YOLO-BASED RECOGNITION OF SOME CROP CATEGORIES FROM REAL-WORLD AERIAL IMAGES

Salvatore Calcagno, Erika Scaletta, Emiliano Tramontana, Gabriella Verga

33rd IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2025)

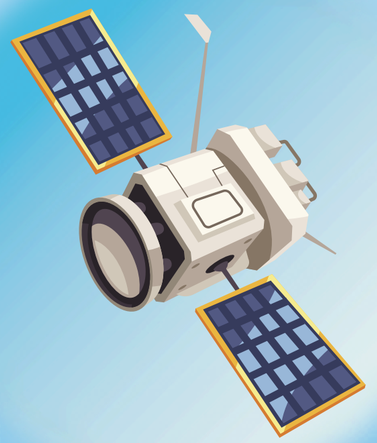
July 23rd-25th, 2025 - Catania, Italy

WHY DETECT CROPS AND OBJECTS FROM ABOVE?

- **Crop planning and monitoring**
  Useful for estimating gains, planning harvests, and managing costs

- **Crop health and yield forecasting**
  Detects plant stress, predicts productivity, and optimizes irrigation

- **Environmental impact**
  Enables mapping of deforestation and land use changes



CHALLENGES IN AERIAL IMAGE ANALYSIS

- Complex to analyze
- Difficult to find
- Computationally costly

3D IMAGE

SATELLITE IMAGE

OUR OBJECTIVE

Propose a solution to automatically detect crop categories and man-made objects in such images.

YOLO (You Only Look Once) is a family of **real-time object detection models** that process images in a single pass, making them extremely fast and efficient. YOLO models predict **bounding boxes** and **class probabilities** directly from full images, enabling quick and accurate object localization.

- **The model**
  - YOLOv11 → **Latest version** of the YOLO family
  - YOLOv11-seg → Enhanced for **image segmentation**

↓

Detects complex shapes with **pixel-level precision**

OUR CONTRIBUTIONS

1. **Creating** and **labeling** a real **dataset** based on aerial images extracted from Google Maps

2. **Applying YOLOv11-seg model for the first time** for real-world aerial image object detection, which represents a significant innovation, and allowing the identification of **objects having polygonal shapes** (whereas previous work only identified rectangular shapes)

3. The **complete recognition of objects** within an image

IMAGE FEATURES

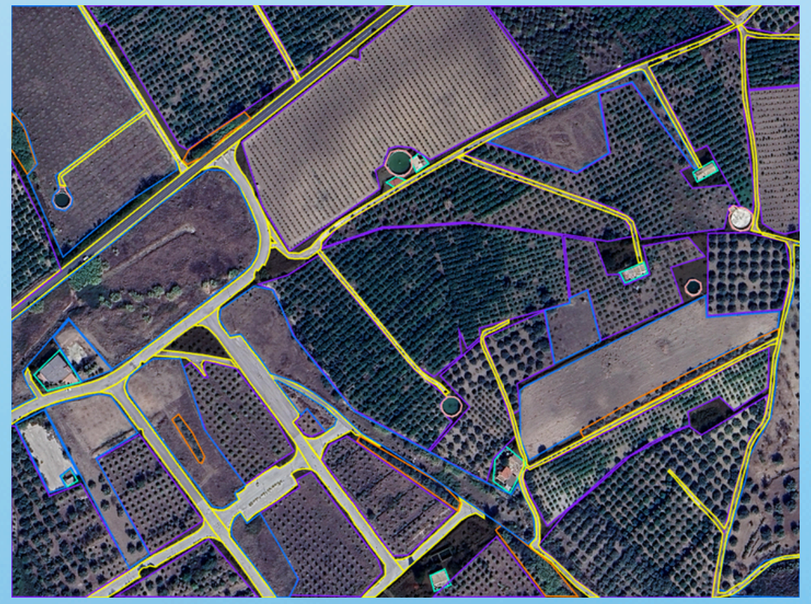
- Image source: **Google Maps**
- Geographical area: **Sicily**, in particular the provinces of Catania and Agrigento
- Resolution: **10240 x 7680 pixel** split into several parts
- Format: **PNG** with **RGB** color channels
- Source rationale: Google imagery offers **high resolution**, useful for visual analysis, and is easily accessible for covering large areas

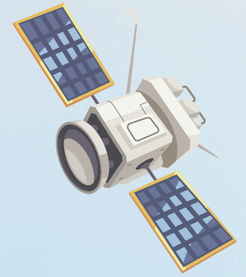
IMAGE LABELING

The dataset consisting of aerial images was labeled according to the YOLO standard and used to train the model

