Solar Panel Segmentation on Aerial Images using Color and Elevation Information

Gerrit Luimstra and Kerstin Bunte

University of Groningen - Faculty of Science and Engineering

Abstract. The automatic detection of solar panels from aerial imagery is highly desirable for energy planning and urban development in the Netherlands, where such data has not been extensively explored. To address this gap, we publicise a new annotated dataset, tailored for the Dutch landscape, and compare several state-of-the-art semantic segmentation models. While traditional approaches primarily utilize RGB data, we incorporate elevation and angle information in the model to analyse its potential benefit. We achieved satisfactory performance for automated solar panel detection and segmentation, with surface estimates only diverging by 1m within a 900m^2 area. The additional elevation information does not improve the performance significantly, but is more robust in certain cases.

1 Introduction

The use of renewable energy is critical for energy planning and urban development [1]. To assess the adoption of solar energy in Emmen in the Netherlands the municipality is investigating the automatic detection of solar panels from aerial imagery. A system detecting solar panels from aerial imagery would allow for evaluation of solar panel installation trends over time, supporting better policy development for future sustainable energy initiatives. While the automatic detection of objects, such as roads and buildings, in aerial imagery has gained significant attention in recent years [2, 3], the segmentation of solar panels is a relatively new field, that nevertheless has produced promising results [4, 5].

Initial work in solar panel detection include Support Vector Machine classifiers [6] applied to regions of interest extracted from satellite images using the Maximally Stable Extremal Regions algorithm. Since these methods only identify the presence of solar panels, subsequent approaches used random forests for pixel segmentation to obtain more accurate size and shape estimates [7]. Convolutional Neural Networks (CNNs) began replacing traditional classifiers [8] with methods such as AlexNet, and SegNet [9] learning features automatically end-to-end, capturing multi-stage information from aerial images. Recent architectures like U-Net, DeepLabV3, and the Feature Pyramid Network (FPN) [10, 11, 12] were specifically designed for semantic segmentation tasks and became the standard for generating high resolution segmentation masks from aerial images as they can process entire images, rather than smaller patches, which was a limitation of earlier methods.

In this study, we extend solar panel segmentation to incorporate both RGB and elevation data. The aerial imagery dataset[1], captured at a resolution of 7.5 cm per pixel, is supplemented with elevation data to provide height and tilt

[1] Annotated aerial imagery dataset available at https://doi.org/10.5281/zenodo.14860030

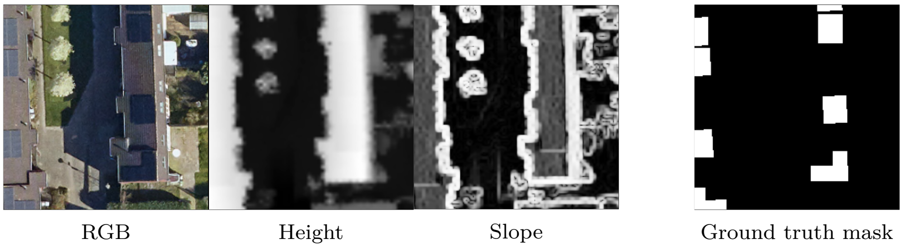

RGB Height Slope Ground truth mask

Fig. 1: An example input image (left) and the output ground truth mask (right).

information for each pixel. We analyse the effectiveness comparing the modified models with baseline methods on our real-world data set captured in the north of the Netherlands, to evaluate whether incorporating elevation and tilt data improves segmentation performance.

2 Data

USA datasets for solar panel segmentation [13, 14] cannot be used directly, due to large differences in landscape and foliage of the Netherlands. The government of Emmen produces an aerial image taken from a small plane every year under similar conditions. The spatial resolution of an image taken by a plane is much higher compared to a satellite (7.5cm per pixel instead of 30cm), resulting in sharper and more detailed images. The area Emmen spans 346.26km^2 of which four regions were selected including different types of buildings and vegetation totalling 18.55km^2. These areas were subdivided in cells of 30 by 30 meters, resulting in 20.618 squares of 900m^2. The solar panels in these areas were manually labelled with polygons, resulting in 4389 unique objects. The ratio of solar panel surface versus total area is very small. Therefore we only include 224 by 224 pixel RGB images from cells that either contain or are close to a cell with solar panels, to avoid large class imbalance. The resulting dataset contains a total of N=5327 annotated images of which 1743 contain solar panels.
We incorporate vertical placement and angle (tilt) from the Actueel Hoogtebestand Nederland (AHN) dataset[2]. It includes high resolution digital elevation models (DEMs) and LiDAR-derived point cloud data, offering precise elevation measurements for every half-meter land parcel relative to the Normaal Amsterdams Peil (NAP). The AHN covers the entire country with a minimum of 10 measurements per square meter obtained by plane-mounted laser scanning transformed into 3D point clouds and grids. The current version AHN4 was collected over the years 2020, 2021, and 2022. The AHN4 DTM ground level grid is generated using a Squared IDW method. To complement the height information the tilt is calculated using a 3 × 3 sliding window according to $\theta = \tan^{-1}\left(\sqrt{({}^{dz}/_{dx})^2 + ({}^{dz}/_{dy})^2}\right)$, where $\frac{dz}{dx}$ and $\frac{dz}{dy}$ denote the rate of change in the horizontal and vertical directions from the centre cell to each adjacent cell. The dataset[3] is augmented during training at the beginning of a new epoch with a probability p_{aug} for flipping, rotation or a brightness change. In case of a rotation the degree is determined uniformly between $[-30, 30]$.

