

Deep Learning Method for Heliostat Instance Segmentation

Benjamin Liu¹[\[https://orcid.org/0000-0003-2327-6117\]](https://orcid.org/0000-0003-2327-6117), Alexander Sonn¹, Anthony Roy¹, and Brian Brewington¹

¹ Heliogen. 130 W Union St. Pasadena, CA 91103, USA.

Abstract. Heliostat instance segmentation (HST-IS) is a crucial component of the heliostat tracking system at Heliogen’s Lancaster test facility. The system estimates the mirror normal of each heliostat by performing a nonlinear optimization-based fitting strategy using approximations of the non-shaded, non-blocked sunlit pixels on each heliostat, and the tracking system uses these estimates to improve performance.

HST-IS is fundamentally challenging due to variability in lighting conditions and heliostat size relative to the capturing camera. Deep learning-based convolutional neural networks (CNN) have emerged in recent years by demonstrating noteworthy precision in tasks such as object recognition, detection, and segmentation. CNN-based methods offer a robust augmentation to HST-IS methods as they capture a context-less hierarchy of image features.

In this study, we developed deep learning models to automatically segment heliostat instances from elevated images taken from the field. We study various image parameters and architectural customizations to optimize for scalability, robustness, and accuracy in our predictions. We perform robust evaluations of our best model to quantify gaps between model development and real-world deployment and provide evidence for utility in the field.

Keywords: Deep Learning, Instance Segmentation, Heliostat, Tracking

1. Introduction

1.1 Related Work

HST-IS is useful in many concentrated solar power measuring contexts that require a pixel-precision understanding of heliostat positioning, particularly in soiling detection and calibration [1, 2, 3]. In many of these downstream applications, the need for automation and consistency in recalibration cadence are crucial. This is due to the tendency for field conditions to change key parameters such as soiling conditions and heliostat orientation [4].

Several previous studies leverage computer vision techniques to study the properties of heliostats. Roger et al. developed an edge detection technique that extracts heliostat vertices to calculate the surface normal, demonstrating results comparable to manual photogrammetry at a fraction of the required time [5]. Coventry et al. leveraged computer vision thresholding methods to determine the soiling levels of heliostat mirrors from aerial imagery by segmenting mirrors from background pixels [6]. Ydrissi et al. designed and trained a convolutional neural network on images of soiled mirror images to optimize the task of

predicting reflectivity, dust density, and cleanliness values [7]. Computer vision-based approaches for HST-IS frequently assume inputs of images that feature a single heliostat, which are difficult to collect at scale [5, 6, 7]. This limitation highlights the value of generalizable, automated HST-IS approaches that require minimal context.

