Geomatic Techniques to Support Phytosanitary Products Tests whithin the EPPO Standard Framework

Samuele Bumbaca

University of Turin

August 28, 2025

The Traditional Approach to Agricultural Trials

Block 3	R	C	T
Block 2	T	R	C
Block 1	C	T	R

C Control
T Tested Product
R Reference Product

ANOVA Model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Where:

- y_{ij} = response
- μ = overall mean
- α_i = treatment effect
- β_j = block effect
- ε_{ij} = random error

Note:

This is the additive model. Modern approaches may include interaction terms: $\alpha_i \times \beta_j$

Key Assumptions of Traditional ANOVA

Statistical Assumptions:

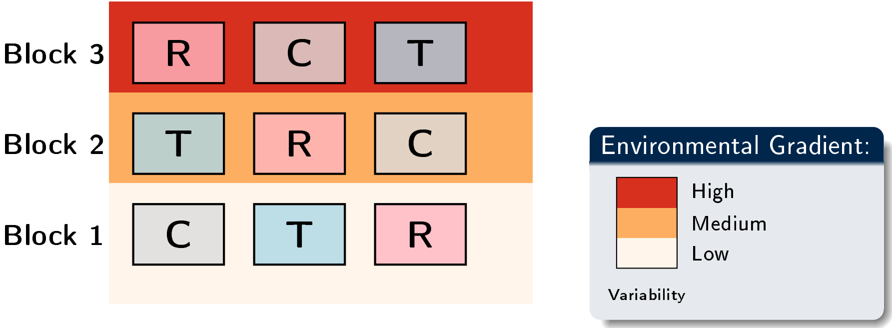
- **Randomization**: Treatments randomly assigned within blocks
- **Replication**: Each treatment appears in each block
- **Independence**: Observations are independent given the design
- **Homoscedasticity** : Equal variances across treatments
- **Normality**: Residuals follow normal distribution

Consequences of Assumption Violations:

- **Invalid conclusions of parametric tests**: Need for non-parametric tests leading to reduced statistical power

Based on R. A. Fisher, Statistical Methods for Research Workers, in S. Kotz & N. L. Johnson (eds.), Breakthroughs in Statistics: Methodology and Distribution, pp. 66–70, Springer, New York, 1992.

The Right Blocking: Capturing Environmental Variability

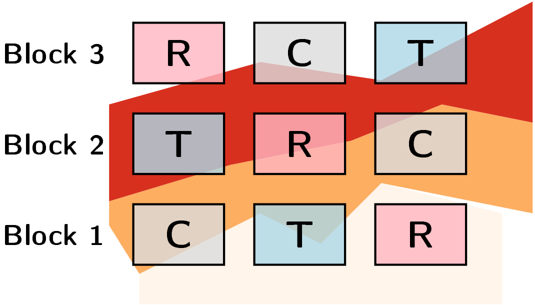


C Control
T Tested Product
R Reference Product

Success of Blocking Strategy:

- **Within-block homogeneity**: Treatments compared under similar conditions
- **Between-block heterogeneity**: Environmental gradient captured by block effects

The Wrong Blocking: Assumption Violation



Environmental Gradient:

High
Medium
Low

Variability

C Control
T Tested Product
R Reference Product

Heteroscedasticity Assumption Violation Problem:

- **Blocks fail to capture environmental variability**: Treatments compared under different conditions
- **Invalid parametric test**: Residual variance differs across treatments

Current Limitations in Statistics for Agricultural Trials

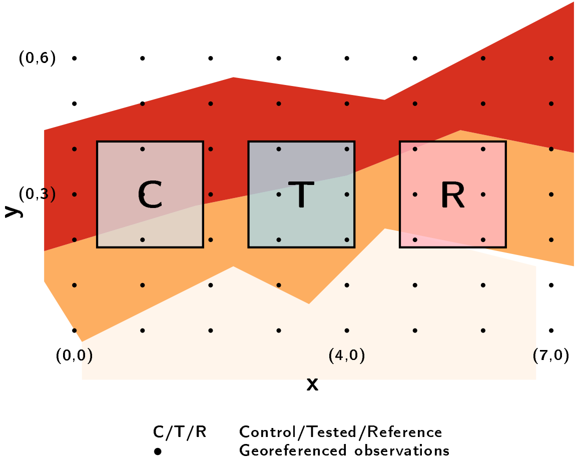
Traditional Approach Issues:

- **Human-dependent blocking**: Environmental variability assessment relies on experimenter experience
- **A priori identification**: Must identify variance sources BEFORE data collection

The Challenge:

How can we capture environmental variability mathematically rather than through human judgment?

Geostatistical Approach: Spatial Linear Mixed Models



Spatial LMM:

$$y(s_i) = \mu + \alpha_j + f(s_i) + \varepsilon_i$$

Where:

- $y(s_i)$ = response at s_i
- μ = overall mean
- α_j = treatment effect
- $f(s_i)$ = spatial random field
- ε_i = error
- $s_i = (x_i, y_i)$ = coordinates

Benefits:

- **No blocking**: Spatial correlation captures variability
- **Post-hoc**: No a priori variance identification
- **Homoscedasticity**: Assumption satisfied in more cases in respect blocking

Statistical Methods Comparison: Introduction

Comparison Objective:

Evaluate the performance of **traditional RCBD** versus **spatial geostatistical methods** (SpATS) in capturing environmental variability and estimating treatment effects.

Synthetic Dataset:

- **54 observations**(6×9 grid)
- **3 treatments**: Control (0 t/ha), Reference (0.5 t/ha), Test (1.0 t/ha)
- **3 blocks**(18 plots each)
- **Environmental zones**: Low (-1.5 t/ha), Medium (0 t/ha), High (+1.5 t/ha)

Tested Models:

1. **RCBD Model**: Linear Mixed Model with random block effects

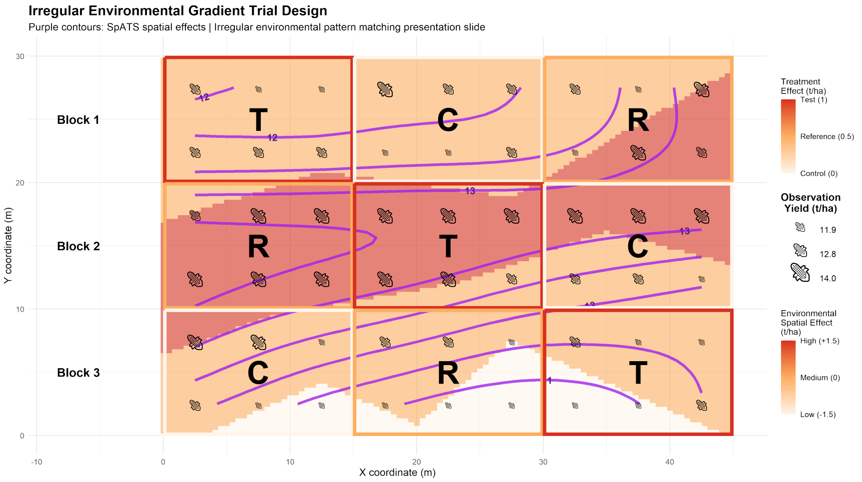
$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

2. **SpATS Model**: Spatial model with PSANOVA splines

$$y(s) = \mu + \alpha_i + f(s) + \varepsilon(s)$$

Where: α_i = treatment effects, β_j = block effects, $f(s)$ = spatial smooth

Statistical Methods Comparison: The Field Trial Design



Statistical Methods Comparison: Results

Model Performance (Mean Absolute Errors tonn/ha):

Model	Treat. Error	Env. Error
RCBD Model	0.13	0.62
SpATS Spatial	0.03	0.45

Treatment Effect Estimation (tonn/ha):

Treatment	True	RCBD	SpATS
Control	0.00	0.00	0.00
Reference	0.50	0.40	0.45
Test	1.0	0.69	0.94

Key Findings:

- **Both models satisfied assumptions**
- **SpATS outperformed RCBD**:
  - 3.8× better treatment effect estimation
  - 1.4× better environmental effect estimation
- **RCBD underestimated** by 20-31%
- **SpATS <6% error**

Implications:

Even when traditional RCBD meets statistical assumptions, **spatial modeling provides superior accuracy** in treatment effect estimation by properly accounting for environmental spatial variability.