[2] https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn

[3] All data and models are public at https://doi.org/10.5281/zenodo.14860030

3 Methods

We use EfficientNet [15] as the encoder backbone for its beneficial trade-off between computational cost and performance. Concretely, EfficientNet-B0 serves as the baseline, while B2 and B4 variations are explored for segmentation performance improvements. Pre-trained weights are utilized to leverage transfer learning. For segmentation, the U-Net [10], DeepLabV3 [11], and FPN [12] are used as decoders. U-Net incorporates skip connections to merge features from different stages of the encoder, enabling accurate boundary predictions, making it particularly effective for tasks requiring precise delineation of small or irregular objects such as solar panels. DeepLabV3 applies Atrous Spatial Pyramid Pooling (ASPP) to capture contextual information at multiple scales, which is advantageous for handling the varying sizes and spatial arrangements of solar panels within complex urban environments. FPN constructs a feature pyramid to extract and leverage information from multiple resolutions, improving the detection of small objects while maintaining the performance on larger features. All decoder outputs are upsampled to match the original resolution.

The standard input to these architectures is an RGB tensor $\mathbf{X}_i \in \mathbb{R}^{\langle H_i, W_i, C_i \rangle}$, where H_i, W_i and C_i represent the height, width and channels. Explicitly in our case with RGB this leads to $\langle 224, 224, 3 \rangle$. To incorporate the two additional channels for elevation information we employ data fusion [16] by expanding the input tensor to shape $\langle 224, 224, 5 \rangle$. This fusion enables the model to learn with additional spatial terrain features, while maintaining computational efficiency. Separate pathways for RGB and elevation data would lead to excessive model complexity and much slower training times, and an preliminary experiment suggested that after 200 epochs there was no discernable difference between the two approaches. Instead, the architecture is only minimally modified by replacing the first convolutional filter $\mathcal{F}_1$, originally designed for 3-channel input, by one with five channels. To achieve this, the pre-trained weights were first loaded into the model, and the weight kernel corresponding to the red channel was copied over an extra two times to handle the new channels. The rest of the pre-trained encoder remains unchanged and the entire model is trained as normal. In the experiments we evaluate four model variations: models trained on RGB data only, models including height information, models incorporating slope information, and models using both height and slope data.

4 Experiments

In this work, we use binary cross-entropy (BCE) loss to train the model for the pixel-wise binary classification task of image segmentation. BCE computes the average dissimilarity between the predicted segmentation mask and the ground truth. For N samples, it is defined as:

$$\mathcal{L}_{\text{mask}} = -\frac{1}{N \cdot 224^2} \sum_{i=1}^{N} \sum_{x,y} \Big[\mathbf{S}_i[x,y] \log(\mathbf{P}_i[x,y]) + (1 - \mathbf{S}_i[x,y]) \log(1 - \mathbf{P}_i[x,y]) \Big]$$

where $\mathbf{S}_i[x,y] \in \{0,1\}$ indicates the ground truth label for pixel (x,y), and $\mathbf{P}_i[x,y] \in [0,1]$ is the predicted probability.

In order to comprehensively evaluate the performance, we employ 5-fold stratified cross-validation and assess the models using three key metrics: accuracy, mean Intersection over Union (mIoU) and mean absolute area error (MAAE):

$$\text{mIoU} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{S}_i \cap \mathbf{P}_i}{\mathbf{S}_i \cup \mathbf{P}_i} \ , \qquad \text{MAAE} = \frac{1}{N}\sum_{i=1}^{N}\left|A_i - \hat{A}_i\right| \ ,$$

where A_i and $\hat{A}_i$ denote the area and estimated area respectively. The mIoU measures spatial accuracy by evaluating the overlap between predicted and ground truth masks, ensuring the model captures the target region while penalizing false positives and negatives. MAAE quantifies size precision by focusing on the absolute difference in object area between true A_i and predicted $\hat{A}_i$, making it particularly useful for tasks like solar panel size estimation. Additionally, accuracy measures whether the model correctly identifies the presence of any solar panel in an image, providing a binary image-level evaluation. In the experiments we trained the models for 200 epochs, used a fixed learning rate of $\eta = 0.0005$ and batch size of 64, as this showed good convergence and performance for all models tested. For the ADAM optimizer we used $\rho_1 = 0.9$ and $\rho_2 = 0.999$ and set the augmentation probability to $p_{\text{aug}} = 0.5$. For each model, we chose the epoch with the highest training mIoU and reported the metrics for this epoch.

