

What Improves UAV RGB Segmentation of Maize and Weeds? An Ablation Study of U-Net Components

Massimo Martinelli
0000-0001-7419-5099
Institute of Information Science and
Technologies, National Research
Council, Pisa, Italy
Email: massimo.martinelli@isti.cnr.it

Leonardo Ercolini, Nicola Grossi, Nicola Silvestri,
0000-0002-7463-7632, 0000-0002-6997-1599,
0000-0003-3430-8911
Department of Agricultural, Food and Agro-Environmental Sciences,
University of Pisa, Italy
Email: {name.surname@unipi.it}

Andrea Berton
0000-0002-8798-9469
Institute of Geosciences and Earth
Resources, National Research
Council, Pisa, Italy
Email: andrea.berton@cnr.it

Davide Moroni
0000-0002-5175-5126
Institute of Information Science and Technologies, National
Research Council, Pisa, Italy
Email: davide.moroni@isti.cnr.it

Abstract—The use of RGB images acquired by unmanned aerial vehicles (UAVs) is increasingly common in precision agriculture, for monitoring crops, weeds, and field conditions. However, large variations in plant size, illumination, shadows, and background appearance make automatic segmentation difficult. A major challenge is discriminating crop plants from weeds that may compete for resources and identifying conditions that may favor the development of plant diseases. In this study, we evaluated U-Net-based semantic segmentation models on 142 UAV RGB images from two maize fields using grouped 5-fold cross-validation to segment maize, weeds, and bare soil. The full baseline and seven one-component ablations were each evaluated with three output modes, yielding 24 run-level configurations and 120 fold-level evaluations. Across the five folds, `no_deep_supervision` with `pure_dl` achieved the highest mean F1 (0.7680 ± 0.0250), `mIoU_no_bg` (0.6436 ± 0.0267), aggregate score (0.9593 ± 0.0245), and weed IoU (0.4657 ± 0.0531), while also yielding the lowest mean HD95 (60.5588 ± 12.6210).

ExG-guided refinement remained close to `pure_dl` but did not improve mean performance consistently, whereas hard-threshold reconstruction sharply degraded both overlap and distance metrics. The highest individual fold-level weed IoU was 0.5222 for `no_sigreg` with `pure_dl` in fold 2; however, `no_deep_supervision` with `pure_dl` remained the best configuration when performance was averaged across all folds. Annotation examples also showed that isolated weed-label outliers near image borders can disproportionately increase distance-based metrics even when the predicted mask is visually plausible. Overall, the results favor a simpler pipeline for this dataset, direct multiclass inference, and cautious interpretation of weed HD95. The study does not propose a new architecture; instead, it provides a controlled ablation protocol and practical evidence that additional U-Net components should be validated rather than assumed beneficial in small-data UAV RGB segmentation.

Index Terms—image segmentation, U-net, maize, weeds, UAV RGB imagery, precision agriculture, boundary accuracy, annotation quality.

I. INTRODUCTION

WEED control is one of the most challenging problems in precision agriculture, especially in the early stages, when crops are most vulnerable to competition. Using drones to take RGB images is now common, but the real challenge remains in image analysis: it's difficult to distinguish maize from weeds because they often have a similar color and shape, moreover shadows and camera angles often obscure their true contours. Recent studies using complex modular systems with attention blocks or specialized loss functions [1]–[11] often require large datasets to achieve good results and can be ineffective for smaller datasets. In this work, we analyze a modular system component by component for segmenting a dataset to understand which components improve performance by comparing performance on maize and weeds separately and assessing how annotation quality affects the interpretation of distance-based metrics. U-Net remains one of the most widely used architectures for semantic segmentation because it combines the contextual information and the fine spatial detail [1]. Several studies have been proposed to improve performance, including attention mechanism on skip connections, deep supervision of intermediate decoder features, boundary-aware losses, and regularization terms aimed at producing smoother masks [2][3]. These components can be beneficial in some settings, especially for small or highly fragmented objects to be identified. but, in small datasets are not guaranteed. Transformer-based segmentation models are becoming increasingly common [4–6], although recent studies show that U-Net-based systems remain competitive in agricultural image analysis. However, in-field performance may degrade because acquisition conditions change with seasons, locations, or lighting conditions [11]. Many are the works adopting additional regularization terms trying to obtain smoother masks or more plausible class boundaries: among these, formulations based on the phase field and partial differential equations are useful

