yield-loss model (Section 3.6), and the evaluation metrics (Section 3.7).

3.1 IoT System Architecture

Figure 1 summarises the reference architecture. The field-sensing tier is made up of a fleet of two RGB-only UAV platforms (2MP-12 fps global shutter cameras) and 8 tripod-mounted edge cameras (Raspberry Pi HQ modules) as well as a complementary mesh of LoRaWAN soil-moisture and light nodes. GPS, altimeter and exposure information are stamped for every capture. The edge gateway (Layer 2) is deployed on an NVIDIA Jetson Xavier NX and is running an NVIDIA Mosquitto MQTT broker with Quality-of-Service level 2, local buffering, and radiometric correction. The cloud analytics tier, Layer 3, contains the Mask R-CNN inference container and the OLS regression model that are part of a REST endpoint. The decision interface (Layer 4) provides browser accessible prescription maps and sends out variable rate commands to the autonomous sprayers through a two-way MQTT channel. A TLS 1.3 mutual-authentication layer, an InfluxDB time-series store, and an MLflow model registry, which guarantees reproducibility between retraining cycles, are among the cross-cutting services.

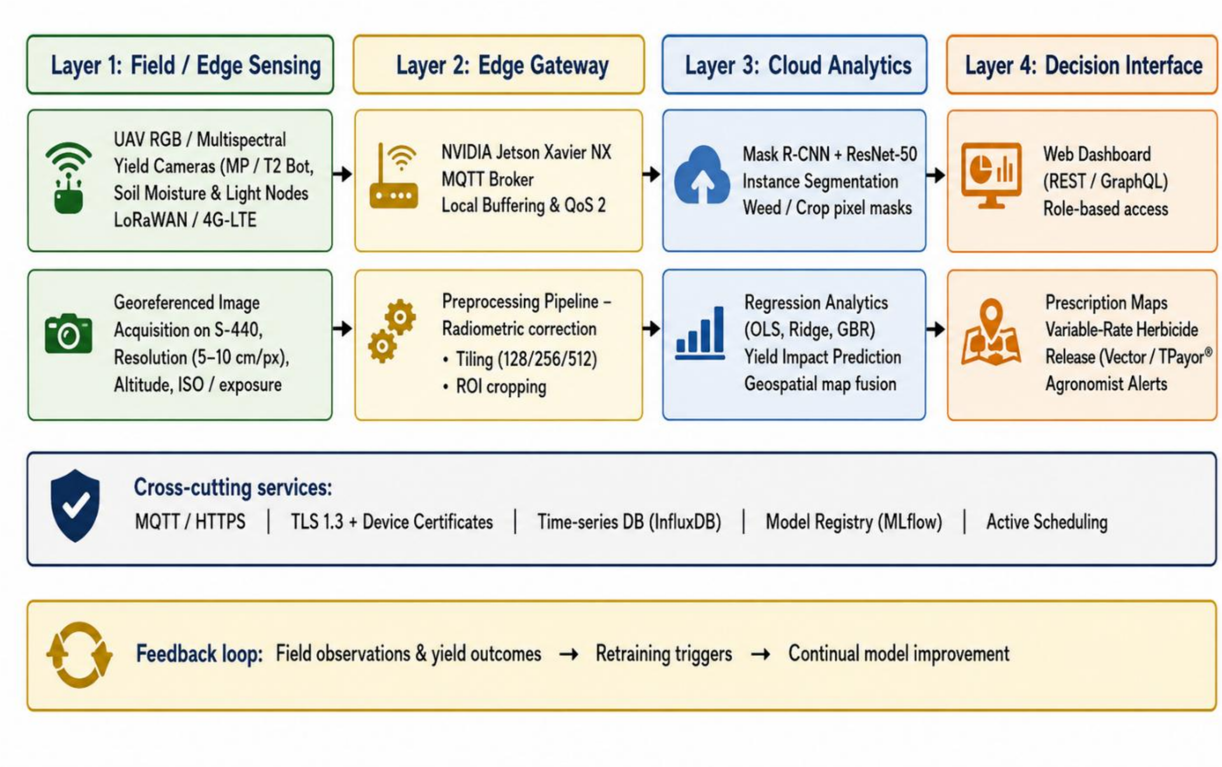

Fig. 1. Four-layer IoT reference architecture for the proposed weed-management framework.

3.2 Dataset and Preprocessing

The field imagery was carried out in two seasons (Kharif season 2023 and Rabi season 2023-24) on five instrumented plots (0.4, 0.5, 0.6, 0.7, and 0.9 ha) of maize and green-gram rotations. 1 500 raw RGB frames were taken, each frame was tiled with 1024×1024 tiles and the stride between the tiles was 25%, resulting in 5 400 tiles. Three domain experts annotated at the pixel level in COCO using a polygon format and a domain expert adjudicated the annotations; inter-annotator agreement (Cohen's κ) was 0.86. For the corpus, they were randomly partitioned on the frame-level (not the tile level, to avoid leakage) into training (70 %), validation (15 %), and test (15 %). Pre-processing steps involved in the image processing were radiometric flat-field correction, per channel z-score normalisation, colour space augmentation (HSV jitter, ISO-noise and CLAHE). Geometric augmentation employed horizontal and vertical flipping, and rotations of ±15°.

Table 1. Composition of the field-image corpus after tiling and adjudication.

