Point-Based Counting of Cereals and Legumes in Intercropped Fields

Baptiste Pras*

INRIA, Université Paris-Saclay

Abstract. Accurately estimating per-species plant density in intercropped fields, here for cereal (wheat) and legume (pea) mixtures, is critical for evaluating low-input farming practices. However, manual counting remains labor-intensive and impractical across hundreds of micro-plots. In this work, we adapt PET, a point-query crowd-counting transformer, to count wheat and pea seedlings from close-range smartphone images, using leaf-tip point annotations for wheat. We evaluate both counting accuracy (MAPE) and spatial localization (point-level F1-score) across various input resolutions and edge-border filtering strategies. Our best configuration achieves a 4.7% MAPE with a 74% F1-score on wheat, and a 7.8% MAPE with an 86% F1-score on pea.

Keywords: Plant counting · Point-based detection · Intercropping

1 Introduction

Intercropping a cereal (here wheat) with a legume (here pea) improves soil health and reduces reliance on chemical inputs. However, evaluating the agronomic performance of these systems requires per-species plant density estimates across hundreds of micro-plots, which cannot realistically be obtained by manual counting. While automated computer vision offers a scalable alternative, it faces major challenges such as soil background clutter, inter-plant occlusions, wide density variations across plots, and the non-rigid, evolving morphology of wheat plants.

In this work, we tackle these challenges using close-range smartphone images, where wheat and pea seedlings are clearly visible. We adapt PET [1], a point-query crowd-counting transformer, to localize and count wheat leaf tips and pea plants, scoring models on both count accuracy and spatial localization. Our contributions are threefold:

1. We demonstrate that PET effectively performs per-species plant counting in cereal and legume intercropped fields;
2. We evaluate strategies for giving high-resolution field images to the model, identifying the optimal input scale to preserve small plant structures;
3. We quantify the impact of edge-border filtering to mitigate boundary artifacts and evaluate the resulting trade-off on count bias and localization.

* Our complete implementation is available on GitHub: https://github.com/baptistepras/Point-Based-Counting-of-Cereals-and-Legumes-in-Intercropped-Fields

2 Method

Data. We use close-range smartphone images from the INRAE STICSMIX field trial. While pea plants are labeled at their individual center, wheat is annotated at leaf tips: varying leaf orientations, shapes, and severe overlaps make whole-plant centroids ambiguous, whereas leaf tips offer invariant, well-localized targets that easily map to plant counts given the uniform leaf number per seedling at a given growth stage. The dataset contains 116 wheat and 63 pea images, split at the plot level into train, validation, and test sets to prevent data leakage.

Model and metrics. PET is a VGG16-bn point-query transformer; we finetune it on our dataset from a crowd checkpoint and report global count metrics: MAPE, R^2, and signed percentage bias (positive for overcounting, negative for undercounting), as well as point-level localization quality. Predicted and ground-truth points are greedily matched within a distance threshold chosen to give an acceptable tolerance to compute true positives, false positives, and false negatives, yielding F1-score. Training checkpoints are selected on validation MAE to minimize the count, F1-Score is used afterwards as a sanity check.

Inference configurations. We evaluate two test-time strategies: (i) **Input resolution:** original images (4000×3000 px) are rescaled along their long edge across $\{1500, 2048, 2500, 3000, 4000\}$ px to assess how plant scale impacts detection performance. (ii) **Edge-border filtering:** partially visible plants truncated by image boundaries introduce counting ambiguities; we evaluate a filtering strategy where both ground-truth and predicted points within a fixed border margin wide enough to span one plant are excluded prior to scoring.

3 Results

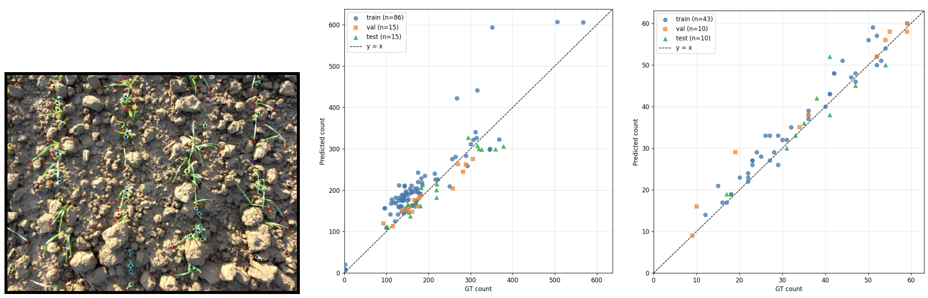
Table 1 reports test-set performance across input resolutions and edge-border filtering settings for both wheat and pea.