- Tool used: **Roboflow** (for **polygon labeling**)

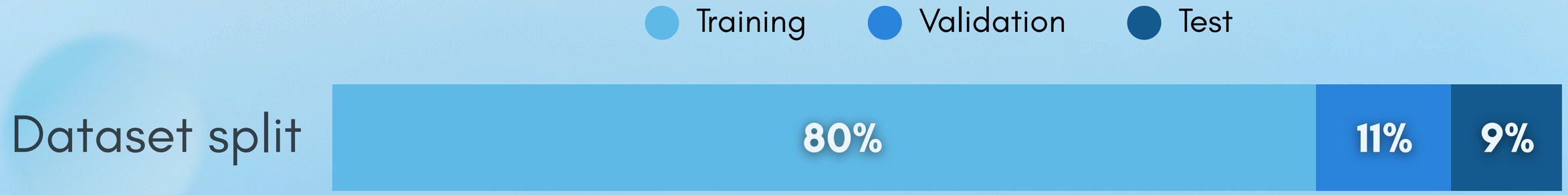
- Labeled categories:
  - Orange/lemon groves
  - Olive Groves
  - Houses
  - Roads
  - Trees
  - Lawn
  - Soil
  - Wells

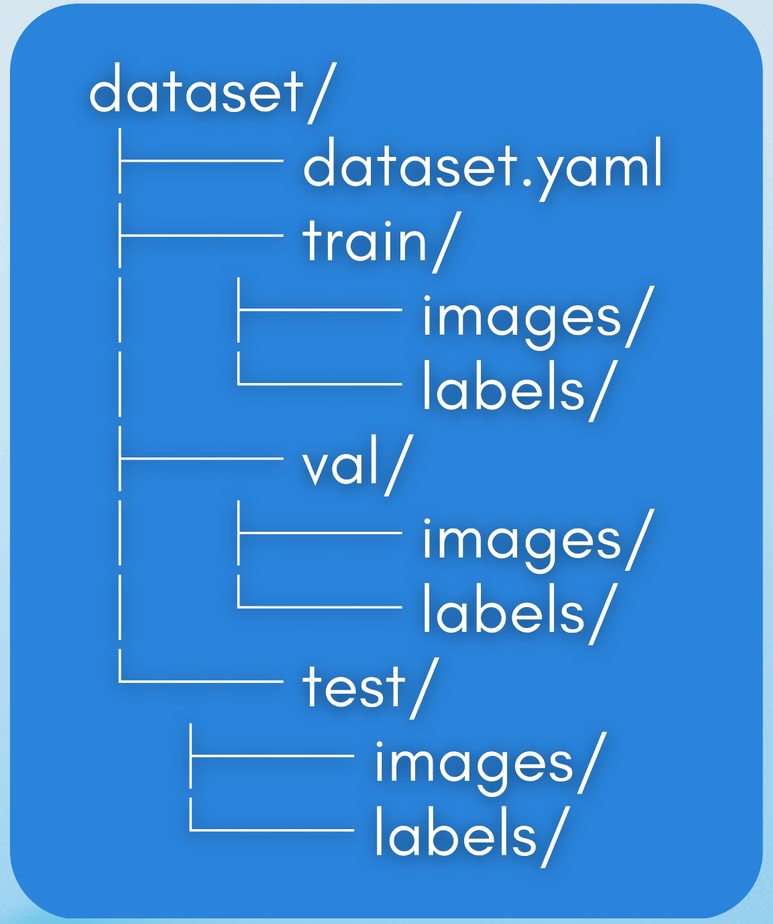




DATA SET DOWNLOAD

The images were downloaded using the **RoboFlow interface**.





The Dataset Files:

- **YAML file**
  - Includes paths for training, validation, and test
  - Contains list of category names

- Each **folder (train, val, test)** contains:
  - **images** → image files
  - **labels** → .txt files with polygon coordinates + class ID

DATASET CLEANING

- **Model requirements** → YOLOv11x-seg requires polygons with ≥ 5 points

- Custom **script** for validation:
  - Checks all labeled polygons
  - **Removes non-compliant** annotations/images

MODEL TRAINING

- **Tools & Framework** → Python + **Ultralytics** library

Training Workflow:

1. Model selection → **YOLOv11x-seg**
2. Training on our dataset
3. Object detection on new images

```
# Load a model
model = YOLO("yolov11x-seg.pt")
# Train the model
trainingResults = model.train(data="data.yaml", epochs=100,
      imgsz=640, device="mps")
# Perform object detection on an image
results = model("image1.jpg")
```