because they allow more direct control over fragmented or unstable interfaces [12-13]. These observations motivate a careful evaluation of model components under constrained data conditions. Instead of assuming that complex architectures are always needed, we investigate which specific elements of a U-Net-based pipeline actually improve segmentation on a small UAV RGB dataset depicting maize crops and co-occurring weeds. This paper makes three key contributions: (1) it presents a controlled ablation study of architectural, loss-function, and inference components on a comprehensive baseline; (2) it isolates and compares the effects of these components on maize versus weed segmentation; and (3) it analyzes how annotation quality impacts the reliability of distance-based evaluation metrics (e.g., HD95) for weed detection.

II. MATERIALS AND METHODS

A. Dataset and Annotation

We conducted a UAV survey of two maize fields during the pre- and post-emergence stages, resulting in a dataset of 142 RGB images. Three annotators annotated the images at the pixel level for three classes: maize, weeds, and bare soil. To reduce subjectivity, each annotator produced segmentation masks independently, and the final reference mask was obtained via pixel-level majority voting. Image annotation was particularly challenging because of the difference in plant size between the two stages, ill-defined weed boundaries compared to uniform maize leaves, the uneven terrain, and the occlusions caused by overlapping foliage and shadows.

B. Model Architecture and Training

The baseline model was a three-class U-Net with a ResNet-101 encoder pretrained on ImageNet. We implemented the code using PyTorch, then we ran the experiments on a workstation equipped with two NVIDIA RTX 5050 GPUs. We have considered two architectural variants: an Attention U-Net with gated skip connections and a standard U-Net without attention. Deep supervision was implemented via three auxiliary decoder heads (auxiliary outputs) with weights of 0.20, 0.10, and 0.05.

RGB images were resized to 512×512 . For optimization, we used AdamW with cosine annealing (learning rate 1×10^{-4} and a weight decay of 1×10^{-4}). Training ran for up to 70 epochs with a batch size of 4 and an early stopping patience of 12 epochs. Data augmentation included horizontal and vertical flips, shift/scale/rotation, brightness and contrast changes, hue/saturation variation, color jitter, blur, Gaussian noise, and gamma augmentation. Model selection and early stopping were based exclusively on the validation performance of the unprocessed network predictions.

C. Loss Function and Regularization

The main segmentation loss combined the Dice, the weighted focal, and the boundary ones, defined as:

$$L_{seg} = w_1 L_{Dice} + w_2 * L_{Focal} + w_3 * L_{Boundary}$$

where

L_{seg} : Resulting total segmentation loss;
 L_{Dice} : Term related to Dice Loss (overlap coefficient);
 L_{Focal} : Term related to Focal Loss (handling difficult sampling);

$L_{Boundary}$: Term related to Boundary Loss (edge precision) and $w_1 = 0.60$, $w_2 = 0.30$, $w_3 = 0.15$

These coefficients were used as relative weighting factors and were not constrained to sum to one. The focal term used $\gamma = 2.0$ and class weights (1.0, 1.0, 4.5), thus increasing the weight of the weed class. The boundary loss compares the gradient magnitudes (amplitudes) of the softmax probability maps and the one-hot encoded target masks; we used this strategy based on edge sharpness and shape consistency to reward clear plant delineation and improve the accuracy of the model in recognizing maize. This process was complemented by an additional mathematical rule, inspired by the physical laws of thermodynamics, known as the Ginzburg-Landau / Allen-Cahn energy. This rule, applied with $\epsilon = 0.5$ only to the maize class within a boundary band computed from the ground-truth mask, prevents the computer from generating random spots or excessively irregular shapes, ensuring that the identified area is compact and realistic. The influence of this physics-inspired term was introduced gradually: after an initial 8-epoch warm-up phase, it underwent a 12-epoch ramp-up until it reached its maximum strength of $\lambda_{max} = 10^{-3}$. This warm-up period allowed the network weights to stabilize to comply with the more stringent geometric constraints. The full baseline also included SIGReg latent-space regularization; the no_sigreg ablation disabled only this term while leaving the remaining training and inference settings unchanged.

