Segment-then-Classify: Few-shot instance segmentation for environmental remote sensing

Yang Hu, Kelly Caylor, Anna Boser
UC Santa Barbara
Santa Barbara, CA 93106
yanghu, caylor, annaboser @ucsb.edu

Abstract

Instance segmentation is pivotal for environmental sciences and climate change research, facilitating important tasks from land cover classification to glacier monitoring. This paper addresses the prevailing challenges associated with data scarcity when using traditional models like YOLOv8 by introducing a novel, data-efficient workflow for instance segmentation. The proposed Segment-then-Classify (STC) strategy leverages the zero-shot capabilities of the novel Segment Anything Model (SAM) to segment all objects in an image and then uses a simple classifier such as the Vision Transformer (ViT) to identify objects of interest thereafter. Evaluated on the VHR-10 dataset, our approach demonstrated convergence with merely 40 examples per class. YOLOv8 requires 3 times as much data to achieve the STC's peak performance. The highest performing class in the VHR-10 dataset achieved a near-perfect mAP@0.5 of 0.99 using the STC strategy. However, performance varied greatly across other classes due to the SAM model's occasional inability to recognize all relevant objects, indicating a need for refining the zero-shot segmentation step. The STC workflow therefore holds promise for advancing few-shot learning for instance segmentation in environmental science.

1 Introduction

Instance segmentation techniques have emerged as transformative tools in environmental sciences and climate change research. These techniques have enabled climate change research at previously impossible scales, performing tasks such as the delineation of individual coral reef colonies [Talpaert Daudon et al., 2023], identification of individual tree crowns [Sun et al., 2022], mapping agricultural fields [Chen et al., 2023a], pollution tracking [Temitope Yekeen et al., 2020], and glacier monitoring [Heidler et al., 2023]. However, environmental scientists often encounter a scarcity of labeled data needed to train instance segmentation models effectively [Sun et al., 2021].

State-of-the-art instance segmentation models such as YOLOv8 are known to require a substantial volume of data for training [TAO et al., 2023], hampering their use in environmental science research. Such models rely on a "Detect-then-Segment" strategy which starts by finding objects using bounding boxes and then uses these boxes as a reference to segment each object individually [Chen et al., 2019]. This approach relies on extensive instance mask datasets that require labor-intensive human labeling [Potlapally et al., 2019]. Additionally, the pre-trained versions of such models are typically not optimized for remote sensing data, further increasing the burden of generating large labeled datasets for remote sensing applications [Wang et al., 2023].

We introduce a novel instance segmentation workflow designed to diminish training data requirements. Instead of the conventional Detect-then-Segment workflow, our Segment-then-Classify (STC) method leverages the zero-shot capabilities of the Segment Anything Model (SAM) to segment all objects in a remote sensing image [Chen et al., 2023b] before relying on a simple classifier, such as ViT, to

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2023.

retain the masks of interest. Pre-trained on a vast dataset of over one billion masks[Kirillov et al., 2023], SAM has shown remarkable generalization capabilities that allow it to adapt to the diverse and dynamic nature of remote sensing data [Osco et al., 2023] without needing additional training. Because image classifiers require comparatively small amounts of data for training as compared to instance segmentation models, STC has the potential to drastically reduce training data requirements and present a more widely accessible instance segmentation workflow for earth observation data.

To test the performance of the proposed workflow across different dataset sizes and remotely sensed classes, we use the VHR-10 dataset, composed of satellite images labeled with ten different classes. The ViT classifier, and therefore the STC strategy, reaches convergent performance after a mere 40 examples per class, highlighting the small number of labels required for this workflow to reach its peak performance. By comparison, YOLOv8 requires up to three times the amount of labeled data to achieve comparable performance. Certain classes achieved mAP@0.5 scores of up to 0.99, an impressive near-perfect score given the small amount of data required for training. However, the mAP@0.5 of the worst performing class was a dissatisfactory 0.08, suggesting that the ability to fine-tune the segmentation step would lead to a more universally applicable approach [Mattjie et al., 2023, Ji et al., 2023].

This work represents a new paradigm for instance segmentation that has the potential to drastically reduce the amount of data required for training. By relaxing this reliance on labeled data, STC has the potential to greatly expand the use of instance segmentation in areas such as remote sensing for climate change research, where a lack of labeled data hampers current use.

2 The Segment-then-Classify Workflow

The Segment-then-Classify strategy employs two steps (Figure 1). First, SAM's "everything" mode is used to automatically generate instance masks across the remote sensing RGB image. Second, a classifier, for example ViT, filters and labels the instance masks.