The Missing Link: Spatial Coordinates

Geostatistical Methods Advantages:

- ✓ **Mathematical modeling** of environmental variability
- ✓ **Post-hoc analysis** - no need for prior knowledge of the environment variables and of their distribution
- ✓ **Superior performance** in handling spatial heterogeneity
- ✓ **EPPO recognized** approach (PP1/152(4) - Design and analysis of efficacy evaluation trials)

Current Barrier:

- × **Requires spatially referenced observations**
- × **Traditional manual assessments lack coordinates**
- × **Implementation gap** in practical field trials

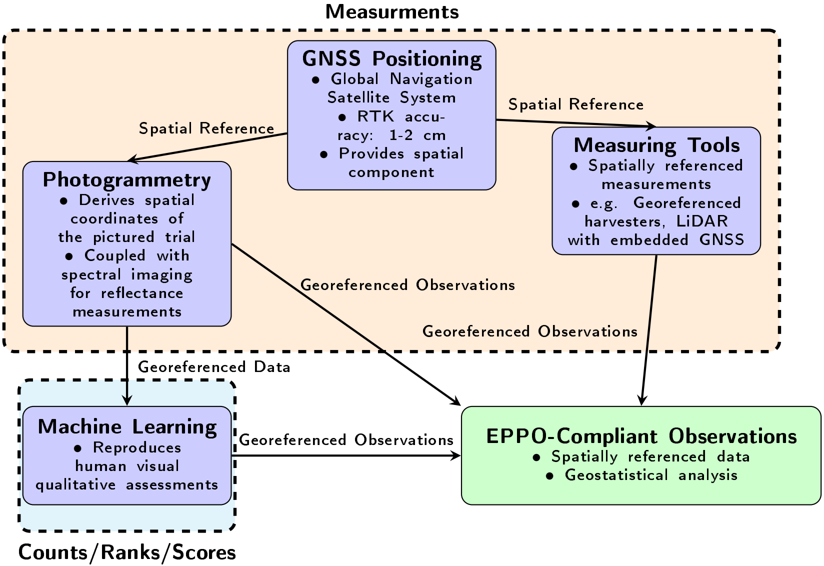
Central Research Question

Can geomatics technologies provide spatially referenced observations that enable geostatistical analysis within EPPO-compliant Plant Protection Product trials?

Specific Objectives:

1. Establish which geomatics technologies can be used to collect spatially referenced observations
2. Demonstrate the feasibility of collect spatially referenced observations in compliant with EPPO standards
3. Validate performance against traditional methods
4. Provide practical implementation guidelines

Geomatic Technologies: Workflow for Spatially Referenced Observations



Georeferencing EPPO Standard Assessments

Table: EPPO's types of variables

Type of Variable	Measurement	Ranking	Scoring
Continuous not limited	X		
Continuous limited	X		
Discrete	X		
Ordinal		X	X
Nominal			X
Binary			X

Summary from EPPO PP 1/152: Design and analysis of efficacy evaluation trials

Current State of Georeferencing in Agricultural Trials:

Tool-based measurements (e.g., yield harvesters) can be easily georeferenced by integrating GNSS receivers on the tool. For visual assessments as counting, scoring or ranking, a method to transform georeferenced data to georeferenced observations is needed.

Machine Learning Inference on Georeferenced Data

Machine Learning for Visual Assessments:
Machine Learning (ML) can reproduce human visual assessments, allowing for spatially referenced observations.

Machine Learning Limitations:
ML needs training data to reproduce human assessments.
Understand data requirment to fit EPPO standard is critical.

Application Case Studies:
This thesis provides a case study to prove the possibility to get georeferenced observations for each variable type that need for a ML step.

EPPO ML integration

EPPO PP 1/333(1): Digital Technologies in PPP Trials

ML integrated assessments must meet the same quality standards as manual assessments and require validation through comparison with manual assessments (golden sample).

Validation Benchmarks[1]

[1] Based on EPPO PP 1/333(1): Use of digital technologies in efficacy and selectivity trials

- **Continuous/Discrete**: $R^2 > 0.85$ (1:1 relationship)
- **Ordinal/Nominal**: Cohen's $\kappa > 0.7$
- **Binary**: Accuracy > 0.85

Georeferencing Gap in EPPO Standard Assessments

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
→	Discrete	X		
→	Ordinal		X	X
→	Nominal			X
→	Binary			X

Case Studies:

This thesis aim to prove the reliability of georeferencing every EPPO standard assessmentEach case study addresses a specific variable type as defined in the EPPO standards

- **Discrete (Counts)** : Plant counting
- **Ordinal** : Phytotoxicity scoring
- **Nominal** and **Binary** : Disease detection

Georeferencing Counts (Discrete Variable)

	Type of Variable	Measurement	Ranking	Scoring
	Continuous not limited	X		
	Continuous limited	X		
→	Discrete	X		
	Ordinal		X	X
	Nominal			X
	Binary			X

Georeferencing Counts:

- **Counts** are discrete variables required for measuring density of individuals (e.g. plant density in PP1/46 (3) - Wireworms).
- the **Case Study**: Counting plants from georeferenced photogrammetric orthomosaics by ML Object Detection.
- this study is discussed in the scientific article **Bumbaca, S.; Borgogno-Mondino, E.C. On the Minimum Dataset Requirements for Fine-Tunining an Object Detector for Arable Crop Plant Counting: A Case Study on Maize Seedlings. Remote Sens. 2025, 17, 2190. DOI: 10.3390/rs1713219061**

Arable Crop Plant Counting by Object Detection

The Critical Need after EPPO Assessments:

- Plant counting is **fundamental** also in precision agriculture and plant breeding
- Traditional manual counting is **time-consuming** and bring **human error** risks
- **Computer vision** offers a solution, but requires **dataset size and quality** characterization to prove the reliability for this task.

EPPO Benchmark Standards:

Coefficient of determination (R^2) ≥ 0.85 of ML method w.r.t. manual counting (no bias nor slope linear first order relation)

Research Gap:

What are the minimum dataset requirements to achieve this benchmark across different inference datasets?