D. Validation and Ablation Study

Starting from a full baseline containing all components, one component was removed at a time to assess its contribution. The seven ablations were: no attention, no deep supervision, no weighted focal term, no boundary term, no physics-inspired term, no SIGReg latent-space regularization, and no test-time augmentation. Together with the full baseline, eight training configurations were evaluated with three output modes, producing 24 run-level configurations.

Each configuration was assessed over five grouped folds, for a total of 120 fold-level evaluations.

To reduce leakage between correlated images, we have used a strictly grouped 5-fold cross-validation protocol using GroupKFold, where groups were derived from filename identifiers encoding acquisition context, ensuring that images from the same acquisition group (local scene) were not split across training and validation folds.

During inference, test-time augmentation (TTA) was optionally applied via horizontal inversion. Three output modes were compared:

1. pure_dl: standard multiclass inference obtained by taking the argmax of the network logits;

2. refinement: a soft boost of the weed logit on residual background pixels using the ExG percentile rule ($\text{exg_threshold} = 8$, $\text{exg_percentile} = 70$, and $\text{refinement_alpha} = 1.0$);

3. hard_threshold: reconstruction of the weed class only on residual background pixels using the same ExG percentile settings.

III. EVALUATION METRICS

We used overlap-based and distance-based metrics: precision, recall, F1, mean IoU without background, mean Dice without background, boundary F1 (with a tolerance equal to 2% of the image diagonal), and HD95. Maize and weed performance were analyzed separately because the components can affect the two classes differently. For compact comparison, we also used the aggregate score shown below. The score uses a coefficient of 1.0 for mIoU_no_bg ; the main conclusions were also checked against the individual metrics.

$$\text{Score} = \text{mIoU_no_bg} + 0.5 \text{BF1} - 0.002 \text{mHD95}$$

IV. RESULTS

Relative to the full baseline, removing deep supervision improved pure_dl mIoU_no_bg from 0.6381 ± 0.0237 to

0.6436 ± 0.0267 , weed IoU from 0.4570 ± 0.0461 to 0.4657 ± 0.0531 , and aggregate score from 0.9435 ± 0.0301 to 0.9593 ± 0.0245 . The same configuration also produced the lowest mean HD95, decreasing it from 64.6014 ± 20.8353 to 60.5588 ± 12.6210 . The maximum individual weed IoU, 0.5222, occurred for no_sigreg with pure_dl in fold 2, but its cross-fold mean was 0.4576 ± 0.0414 and therefore did not exceed the no_deep_supervision mean. Refinement remained very close to pure_dl and was usually slightly worse, whereas every hard_threshold variant produced a large loss in weed overlap and a marked increase in HD95. Figure 1 summarizes the aggregate score, weed-class IoU, and mean HD95.

The 24 run-level configurations can be summarized as follows:

- no_deep_supervision + pure_dl achieved the best cross-fold means for F1, mIoU_no_bg , score, class2_F1, and class2_IoU, and the lowest mHD95;
- the full baseline, no_tta, and no_focal formed the next group under pure_dl, with aggregate scores of 0.9435, 0.9357, and 0.9325, respectively;
- no_sigreg + pure_dl produced the highest single-fold class2_IoU (0.5222 in fold 2), but its mean score (0.9194 ± 0.0253) and mean class2_IoU (0.4576 ± 0.0414) were below those of no_deep_supervision;

TABLE I. QUANTITATIVE SUMMARY OF THE 24 RUN-LEVEL ABLATION CONFIGURATIONS.

All values are mean $\pm$ sample standard deviation across the five grouped folds. Higher is better except for mHD95, for which lower is better. The best mean value in each column is bolded.