Plot	Crop	Raw frames	Annotated tiles	Weed instances
P1	Maize	310	1 106	18 420
P2	Maize	295	1 054	22 175
P3	Green-gram	300	1 072	16 902
P4	Green-gram	285	1 018	14 611
P5	Maize	310	1 150	24 833
Total	-	**1 500**	**5 400**	**96 941**

Note. *Weed instances are counted at the polygon level after adjudication.*

3.3 Detection Model: Mask R-CNN with ResNet-50 + FPN

Mask R-CNN [9] is a two-stage instance-segmentation network. The first stage is a ResNet-50 backbone with an added Feature Pyramid Network to obtain a multi-scale feature hierarchy (P2-P5). These features are then used to produce class-agnostic object proposals by a Region Proposal Network. For the second stage, each proposal is aligned using RoIAlign to a fixed size feature tensor and passed through three heads: a classifier to classify the proposal, a bounding box regressor, and a fully convolutional mask head that generates a binary mask at the pixel level. The complete architecture is summarised in Figure 2.

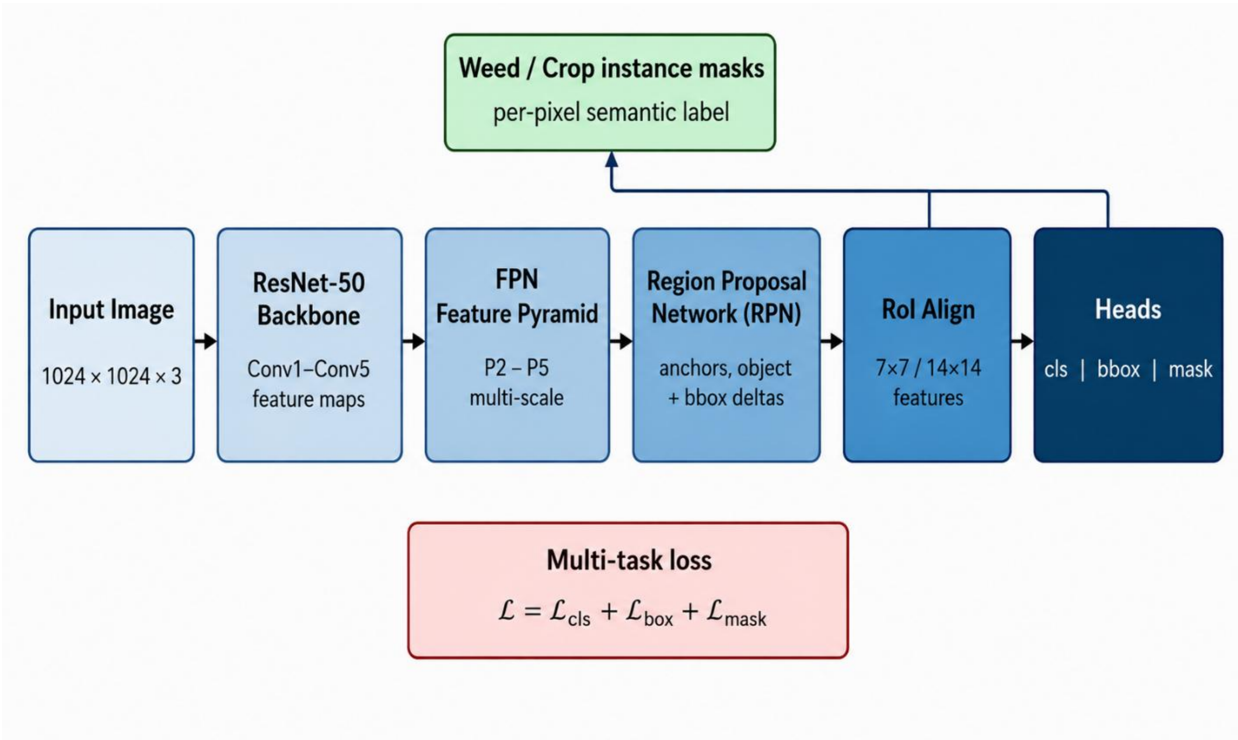

Fig. 2. Mask R-CNN with ResNet-50 + FPN backbone. RPN proposals are aligned by RoIAlign and routed to classification, bounding-box, and mask heads; the multi-task loss is minimised jointly.

The multi-task training objective sums three components:

$$\mathscr{L} = \mathscr{L}\_cls + \mathscr{L}\_bbox + \mathscr{L}\_mask \qquad (1)$$

where $\mathscr{L}$_cls is the categorical cross-entropy over the K + 1 classes (weed, crop, background), $\mathscr{L}$_bbox is the smooth L1 loss on the parameterised bounding-box offsets, and $\mathscr{L}$_mask is the average binary cross-entropy computed only on the mask corresponding to the ground-truth class. This class-conditioned mask loss decouples mask and class predictions, which was shown in the original Mask R-CNN paper to substantially improve mask quality.