Photogrammetric Orthomosaics for Plant Counting

Advantages over other kind of data:

- **Geographical coordinates**: Suitable for spatial analysis
- **Fixed scale and orientation images**: Eliminate perspective inconsistencies
- **Achievable High-resolution**: From low altitude nadiral overlapping images[1]

[1] Kraus, K. Photogrammetry: Geometry from Images and Laser Scans. De Gruyter: Berlin, Germany, 2011

Limitations:

- **Occlusions**: Overlapping vegetation canopy issues [1] -> Target crop and phenological stage selection
- **Georeferencing errors**: Due to low-quality/insufficient GNSS embedded systems or Ground Control Points (GCPs) [2] -> Hardware requirements
- **Computational demand**: Processing time constraints for large-area orthomosaics -> Not real-time suitable

---

[1] Habib et al. Automated Ortho-Rectification of UAV-Based Hyperspectral Data over an Agricultural Field Using Frame RGB Imagery. *Remote Sens.*, 8:796, 2016

[2] Pugh et al. Comparison of image georeferencing strategies for agricultural applications of small unoccupied aircraft systems. *Plant Phenome J.*, 4:e20026, 2021

Plants Occlusion Solution: Maize Seedlings at BBCH 13-15 Stage

- **Optimal detection conditions**: Regular planting pattern, minimal plant overlapping at BBCH 13-15 stage [1]
- **Data availability**: Most represented plant in scientific [2] [3] and public datasets
- **Economic importance**: World's highest-production crop [4]
- **Rappresentative crop**: Findings applicable to other row crops [5] (e.g. Sunflower, Sugar beet)

---

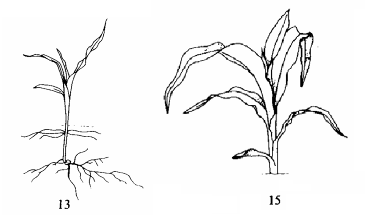
[1] Meier et al. The BBCH system to coding the phenological growth stages of plants. *J. Für Kult.*, 61:41–52, 2009

[2] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[3] Liu et al. IntegrateNet: A deep learning network for maize stand counting from UAV imagery. *IEEE Geosci. Remote Sens. Lett.*, 19:6512605, 2022

[4] FAO. *Agricultural Production Statistics 2010–2023*. FAOSTAT, Rome, Italy, 2024

[5] Torres-Sánchez et al. Early Detection of Broad-Leaved and Grass Weeds in Wide Row Crops Using Artificial Neural Networks and UAV Imagery. *Agronomy*, 11:749, 2021



Suitable Hardware and Photogrammetric Picturing

- **UAV Platform**: Phantom 4 Pro v2.0 (DJI, Shenzhen, China)
- **Camera**: Default series RGB camera
- **Flight Altitude**: 10 m above ground level
- **Original GSD**: 2.7 mm/pixel
- **GNSS Mode**: VRS-NRTK for GCP surveying
- **Bundle Adjustment Error**: 38 mm
- **Final Orthomosaic GSD**: 5 mm/pixel
- **Reference System**: WGS84/UTM 32 N

Key Processing Steps:

1. Nadiral image capture with 70%-80% overlapping patterns
2. Ground Control Points (GCPs) surveyed with high-precision GNSS
3. Photogrammetric bundle adjustment and orthomosaic generation
4. Georeferenced orthomosaic output ready for spatial analysis

Plant Counting - Object Detection Paradigms

The State-Of-The-Art (SOTA) methods to count plants by orthomosaics rely on object detectors. Historically, for most of the tasks, object detectors increased their performance in this order:

- **Traditional Methods**: Handcrafted features (hardcoded) - untill 1990s
- **Machine Learning Approaches**: 1990s to 2010s
- **Deep Learning Approaches**: Convolutional Neural Networks (CNNs) - 2010s to 2020s
- **Transformer Architectures**: Attention mechanisms introduction - 2020s
- **Data-Efficient Methods**: Few-Shot and Zero-Shot Detection - Now

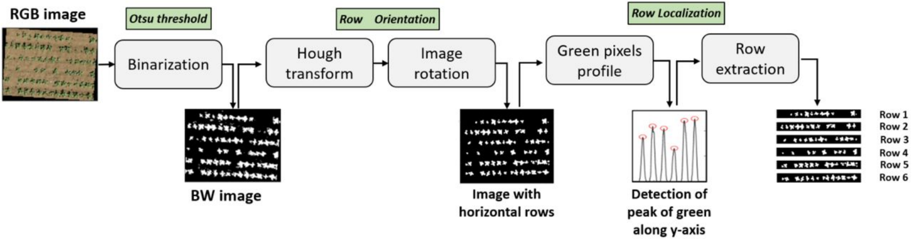
Plant Counting - Classic Object Detection Methods

Handcrafted Methods (HC):

- **Traditional approach**: Still used in agricultural applications [1] [2]
- **Explicit programming**: Color thresholding, edge detection, morphological operations
- **Advantages**: Domain expertise, interpretability, computational efficiency

[1] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[2] García-Martínez et al. Digital count of corn plants using UAVs and cross correlation. *Agronomy*, 10:469, 2020



Plant Counting - Classic Object Detection Methods

Machine (Deep) Learning Approaches:

Convolutional Neural Networks [1] Grid-based convolutions

- **Faster R-CNN** [2]
- **YOLO variants** for faster inference

[1] LeCun et al. Deep learning. *Nature*, 521:436–444, 2015

[2] Faster R-CNN: Towards real-time object detection with region proposal networks

Transformer Architectures [3] Image patches processing (attention-based)

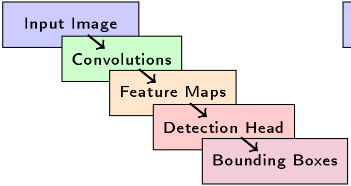
- **DETR** [4]
- **Hybrid approaches** with convolutions and attention

Superior with scarce data [5]

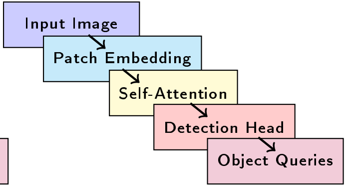
[3] Vaswani et al. Attention is all you need. *NIPS*, 2017

[4] Carion et al. End-to-end object detection with transformers. *arXiv:2005.12872*, 2020

[5] Rekavandi et al. Transformers in small object detection: A benchmark and survey. *arXiv:2309.04902*, 2023



CNN-based

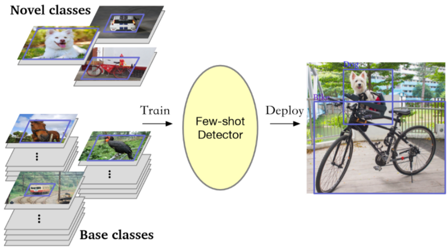


Vision Transformer

Plant Counting - Data-Efficient Detection Methods

Few-Shot Detection:

- **Learning from minimal examples**: 1-30 annotated instances
- **Meta-learning** approach [1]
- **Advantage**: Reduce annotation burden for new classes
- **Limited studies**: Only two for maize seedlings [2] [3]



4

[1] Li et al. Meta-SGD: Learning to Learn Quickly for Few-Shot Learning. *arXiv*, arXiv:1707.09835, 2017

[2] Karami et al. Automatic Plant Counting and Location Based on a Few-Shot Learning Technique. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 13:5872–5886, 2020

[3] Wang et al. Advancing Image Recognition: Towards Lightweight Few-shot Learning Model for Maize Seedling Detection. *SCIS*, 2024

[4] https://si-analytics.tistory.com/

Plant Counting - Data-Efficient Detection Methods

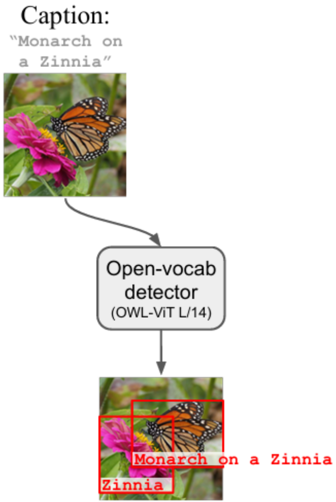
Zero-Shot Detection:

- **No labeled examples**: Detect novel objects without training data
- **Semantic relationships**: Exploit contextual information [1]
- **State-of-the-art**: OWLv2 [2], Grounding DINO [3]
- **Agricultural gap**: No studies for maize seedling counting

[1] Bansal et al. Zero-shot object detection. *ECCV*, 2018

[2] Minderer et al. Scaling Open-Vocabulary Object Detection. *Adv. Neural Inf. Process. Syst.*, 36:72983–73007, 2023

[3] Liu et al. Grounding DINO: Marrying DINO with Grounded Pre-training for Open-Set Object Detection. *ECCV*, 2024



[4]

[4] Minderer et al. Scaling Open-Vocabulary Object Detection. *Adv. Neural Inf. Process. Syst.*, 36:72983–73007, 2023

Plant Counting - Research Gap

Critical Research Gaps:

- **Minimum dataset requirements** None of the studies taken into account systematically tested[1] the minimum dataset requirements for robust plant detectors (EPPO benchmark).
- **In-domain vs. out-of-distribution data** Despite some author already studied the impact[2] [3] none did it in a systematic way.
- **Architecture influence** Many studies compared different architectures, someone claiming few-shot performances [4] [5], but none systematically tested the minimum dataset capability and none tested the zero-shot possibility.