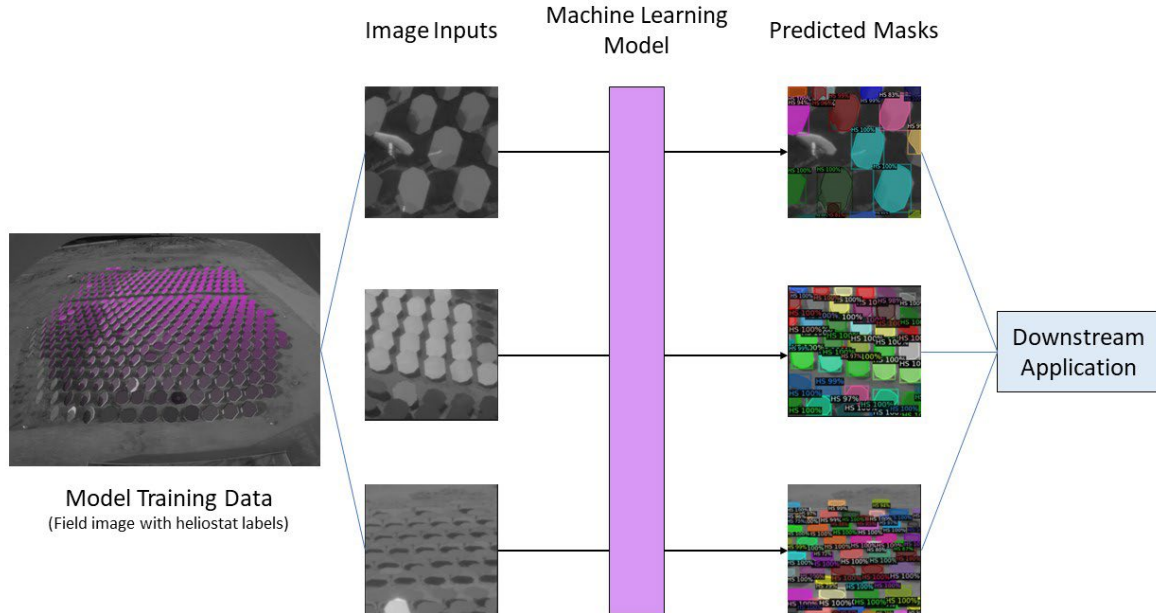


Figure 1. Graphical summary illustrating process of deep learning HST-IS method.

1.2 Heliostat Instance Segmentation

Heliostat instance segmentation (HST-IS) is a significant underlying process in the Heliostat tracking system at the Lancaster test facility. This system includes a module called SOHOT [1] (“System for Observing Heliostat Orientations While Tracking”) that estimates the mirror normal of each heliostat by performing a nonlinear optimization to find the mirror orientations that best fit radiance-proportional values for regions of interest (ROIs) corresponding to heliostats in images. These estimates are used to train per-heliostat kinematic models.

HST-IS from SOHOT imagery is fundamentally challenging for two main reasons. Firstly, shading and blocking patterns are inconsistent since heliostats appear in a variety of lighting conditions (Fig. 1). With these inconsistencies, naive patterns and shapes cannot be used to distinguish ROI pixels from non-ROI pixels. Secondly, sizes of heliostats in each image vary drastically due to variable distances from the capturing camera. These differences in scale present significant challenges to many standard computer vision methods that do not have parameters corresponding directly to an image’s receptive field. In SOHOT, error in ROIs contribute to error in normal estimates, and existing ROI calculation techniques are susceptible to subtle errors in camera calibration and positional assumptions. These sources of error can render inconsistent or inaccurate ROI calculation results.

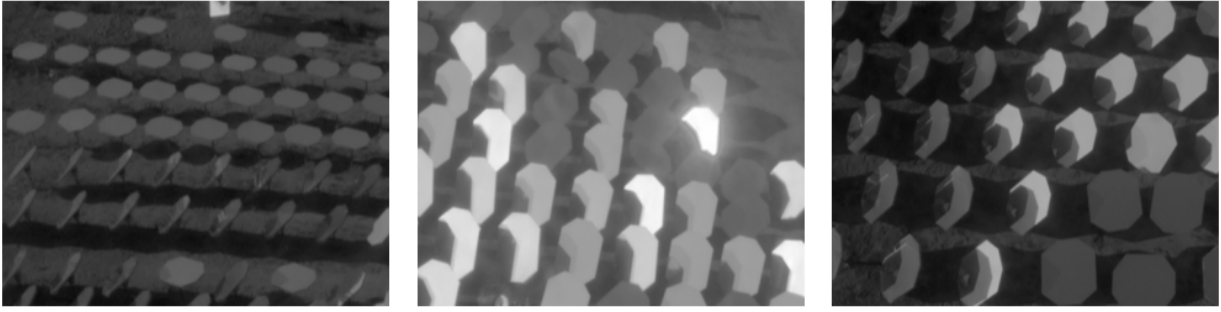


Figure 2. Examples of images in which heliostat segmentation proves challenging due to lightning conditions (right) and differences in heliostat scale and orientation (left, middle).

1.3 Deep Convolutional Neural Networks

Deep neural networks have advanced significantly in recent years, demonstrating success in many real-world prediction tasks and capturing the attention of several diverse scientific fields [8]. Neural networks are parameter-dense functions that are structured in a progressive hierarchy of layers to model complex sets of features from underlying inputs. Convolutional neural networks (CNNs), which are the basis for the models leveraged in this study, are neural networks designed specifically for image data. Parameters of CNN's correspond directly to progressively wider areas of an image's receptive field and learn low and high-level features from local structures in the image to form a more linearly interpretable set of feature representations. CNN's have been applied successfully in many real-world contexts to perform diverse tasks such as object tracking, instance segmentation, and image classification [9, 10, 11].

