International Scientific and Practical Conference Samarkand Institute of Economics and Service

Vol. 1 Issue 1 | pp. 163-165 **| ISSN:** XXXX-XXXX

Available online @ https://globalconf.net/index.php/samisi

Article

Evaluating Maize Weed Detection on Unseen Drone Frames

Mukhammedali Beriktassuly *[1], Gaini Mukhanova[2]

1. 1st year student of the «IT Enterpreneurship» educational program, Astana IT University
2. PhD in Economics, Professor of the School of Digital Public Administration , Astana IT University

* Correspondence: -

Abstract: Two tiles cut from the same drone frame can contain the same plant. A split that puts these tiles in training and test can therefore make weed detection look easier. The experiment on USU-Corn-WeedDB compared two FCOS detectors on the same 84 test tiles. The training sets shared 429 of their 515 tiles. The remaining 86 came from test frames for one detector and from separate frames for the other. At a budget of 100 epochs, test mAP50–95 was 36.03% with shared frames and 31.61% with separate frames. The difference was 4.41 percentage points on this split. For tests on unseen frames, the frame should be the unit of splitting.

Keywords: weed detection, UAV imagery, precision agriculture, data leakage, FCOS

Citation: Mukhammedali Beriktassuly, et all Evaluating Maize Weed Detection on Unseen Drone Frames. International Scientific and Practical Conference Samarkand Institute of Economics and Service 2026, 1(1), 163-165.

Received: 12th Aug 2026
Revised: 26th Sept 2026
Accepted: 8th Oct 2026
Published: 22th Nov 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

Introduction

A weed near the edge of one image tile can appear again in the next tile cut from the same drone frame. If the first tile goes into the training set and the second into the test set, the detector can be scored on a plant it has already seen. Checking that the file names differ will miss this. A detector used to produce weed maps for targeted treatment [1] needs to work on new survey images.

I measured how much this overlap changes the FCOS score on USU-Corn-WeedDB [2]. The audit counted shared image area and nearly coincident weed boxes. The two training runs used different sources for 86 tiles and shared both the validation and test sets. This lets the two detectors be compared on the same plants.

Related Work

Beery et al. found that animal recognition was harder at new camera-trap locations than at cameras seen during training [3]. Photos from the same camera have much the same background. The WILDS benchmark controls another source of variation in some of its comparisons by fixing the test set and changing the training data [4]. Closer to this experiment, Kattenborn et al. reported higher scores for tree-species segmentation when drone-image tiles were split at random than when they were held out in spatial blocks [5]. Here the unit of interest is a single drone frame, and the tile offsets allow repeated image content to be counted directly.

Methodology

USU-Corn-WeedDB has 800 labelled RGB tiles of 640 × 640 pixels from a forage maize field in Cache Valley, Utah [2]. Its 10,539 boxes mark three species: common lambsquarters (Chenopodium album), redroot pigweed (Amaranthus retroflexus) and green foxtail (Setaria viridis).

Frame IDs and pixel offsets locate each tile in its original image. A test tile counted as overlapping if it shared any pixels with a training tile from the same frame. The box audit

used coordinates in the original frame. A test box counted once if at least one training box of the same class reached an intersection over union (IoU) of 0.80 or higher. The same checks were used on the published split and on the two splits in the paired experiment. The test set contained 84 tiles from 13 frames and 974 annotated boxes. The validation set used 115 tiles from separate frames. Each training set contained the same 429 tiles plus 86 more (Figure 1). In the overlapping set those 86 came from the test frames, with the test tiles themselves held out. In the frame-disjoint set they came from other frames and were chosen to give similar class counts.

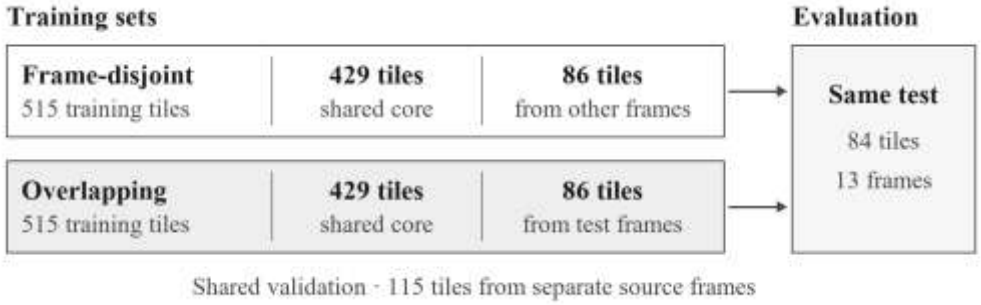


Figure 1. The two training sets have 429 tiles in common. Both detectors were scored on the same test tiles.

Both FCOS detectors with a ResNet-50 feature pyramid [6] started from random weights. Each run used seed 0, batches of four and 128 updates per epoch. The learning rate reached 0.005 after a 300-update warm-up and dropped after epochs 67 and 92. For each run and budget of 24, 50 or 100 epochs, the test evaluation used the checkpoint with the highest validation mAP50–95 within that budget. The reported metrics are mean average precision at IoU 0.50 and mean average precision across IoU thresholds from 0.50 to 0.95 in steps of 0.05.