[1] Sun et al. Revisiting unreasonable effectiveness of data in deep learning. *arXiv:1707.02968*, 2017

[2] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[3] Andvaag et al. Counting canola: Toward generalizable aerial plant detection. *Plant Phenomics*, 6:0268, 2024

[4] Wang et al. Advancing Image Recognition: Towards Lightweight Few-shot Learning Model for Maize Seedling Detection. *SCIS*, 2024

[5] Karami et al. Automatic Plant Counting and Location Based on a Few-Shot Learning Technique. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 13:5872–5886, 2020

Study Aim

Primary Objective:

Establish the **minimum dataset requirements** for accurate maize seedling detection (EPPO benchmark) in georeferenced orthomosaics across different object detection paradigms

Key Definitions:

- **Dataset size**: Amount of annotated images in training set
- **Dataset quality**: Accuracy of annotations (percentage of correct annotations relative to ground truth)

Specific Research Questions:

1. How does training data source (in-domain vs. out-of-distribution) affect required dataset size and quality?
2. Untill which extent different architectures affect training dataset requirements?

Plant Counting - Material and Methods - Research Methodology

- **Objective**: Investigate minimum dataset size and quality for robust object detection
- **Classic Object Detectors requirements**:
  - with out-of-distribution (OOD) training datasets.
  - with in-domain (ID) training datasets.
- **Data-efficient Methods**:
  - Test Few-Shot Detector requirements with ID training samples.
  - Test Zero-Shot Detector performances.
- **Empirical Modeling Approach**:
  - Analyze the relationship between dataset size/quality and model performance
  - Fit empirical functions to characterize this relationship
  - Use fitted functions to predict performance with varying dataset size/quality

Plant Counting - Material and Methods - Dataset

Dataset Classification:

- **Out-of-Distribution (OOD)**: Training datasets from different sources than inference target
- **In-Domain (ID)**: Training datasets from same source/distribution as testing dataset

OOD Scientific Datasets:

Source: Scientific literature

OOD Internet Datasets:

Source: Internet repositories

ID Datasets:

Source: Collected by the author

Key Processing Parameters:

All dataset preprocessed to get standard **tile size**: 224×224 pixels (1.12×1.12 m field coverage for georeferenced)

Plant Counting - Material and Methods - Datasets

Dataset	Phenological Stage	Train Size	Test Size
OOD Scientific			
DavidEtAl.2021 [1]	V3	182 tiles	N/A*
LiuEtAl.2022 [2]	V3	596 tiles	N/A*
OOD Internet			
OOD_int_1 [3]	V3	216 tiles	N/A*
OOD_int_2 [4]	V5	174 tiles	N/A*
ID [5]			
ID_1	V3	150 tiles	20 tiles
ID_2	V3	150 tiles	20 tiles
ID_3	V5	150 tiles	20 tiles

* N/A indicates that these datasets were used only for training purposes and do not have separate test sets in this study.

[5] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*, 2021

[5] Liu et al. IntegrateNet: A deep learning network for maize stand counting from UAV imagery. *IEEE Geosci. Remote Sens. Lett.*, 19:6512605, 2022

[5] Maize_seeding dataset. `https://universe.roboflow.com/objectdetection-hytat/maize_seeding`

[5] Maize-seedling-detection dataset. `https://universe.roboflow.com/fyxdds-icloud-com/maize-seedling-detection`

[5] Bumbaca, Samuele. 'The Original Dataset for the Paper "on the Minimum Data Set Requirements for Fine-tuning an Object Detector for Arable Crop Plant Counting: A Case Study on Maize Seedlings"'. Zenodo, 17 April 2025. https://doi.org/10.5281/zenodo.15235602.

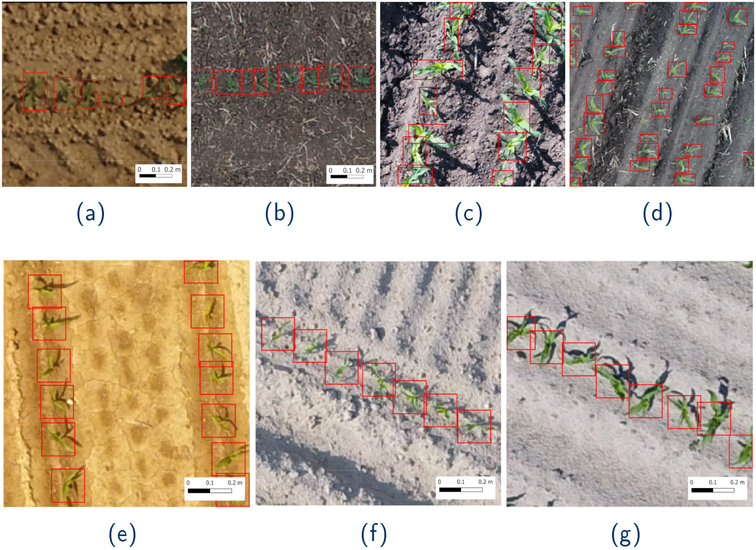


Figure: Image examples taken from each dataset, ground truth bounding boxes are shown in red. (**a**) DavidEtAl.2021, (**b**) LiuEtAl.2022, (**c**) Internet Maize stage V3, (**d**) Internet Maize stage V5, (**e**) ID_1, (**f**) ID_2, (**g**) ID_3.

Primary Evaluation Metrics:

Performance assessed using R^2 and mAP for counting and detection respectively

Counting Metric:

Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Detection (Spatial) Metric:

Mean Average Precision:

$$mAP = \frac{1}{|IoU|} \sum_{t \in IoU} AP_t$$

Metric Interpretation:

R^2: 1 = perfect, 0 = mean prediction, negative = worse than mean | mAP: IoU threshold 0.5

Plant Counting - Material and Methods - Training Protocols

Training Dataset Configuration:

- **Many-shot models**: 90% training / 10% validation split
- **Few-shot models**: Number of shots determines training samples
- **Zero-shot learning**: Natural language descriptions only

Dataset Size Evaluation:

- **Many-shot**: 10 to 150 images (15 steps of 10)
- **Few-shot**: 1, 5, 10, 30, and 50 shots
- **Zero-shot**: Multiple text prompt variations

Quality Assessment:

- **Annotation reduction**: 100% to 10% (10 steps)
- **Constant dataset size**: During quality evaluation
- **OOD vs ID influence**: Same experimental protocol

Plant Counting - Material and Methods - Predictive Modeling

Empirical Function Testing:

