# Point-Based Counting of Cereals and Legumes in Intercropped Fields

**Baptiste PRAS** · M2 Student, Master in Artificial Intelligence, Université Paris-Saclay

**Advisors:** François Landes · Timothée Flutre · Sylvain Chevallier · Jérôme Enjalbert · Jean-Marc Gilliot

*TAU Team, Inria Saclay · LISN, Université Paris-Saclay*

JUNIOR CONFERENCE ON DATA SCIENCE AND ENGINEERING

**11TH JUNIOR CONFERENCE ON DATA SCIENCES AND ENGINEERING · 24-25 SEPTEMBER 2026 · UNIVERSITÉ PARIS-SACLAY**

## 1. Introduction & Motivation

- **Context.** Intercropping wheat (cereal) with pea (legume) improves soil health and reduces chemical inputs.
- **Problem.** Evaluating it needs per-species plant density across hundreds of micro-plots: manual counting does not scale.
- **Research question.** Can a point-based model localize and count each species separately from smartphone photos?
- **Contributions.**
  - Adapt PET to per-species wheat/pea counting in intercropping
  - Identify the optimal input resolution for small plant structures
  - Quantify edge-border filtering trade-offs (bias vs. localization)

## 2. Proposed Approach

- **Step 1: Data.** INRAE STICSMIX field trial; wheat labeled at leaf tips (orientation-invariant), pea at plant center. 116 wheat / 63 pea images, split at the plot level (train/val/test).
- **Step 2: Model.** PET [1], a VGG16-bn point-query transformer, finetuned from a crowd-counting checkpoint.
- **Step 3: Test-time levers.** Input resolution (1500-4000 px) and edge-border point filtering.

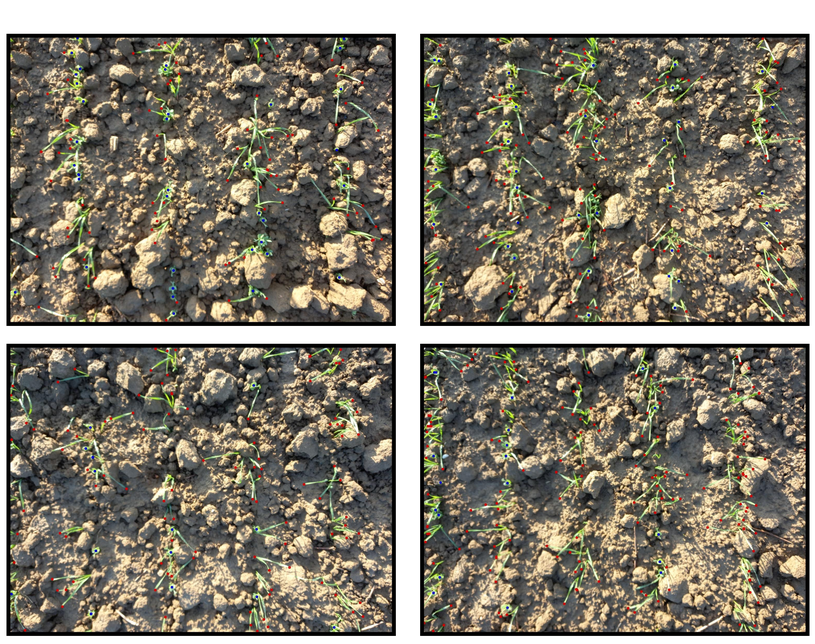

**Fig. 1.** Point annotations on two mixed wheat-pea plots, two photos each (red: wheat leaf tips, blue: pea plant centers; 1500 px, border filtering on).

## 3. Experimental Setup

- **Metrics** (greedy point matching, fixed distance tolerance):
  - MAPE: count error (%) ↓
  - $R^2$: count variance explained ↑
  - Signed bias (%): + over-counting, − under-counting, closer to 0 is better
  - F1 (%): point-level localization ↑
- **Grid.** 5 resolutions × {with, without} edge-border filtering, per species.
- **Implementation.** VGG16-bn backbone; finetuned from a crowd-counting checkpoint; model selected on validation MAE.

## 4. Results

- **Best trade-off:** wheat 1500 px+border → **4.7% MAPE, 74% F1** · pea 2048 px, no border → **7.8% MAPE, 86% F1**.
- MAPE can mislead: the best-MAPE resolution is not always the best-localized one (FP/FN cancellation).
- Beyond 2048 px, both species degrade: pea detection **collapses at 4000 px** ($R^2$=−4.33).

| | | 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** | **4.69** $_{-4.34}$ | **0.97** $_{+0.07}$ | **-0.48** $_{-7.01}$ | **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** | **8.59** $_{+0.76}$ | 0.90 $_{+0.08}$ | 4.00 $_{+3.30}$ | **86.5** $_{+0.3}$ |
| | 1500 | 9.65 | 0.68 | -7.07 | 82.6 | 9.66 $_{+0.01}$ | 0.80 $_{+0.12}$ | **-1.59** $_{-5.48}$ | 84.9 $_{+2.3}$ |
| | 2500 | 9.56 | **0.87** | **-0.13** | 84.5 | 9.82 $_{+0.26}$ | **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 $_{+2.7}$ |

**Table 1.** Test-set performance across input resolutions, with/without edge-border filtering. Subscripts: Δ vs. no-border (green = gain, red = loss). **Bold** = best per column within each block.

## 5. Discussion

- Border filtering drops truncated points from *both* GT and predictions ⇒ fewer boundary false negatives (F1↑), but fewer counted plants ⇒ over-counting bias.
- At native resolution, pea plants occupy far more pixels than PET's crowd-counting pretraining scale, a likely driver of the 4000 px collapse.
- F1 alongside MAPE is essential: MAPE alone can hide false-positive/negative cancellation.
- **Limits:** single trial, single growth stage.

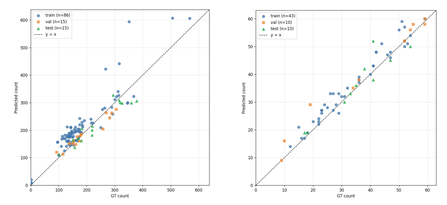

**Fig. 2.** Predicted vs. ground-truth counts, wheat (left) / pea (right), across train/val/test splits (dashed: $y=x$).

## 6. Conclusion & Future Work

- **Take-home.** PET reliably counts wheat and pea separately from smartphone photos, reaching single-digit MAPE and 74-86% F1 at the right input scale.
- **Field gap.** To our knowledge, no general baseline exists yet for per-species point-based counting in intercropped fields.
- **Next steps.**
  - Count both species jointly in a single pass, instead of per-species models
  - Cross-reference counts with post-harvest yield data

## References & Acknowledgments

[1] Liu, C., Lu, H., Cao, Z., Liu, T.: Point-query quadtree for crowd counting, localization, and more. ICCV (2023).

**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 the field data.

paper / code

**Contact:** baptiste.pras@universite-paris-saclay.fr
TAU Team, Inria Saclay · LISN, Université Paris-Saclay · jdse2026.github.io