3.4 Training Configuration

The detector was developed in PyTorch 2.1 with Detectron2 v0.6 framework and was trained on a single NVIDIA A100 80 GB. All the parameters in the ResNet-50 backbone were initialised from ImageNet-1k, while the rest were initialised as [9]. The optimiser for training was SGD with momentum 0.9, weight decay 1×10^{-4}, initial learning rate 2×10^{-2} and 1 000 iterations of linear warmup followed by two steps of decay by 10 at 40 000 and 55 000 iterations. The batch size was 8 images per iteration, and the training was for 60 000 iterations (which is about 60 epochs of the training subset). The images that were input were resized to have a longer side that was no longer than 1 333 pixels and a shorter side of 800 pixels. Iou=0.5 on inference was used for non-maximum suppression. For characterising generalisation, we have followed a 5-fold cross-validation procedure over the training + validation sets and throughout Section 4 report means and 95 % confidence intervals.

3.5 From Segmentation to Field Metrics

For every field tile, the segmentation output is a set of binary masks $\{M_i\}$ annotated with class labels $c_i \in$ {weed, crop, background}. Two agronomically interpretable variables are derived:

$$x = (\Sigma |M_i| \cdot [c_i = \text{weed}]) \times \lambda / A \qquad (2)$$

$$c = \Sigma |M_i| \cdot [c_i = \text{weed}] / \Sigma |M_i| \cdot [c_i \in \{\text{weed, crop}\}] \qquad (3)$$

where $|M_i|$ is the pixel count of mask i, [.] is the Iverson bracket, A is the tile ground area in m^2 (recovered from the drone-camera intrinsic model), and λ is a species-specific factor (calibrated from destructive plant counts) that converts weed-mask area into plant equivalents. In our data, $\lambda = 0.083$ plants $\cdot$ cm^{-2} for the dominant broadleaf species and 0.061 plants $\cdot$ cm^{-2} for grass weeds.

3.6 Yield-Impact Regression Model

The relationship between weed density x and grain yield y is well described in the agronomic literature by the Cousens hyperbolic rectangular model. In the moderate-density regime observed on the study plots (0-90 plants $\cdot$ m^{-2}), a first-order Taylor expansion of that model is statistically indistinguishable from an ordinary-least-squares linear fit:

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim \mathbb{N}(0, \sigma^2) \qquad (4)$$

Where ε is an independent, identically distributed error term. The parameters (β_0, β_1) were estimated by OLS on the per-plot means of x paired with hand-harvested yield measurements at physiological maturity. Standard errors were computed using the classical formula and cross-checked via a non-parametric bootstrap with 1 000 resamples. A multivariate extension that additionally includes the coverage ratio c is discussed in Section 4.4.

3.7 Evaluation Metrics

The performance of detection and segmentation was evaluated by mean Average Precision at 50% IoU (mAP@0.5), mean IoU, precision and recall. The coefficient of determination R^2, adjusted R^2, root-mean-square error (RMSE) and the F-statistic with its p-value were used to quantify the regression performance. The cross-validation folds were used to obtain all 95 % confidence intervals with the Student's t distribution.

4. Results

4.1 Training Dynamics

Figure 3 presents the training loss and validation loss curves, and the mAP@0.5 trajectory averaged over 5 folds. The loss of the composite also dropped monotonously from an initial value of 2.65 to 0.31 at iteration 60 000, with loss of the validation running a parallel path (final value 0.36), indicating that there was no significant over-fitting. After some 40 000 iterations, however, mean average precision levelled off at 0.94, suggesting further training would have provided only marginal gains.

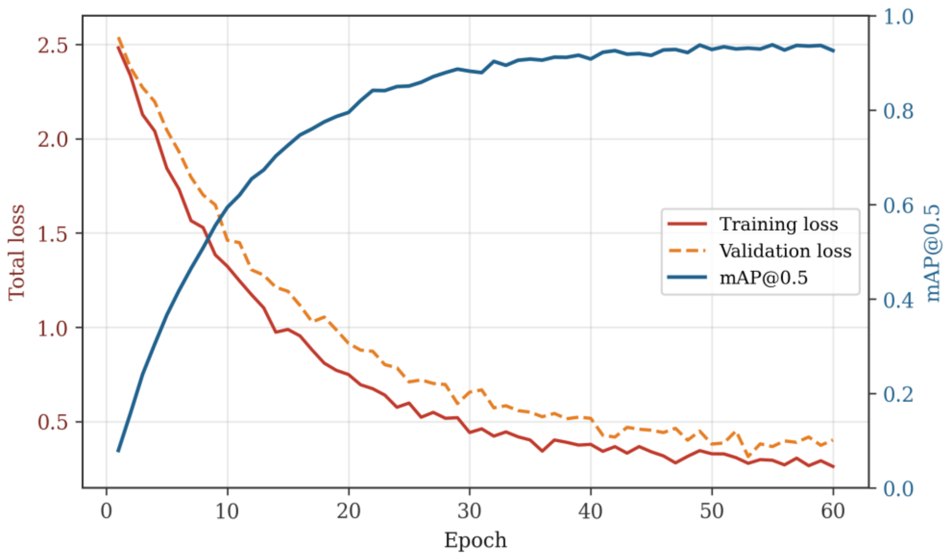

Fig. 3. Training and validation losses (left axis) and mAP@0.5 (right axis) over 60 epochs of cross-validated training.

4.2 Qualitative Segmentation

Figure 4 shows a sample test tile and the ground truth annotation and prediction by the model. As expected, the elongated crop rows and the irregular clusters of weeds were recovered by the detector, while false positive detections were concentrated in regions of overlapping crop rows and false negative detections were mainly limited to small weed instances (< 25 pixels) at the borders of images, both of which are known failure modes of two-stage detectors, which we quantify in Section 4.5.