Experiment	Mode	F1	mIoU_no_bg	Score	class2_F1	class2_IoU	mHD95
Baseline full	pure_dl	0.7634 $\pm$ 0.0219	0.6381 $\pm$ 0.0237	0.9435 $\pm$ 0.0301	0.6262 $\pm$ 0.0432	0.4570 $\pm$ 0.0461	64.6014 $\pm$ 20.8353
	refinement	0.7624 $\pm$ 0.0219	0.6371 $\pm$ 0.0237	0.9411 $\pm$ 0.0295	0.6243 $\pm$ 0.0433	0.4550 $\pm$ 0.0460	65.1223 $\pm$ 20.3869
	hard_threshold	0.4753 $\pm$ 0.0112	0.4224 $\pm$ 0.0086	0.5500 $\pm$ 0.0534	0.0500 $\pm$ 0.0182	0.0257 $\pm$ 0.0097	119.6765 $\pm$ 26.0586
No attention	pure_dl	0.7570 $\pm$ 0.0232	0.6305 $\pm$ 0.0258	0.9223 $\pm$ 0.0503	0.6161 $\pm$ 0.0407	0.4462 $\pm$ 0.0421	69.3409 $\pm$ 21.4556
	refinement	0.7556 $\pm$ 0.0239	0.6291 $\pm$ 0.0264	0.9174 $\pm$ 0.0489	0.6132 $\pm$ 0.0421	0.4432 $\pm$ 0.0435	70.7845 $\pm$ 19.7058
	hard_threshold	0.4748 $\pm$ 0.0062	0.4207 $\pm$ 0.0076	0.5780 $\pm$ 0.0418	0.0516 $\pm$ 0.0069	0.0265 $\pm$ 0.0036	103.5055 $\pm$ 15.8579
No deep supervision	pure_dl	0.7680 $\pm$ 0.0250	0.6436 $\pm$ 0.0267	0.9593 $\pm$ 0.0245	0.6340 $\pm$ 0.0499	0.4657 $\pm$ 0.0531	60.5588 $\pm$ 12.6210
	refinement	0.7672 $\pm$ 0.0256	0.6428 $\pm$ 0.0273	0.9574 $\pm$ 0.0262	0.6325 $\pm$ 0.0510	0.4642 $\pm$ 0.0542	61.0436 $\pm$ 13.2494
	hard_threshold	0.4781 $\pm$ 0.0075	0.4247 $\pm$ 0.0063	0.5388 $\pm$ 0.0892	0.0542 $\pm$ 0.0133	0.0279 $\pm$ 0.0070	126.6113 $\pm$ 42.2096
No focal	pure_dl	0.7623 $\pm$ 0.0200	0.6347 $\pm$ 0.0241	0.9325 $\pm$ 0.0638	0.6309 $\pm$ 0.0329	0.4615 $\pm$ 0.0351	67.3384 $\pm$ 15.5115
	refinement	0.7621 $\pm$ 0.0203	0.6345 $\pm$ 0.0244	0.9324 $\pm$ 0.0654	0.6306 $\pm$ 0.0332	0.4612 $\pm$ 0.0355	67.3305 $\pm$ 15.8355
	hard_threshold	0.4545 $\pm$ 0.0051	0.4078 $\pm$ 0.0096	0.2521 $\pm$ 0.2727	0.0154 $\pm$ 0.0074	0.0077 $\pm$ 0.0038	227.8918 $\pm$ 123.0894
No boundary	pure_dl	0.7487 $\pm$ 0.0254	0.6209 $\pm$ 0.0262	0.9166 $\pm$ 0.0318	0.6036 $\pm$ 0.0519	0.4338 $\pm$ 0.0540	66.3164 $\pm$ 14.5570
	refinement	0.7472 $\pm$ 0.0257	0.6194 $\pm$ 0.0264	0.9140 $\pm$ 0.0320	0.6006 $\pm$ 0.0524	0.4309 $\pm$ 0.0543	66.4767 $\pm$ 13.9417
	hard_threshold	0.4753 $\pm$ 0.0106	0.4186 $\pm$ 0.0070	0.5256 $\pm$ 0.0597	0.0568 $\pm$ 0.0194	0.0293 $\pm$ 0.0101	118.2201 $\pm$ 24.6862
No physics	pure_dl	0.7579 $\pm$ 0.0168	0.6312 $\pm$ 0.0182	0.9146 $\pm$ 0.0411	0.6182 $\pm$ 0.0316	0.4480 $\pm$ 0.0326	71.2147 $\pm$ 19.6802
	refinement	0.7567 $\pm$ 0.0175	0.6299 $\pm$ 0.0190	0.9062 $\pm$ 0.0358	0.6156 $\pm$ 0.0329	0.4453 $\pm$ 0.0338	74.2497 $\pm$ 19.0162
	hard_threshold	0.4712 $\pm$ 0.0046	0.4187 $\pm$ 0.0063	0.5372 $\pm$ 0.0578	0.0448 $\pm$ 0.0044	0.0229 $\pm$ 0.0023	119.7794 $\pm$ 28.1690
No SIGReg	pure_dl	0.7563 $\pm$ 0.0216	0.6263 $\pm$ 0.0262	0.9194 $\pm$ 0.0253	0.6270 $\pm$ 0.0385	0.4576 $\pm$ 0.0414	69.4164 $\pm$ 18.6243
	refinement	0.7553 $\pm$ 0.0217	0.6253 $\pm$ 0.0260	0.9188 $\pm$ 0.0228	0.6251 $\pm$ 0.0390	0.4556 $\pm$ 0.0419	69.1284 $\pm$ 18.3193
	hard_threshold	0.4689 $\pm$ 0.0102	0.4110 $\pm$ 0.0151	0.5152 $\pm$ 0.0854	0.0523 $\pm$ 0.0105	0.0269 $\pm$ 0.0055	129.5775 $\pm$ 44.4243
No TTA	pure_dl	0.7576 $\pm$ 0.0221	0.6311 $\pm$ 0.0229	0.9357 $\pm$ 0.0335	0.6177 $\pm$ 0.0441	0.4481 $\pm$ 0.0455	61.7184 $\pm$ 10.0526
	refinement	0.7569 $\pm$ 0.0215	0.6303 $\pm$ 0.0222	0.9330 $\pm$ 0.0307	0.6163 $\pm$ 0.0429	0.4465 $\pm$ 0.0442	62.4175 $\pm$ 10.5257
	hard_threshold	0.4726 $\pm$ 0.0072	0.4193 $\pm$ 0.0030	0.5605 $\pm$ 0.0412	0.0476 $\pm$ 0.0166	0.0244 $\pm$ 0.0087	108.2814 $\pm$ 15.5142