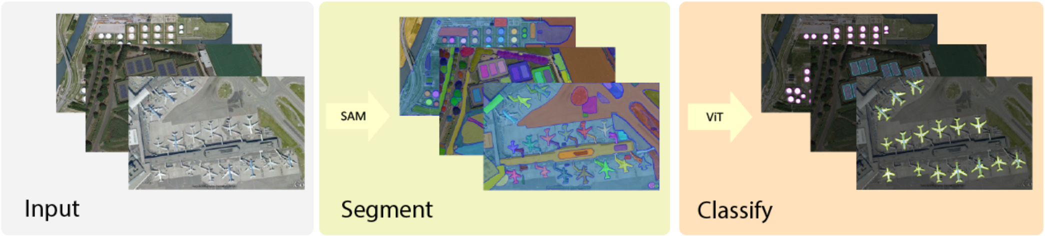

Figure 1: Overview of the Segment-then-Classify Workflow.

3 Experiments

Dataset We experiment on the NWPU VHR-10 dataset, a benchmark for geospatial object instance segmentation in remote sensing data. Comprising 800 very-high-resolution (VHR) optical remote sensing images and 3,775 instance labels across 10 distinct categories, the dataset offers a diverse range of spatial resolutions, from 0.08m to 2m. Some images are sourced from Google Earth, while others are pan-sharpened color infrared images from the Vaihingen dataset [Su et al., 2019]. This diversity makes NWPU VHR-10 an ideal testbed for assessing our strategy under various challenging conditions. We partition the dataset into training, validation, and test sets, maintaining a 70-20-10 split while ensuring a similar distribution of instance categories across each set.

Step 1: Automated instance segmentation The automatic generation of instance masks using SAM is accomplished by uniformly scattering points across the image and treating each point as a prompt for segmentation. For each image, we employ a 32x32 grid of points as prompts and apply an initial noise-cleaning step to remove disconnected mask regions or holes with an area less than 0.2% of the total image size. We also employ Non-Maximum Suppression to eliminate duplicate masks, keeping only the most confident predictions and discarding others with high overlap.

Step 2: Classification After obtaining the masks for all objects we clip the image to the extent of each mask and employ a classifier to filter out extraneous objects, retaining and labeling those that are of interest. We opt for a Vision Transformer (ViT) model as our classifier. While models like CLIP can be employed in a zero-shot learning regime [Radford et al., 2021], custom-trained classifiers such as the ViT tend to outperform these in remotely sensed imagery [Xu et al., 2022]. To this end, we construct an 11-class classification dataset by extracting cropped images by the bounding boxes of instance masks from the NWPU VHR-10 dataset, plus an "other" class gathered from the SAM results to account for those miscellaneous instances that we aim to get rid of.

Evaluation We evaluate model performance using the mean Average Precision (mAP) score, which is calculated based on the Intersection over Union (IoU) between the predicted and ground truth bounding boxes. Specifically, we employ mAP at an IoU threshold of 0.5, denoted as mAP@0.5, which means that a predicted bounding box is considered a true positive only if it has an overlap of 50% or more with the ground truth bounding box.

4 Results

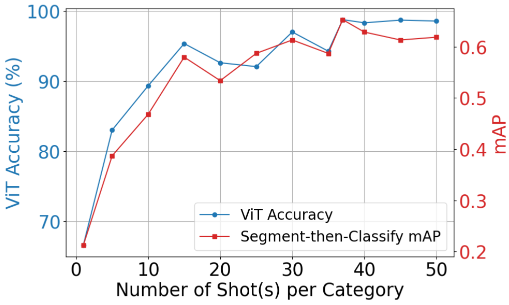

Figure 2: ViT accuracy and Segment-then-Classify overall performance in few-shot training

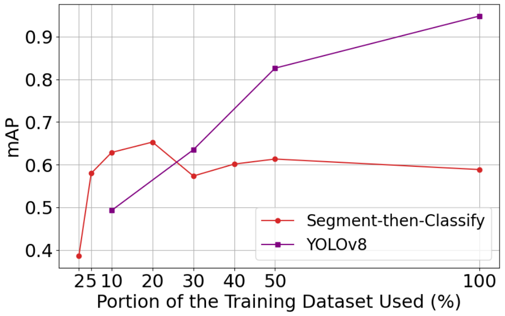

Figure 3: Comparison of Segment-then-Classify and YOLOv8 across training set sizes

Data efficiency: The overall performance of our method achieves an mAP@0.5 score of 0.65. The ViT classifier is the only step that requires training in the STC workflow and it converges to its peak performance of 98% after training on only 40 object images, or "shots" per class (Figure 2). Because the performance of STC closely follows that of the ViT classifier, its mAP performance also converges around 40 shots. For comparison, we evaluate STC against the YOLOv8 model using different training set sizes (Figure 3). The performance of YOLOv8 continues to enhance with the incorporation of more data. However, to attain an mAP score comparable to what STC achieves with only 10% of the training set, YOLOv8 requires 30% of the training dataset, equivalent to 1,059 mask labels.

Dataset analysis: When we break up the performance of the STC strategy across the ten classes present in the VHR-10 dataset, we find a very large variation in performance (Figure 4).