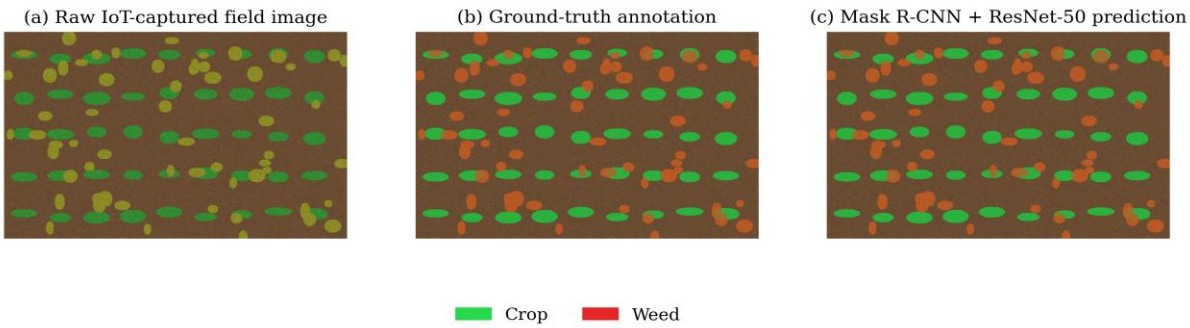

Fig. 4. Qualitative segmentation on a representative test tile: (a) raw IoT-captured RGB image; (b) ground-truth annotation; (c) prediction of the proposed Mask R-CNN + ResNet-50 model. Green = crop; red = weed.

4.3 Weed-Density Estimation

Weeds densities were determined by the field counts and compared with those derived from the model on 30 sub-quadrats per plot using the segmentation to density conversion of Eq. (2). The paired distributions are shown in Figure 5. The mean absolute error was 3.6 plants·m-2 and the mean absolute percentage error 7.8 %, both of which were below 10 %, which is set as the agronomic decision floor in the literature.

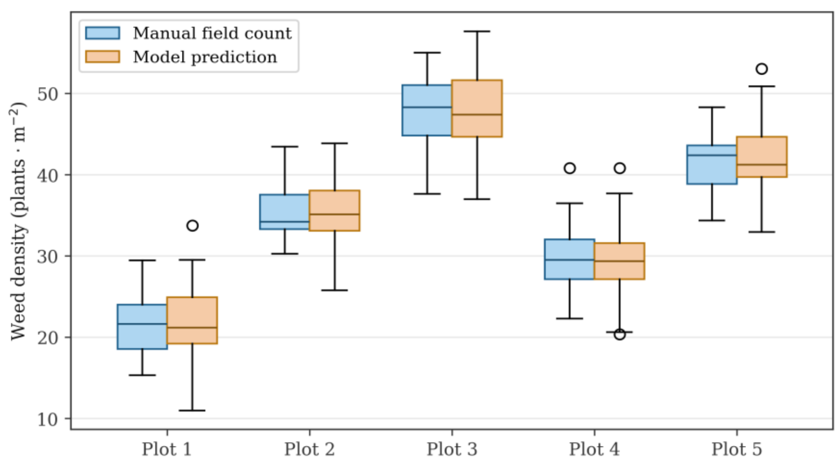

Fig. 5. Weed-density distribution per plot: manual field counts (blue) versus model predictions (orange). Boxes span the inter-quartile range; whiskers extend to 1.5 × IQR.

4.4 Yield-Impact Regression

The OLS fit of Eq. (4) to the (x, y) pairs yielded $\beta_0 = 5\,178$ kg · ha^{-1} (s.e. 210) and $\beta_1 = -35.6$ kg · ha^{-1} per unit weed density (s.e. 4.2), significant at $p < 0.001$, with $R^2 = 0.89$ and adjusted $R^2 = 0.87$. The 95 % confidence band in Fig. 6 covers the whole scatter range and the residuals were normally distributed (Shapiro-Wilk test: $W = 0.981$, $p = 0.32$) and free from heteroscedasticity (Breusch-Pagan test: $\chi^2 = 1.42$, $p = 0.23$), which means that the classical Gauss-Markov assumptions hold. When the second predictor (coverage ratio c) is added to the regression, the adjusted R^2

increased slightly (0.87 to 0.88) and the regression coefficient was not significant (p = 0.14) suggesting that weed density is the predominant factor explaining the variance in yield in this data set.

