S³AD: Semi-supervised Small Apple Detection in Orchard Environments

Robert Johanson, Christian Wilms, Ole Johannsen, and Simone Frintrop
University of Hamburg {firstname.lastname}@uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Motivation

- Apple detection for agricultural applications in orchards is challenging: Little data, small apple size, occlusions, ...
- **Idea 1**: Semi-supervised apple detection + new dataset
- **Idea 2**: Selective tiling for improved/efficient small apple detection

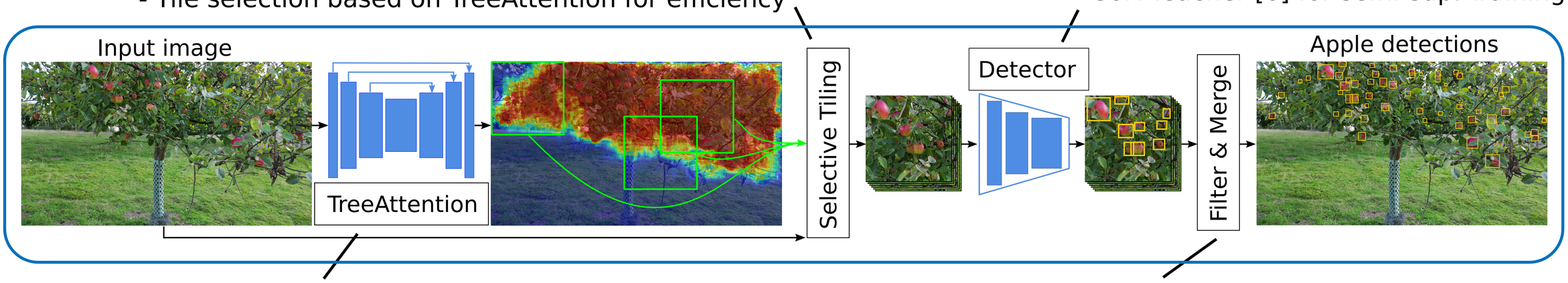
We propose...
... the Monastery Apple Dataset for semi-sup. apple detection
... S³AD as an attention-based semi-sup. apple detection system

Dataset: MAD

- One apple tree part per image
- 14,667 manually annotated apples across 105 images
- 4,440 unlabeled images
- 4k resolution

S³AD: Semi-supervised Small Apple Detection

- Tiling for improved small apple detection
- Tile selection based on TreeAttention for efficiency



- Locate tree crown with learned attention
- 3-stage U-Net predicts per-pixel attention

- Apple detection per tile
- Faster R-CNN with FPN [3]
- Soft Teacher [6] for semi-sup. training

- Remove detections along border
- Merge results and apply NMS

Results on MAD

Comparison to fully-sup. (small) object detection systems on MAD

System	Backbone	Semi-sup.	AP↑
SNIPER [1]	R-101		0.387
AutoFocus [2]	R-101		0.417
Faster R-CNN [3]	R-50		0.368
PANet [4]	R-50		0.378
Deformable DETR [5]	R-50		0.385
S³AD w/o tiling (ours)	R-50	X	0.408
S³AD (ours)	R-50	X	**0.423**

We outperform all other systems!

S³AD (ours) semi-sup.

Faster R-CNN fully-sup.

Conclusion

- Large-scale apple detection dataset (4545 img.)
- First semi-supervised apple detection system
- Selective tiling based on TreeAttention for improved/efficient small apple detection
- Outperforming fully-supervised systems

[1]: Singh et al.: SNIPER: Efficient multi-scale training. NeurIPS'18
[2]: Najibi et al.: AutoFocus: Efficient multi-scale inference. ICCV'19
[3]: Lin et al.: Feature pyramid networks for object detection. CVPR'17
[4]: Liu et al.: Path aggregation network for instance segmentation. CVPR'18
[5]: Zhu et al: Deformable DETR: Deformable transformers for end-to-end object detection. ICLR'20
[6]: Xu et al.: End-to-end semi-supervised object detection with soft teacher. ICCV'21