Three mathematical functions tested to model dataset size/quality vs performance relationships

Logarithmic:

$$f(x) = a\ln(x) + b$$

Behavior: Diminishing returns pattern **Theory**: Asymptotic performance approach

Arctangent:

$$f(x) = a\arctan(bx) + c$$

Behavior: Saturating performance **Theory**: Bounded metrics plateau

Algebraic Root:

$$f(x) = ax^{1/b} + c$$

Behavior: Power-law relationships **Theory**: Flexible scaling dynamics

Model Selection Criteria:

- **Goodness-of-fit**: $GoF = R^2_{fit}$ for function selection
- **Best predictor**: Highest fit determines model-metric combination
- **Practical guidance**: Annotation planning through interpolation/extrapolation

Plant Counting - Materials and Methods - Classic Detectors Architecture

YOLOv5 - Baseline CNN Architecture:

- **Backbone**: CSP (Cross Stage Partial) with PANet neck
- **Agricultural dominance**: Most widely adopted in crop monitoring [1]
- **Reference point**: Well-established baseline for dataset requirements

[1] Badgujar et al. Agricultural object detection with YOLO algorithm. *Comput. Electron. Agric.*, 223:109090, 2024

YOLOv8 - Improved CNN Architecture:

- **Backbone improvement**: C2f blocks for enhanced efficiency
- **Detection head**: Anchor-free design with decoupled heads
- **Performance**: Superior accuracy-speed trade-offs [2]

[2] Terven et al. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.*, 2023

CNN Architecture Benefits:

Computational efficiency, **proven agricultural performance**[3] [4], and **established baseline** for dataset requirement comparison

[3] Kitano et al. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.*, 16:1–5, 2019

[4] Barreto et al. Automatic UAV-based counting of seedlings. *Comput. Electron. Agric.*, 191:106493, 2021

Plant Counting - Materials and Methods - Classic Detectors Architecture

RT-DETR - CNN+Transformer Hybrid:

- **Architecture**: CNN backbone + Transformer decoder
- **Attention mechanism**: Deformable attention for adaptive feature sampling
- **Global relationships**: Models object interactions across entire image
- **Real-time performance**: Parallel prediction heads
- **Agricultural proven**: Superior inference performances in respect pure-CNN YOLOs [1]

[1] Zhao et al. DETRs beat YOLOs on real-time object detection. *arXiv:2304.08069*, 2024

YOLO11 - Transformer-mixed:

- **Key innovation**: Multi-scale deformable attention mechanisms (for small object detection)
- **Hybrid approach**: YOLO backbone + Transformer attention

Research Question:

Do Transformer-mixed improvements affect minimum dataset requirements for small object detection compared to pure CNN approaches?

Plant Counting - Materials and Methods - Classic Detectors Architecture

Unified Training Configuration and Implementation:

- **Library**: Ultralytics open-source implementation [1]
- **Consistency**: Same framework enables fair architectural comparison
- **Hardware**: Intel Xeon E5-2670 v3, 64GB RAM, NVIDIA RTX A5000 (24GB VRAM)

[1] Jocher, Glenn; Qiu, Jiarui; Chaurasia, Anil. GitHub Ultralytics YOLO. 2023. Available online: `https://github.com/ultralytics/ultralytics`

Training Hyperparameters:

- Batch size: 16
- Max epochs: 200
- Early stopping: 15 epochs without improvement

Data Augmentation Protocol:

- **Geometric**: Random scaling, Translation
- **Photometric**: HSV augmentation
- **Composition**: Mosaic augmentation, Horizontal flip

Excluded Alternatives:

Faster R-CNN: Computational overhead[2] | **Pure DETR**: Prohibitive training requirements for small datasets [3]

[2] Velumani et al. Estimates of Maize Plant Density from UAV RGB Images Using Faster-RCNN Detection Model. *Plant Phenomics*, 2021:9824843, 2021

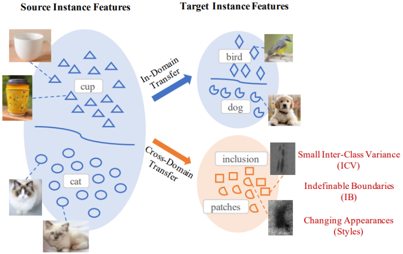
[3] Carion et al. End-to-end object detection with transformers. *arXiv:2005.12872*, 2020

CD-ViTO - Cross-Domain Vision Transformer:

- **Paradigm**: Cross-domain prototype matching approach
- **Training data**: Small set of annotated examples (shots) as class prototypes
- **State-of-the-art**: Leading performance in few-shot detection

Implementation Details:

- **Shot definition**: 1 image with single annotated plant
- **Tested shots**: 1, 5, 10, 30, and 50 shots



[1]

[1] Fu et al. Cross-Domain Few-Shot Object Detection via Enhanced Open-Set Object Detector. *arXiv*, arXiv:2402.03094, 2024

Plant Counting - Materials and Methods - Data-Efficient Detection

OWLv2 - Open-Vocabulary Detection:

- **Paradigm**: Object detection based solely on text prompts
- **State-of-the-art**: Leading performance in open-vocabulary detection [1] [2]

[1] Minderer et al. Scaling Open-Vocabulary Object Detection. *Adv. Neural Inf. Process. Syst.*, 36:72983–73007, 2023

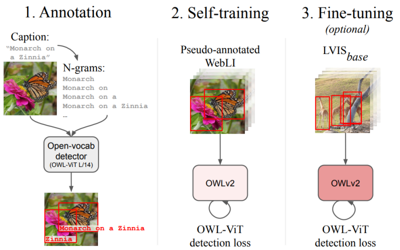
[2] Liu et al. Grounding DINO: Marrying DINO with Grounded Pre-training for Open-Set Object Detection. *ECCV*, 2024

Model Variants Tested:

- **Encoder sizes**: ViT-B/16, ViT-L/14
- **Base models**: Self-supervised OWL-ST method training
- **Fine-tuned models**: Further trained on human-annotated datasets
- **Ensemble models**: Multiple weight combinations for balanced performance

Text Prompt Strategy:

Prompt variety: Simple terms ("maize", "seedling") to descriptive phrases ("aerial view of maize seedlings") **Systematic evaluation**: 11 different prompts tested, best-performing reported



Plant Counting - Materials and Methods - Architecture Summary

Table: Summary of tested architectures and model sizes (millions of parameters)

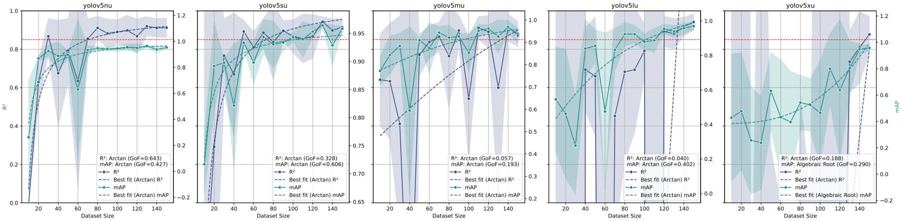
Architecture	Shots	n	s/S	m/B	l/L	x
YOLOv5	many	1.9	7.2	21.2	46.5	86.7
YOLOv8	many	3.2	11.2	25.9	43.7	68.2
YOLO11	many	4.0	12.5	28.0	50.0	75.0
RT-DETR	many	-	-	-	60.0	80.0
CD-ViTO	few	-	22.0	86.0	307.	-
OWLv2	zero	-	-	86.0	307.0	-

n: nano, **s**: small, **m**: medium, **l**: large, **x**: extra-large **S**: ViT-S (Small) backbone **B**: ViT-B (Base) backbone **L**: ViT-L (Large) backbone