COMPUTATIONAL RESOURCES

Training Setup

- RAM: 32 GB
- CPU: Intel i7 12700H
- GPU: NVIDIA RTX 3050 Ti

Model Parameters

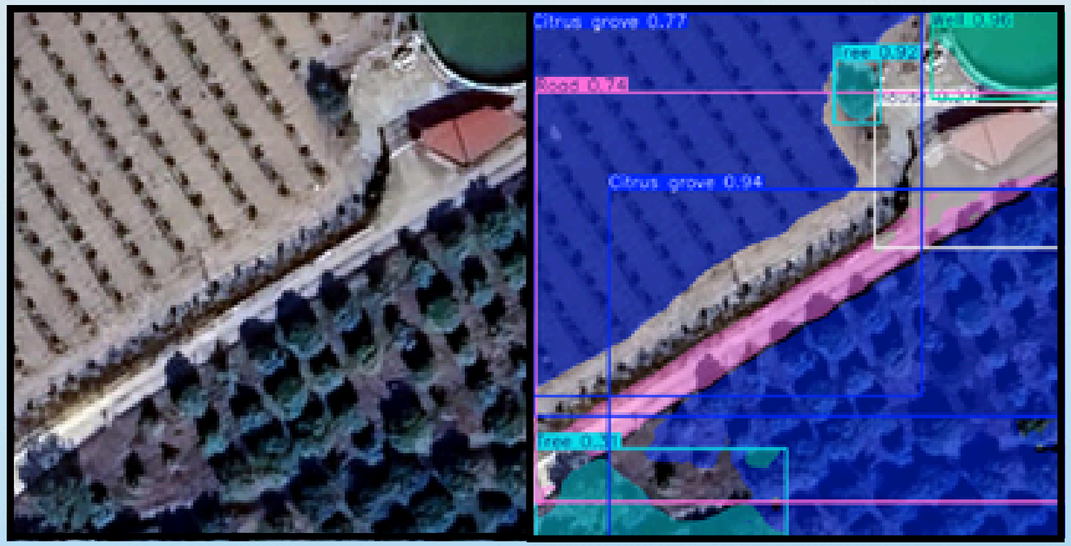
- device=cuda → GPU acceleration
- batch=5 → Prevents memory overload

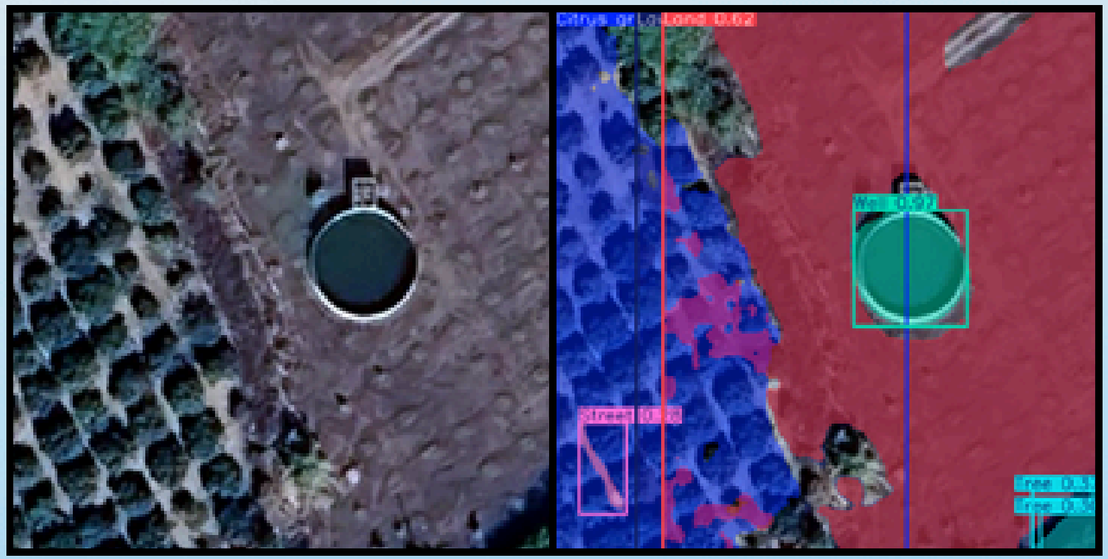
```
yolo detect train data=data.yaml model=yolo11x-seg.pt epochs=200
      imgsz=640 batch=5 device=cuda
```