Note to Table I—The highest cross-fold mean class2_IoU was 0.4657 ± 0.0531 , obtained by no_deep_supervision + pure_dl. The highest individual fold value was 0.5222, obtained by no_sigreg + pure_dl in fold 2. Model selection is based on the cross-fold mean, not on the maximum value observed in a single fold.

- no_attention, no_boundary, and no_physics did not improve the cross-fold means relative to the full baseline;
- refinement closely tracked pure_dl across all training configurations but provided no consistent mean gain;
- hard_threshold was the weakest output mode in every ablation, with class2_IoU means between 0.0077 and 0.0293 and substantially larger mHD95.

In summary, no_deep_supervision with pure_dl was the strongest cross-validated configuration. The isolated class2_IoU value above 0.52 belongs to one no_sigreg fold and should not be interpreted as the best model-level mean.

Figure 2 shows an annotation-quality issue in which isolated weed (Class 2) pixels appear near the image border in the reference mask, whereas the model does not predict weed at that location. The raw image does not support a plausible weed region at those pixels, so the model is penalized for not reproducing a likely annotation artifact. Because HD95 is distance based, a small external component can create a large error even when the predicted segmentation is visually plausible. Weed distance metrics should therefore be interpreted together with overlap scores and visual inspection.

Figure 3 shows another uncertain annotation: some pixels were labeled as maize even though they lie outside the crop row. The model suggests that they may be weeds, but the RGB resolution is insufficient to determine the class with certainty. This example illustrates the residual ambiguity that remains even after multi-annotator majority voting.

V. DISCUSSION

From an agronomic perspective, an important factor to be considered when interpreting these results is the biological and spatial variability of weed flora under real field conditions.

Weed species differ substantially in morphology, spectral response, and growth pattern, which helps explain the fold-to-fold variability of the weed metrics. The 24-configuration ablation study indicates that deep supervision was the only removal that improved all reported cross-fold means simultaneously under pure_dl. Its advantage was not caused by post-processing, because pure_dl was also the best output mode for that training configuration. The fold-level class2_IoU maximum of 0.5222 for no_sigreg is useful evidence that the model can exceed 0.52 on a favorable split, but the cross-fold mean and variability remain the appropriate basis for model selection.