Deep CNN-based models for instance segmentation have emerged in recent years with the development of foundational image models trained on several thousands of images of commonly encountered objects such as cars, balloons, pets, and houses [12]. These models have been applied in many challenging problem contexts. For example, models such as UNet [13] and Yolo-V3 [14] have been adapted for real-time segmentation of pedestrians and vehicles to service autonomous self-driving tasks [15].

In this paper, we design, develop, and test a novel deep learning-based approach to perform HST-IT from images taken of our Lancaster test facility by adapting the Mask R-CNN architecture proposed by He et al. [16]. We perform studies on the effect of datacentric parameters and architectural components on model performance and perform a stratified analysis of our best model with respect to object instance size to quantify gaps between model development and model deployment. Finally, we discuss the implications of our results on improving the robustness and scalability of HST-IS approaches using deep learning-based computer vision techniques and offering a feasible solution to low-stakes, automated heliostat monitoring needs.

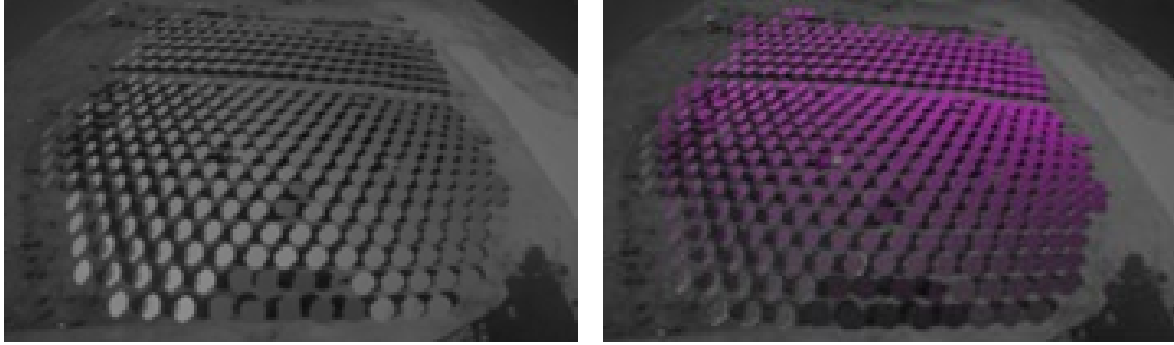


Figure 3. Example of an HDR-stitched full-field image (left) and its mask-overlay counterpart (right).

2. Methods

2.1 Dataset

Images of the full field were taken at six different camera positions in four distinct exposures for a collection of timestamps. For a given timestamp, images taken at different exposures for the same camera position were registered using phase correlation and subsequently stitched together into a resultant HDR image. Heliostat instance mask labels for each HDR image were generated using SOHOT's ROI generation process (Fig. 2) and used for the training and evaluation results in this paper [1]. Ablation studies were performed on several image-specific parameters relative to HST-IS performance, such as bit depth, image stitching technique, and sampling of seasonal, time-of-day periods. For our best model, HDR images from all six cameras were randomly sampled from a full-year period and full-day distribution of times.

2.2 Model Experiments

2.2.1 Neural Network Architecture and Training

To address the unique set of challenges in HST-IS, we conducted ablation studies to tune hyperparameters and customize model subcomponent architectures in Mask R-CNN. Specifically, we study the choice of CNN backbone, depth of CNN backbone, choice of pretrained baseline, region proposal network (RPN) loss function, RPN thresholding parameters, RPN anchor parameters, number of linear layers in projection heads, random data augmentation strategies, and image preprocessing and postprocessing strategies. Each ablation was studied in the context of multiple evaluation benchmarks stratified by distinct heliostat size categories and separated between bounding box and mask generation performance. Average precision metrics at average recalls above 90% were evaluated alongside qualitative observations of predictions to make conclusions. Our best model leverages a ResNet-50 feature proposal network (FPN) backbone with pretrained ImageNet weights and finetuned on 50K image tiles with a batch size and learning rate of 8 and 0.005, respectively. We use a momentum and weight decay of 0.9 and 0.0001, respectively, with a multistep learning-rate warmup scheduler. Furthermore, we leverage the multi-task training loss function defined by the original Mask R-CNN paper [16], which features a binary cross-entropy classification loss, binary cross-entropy mask loss, and smooth L1 bounding-box localization loss.