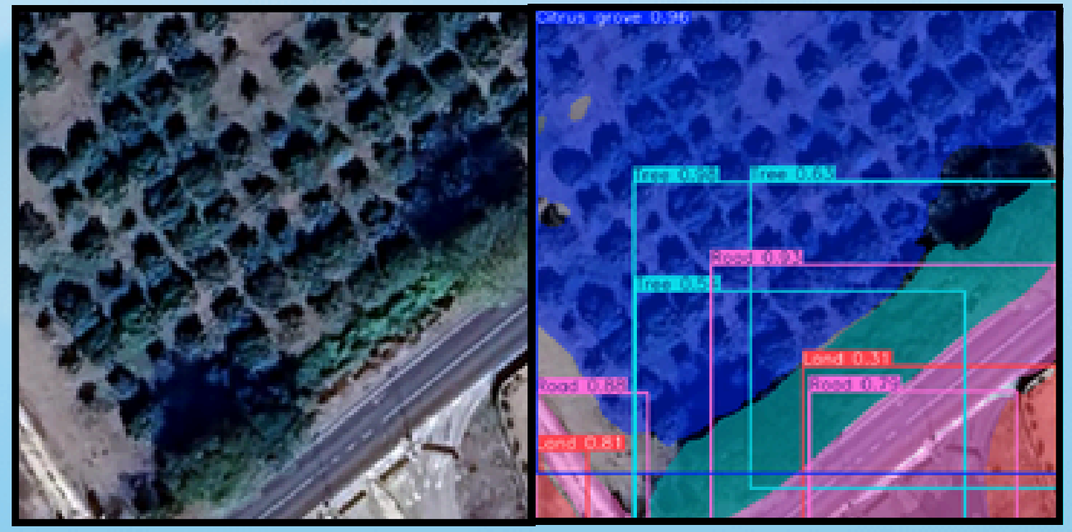
Performance

- Training Time: 20 hours on 150 images
- Inference Time: a few milliseconds per image

RESULTS


Citrus grove 0.77
Citrus grove 0.94

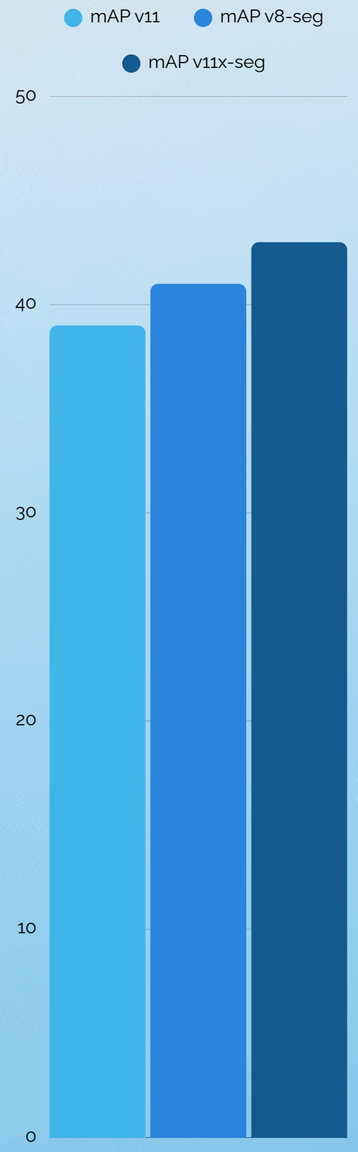

Land 0.62


Citrus grove 0.96
Land 0.51
Land 0.81

METRICS

Mean Average Precision values for each category given by several YOLO models trained on the prepared dataset of images

Category	mAP v11	mAP v8-seg	mAP v11x-seg
citrus grove	0.824	0.778	0.749
tree	0.291	0.286	0.275
house	0.242	0.319	0.269
well	0.446	0.584	0.68
lawn	0.374	0.42	0.319
road	0.304	0.258	0.365
land	0.243	0.233	0.283
olive grove	0.453	0.472	0.487
all classes	**0.397**	**0.419**	**0.428**



- **accuracy** = 0.823
- **precision** = 0.867
- **recall** = 0.942

CONCLUSIONS

- Aerial imagery + Deep Learning
- Custom dataset of labeled images
- Polygon-based object detection with YOLOv11x-seg
- High performance on real-world data
- Successful detection of crops and man-made objects

Thank You!

Erika Scaletta
University of Catania
erika.scaletta@phd.unict.it