Results and Discussion

In the split published with USU-Corn-WeedDB [2], 65 of 70 test tiles shared pixels with a training tile from the same frame. Of the 946 test boxes, 164 had a training box of the same class at IoU 0.80 or higher. The paired FCOS experiment used a separate test set of 84 tiles. In the paired experiment, 65 of 84 test tiles shared pixels with a training tile when training included parts of the test frames. Of the 974 test boxes, 107 had a training box of the same class at IoU 0.80 or higher. Both counts were zero when whole frames were kept separate. At the 100-epoch budget, the detector trained on parts of the test frames scored 36.03% mAP50–95. The detector trained on separate frames scored 31.61%. The gap was 4.41 percentage points before rounding, compared with 0.42 at 24 epochs and 1.10 at 50 (Figure 2). The checkpoints for the 100-epoch comparison came from epoch 70 with overlap and epoch 79 without overlap. Validation scores determined both choices.

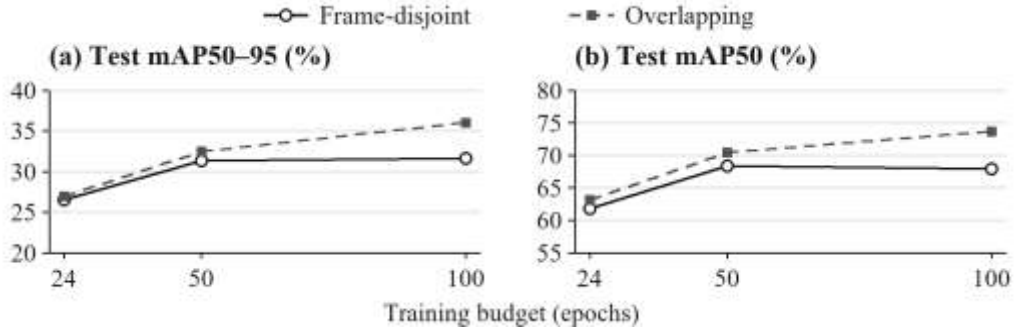


Figure 2. Test scores at three training budgets. Each point represents the best validation checkpoint within its budget. The panels use different vertical scales.

With the frames kept separate, mAP50–95 rose by 4.84 points from 24 to 50 epochs and by 0.26 points from 50 to 100. The added epochs made little difference in that run. At the largest budget the overlapping split gave a higher AP50–95 for all three weeds (Figure 3).

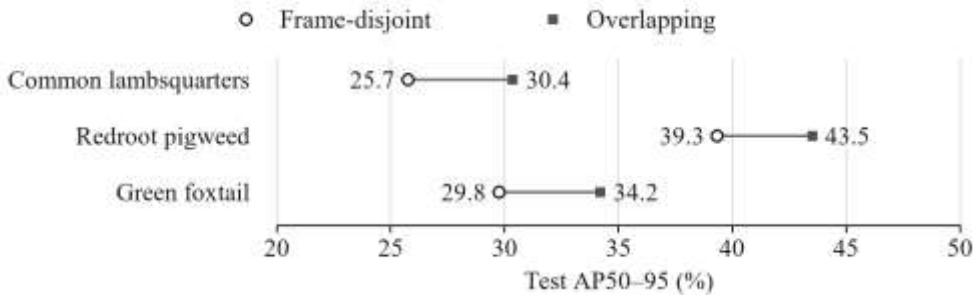


Figure 3. Test AP50–95 by species at the 100-epoch budget. Each line connects the two scores for one species.

The 4.41-point gap comes from one split and one seed. Matching the class counts still leaves differences in how hard the replacement tiles may be. Nearby drone frames can also cover the same ground. All images came from one field, so measuring performance on a new farm would require a test on another field.

Conclusions and Recommendations

For a test on unseen frames, split the original photographs first and cut the tiles afterwards. If the tiles are already cut, keep every tile with its source frame. Report the number of test frames as well as the number of tiles. For an assessment on new flights, apply the same rule to flights, as in CamelinaWeed [7].

REFERENCES

[1] N. Nikolić, P. Mattivi, S. E. Pappalardo, M. De Marchi and R. Masin. From detection to action: Creating operational prescription maps for weed management using low-cost UAVs. Italian Journal of Agronomy, 20, 100047, 2025. doi:10.1016/j.ijagro.2025.100047.

[2] U. Bhandari, S. Burlakoti, R. Miller, S. Young, E. Westra and A. Etienne. USU-Corn-WeedDB: A UAV RGB Image Dataset for Multi-Species Weed Detection in Forage Corn. arXiv:2606.06709, 2026.

[3] S. Beery, G. Van Horn and P. Perona. Recognition in Terra Incognita. European Conference on Computer Vision, 2018. doi:10.1007/978-3-030-01270-0_28.

[4] P. W. Koh et al. WILDS: A Benchmark of in-the-Wild Distribution Shifts. Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 5637–5664, 2021.

[5] T. Kattenborn, F. Schiefer, J. Frey, H. Feilhauer, M. D. Mahecha and C. F. Dormann. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open Journal of Photogrammetry and Remote Sensing, 5, 100018, 2022. doi:10.1016/j.ophoto.2022.100018.

[6] Z. Tian, C. Shen, H. Chen and T. He. FCOS: Fully Convolutional One-Stage Object Detection. IEEE/CVF International Conference on Computer Vision, 2019. doi:10.1109/ICCV.2019.00972.

[7] I. N. Rallis, K. Kempapidis, P. Raptis, E. K. Raptis and A. Ch. Kapoutsis. CamelinaWeed: an expert-agronomist-annotated UAV RGB and multispectral dataset for weed and crop monitoring in Camelina sativa. Data in Brief, 68, 113135, 2026. doi:10.1016/j.dib.2026.113135.