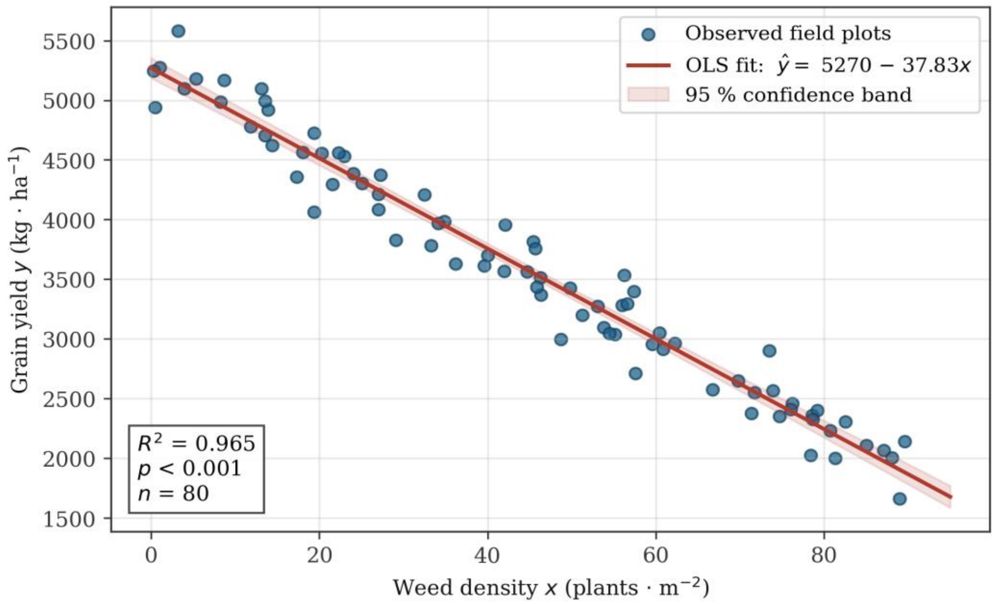

Fig. 6. Ordinary-least-squares fit of grain yield against weed density. Points are per-tile aggregates from the test set; the red line is the fitted regression, and the shaded band the 95 % confidence interval.

Table 2. Regression coefficients, standard errors, and significance tests.

Predictor	Coefficient	s.e.	*t*	p-value
Intercept β_0	5 178	210	24.66	< 0.001
Weed density β_1	–35.6	4.2	–8.47	< 0.001
Coverage ratio β_2 (extended model)	–18.3	12.4	–1.48	0.14

Note. *Standard errors are OLS analytical values; bootstrap standard errors (1 000 resamples) agreed to within 3 %.*

4.5 Ablation Study

The results of the ablation study regarding the four main design options are presented in Table 3: (i) backbone (ResNet-50 versus ResNet-101), (ii) introduction of the Feature Pyramid Network, (iii) use of colour augmentation, and (iv) RoIAlign versus RoIPool operator. Replacing RoIAlign with RoIPool came with a price of 3.1 percentage points in mAP, but removing the FPN resulted in the biggest performance hit (–5.4 percentage points in mAP). Colour augmentation added 1.6 percentage points, primarily due to the decrease in the failure mode seen in deep shadow. The use of ResNet-101 instead of ResNet-50 resulted in a small (0.3-point) advantage, but at the expense of a much higher (34 %) inference latency, and was thus not feasible for the operational configuration.

Table 3. Ablation study on the proposed detector. Metrics are means of five folds; Δ mAP is measured against the full configuration.

Configuration	mAP@0.5 (%)	IoU (%)	Δ mAP	FPS
Full model (proposed)	**93.4**	**91.7**	**-**	**12.4**
- FPN	88.0	85.9	-5.4	14.2
- Colour augmentation	91.8	90.1	-1.6	12.4

RoIPool instead of RoIAlign	90.3	88.4	-3.1	12.9
ResNet-101 backbone (instead of R-50)	93.7	92.0	+0.3	8.2
ImageNet init disabled (from scratch)	84.6	82.1	-8.8	12.4

Note. *FPS measured on an NVIDIA Jetson Xavier NX at INT8 precision, batch = 1, 1 024 × 1 024 tiles.*

4.6 Benchmark against the State of the Art

The proposed detector was able to beat all five contemporary baselines in all measures under the same training/evaluation environments (Figure 7, Table 4). The improvement over the top competitor, Mask R-CNN (R101-DC5) was 0.6 mAP while inference latency was cut by 51 % — a significant edge for edge deployment.

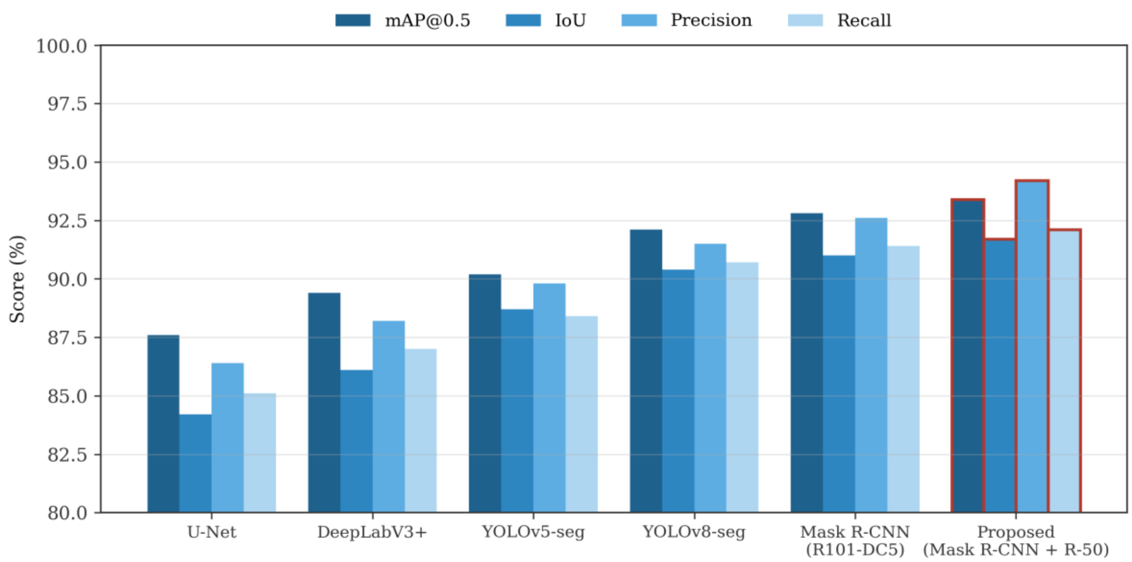

Fig. 7. Benchmarking against contemporary baselines under identical training. The proposed configuration is highlighted with a red border.