Three main findings emerge. First, count accuracy and spatial localization do not necessarily peak at the same scale: on wheat, 2048 px yields the best MAPE (8.74%) but lower localization (62.3% F1) compared to 1500 px (72.3%). This highlights the risk of relying solely on MAPE, which can be artificially lowered by false-positive and false-negative cancellation. Second, scaling resolution beyond 2048 px steadily degrades performance for both species, even causing pea detection to collapse at 4000 px ($R^2 = -4.33$). Third, edge-border filtering consistently improves F1-score by removing boundary false negatives (edge plants missed without filtering), though discarding these points naturally shifts the count bias toward positive values. Overall, the best trade-offs are achieved by 1500 px with border filtering for wheat (4.69% MAPE, 74.2% F1) and 2048 px without border filtering for pea (7.83% MAPE, 86.2% F1).

Figure 1 presents qualitative detections on a mixed plot together with predicted versus ground-truth scatter plots for both species across train, validation, and test splits.

Table 1: Test set performance across input resolutions, with and without edge-border filtering. Subscripts indicate signed differences vs. no-border baseline (green: gain, red: loss). Best results per column within each block in bold.

		Without border				With edge border			
Species	**Res.**	**MAPE**	R^2	**Bias (%)**	**F1 (%)**	**MAPE**	R^2	**Bias (%)**	**F1 (%)**
Wheat	2048	**8.74**	**0.90**	-1.65	62.3	$12.18_{+3.44}$	$0.81_{-0.09}$	$11.30_{+9.65}$	$65.9_{+3.6}$
	1500	9.03	**0.90**	-7.49	**72.3**	$\mathbf{4.69}_{-4.34}$	$\mathbf{0.97}_{+0.07}$	$\mathbf{-0.48}_{-7.01}$	$\mathbf{74.2}_{+1.9}$
	2500	9.73	0.87	**0.28**	62.1	$10.50_{+0.77}$	$0.91_{+0.04}$	$9.53_{+9.25}$	$64.4_{+2.3}$
	3000	12.54	0.79	4.25	53.4	$18.11_{+5.57}$	$0.75_{-0.04}$	$16.76_{+12.51}$	$55.9_{+2.5}$
	4000	13.77	0.63	-13.02	63.7	$7.15_{-6.62}$	$0.89_{+0.26}$	$-4.14_{-8.88}$	$66.7_{+3.0}$
Pea	2048	**7.83**	0.82	0.70	**86.2**	$\mathbf{8.59}_{+0.76}$	$0.90_{+0.08}$	$4.00_{+3.30}$	$\mathbf{86.5}_{+0.3}$
	1500	9.65	0.68	-7.07	82.6	$9.66_{+0.01}$	$0.80_{+0.12}$	$\mathbf{-1.59}_{-5.48}$	$84.9_{+2.3}$
	2500	9.56	**0.87**	**-0.13**	84.5	$9.82_{+0.26}$	$\mathbf{0.93}_{+0.06}$	$6.88_{+6.75}$	$85.6_{+1.1}$
	3000	13.15	0.78	1.31	76.4	$15.33_{+2.18}$	$0.81_{+0.03}$	$7.37_{+6.06}$	$78.1_{+1.7}$
	4000	67.94	-4.33	-67.70	33.5	$64.11_{-3.83}$	$-2.95_{+1.38}$	$-64.11_{-3.59}$	$36.2_{+2.7}$



(a) Joint detections (b) Wheat scatter plot (c) Pea scatter plot

Fig. 1: (a) Qualitative detections (red: wheat tips, blue: pea) (resolution 1500, borders on); (b, c) predicted vs. ground-truth counts across splits ($y = x$ dashed).

4 Conclusion

PET reliably estimates per-species wheat and pea density from smartphone imagery with single-digit MAPE and solid spatial localization. Evaluating point-level F1-score alongside count metrics is critical to reveal false-positive and false-negative cancellations, while edge-border filtering trades a positive bias for improved localization precision. **Acknowledgments:** We thank François Landes, Timothée Flutre, Sylvain Chevallier, Jérôme Enjalbert, and Jean-Marc Gilliot for their supervision, Romain Darras for the wheat tip annotations, and the INRAE STICSMIX project team for providing field data.

References

1. Liu, C., Lu, H., Cao, Z., Liu, T.: Point-query quadtree for crowd counting, localization, and more. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1676–1685 (2023)