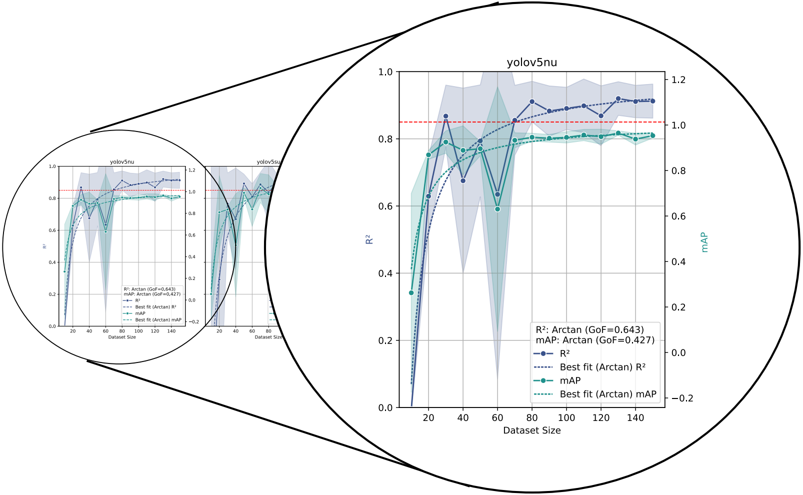
Architecture Selection Strategy:

Parameter count affects dataset requirements: larger models may need more data for training but offer better feature extraction capabilities for complex tasks

Plant Counting - Results - YOLOv5 dataset size

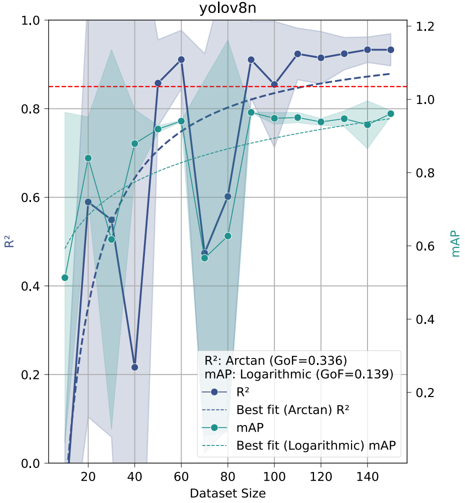


Plant Counting - Results - YOLOv5 dataset size



YOLOv5 demonstrates that traditional CNN architectures with low parameter amount are sufficient even with only 130 samples

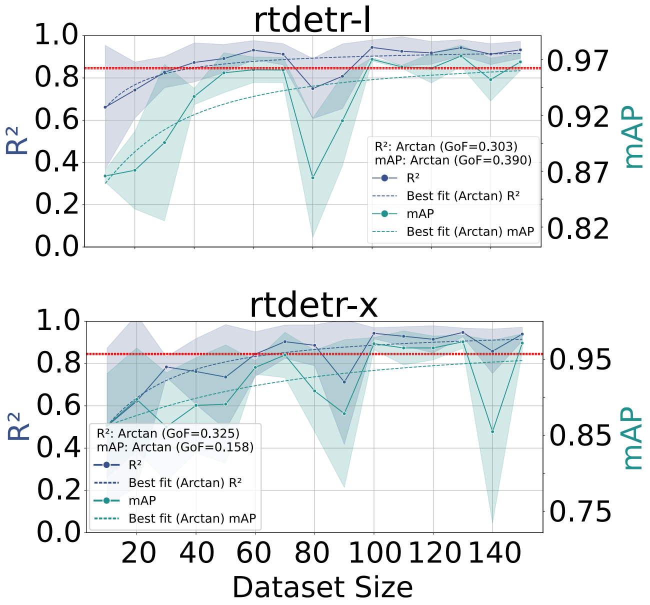
Plant Counting - Results - YOLOv8 dataset size



Evolution Impact:

- Reductions in annotation burden (110 images)
- Only low amount of parameters succeded like in YOLOv5

Plant Counting - Results - RT-DETR dataset size



Plant Counting - Results - Dataset Size

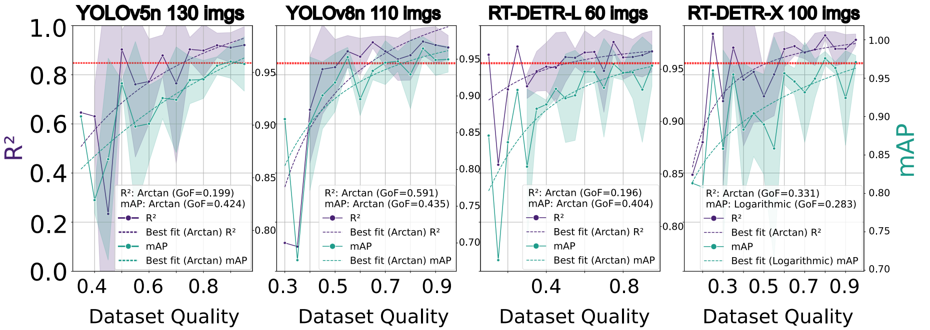
Architecture	Parameters	Dataset Size
YOLOv5	1.9 (n)	130
YOLOv8	3.2 (n)	110
RT-DETR	60 (l)	60
RT-DETR	80 (x)	100

Transformer-Mixed Superiority at a higher parameters price:

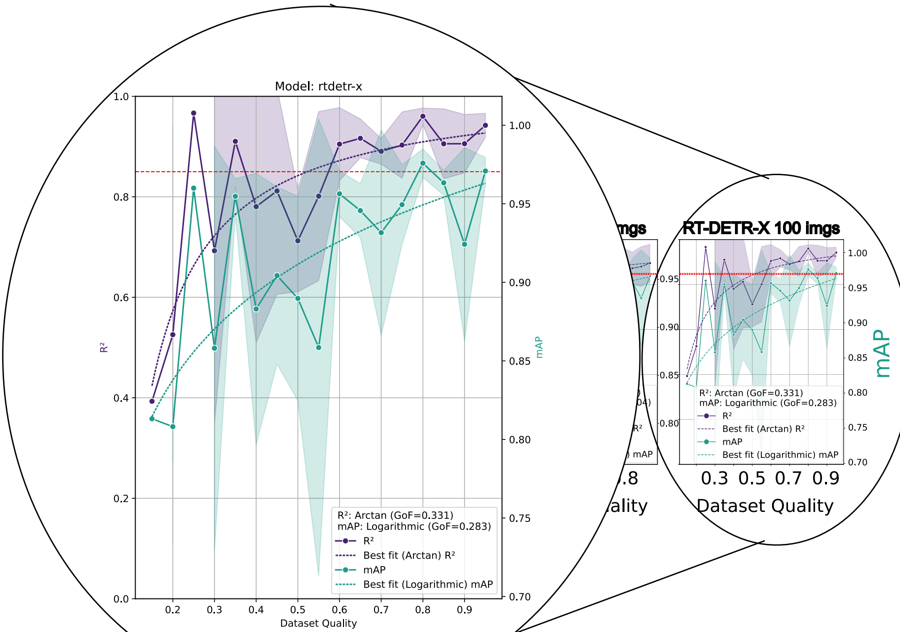
RT-DETR demonstrates reduced dataset requirements in respect pure-CNN counter parts.
YOLO11 did not succeded to reach the benchmark with any dataset and parameter size.