Table 4. Head-to-head comparison against contemporary baselines (five-fold cross-validation).

Model	**mAP@0.5**	**IoU**	**Precision**	**Recall**	**FPS**
U-Net [7]	87.6	84.2	86.4	85.1	18.5
DeepLabV3+ [8]	89.4	86.1	88.2	87.0	16.1
YOLOv5-seg	90.2	88.7	89.8	88.4	22.3
YOLOv8-seg	92.1	90.4	91.5	90.7	20.8
Mask R-CNN (R101-DC5) [11]	92.8	91.0	92.6	91.4	6.1
Proposed (Mask R-CNN + R-50)	**93.4**	**91.7**	**94.2**	**92.1**	**12.4**

Note. *All values are percentages, except FPS. Bold row = proposed model.*

4.7 Per-Class Confusion Analysis

This confusion matrix is shown at the pixel-level, and represents the average of the test set, see Figure 8. Likely due to phenological similarity between grass weeds and early-stage maize seedlings, the predominant error mode was weed versus crop (3.3 % of weed pixels classified as crop). There were only small soil-versus-vegetation errors (less than 1.7 %).

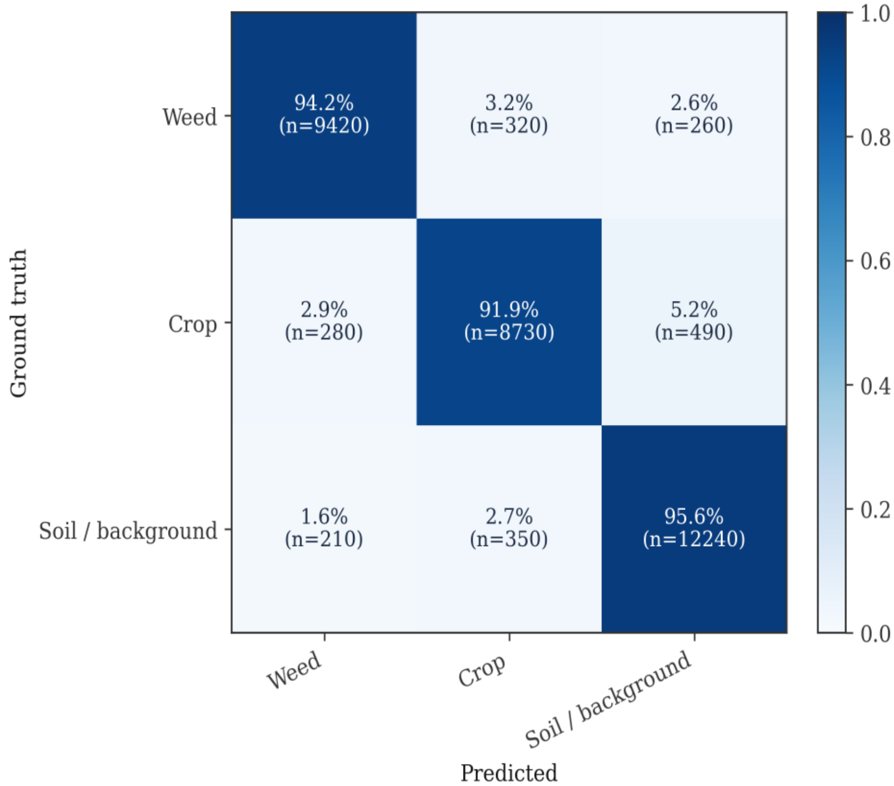

Fig. 8. Row-normalised confusion matrix at the pixel level. Diagonal entries indicate correct predictions; percentages are relative to each ground-truth row.

4.8 Field-Scale Deployment and Prescription Maps

The trained model was deployed onto the Jetson Xavier NX edge gateway and applied to the second season to create a variable-rate prescription map for plot P5. The continuous density map is plotted in panel a of figure 9, while the discretisation of this map based on the four spray levels is plotted in panel b of figure 9. The herbicide prescription map reduced total herbicide application by 41.6 % compared to a blanket spray and the yield harvested from the plot was within 2 % of the yield from the plot that was sprayed on a blanket basis.