5 Results and Discussion

Results in term of key metrics are summarized in Table 1 and demonstrate the effectiveness of the semantic segmentation models for detecting solar panels in aerial imagery. The U-Net and FPN architectures outperformed DeepLabV3, which is hence omitted from the table. It struggled with the data, achieving a lower mIoU of 0.740-0.793 and a higher MAAE of 1.726-1.299, with the best values being achieved by the EfficientNet-B4 variant with slope information (S). We could not find an obvious reason and increasing the number of parameters did not improve the scores. The U-Net and FPN models consistently achieved high mIoU values above 0.8 and an MAAE below 1.3 square meters per 900m^2 image. For all models tested, the performance is very robust with between fold variation within [0.006,0.021]. The choice of decoder architecture is more significant than the encoder used, across all the metrics. While the U-Net achieves the highest evaluation performance, the improvement over the FPN is not significant. Similarly to DeepLab increasing the depth of the EfficientNet backbone (from B0 to B4) did not return significantly higher performance. Concretely, for the baseline model with U-Net decoder using the EfficientNet-B4 with 17.1 million parameters compared to the EfficientNet-B0 model with 4.1 million parameters only improved the mIoU from 0.825 to 0.838. As the former requires about 10 times as much FLOPS, the 1.57% performance increase does not justify the cost. The inclusion of elevation information, whether height (H), slope (S), or both (HS), almost always improved performance over the baseline RGB-only models. Across all model variations, those utilizing slope information performed best on average, achieving the highest mIoU and the lowest MAAE in most configurations. While these improvements are consistent, they are on average not as substantial as one might have expected from adding complementary information.

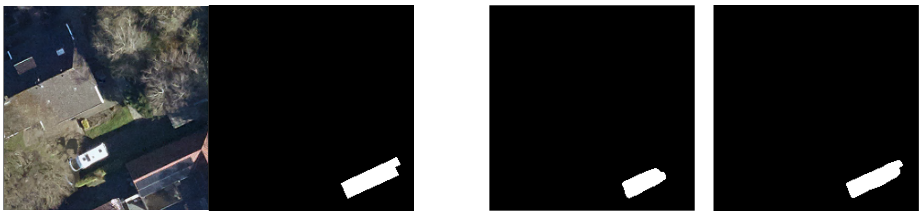
Table 1: Summary performance of different architectures and encoders. Abbreviations indicate whether the baseline (B), height (H), slope (S), or both height and slope (HS) were used to incorporate the AHN information. Arrows indicate the optimum.

	Architecture	Train metrics			Validation metrics		
		Acc↑	mIoU↑	MAAE↓	Acc↑	mIoU↑	MAAE↓
U-Net	B - EfficientNet-B0	0.979	0.912	0.366	0.938	0.825	1.154
	H - EfficientNet-B0	0.985	0.913	0.361	0.950	0.828	1.140
	S - EfficientNet-B0	0.979	0.913	0.382	0.945	0.833	1.139
	HS - EfficientNet-B0	0.977	0.912	0.370	0.943	0.830	1.117
	B - EfficientNet-B2	0.992	0.919	0.332	0.959	0.826	1.129
	H - EfficientNet-B2	0.989	0.919	0.340	0.958	0.830	1.180
	S - EfficientNet-B2	0.989	0.921	0.330	0.961	0.838	1.055
	HS - EfficientNet-B2	0.991	0.918	0.337	0.960	0.833	1.096
	B - EfficientNet-B4	0.995	0.929	0.275	0.965	0.838	1.031
	H - EfficientNet-B4	0.995	0.929	0.282	0.967	0.839	1.056
	S - EfficientNet-B4	0.996	**0.931**	0.273	**0.970**	**0.843**	**1.025**
	HS - EfficientNet-B4	**0.997**	0.929	**0.277**	0.968	0.838	1.037
FPN	B - EfficientNet-B0	0.988	0.889	0.430	0.954	0.799	1.233
	H - EfficientNet-B0	0.985	0.890	0.446	0.952	0.808	1.208
	S - EfficientNet-B0	0.984	0.892	0.436	0.955	0.813	1.187
	HS - EfficientNet-B0	0.986	0.889	0.448	0.956	0.806	1.292
	B - EfficientNet-B2	0.992	0.903	0.371	0.960	0.815	1.212
	H - EfficientNet-B2	0.991	0.903	0.376	0.956	0.816	1.198
	S - EfficientNet-B2	0.987	0.904	0.388	0.953	0.821	1.169
	HS - EfficientNet-B2	0.990	0.904	0.385	0.962	0.821	1.157
	B - EfficientNet-B4	0.997	0.919	0.301	0.968	0.827	1.099
	H - EfficientNet-B4	0.996	0.919	0.296	0.972	0.829	1.061
	S - EfficientNet-B4	0.996	0.921	0.293	0.967	0.837	1.055
	HS - EfficientNet-B4	0.997	0.919	0.296	0.969	0.829	1.052