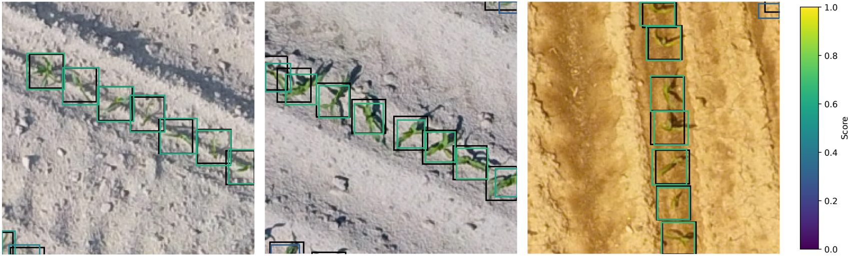
RT-DETR L predictions trained on 60 images

Plant Counting - Results - Dataset Quality Requirements



Plant Counting - Results - Dataset Quality Requirements





RT-DETR X predictions with 35% quality reduction

Quality vs Quantity Trade-offs:

Strategic insight: RT-DETR architectures offer flexibility - achieve benchmark with either minimal high-quality data (60 images, 100% quality) or more abundant medium-quality data (100 images, 65% quality)

The Dataset Requirements Challenge in Agriculture

Known Performance Factors:

- **Dataset size**: Performance directly related to training data amount [1]
- **Data quality**: Annotation accuracy critically affects model performance [2]
- **Model architecture**: Different models require different dataset sizes for same performance [3] [4]

[1] Sun et al. Revisiting unreasonable effectiveness of data in deep learning. *arXiv:1707.02968*, 2017

[2] Alhazmi et al. Effects of annotation quality on model performance. *ICAIIC*, 2021

[3] Nguyen et al. An evaluation of deep learning methods for small object detection. *JECE*, 2020

[4] Brigato et al. Close look at deep learning, 2020

Domain-Specific Factors:

- **Backbone importance** [5]
- **Domain-specific pre-training** benefits [6]
- **Data augmentation** strategies [7]

[5] Du et al. SpineNet: Learning scale-permuted backbone. *CVPR*, 2020

Most Critical Factor:

Dataset source [8]: In-domain vs. out-of-distribution dramatically affects accuracy and dataset size requirements [9] [10]

[8] Sun et al. Revisiting unreasonable effectiveness of data in deep learning. *arXiv:1707.02968*, 2017

[9] David et al. Plant detection and counting from high-resolution RGB images acquired from UAVs. *bioRxiv*,

Deep Learning Architectures Comparison

Architecture Analysis:

- **Trade-off**: High precision on subset vs. limited generalizability

CNN-based Models:

Convolutional Neural Networks [1]

- **YOLO family**: YOLOv5, YOLOv8
- **Faster R-CNN** [2]
- Grid-like image processing
- Efficient for small images

[1] LeCun et al. Deep learning. *Nature*, 521:436–444, 2015

[2] Faster R-CNN: Towards real-time object detection with region proposal networks

Transformer-based:

Attention mechanisms [3]

- **DETR** [4]
- **RT-DETR**: Real-time performance
- Sequence of patches processing
- Superior with scarce data [5]

[3] Vaswani et al. Attention is all you need. *NIPS*, 2017

[4] Carion et al. End-to-end object detection with transformers. *arXiv:2005.12872*, 2020

[5] Rekavandi et al. Transformers in small object detection: A benchmark and survey. *arXiv:2309.04902*, 2023

Key Insight:

Representative Architectures: YOLO vs RT-DETR

YOLO Family (Pure CNN):

Why chosen as representatives:

- **Large adoption** in agriculture [1]
- **Good precision** and low dataset requirements vs. other CNNs [2] [3]
- **Speed-accuracy** optimizations

Variants tested:

- **YOLOv5**: CSPDarknet53 backbone
- **YOLOv8**: Improved architecture with decoupled head

[1] Badgujar et al. Agricultural object detection with YOLO algorithm. *Comput. Electron. Agric.*, 223:109090, 2024

[2] Tan et al. EfficientDet: Scalable efficient object detection, 2020

[3] Zhang et al. Comparison of YOLO-based sorghum

RT-DETR (Transformer-mixed):

Why chosen:

- **Outperforms** YOLOv5 and YOLOv8 [4]
- **Real-time** transformer detector
- **Hybrid approach**: CNN backbone + transformer decoder

Architecture advantages:

- Enhanced feature extraction
- Better handling of spatial relationships
- Improved performance with limited data

[4] Zhao et al. DETRs beat YOLOs on real-time object detection. *arXiv:2304.08069*, 2024

Object Detection Paradigms for Plant Counting

Many-Shot Models:

- **CNN-based**: YOLOv5, YOLOv8
- **Transformer-mixed**: RT-DETR, YOLO11
- Require extensive labeled datasets
- State-of-the-art performance

Zero-Shot Models:

- **OWLv2**: Open-vocabulary detection
- Vision-language foundation models
- Text prompt-based detection
- No training data required

Few-Shot Models:

- **CD-ViTO**: Cross-domain adaptation
- Meta-learning approaches
- 1-50 training examples
- Promising but unvalidated for agriculture

Handcrafted Methods:

- **Color thresholding** + agronomic knowledge
- High precision in constrained scenarios
- Still used as baseline/annotation tool
- Limited generalizability

Study Objectives and Research Questions

Primary Objective:

Determine minimum dataset size and quality required to achieve EPPO benchmarks ($R^2 \geq 0.85$) for maize seedling detection across different object detection paradigms.

Specific Research Questions:

1. What is the impact of **dataset source** (in-domain vs. out-of-distribution)?
2. How do **model architectures** affect dataset requirements?
3. What is the minimum acceptable **annotation quality**?
4. Can **few-shot/zero-shot** approaches meet agricultural benchmarks?
5. What role do **handcrafted methods** play in the DL era?

Case Study Focus:

Maize seedlings (Zea mays L.) at V3-V5 growth stage from georeferenced orthomosaics

Dataset Collection and Preparation

Dataset Sources:

Out-of-Distribution (OOD):

- Scientific literature: 778 tiles
- Internet repositories: 390 tiles
- Pre-annotated datasets

In-Domain (ID):

- 3 study sites: 450 training + 60 test tiles
- Phantom 4 Pro v2.0 @ 10m AGL
- Bundle adjustment error: 38mm (GNSS VRS-NRTK)

Technical Specs:

- **Resolution**: 5 mm/pixel
- **Tile size**: 224×224 pixels
- **Coverage**: 1.12×1.12 meters
- **Content**: 2 maize rows per tile
- **Annotation**: Squared bounding boxes centered on stems

Key Insight:

Tile size optimized for row pattern identification and model compatibility

Handcrafted Object Detector: Two-Stage Pipeline

Stage 1 - HC1 (Detection):

1. **Color thresholding** in HSV space
2. **Connected components** analysis
3. **Size filtering** based on leaf area
4. Outputs: Potential plant regions

Result: High recall, many false positives

Stage 2 - HC2 (Verification):

1. **RANSAC** line fitting for row detection
2. **Row spacing** validation
3. **Plant count** verification per row
4. **Agronomic knowledge** application

Result: High precision, limited coverage

Algorithm Performance:

Dataset	R^2	Coverage
ID_1	0.95	7.8%
ID_2	0.93	4.2%
ID_3	0.87	1.8%

Trade-off:

Excellent accuracy on subset of data vs. limited generalizability

Deep Learning Model Configuration

Many-Shot Models (Ultralytics Implementation):

CNN-based:

- YOLOv5 (n, s, m, l, x)
- YOLOv8 (n, s, m, l, x)

Transformer-mixed:

- YOLO11 (n, s, m, l, x)
- RT-DETR (l, x)