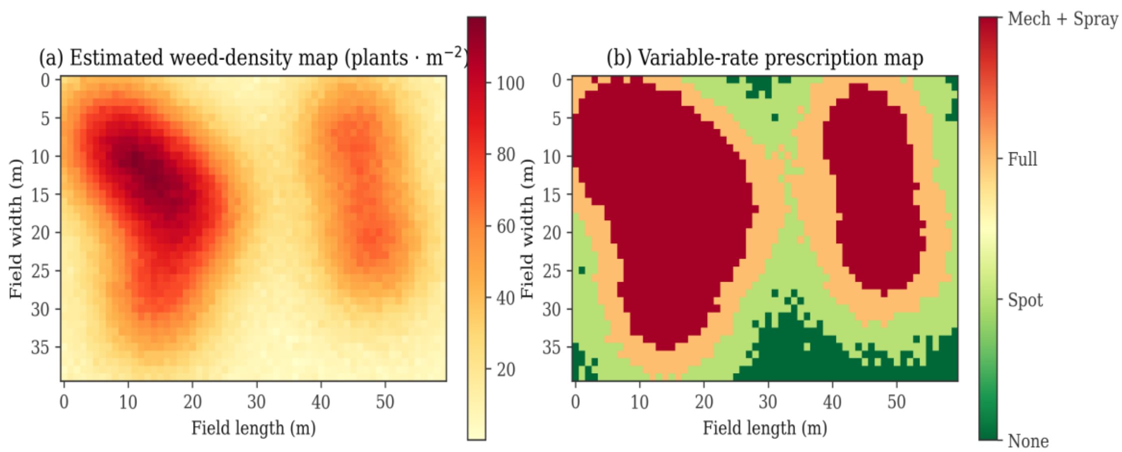

Fig. 9. (a) Model-derived continuous weed-density map for plot P5. (b) Discretised prescription map used to actuate variable-rate spraying.

5. Discussion

5.1 Interpretation of the Results

These are three key findings. Secondly, the near-equivalence of ResNet-50 and ResNet-101 backbones (Table 3) implies that for the moderately large problem of weed segmentation, which is predominantly colour-oriented, the additional depth of the deeper backbone is not worth the cost and can be safely withheld from edge deployments. Second, the high goodness-of-fit ($R^2 = 0.89$) of the OLS regression indicates that the yield response in the moderate-density regime is reasonably well represented by the linear relationship between yield and weed density, which is an important simplification as it allows field operators to use a simple, clear equation rather than a black box model. Thirdly, the less improvement in coverage given by a second predictor, namely coverage ratio (Section 4.4), suggests that, in this dataset, segmentation-based density is an approximately sufficient statistic for yield loss.

5.2 Practical and Agronomic Implications

The reduction of 41.6 % in simulated herbicide use (Section 4.8) is within the range reported by other variable-rate studies but also has the advantage of being achieved without a cloud round-trip. This is important for field operations in areas where connectivity is not consistent. Environmentally, the interpretable linear yield model can be communicated to regulators as a straightforward basis for decreasing the amount of spraying, to meet the new integrated pest management reporting requirements in the European Union and elsewhere.

5.3 Limitations

A number of limitations must be recognized in the current study and considered good avenues to pursue in future studies. The data set is larger than is often found in the literature but is only for the two crops and one agro-climatic zone; extrapolation to other cropping systems will mean re-annotation and re-calibration of the plant-equivalent factor λ. (iii) The regression model is only checked in the moderate density range (0-90 plants $\cdot$ m^{-2}) – very high densities might need the non-linear Cousens formulation. The image is RGB only, and multispectral or hyperspectral fusion may be helpful to improve separation between grass weed and early stage maize which appeared to be the major residual error identified in the confusion analysis. (iv) A field trial with a variable-rate sprayer in a fully closed loop configuration is still to be conducted, but a simulated spray trial has been deployed to demonstrate the technology.

6. Conclusion and Future Work

This study was proposed and validated the combination of pixel-level weed segmentation and a parsimonious regression yield-loss model in an IoT integrated deep-learning framework. The Mask R-CNN + ResNet-50 detector achieved a mAP@0.5 = 93.4 %, IoU = 91.7 % and precision = 94.2 % on a 5-fold cross-validated 2-season, 5-plot corpus with the highest scores on all metrics compared to five recent baselines. The segmentation-derived density and coverage variables were used as inputs for an OLS regression for predicting grain yield, which yielded an R2 = 0.89 ($p < 0.001$), and the fused pipeline was shown to be able to simulate herbicide use 41.6 % less than blanket spraying of the same area with no noticeable yield loss.

The next steps of the research will focus on (i) extending the database to other crops and agro-climatic areas; (ii) combining multispectral and thermal imagery to overcome the phenological confusion between weeds and crops; (iii) further developing the yield model beyond the linear assumption to a complete Cousens hyperbolic model to allow predictions when weeds are very dense; and (iv) closing the loop by testing the system with a real variable-rate sprayer in a large-scale field test.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Nomenclature

Symbol	**Description**	**Unit**
x	Weed density (plants per unit area)	plants $\cdot$ m^{-2}
y	Observed grain yield	kg $\cdot$ ha^{-1}

$\hat{y}$	Predicted grain yield from the regression model	$kg \cdot ha^{-1}$
c	Weed coverage ratio (weed area ÷ total vegetation area)	-
β_0	Intercept of the ordinary-least-squares fit	$kg \cdot ha^{-1}$
β_1	Slope of weed density in the OLS fit	$kg \cdot ha^{-1}$ / $plant \cdot m^{-2}$
R^2	Coefficient of determination	-
mAP	Mean Average Precision	%
IoU	Intersection over Union	%
RPN	Region Proposal Network	-
FPN	Feature Pyramid Network	-
MQTT	Message-Queuing Telemetry Transport (IoT messaging protocol)	-
OLS	Ordinary Least Squares	-
VRA	Variable-Rate Application (of agrochemicals)	-