However, models enhanced with elevation data did show greater robustness in certain challenging scenarios. As illustrated in Figure 2 the AHN-based model correctly segmented solar panels under shadow occlusion, where the baseline RGB model struggled. Hence, the AHN model is able to utilize the elevation information as the height and slope information helps deciding that a shadow pixel is still a solar panel due to its location on the roof.

6 Conclusion and Future Work

In this work we demonstrate the effectiveness of semantic segmentation models on our newly introduced aerial imagery dataset from the Netherlands. With a solar panel area estimation error of around 1m per 900m^2, our models are highly suitable for practical applications, such as renewable energy planning and resource assessment. The inclusion of elevation data further enhance robustness in challenging scenarios, emphasizing the potential of integrating spatial information for more reliable segmentation outcomes. In future work we investigate the encoding of shape constraints to capture solar panels even better [17].



RGB Ground truth RGB only RGB + HS

Fig. 2: An example where an AHN model outperforms the RGB only model

Acknowledgements: We thank the Municipality of Emmen for providing computational resources and Harry ten Brink for his insights as a geo-specialist.

References

[1] J. M. Malof, R. Hou, L. M. Collins, K. Bradbury, and R. Newell, "Automatic solar photovoltaic panel detection in satellite imagery," in *2015 ICRERA*, pp. 1428–1431, 2015.

[2] Z. Xu, Y. Liu, L. Gan, Y. Sun, X. Wu, M. Liu, and L. Wang, "RNGDet: Road network graph detection by transformer in aerial images," *IEEE TGRS*, vol. 60, pp. 1–12, 2022.

[3] A. Manno-Kovacs and T. Sziranyi, "Orientation-selective building detection in aerial images," *ISPRS Journal*, vol. 108, pp. 94–112, 2015.

[4] M. A. Wani and T. Mujtaba, "Segmentation of satellite images of solar panels using fast deep learning model," *International Journal of Renewable Energy Research*, 2021.

[5] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states," *Joule*, vol. 2, no. 12, pp. 2605–2617, 2018.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] J. M. Malof, K. Bradbury, L. M. Collins, and R. G. Newell, "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery," *Applied Energy*, vol. 183, pp. 229–240, 2016.

[8] J. M. Malof, L. M. Collins, K. Bradbury, and R. G. Newell, "A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery," in *2016 IEEE ICRERA*, pp. 650–654, 2016.

[9] J. Camilo, R. Wang, L. M. Collins, K. Bradbury, and J. M. Malof, "Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery," 2018.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[11] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *CoRR*, vol. abs/1606.00915, 2016.

[12] T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," *CoRR*, vol. abs/1612.03144, 2016.

[13] H. Jiang, L. Yao, N. Lu, J. Qin, T. Liu, Y. Liu, and C. Zhou, "Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery," *Earth System Science Data*, vol. 13, no. 11, pp. 5389–5401, 2021.

[14] G. Kasmi, Y.-M. Saint-Drenan, D. Trebosc, R. Jolivet, J. Leloux, B. Sarr, and L. Dubus, "A crowdsourced dataset of aerial images with annotated solar photovoltaic arrays and installation metadata," *Scientific Data*, vol. 10, p. 59, Jan 2023.

[15] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th ICML* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *PMLR*, pp. 6105–6114, PMLR, 09–15 Jun 2019.

[16] W. Zhang, H. Huang, M. Schmitz, X. Sun, H. Wang, and H. Mayer, "Effective fusion of multi-modal remote sensing data in a fully convolutional network for semantic labeling," *Remote Sensing*, vol. 10, no. 1, 2018.

[17] L. Castrejón, K. Kundu, R. Urtasun, and S. Fidler, "Annotating object instances with a polygon-rnn," *CoRR*, vol. abs/1704.05548, 2017.