2.2.2 Evaluation Metrics

We evaluate model performance based on average precision metrics at both segmentation mask and bounding box levels at intersection-over-union (IOU) thresholds ranging from 0.5-0.75. The IOU between a prediction and associated ground truth label is calculated by dividing the number of intersecting pixels between both objects by the number of pixels comprising the union of both objects (Fig. 4). We conform to definitions outlined in the Microsoft Coco object detection metric standards [17] in which predictions merged by a non-max suppression

procedure and deemed to be associated with a ground truth label based on an IOU threshold are classified as a correct instance prediction. The Microsoft Coco object detection metric standards are commonly used in deep learning-based object detection and instance segmentation studies [16, 18, 19].

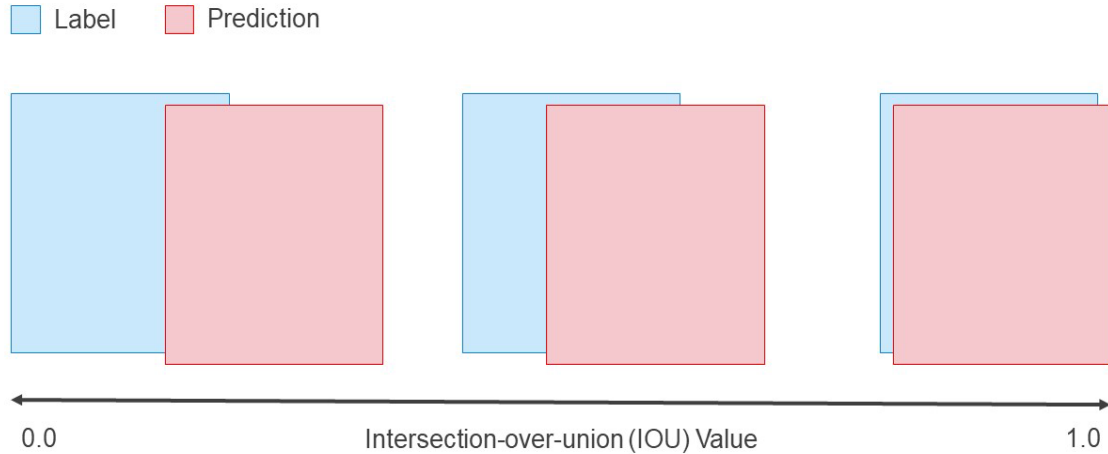


Figure 4. Illustration of prediction-object pairs at various associated IOU thresholds.

2.2.3 Data Preparation and Post-Processing

Full-field images were split into 256 pixel x 256 pixel (256x256) image subcomponents overlapping by a fixed pixel threshold in each direction. Instances contained within each subcomponent were included as labels if their centroid was within a fixed pixel distance from each image boundary. This labeling filter step is included to conform to the modeling task of segmenting fully present heliostat instances. Horizontal and vertical flips were performed during training to improve model robustness. Following model instance segmentation, predicted instances were mapped to their associated trackers with a nearest neighbors-based algorithm using the centroids of predicted instances and centroids of assumed tracker locations. Like our preprocessing labeling filter step, instance predictions whose centroids were not within a fixed pixel distance from image boundaries were discarded to eliminate the presence of duplicate predictions for each tracker.

3. Results

Our best model achieves an average bounding box and segmentation precision of 93.7% and 92.7% averaged over all IOU thresholds, respectively, at a 90% average recall, compared against ROIs generated by the previous technique. Throughout the course of our ablation studies, we discovered unique training paradigms specific to HST-IS. Firstly, we find that training with a wide distribution of region proposal network (RPN) anchor size parameters, which correspond in proportion to the diverse image footprints of captured heliostats, improves robustness to object scale variance. We demonstrate successful segmentation of heliostats that occupy a diverse set of image footprints by tuning the aforementioned RPN anchor size parameters (Fig. 6). Secondly, we find that training our model with a pretrained Resnet-50 FPN backbone results in superior overall performance compared to training with other object detection backbone model architectures. Lastly, we find that aggregating top-performing dataset parameters such as introducing random data augmentation in training, representing our images in a high dynamic range, and selecting from a wide distribution of seasonal, time-of-day periods sees a 10% improvement in average precision compared to model baselines.