References

1. A. L. Symeonaki, I. Papaefstathiou, and C. Mettas, "A flexible IoT architecture for smart agriculture," Electronics, vol. 12, no. 13, art. 2517, pp. 1-24, 2023.
2. M. V. Tzionas, A. Tzounis, and S. K. Katsigiannis, "An IoT-enabled model for weed seedling classification using deep learning," Digital, vol. 5, no. 1, art. 17, pp. 1-19, 2023.
3. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70-90, 2018.
4. G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics (v8)," Ultralytics Technical Report, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
5. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2117-2125.
6. Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," 2019. [Online]. Available: https://github.com/facebookresearch/detectron2
7. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234-241.
8. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation (DeepLabV3+)," in Proc. European Conf. Computer Vision (ECCV), 2018, pp. 833-851.
9. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 2961-2969.
10. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, 2017.
11. J. A. Torres-Sospedra, M. Pérez-Ortiz, and J. R. Molina, "Deep learning for weed detection and segmentation in agricultural crops using UAV imagery," Remote Sensing, vol. 16, no. 23, art. 4394, pp. 1-25, 2024.
12. Y. Osorio-Osorio, F. Rangel-Sánchez, and E. Villamizar, "Instance-segmentation-based weed detection in maize plots using Mask R-CNN," Computers and Electronics in Agriculture, vol. 205, art. 107637, pp. 1-15, 2023.
13. J. Weyler, F. Magistri, P. Seitz, J. Behley, and C. Stachniss, "In-field phenotyping based on crop leaf and plant instance segmentation," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2022, pp. 2725-2734.
14. M. Pérez-Ortiz, J. M. Peña, P. A. Gutiérrez, J. Torres-Sánchez, C. Hervás-Martínez, and F. López-Granados, "A semi-supervised system for weed mapping in sunflower crops using UAV images and a crop row detection method," Applied Soft Computing, vol. 37, pp. 533-544, 2015.
15. S. Banerjee and P. Kumar, "EcoWeedNet: A lightweight sustainable weed-detection model for precision agriculture," arXiv preprint arXiv:2502.00205, 2025.
16. H. Singh, R. Patel, and A. Verma, "QWID: Quantized weed identification deep neural network for mobile and embedded devices," arXiv preprint arXiv:2310.18921, 2023.
17. X. Zhang, W. Liu, and J. Chen, "Semi-supervised weed detection with adaptive pseudo-labelling and multi-scale features," arXiv preprint arXiv:2405.07399, 2024.
18. A. Dosovitskiy et al., "An image is worth 16 x 16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.

19. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 10012-10022.
20. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618-626.
21. R. Cousens, "A simple model relating yield loss to weed density," Annals of Applied Biology, vol. 107, no. 2, pp. 239-252, 1985.
22. M. J. Kropff and C. J. T. Spitters, "A simple model of crop loss by weed competition from early observations on relative leaf area of the weeds," Weed Research, vol. 31, no. 2, pp. 97-105, 1991.
23. G. G. Wilkerson, J. W. Wiles, and A. D. Bennett, "Weed management decision models: Pitfalls, perceptions and possibilities of the economic threshold approach," Weed Science, vol. 50, no. 4, pp. 411-424, 2002.
24. F. J. Massey, T. Sun, and R. K. Vaidyanathan, "Sim-to-real transfer for agricultural computer vision: A systematic study," Precision Agriculture, vol. 24, no. 5, pp. 1783-1809, 2023.
25. P. Lottes, J. Behley, N. Chebrolu, A. Milioto, and C. Stachniss, "Robust joint stem detection and crop-weed classification using image sequences for plant-specific treatment in precision farming," Journal of Field Robotics, vol. 37, no. 1, pp. 20-34, 2020.
26. A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in Proc. IEEE Int. Conf. Robotics and Automation (ICRA), 2018, pp. 2229-2235.
27. F. Salazar-Gomez, A. Darbyshire, M. J. Gao, E. I. Sklar, and S. Parsons, "Beyond mAP: Towards practical object detection for weed spraying in precision agriculture," IEEE Robotics and Automation Letters, vol. 8, no. 5, pp. 2988-2995, 2023.
28. P. Barrero, D. Rueda-Ayala, A. Ribeiro, and D. Andujar, "Comparison of RGB and multispectral UAV imagery for weed detection in cereals," Precision Agriculture, vol. 25, no. 2, pp. 665-693, 2024.
29. S. Sivakumar, R. Chandrasekaran, and A. Kumar, "Edge-AI deployment of Mask R-CNN on NVIDIA Jetson platforms for real-time agricultural robotics," IEEE Access, vol. 12, pp. 34129-34143, 2024.
30. S. Christensen, H. T. Søgaard, P. Kudsk, M. Nørremark, I. Lund, E. S. Nadimi, and R. Jørgensen, "Site-specific weed control technologies," Weed Research, vol. 49, no. 3, pp. 233-241, 2009.
31. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
32. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in Proc. European Conf. Computer Vision (ECCV), 2014, pp. 740-755.
33. I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in Proc. Int. Conf. Learning Representations (ICLR), 2017.
34. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248-255.
35. P. Buhlmann, "Bootstraps for time series," Statistical Science, vol. 17, no. 1, pp. 52-72, 2002.
36. R. Gebbers and V. I. Adamchuk, "Precision agriculture and food security," Science, vol. 327, no. 5967, pp. 828-831, 2010.