The remaining ablations do not support a claim that removing attention, the boundary term, the physics-inspired term, or SIGReg improves the mean model performance. In the present runs, the full baseline exceeded no_attention, no_boundary, no_physics, and no_sigreg in aggregate score under pure_dl. Removing the weighted focal term preserved competitive mean overlap, but it increased mean HD95 relative to the best configuration and produced the most unstable hard-threshold behavior; therefore, its role appears to be more related to robustness than to the highest mean overlap.

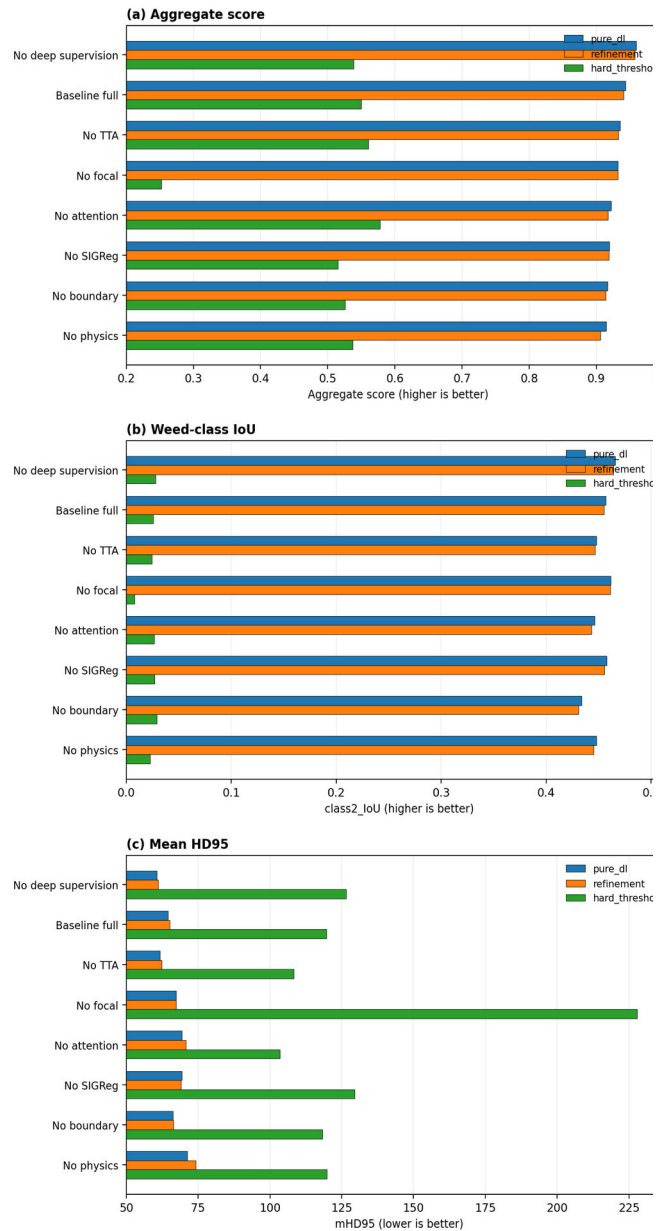


Fig. 1. Mean ablation performance across five grouped folds: (a) aggregate score, (b) weed-class IoU, and (c) mean HD95. No_deep_supervision with pure_dl gives the best mean score and weed IoU and the lowest mean HD95. The bars report fold means; the maximum individual class2_IoU was 0.5222 for no_sigreg with pure_dl in fold 2. Hard_threshold consistently underperforms.

Removing TTA slightly reduced overlap and aggregate score relative to the full baseline, while producing a somewhat lower mean HD95, indicating a mixed and limited inference-time effect. ExG-guided refinement remained almost identical to pure_dl and did not justify additional inference complexity. By contrast, hard-threshold weed reconstruction was consistently unreliable because it converted residual background into many false weed regions. Considering accuracy, stability, and inference simplicity together, no_deep_supervision with pure_dl provides the best balance for this dataset. Future work should evaluate annotation-ro-