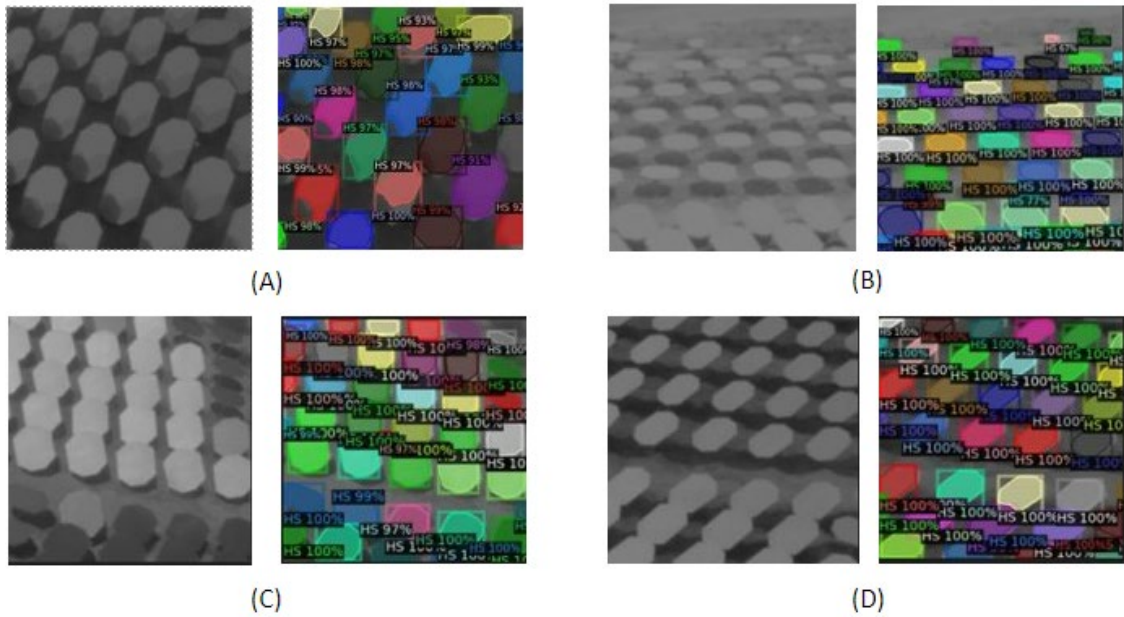


Figure 5. Examples of predicted HST-IS masks and their original counterparts at various positions of the field. As seen, predictions are robust across front (A), middle (C, D), and back (B) areas of the field. Evidence of semantic understanding of shading and blocking constraints is observed as well (A).

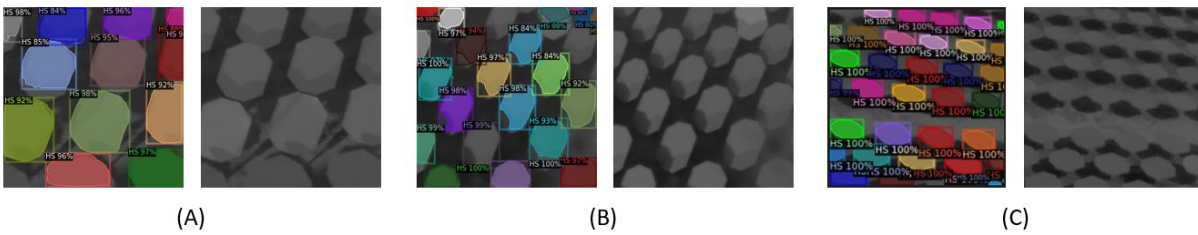


Figure 6. Examples of predicted HST-IS masks and their original counterparts selected from closer (A) to farther (C) distances from the capturing camera. Consistent results provide evidence that our instance segmentation model is robust to variance in object size and scale relative to the component image.