Training Settings:

- Batch size: 16
- Max epochs: 200
- Early stopping: 15 epochs
- Mixed precision training
- Default Ultralytics augmentation

Few-Shot: CD-ViTO

- ViT-S/B/L backbones (22M/86M/307M params)
- 1, 5, 10, 30, 50 shots tested
- Cross-domain adaptation

Zero-Shot: OWLv2

- ViT-B/16, ViT-L/14 encoders
- Base, fine-tuned, ensemble variants
- 11 different text prompts tested

Experimental Design and Evaluation Metrics

Dataset Size Investigation:

- **Many-shot**: 10-150 images in steps of 10
- **Few-shot**: 1, 5, 10, 30, 50 shots
- **Zero-shot**: No training data required

Dataset Quality Investigation:

Annotation quality reduced from 100% to 10% in 10% steps for successful models

Evaluation Metrics:

Counting Performance:

- R^2 (coefficient of determination)
- *RMSE* (root mean square error)
- *MAPE* (mean absolute percentage error)

Detection Performance:

Performance Modeling:

Empirical functions tested:

- $f(x) = a\ln(x) + b$
- $f(x) = a\arctan(bx) + c$
- $f(x) = ax^{1/b} + c$

Best fit selected by R^2 (GoF)

Testing Protocol and Infrastructure

Hardware Configuration:

- **CPU**: Intel Xeon E5-2670 v3 @ 2.30GHz
- **RAM**: 64.0 GB
- **GPU**: NVIDIA RTX A5000 (24GB VRAM)
- **Implementation**: Ultralytics, HuggingFace Transformers

SAHI Testing Method:

1. **Slice**: Test images into overlapping patches
2. **Detect**: Run model on each patch
3. **Merge**: Combine predictions with NMS
4. **Threshold**: Apply confidence score filtering

Rationale: Handles object occlusion at tile boundaries

Confidence Thresholds:

0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.29, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99

Best R^2 selected across all thresholds for each model

Key Principle:

Critical Impact of Dataset Source

Major Finding:

NO out-of-distribution model achieved benchmark performance ($R^2 \geq 0.85$)

OOD Results:

- **Best R²**: < 0.5 (all models)
- **Best MAPE**: 20% (still insufficient)
- **GoF values**: < 0.2 (poor predictability)
- **Dataset size**: Up to 1,168 images tested

Domain Gap Challenge:

- Environmental conditions

ID Success Stories:

Models achieving $R^2 \geq$ 0.85:

- YOLOv5n: 130 samples
- YOLOv5s: 130 samples
- YOLOv8n: 110 samples
- RT-DETR L: 60 samples
- RT-DETR X: 100 samples

GoF values: > 0.3 (high predictability)

Architecture-Specific Dataset Requirements

CNN-based Models (YOLO family):

- **YOLOv5n**: 130 samples (1.9M params)
- **YOLOv5s**: 130 samples (7.2M params)
- **YOLOv8n**: 110 samples (3.2M params)

Pattern: Larger models → more samples needed

Transformer-mixed Models:

- **RT-DETR L**: 60 samples (60M params)
- **RT-DETR X**: 100 samples (80M params)

RT-DETR L most efficient

Key Insights:

- **Transformers more sample-efficient** than CNNs
- **Long-range dependencies** better captured
- **Trade-off**: Higher computational cost

Practical Decision:

CNN approach: Collect 130 samples + lower compute
Transformer approach: Collect 60 samples + higher compute

Performance Predictability:

Logarithmic relationship between dataset size and performance enables resource planning (GoF > 0.3 for successful models)

Dataset Quality Requirements

Quality Tolerance Analysis:

Models achieving benchmark with reduced annotation quality:

Successful Quality Reductions:

- **YOLOv5n**: 85% quality (130 samples)
- **YOLOv5s**: 90% quality (130 samples)
- **YOLOv8n**: 85% quality (110 samples)
- **RT-DETR X**: 65% quality (100 samples)

RT-DETR L Sensitivity:

Failed to maintain benchmark with ANY quality reduction (60 samples baseline)

Quality-Quantity Relationship:

Key Finding: Smaller datasets are more sensitive to annotation errors
RT-DETR L: Minimal dataset (60 samples) $\rightarrow$ Each annotation critical
RT-DETR X: Larger dataset (100 samples) $\rightarrow$ Error redundancy tolerance

Practical Strategy:

Option 1: Perfect annotations + minimal dataset
Option 2: Good quality annotations + larger dataset
Option 3: Semi-automated annotation workflows

Few-Shot and Zero-Shot: Current Limitations

Major Finding:

Neither few-shot nor zero-shot approaches achieved benchmark performance

Few-Shot Results (CD-ViTO):

Best Performance (ViT-B, 50 shots):

- **RMSE**: 3.9 (vs. benchmark 0.39)
- **MAPE**: 25%
- **mAP**: 0.5
- **Error rate**: 1 plant in 4 miscounted

Pattern: Performance plateaus after 30 shots

Zero-Shot Results (OWLv2):

- **R^2**: Always < 0 (worse than mean prediction)
- **RMSE**: 5-25 (extremely high)
- **MAPE**: 40-140%
- **Prompt sensitivity**: High variability across 11 prompts

Why the Poor Performance?

The Value of Hybrid Approaches

Handcrafted Method Performance:

Strengths:

- $R^2 = 0.87$-0.95 (excellent accuracy)
- RMSE $= 0.11$-0.18 (below benchmark)
- Domain knowledge integration
- High precision when applicable

Limitations:

- Coverage: 1.8-7.8% of tiles only
- Color-thresholding bias
- Limited generalizability

Hybrid Strategy Potential:

1. **Bootstrap Training**:
   - HC method generates high-quality annotations
   - Deep learning models trained on HC output
   - Overcomes manual annotation bottleneck
2. **Quality Filtering**:
   - OOD/few-shot/zero-shot occasional good predictions
   - HC2 validates agronomic patterns
   - Reduces color-thresholding bias

Future Work Direction:

Practical Implementation Guidelines

Resource Optimization Strategy:

Step 1: Focus on minimum viable dataset size (60-130 images) **Step 2**: Logarithmic relationship $\rightarrow$ diminishing returns beyond minimum **Step 3**: Quality vs. quantity trade-off consideration

Implementation Pathways:

High-Resource Scenario:

- RT-DETR L + 60 perfect annotations
- Higher computational investment
- Fastest deployment

Medium-Resource Scenario:

- YOLOv8n + 110 good annotations
- Balanced compute/annotation effort
- Robust performance

Low-Resource Scenario:

- YOLOv5n + 130 annotations (85% quality)
- Semi-automated annotation

Critical Success Factors

1. **In-domain data**: Non-negotiable requirement
2. **Architecture choice**: Based on resource constraints
3. **Quality assessment**: Monitor annotation accuracy
4. **Validation protocol**: SAHI testing recommended

Industry Adoption

Conclusions: Evidence-Based Guidelines for Agricultural Object Detection

Core Findings

- **In-domain training data is mandatory** - OOD approaches fail to achieve benchmarks
- **Architecture matters**: Transformer-mixed models (RT-DETR) require 50% fewer samples than CNN-based models
- **Quality tolerance exists**: Models maintain performance with 65-90% annotation quality
- **Current limitations**: Few-shot and zero-shot methods cannot meet precision agriculture requirements

Practical Contributions

- **Minimum dataset requirements established**: 60-130 samples depending on architecture
- **Predictable performance scaling**: Logarithmic relationship enables resource planning
- **Hybrid approach potential**: Handcrafted methods valuable for