# Classification and Segmentation of photovoltaic and solar thermal systems from aerial imagery

Rienk Fidder, Utrecht University.
Supervised by Mang Ning and Itir Önal Ertuğrul (Utrecht University) and Ward de Hond and Marjolein Hordijk (We-Boost)

## 1 Introduction

Photovoltaic (PV) and Solar Thermal (ST) panels mounted on rooftops form a cornerstone in the transition to fully renewable energy generation. However, due to the large gap in data on the number and location of these panels, policy-makers have trouble determining the effectiveness of policies and energy network administrators have trouble building efficient networks. In this study, a model was proposed to automatically classify and segment PV and ST panels from aerial imagery to alleviate this issue. The main contributions of the study are: (1) a novel manually annotated dataset for PV and ST locations on Dutch aerial imagery made publically available [2]; (2) a two-stage pipeline for PV and ST classification and segmentation; (3) an approach to incorporate building information in the model in an attempt to improve results; (4) a novel method for weakly-supervised pseudo-label generation based on greedy CAM refinement and SAM generated segmentations; (5) application of semi-supervised learning in PV and ST segmentation.

## 2 Methodology

A novel dataset of aerial images in the Netherlands, containing image-level and pixel-level annotations of PV and ST panels was build and made publicly available [2]. The dataset contained 50,000 image-level labelled samples, and 4,896 pixel-level annotated samples. This dataset was then used to train both a classification and segmentation model, to build the final 2-stage pipeline. Both models were based on ConvNeXtV2 [6], with a DeepLabV3+ [1] segmentation head used for segmentation.

Masked auto encoder [3] pre-training was employed utilising images from the target domain. For training of the classification model, the image level labels were used for supervised training with the following variants: binary (presence vs. absence of panels) and multi-label (PV and/or ST). Additionally, utilising building polygons as a fourth input channel to the network was experimented with.

For segmentation, a standard fully-supervised approach was tested, as well as two novel approaches utilising semi-, and weakly-supervised methods. In the fully supervised case, pixel level training labels were used only. In the semi-supervised case, the remaining images with only image-level labels were used as unlabelled samples in the CorrMatch framework [5], which was adjusted to use ConvNeXtV2 as a backbone rather than ResNet. Finally, a segmentation network was trained in a weakly-supervised manner (See Figure 1). This was achieved by utilising class activation maps of a greedily retrained classification model to select class agnostic masks generated by the segment anything model [4]. Finally, CorrMatch was again used such that only high-confidence pseudo masks could be incorporated as labelled samples, and the remaining samples could be used as unlabeled samples.

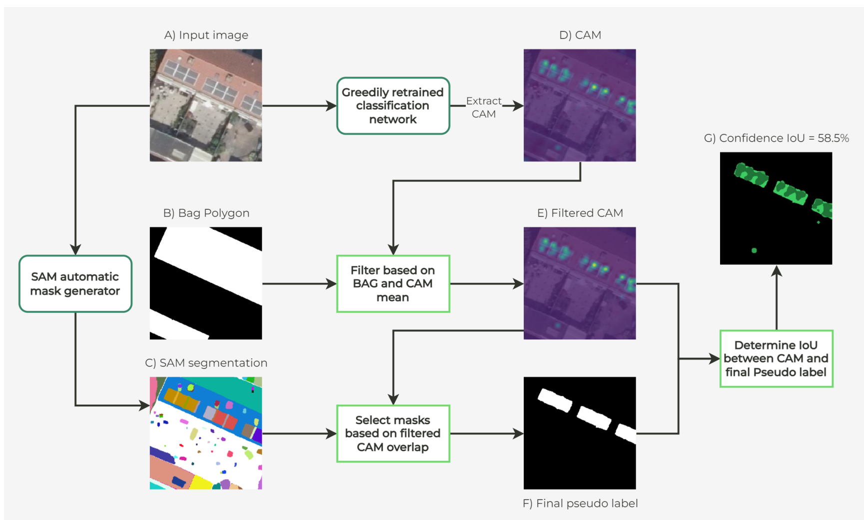


Figure 1: Figure depicting the strategy used to generated pseudo masks for weakly supervised segmentation by utilising class activation maps and the segment anything model.

# 3 Results

The model is shown to exhibit strong classification performance, after finetuning models pre-trained either on ImageNet or Dutch aerial images with F1 scores of 0.954 and 0.946 respectively on the binary task. Performance of fully-, semi-, and weakly-supervised segmentation models is evaluated. It is shown that the best performance is achieved by combining a small set of manually annotated mask labels with a larger set of unlabelled data in a semi-supervised manner. This semi-supervised approach leads to an IoU of 73.3% for binary segmentation, and a class-specific IoU of 77.0% and 37.6% is achieved for the PV and ST classes respectively.

## References


[1] Liang-Chieh Chen et al. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. 2018. arXiv: 1802.02611 [cs.CV]. URL: https://arxiv.org/abs/1802.02611.

[2] Rienk Fidder. *Dutch aerial imagery and building masks with photovoltaic and solar thermal panel annotations*. June 2024. URL: https://public.yoda.uu.nl/science/UU01/NRFYSC.html.

[3] Kaiming He et al. *Masked Autoencoders Are Scalable Vision Learners*. 2021. arXiv: 2111.06377 [cs.CV]. URL: https://arxiv.org/abs/2111.06377.

[4] Alexander Kirillov et al. *Segment Anything*. 2023. arXiv: 2304.02643 [cs.CV]. URL: https://arxiv.org/abs/2304.02643.

[5] Boyuan Sun et al. *CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation*. 2023. arXiv: 2306.04300 [cs.CV]. URL: https://arxiv.org/abs/2306.04300.

[6] Sanghyun Woo et al. *ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders*. 2023. arXiv: 2301.00808 [cs.CV]. URL: https://arxiv.org/abs/2301.00808.