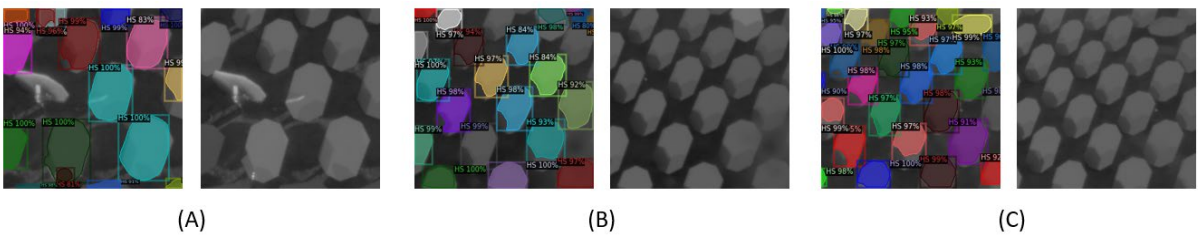


Figure 7. Examples of predicted HST-IS masks and their original counterparts selected from images that contain significant shading and blocking. Our instance segmentation model is able to distinguish blocked heliostat pixels (B, C) and account for non-aiming heliostats (A).

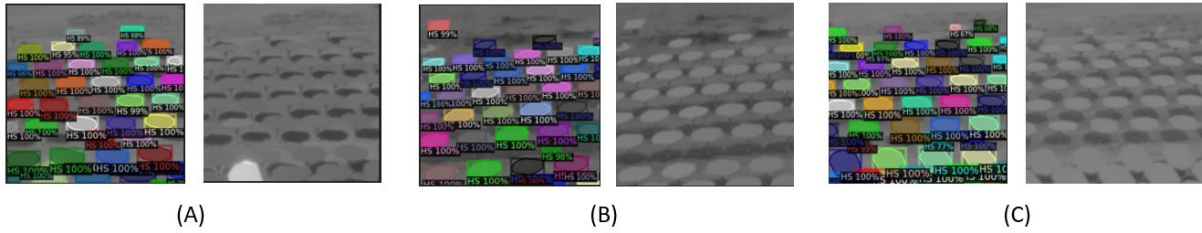


Figure 8. Examples of predicted HST-IS masks and their original counterparts selected from images taken of the farthest sections of the field. Due to loss in image context and fewer pixels making up each heliostat instance at greater distances, performing HST-IS on these sections of the field tends to be more challenging for traditional optimization and computer vision approaches.

4. Discussion and Future Work

In this study, we demonstrate the capability of deep learning models to perform instance segmentation of heliostats from images taken of our Lancaster test facility. Evidence of the approach’s robustness is indicated in observing the high average precision of generated instance bounding boxes and masks relative to ground truth labels produced by our existing SOHOT module. In addition, we qualitatively observe robust instance segmentation performance on images that contain heliostats at varying distances from the capturing camera, alluding to our model’s adaptiveness to object size and scale (Fig. 6). Within our manual inspection of model predictions, we observe evidence of mask generation that distinguishes non-blocked and non-shaded heliostat pixels from blocked or shaded heliostat pixels (Fig. 7). Although the results of this distinction among model predictions are inconsistent, they indicate promise in the modelling approach’s ability to interpret higher-level image features and reconcile a nuanced task.

There are limitations of our study that we hope to address in future work. Firstly, training our deep learning model requires several hundreds of images and associated labels constituting diverse training, validation, and test sets. Acquiring such volumes of imagery and detailed instance labels can be time-intensive, complex, and expensive. In our study, our Lancaster test facility and SOHOT module are configured to naturally satisfy these requirements, which provides a unique advantage. Secondly, the labels used in our training dataset were generated using our SOHOT module, which is known to be associated with a margin of error due to positional and mathematical assumptions. Relative to object human labels, the extents to which these inaccuracies in dataset labels affect the performance of the model are unclear and will be a subject of future work. Despite this limitation, our model offers a standalone solution that can produce similar predictions to those of our SOHOT module without any contextual field information. This capability offers significant value in ROI calibration. Thirdly, the results of our study are closely associated with field conditions in our Lancaster test facility. We anticipate that future challenges will arise for our modelling approach when applied to fields in which heliostats are more distant from the capturing camera, more numerous in count, or positioned more compactly. In future work, we aim to better simulate these conditions and quantify these gaps with the generation of synthetic training data using 3D rendering tools.

