

STAR-YOLO26: Synthetically Trained Amodal Recovery of Broccoli Heads

1. Introduction

- Broccoli are laboriously harvested by humans.
- Underweight broccoli heads cannot be sold in Dutch supermarkets, increasing waste and reducing profits.
- Automated harvesting requires head segmentation for subsequent size and weight estimation.
- These models struggle when a leaf covers the head, leading to underestimation of head area.
- This paper improves head segmentation under occlusion through a synthetically occluded dataset
- This trained model, STAR-YOLO26, is capable of inferring the full (amodal) head of occluded broccoli, while preserving the edge viability of YOLO26.

3. Results

- Tested on a dataset with real occlusions and known ground truth.
- The ORCNN model was pretrained with real occlusions.
- STAR-YOLO26 significantly surpasses the modal ORCNN in all measures except caliper error.
- Amodal ORCNN generally outperforms STAR-YOLO26 except in FPS and Area Bias.
- STAR-YOLO26 performs better under lighter occlusions due to the lack of shape priors.

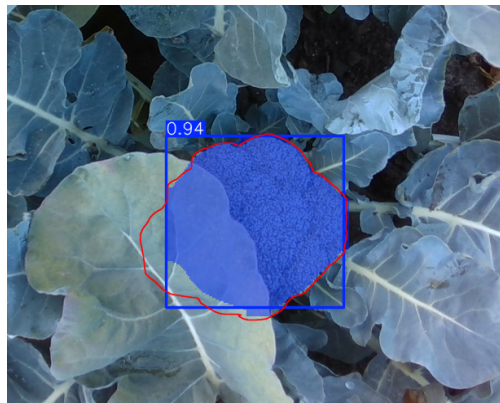


Figure 5: Inference Result on Real Image
Blue: Inference from STAR-YOLO26,
Red: Ground Truth

Model	Area MAPE	Area Bias	Caliper Error	AP@50:95	AR@50:95	FPS
STAR-YOLO26	19.63%	-13.4%	7.92%	42.1%	49.9%	56.34
ORCNN Modal	42.79%	-42.8%	7.46%	11.0%	21.3%	7.93
ORCNN Amodal	18.61%	+16.1%	4.98%	62.4%	66.5%	8.02

Table 1: Results on Real Occlusion Dataset
AP/R 50 = Average Precision/Recall with an Intersection over Union (IoU) of 50%
50:95 Average of all scores with an IoU of 50 to 95 with 5 points per step
Area Mean Absolute Percentage Error (MAPE), Area Bias and Caliper Error were measured with a 25% IoU threshold

2. Methodology: From Modal To Amodal

- There are widely available models that can only detect the visible (modal) part of a broccoli head.
- These models may be able to accurately infer the amodal head, given amodal training data.
- Data required to train the model for amodal segmentation is not widely available as it needs to show different occlusion rates with known ground truth, which is time consuming to collect.
- Synthetic occlusions were utilized instead, which are much faster to create, using a dataset with fully visible broccoli heads.
- Leaves were extracted (fig. 1) and placed on top of a completely visible broccoli head (fig. 2) to create occlusions with known ground truth (fig. 3).
- Six synthetic occlusions were generated per fully visible broccoli head, each with a single leaf.
- The superimposed leaves appear unnatural due to color differences (fig. 3) thus blending was applied, the results (fig. 4) was used to train the model.



Figure 1: Subset of leaves extracted from the dataset



Figure 2: Fully visible broccoli head



Figure 3: Synthetic occlusion before blending



Figure 4: Synthetic occlusion after blending, Red: Amodal Mask, Green: Occluding Leaf

4. Conclusions and Future Work

1. STAR-YOLO26 successfully infers the amodal heads, rather than just tracing visible modal pixels.
2. A synthetically occluded dataset is enough for the model to learn to amodally segment, without necessitating architectural changes thus also preserving the speed of the underlying model (YOLO26).
3. Amodal ORCNN benefits from shape priors in heavy occlusions but systematically overestimates the area.
4. STAR-YOLO26's lack of shape priors makes it superior for lighter occlusions but tends to underestimate the area.