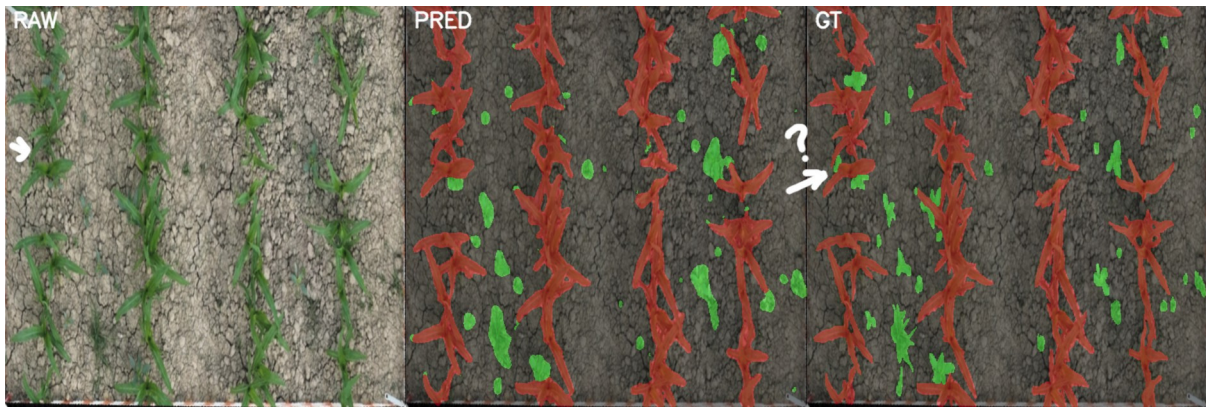


Fig. 2. Annotation-related weed outliers near the image border; isolated labeled pixels can disproportionately affect distance-based metrics.

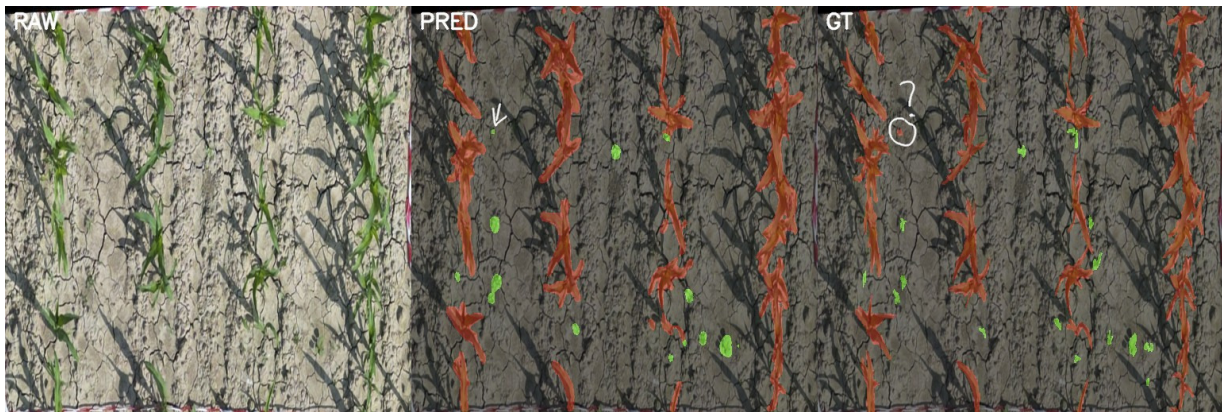


Fig. 3. Uncertain annotation outside the crop row; RGB evidence is insufficient to determine the class with certainty.

bust distance measures and repeat the protocol on larger, multi-site datasets.

Limitations. The findings should be interpreted as design evidence for a limited UAV RGB setting rather than as a universal ranking of U-Net components. The dataset contains 142 images from two maize fields and does not include external validation across additional farms, seasons, sensors, flight altitudes, or weed communities. Some differences among non-thresholded configurations are small relative to fold-level variability, and an isolated maximum from one fold must not be confused with the cross-fold mean. Annotation uncertainty affects both training and evaluation, especially for small weed regions and boundary-based metrics such as HD95. Future work should validate the same ablation protocol on larger multi-site datasets, include stronger contemporary baselines, report parameter counts and inference costs, and evaluate annotation-robust metrics.