5. Conclusion

In this paper, we demonstrate a deep CNN-based approach to performing heliostat instance segmentation on full CSP field images. We study architectural and datacentric parameters to optimize model performance and provide evidence of robustness to object scale variance and semantic understanding of shaded pixel distinction. The method shows promise in being

deployed as an automated, contextless approach for CSP tasks that rely on heliostat instance segmentation.

Competing interests

The authors declare that they have no competing interests.

Data Availability

The data required to reproduce the above findings cannot be shared at this time due to legal reasons.

Author contributions

Benjamin Liu (Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Original Draft, Review, Editing)

Alexander Sonn (Conceptualization, Data curation, Investigation, Review, Editing)

Anthony Roy (Conceptualization, Investigation, Review, Editing)

Brian Brewington (Conceptualization, Investigation, Review, Editing)

References

1. Sonn, Alexander, et al. "Estimating orientations of tracking heliostats using circumsolar radiance." AIP Conference Proceedings. Vol. 2445. No. 1. AIP Publishing LLC, 2022. doi: <https://doi.org/10.1063/5.0087117>
2. Kribus, Abraham, et al. "Closed loop control of heliostats." Energy 29.5-6 (2004): 905-913. doi: [https://doi.org/10.1016/S0360-5442\(03\)00195-6](https://doi.org/10.1016/S0360-5442(03)00195-6)
3. Sattler, Johannes Christoph, et al. "Review of heliostat calibration and tracking control methods." Solar Energy 207 (2020): 110-132. doi: <https://doi.org/10.1016/j.solener.2020.06.030>
4. Bouaddi, S., Ihlal, A., Fernández-García, A., 2017. Comparative analysis of soiling of CSP mirror materials in arid zones. Renew. Energy 101, 437–449. doi: <https://doi.org/10.1016/j.renene.2016.08.067>
5. Röger, Marc, Christoph Prah, and Steffen Ulmer. "Heliostat shape and orientation by edge detection." Journal of solar energy engineering 132.2 (2010). doi: <https://doi.org/10.1115/1.4001400>
6. Coventry, Joe, et al. "A robotic vision system for inspection of soiling at CSP plants." AIP Conference Proceedings. Vol. 2303. No. 1. AIP Publishing LLC, 2020.
7. El Ydrissi, Massaab, et al. "Dust InSMS: Intelligent Soiling Measurement System for dust detection on solar mirrors using computer vision methods." Expert Systems with Applications (2022): 118646. doi: <https://doi.org/10.1016/j.eswa.2022.118646>
8. Deng, Li, and Dong Yu. "Deep learning: methods and applications." Foundations and trends® in signal processing 7.3–4 (2014): 197-387. doi: <http://dx.doi.org/10.1561/20000000039>
9. Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." Progress in Artificial Intelligence 9.2 (2020): 85-112. doi: <https://doi.org/10.1007/s13748-019-00203-0>

10. Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29.9 (2017): 2352-2449. doi: https://doi.org/10.1162/neco_a_00990
11. Sultana, Farhana, Abu Sufian, and Paramartha Dutta. "Evolution of image segmentation using deep convolutional neural network: a survey." *Knowledge-Based Systems* 201 (2020): 106062. doi: <https://doi.org/10.1016/j.knosys.2020.106062>
12. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60.6 (2017): 84-90. doi: <https://doi.org/10.1145/3065386>
13. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
14. Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
15. Yi, Zhang, Shen Yongliang, and Zhang Jun. "An improved tiny-yolov3 pedestrian detection algorithm." *Optik* 183 (2019): 17-23. doi: <https://doi.org/10.1016/j.ijleo.2019.02.038>
16. He, Kaiming, et al. "Mask r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2017.
17. Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.
18. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
19. He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.