VI. CONCLUSION

In this study, we investigated which components of a U-Net-based segmentation pipeline are useful for a small UAV RGB dataset of maize and weed field images. No_deep_supervision with pure_dl was the best cross-validated configuration, achieving the highest mean overlap and aggregate metrics and the lowest mean HD95. Although no_sigreg

with pure_dl reached a weed IoU of 0.5222 in one fold, its cross-fold mean was lower and it was not the best model-level configuration. ExG-guided refinement added little, while hard-threshold weed reconstruction consistently degraded performance. The annotation examples further show that weed HD95 can be dominated by isolated label outliers. These findings support a dataset-specific preference for a simpler, directly inferred model and for evaluation based on cross-fold means, overlap metrics, visual audit, and annotation-robust distance measures.

USE OF AI-BASED WRITING TOOLS

The authors used LLM-based assistance and Grammarly only for language polishing, grammar checking, typo correction, and improving the readability of the manuscript. No AI tool was used to generate the scientific ideas, design the experiments, analyze the data, or draw the conclusions. All AI- and Grammarly-assisted suggestions were reviewed, verified, and edited by the authors, who take full responsibility for the final content.

REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, 2015, pp. 234–241.

- [2] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning Where to Look for the Pancreas," arXiv preprint arXiv:1804.03999, 2018.
- [3] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised Nets," in Proc. 18th Int. Conf. Artificial Intelligence and Statistics (AISTATS), vol. 38, 2015, pp. 562–570.
- [4] Z. Luo, W. Yang, Y. Yuan, R. Gou, X. Li, "Semantic segmentation of agricultural images: A survey", *Information Processing in Agriculture*, Volume 11, Issue 2, 2024, Pages 172-186, ISSN 2214-3173, doi: 10.1016/j.inpa.2023.02.001.
- [5] L. Sandoval-Pillajo, I. García-Santillán, M. Puskás-Chulde, and A. Giret, "Weed detection based on deep learning from UAV imagery: A review," *Smart Agricultural Technology*, vol. 12, 2025, Art. no. 101147. doi: 10.1016/j.atech.2025.101147.
- [6] B. Deng, Y. Lu, and J. Xu, "Weed database development: An updated survey of public weed datasets and cross-season weed detection adaptation," *Ecological Informatics*, vol. 81, Jul. 2024, Art. no. 102546, doi: 10.1016/j.ecoinf.2024.102546.
- [7] J. Cui, F. Tan, N. Bai, and Y. Fu, "Improving U-net network for semantic segmentation of corns and weeds during corn seedling stage in field" *Frontiers in Plant Science*, vol. 15, 2024, Art. no. 1344958. doi: 10.3389/fpls.2024.1344958.
- [8] E. Celikkan, T. Kunzmann, Y. Yeskaliyev, S. Itzerott, N. Klein, and M. Herold, "WeedsGalore: A Multispectral and Multitemporal UAV-Based Dataset for Crop and Weed Segmentation in Agricultural Maize Fields" in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2025, pp. 4767–4777.
- [9] H. M. Sahin, T. Miftahushudur, B. Grieve, H. Yin "Segmentation of weeds and crops using multispectral imaging and CRF-enhanced U-Net" *Computers and Electronics in Agriculture*, Volume 211, 2023, 107956, ISSN 0168-1699, 10.1016/j.compag.2023.107956
- [10] J. Tao, Q. Qiao, J. Song, S. Sun, Y. Chen, Q. Wu, Y. Liu, F. Xue, H. Wu, and F. Zhao, "Deep Learning-Driven Automatic Segmentation of Weeds and Crops in UAV Imagery," *Sensors*, vol. 25, no. 21, Art. no. 6576, 2025.
- [11] Y. Huang and A. Bais, "Unsupervised Domain Adaptation for Weed Segmentation Using Greedy Pseudo-labelling," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2024.
- [12] M. Beneš, V. Chalupský, and K. Mikula, "Geometrical image segmentation by the Allen–Cahn equation," *Applied Numerical Mathematics*, vol. 51, no. 2–3, pp. 187–205, 2004.
- [13] C. Banerjee, K. Nguyen, O. Salvado, T. Tran, and C. Fookes, "PINNs for Medical Image Analysis: A Survey," arXiv preprint arXiv:2408.01026, 2024.