The model exhibits high mAP scores for categories with distinct, regular shapes and clear boundaries, such as "Storage Tanks" (0.99), "Baseball Diamonds" (0.95), and "Ground Track Fields" (0.94). This suggests that our model is particularly adept at recognizing geometrically well-defined objects, making it promising in applications like glacier or pollution monitoring. On the other hand, categories like "Airplanes" (0.893) also perform well, which is indicative of the model's capabilities in handling more complex shapes.

However, model performance decreases significantly for "Basketball Courts" (0.76), "Harbors" (0.64), and "Tennis Courts" (0.52). This can likely be attributed to SAM's difficulty in discerning these objects as separate if they are often found right next to each other. This may prove to be problematic for certain applications such as instance segmentation of agricultural fields, since these are also commonly found closely neighboring one another [Tzepkenlis et al., 2023]. In such cases, fine-tuning the SAM model, for example with PerSAM [Zhang et al., 2023], may be necessary

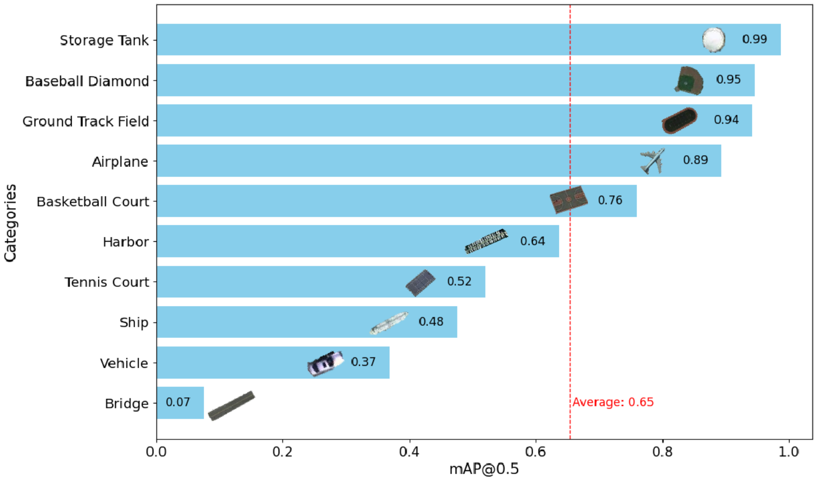

Figure 4: Performance of the Segment-then-Classify workflow by category.

to improve accuracy. "Ships" (0.48), "Vehicles" (0.37), and "Bridges" (0.07) perform the worst, suggesting that relatively small objects on cluttered backgrounds may not be suitable for this workflow without significant amendments to SAM. Prior research indicates that SAM's performance tends to deteriorate with image resolutions exceeding 1 or 2 meters [Osco et al., 2023], which might contribute to the contrasting results for seemingly similar categories like "Ships" and "Airplanes."

5 Conclusion

Remote sensing data are notably underrepresented in instance segmentation, creating challenges for researchers in environmental science, who often lack the resources to label extensive datasets essential for this task. Thus, reducing reliance on such extensive datasets is crucial.

We have demonstrated that the Segment-then-Classify paradigm significantly lessens the dependency on large training datasets. This strategy proves to be highly effective for several classes. However, the wide variation in performance observed among different classes indicates the potential benefit of employing few-shot fine-tuning techniques like PerSAM [Zhang et al., 2023].

Few-shot instance segmentation enabled by STC has substantial potential for numerous applications in environmental science. It enhances researchers' capabilities to observe activities such as pollution or land cover changes contributing to climate change and to monitor the impacts of climate change on entities like coral reefs or glaciers. By doing so, it paves the way for more informed and effective interventions in environmental preservation and restoration.

6 Acknowledgments and Disclosure of Funding

Y. Hu was supported by the Bren Environmental Leadership Program. A. Boser was supported by the National Science Foundation Graduate Research Fellowship Program (1650114).

References

Fen Chen, Haojie Zhao, Dar Roberts, Tim Van de Voorde, Okke Batelaan, Tao Fan, and Wenbo Xu. Mapping center pivot irrigation systems in global arid regions using instance segmentation and analyzing their spatial relationship with freshwater resources. *Remote Sensing of Environment*, 297:113760, November 2023a. ISSN 0034-4257. doi: 10.1016/j.rse.2023.113760. URL `https://www.sciencedirect.com/science/article/pii/S0034425723003115`.

Keyan Chen, Chenyang Liu, Hao Chen, Haotian Zhang, Wenyuan Li, Zhengxia Zou, and Zhenwei Shi. Rsprompter: Learning to prompt for remote sensing instance segmentation based on visual foundation model, 2023b.

Xinlei Chen, Ross Girshick, Kaiming He, and Piotr Dollár. Tensormask: A foundation for dense object segmentation, 2019.

Konrad Heidler, Lichao Mou, Erik Loebel, Mirko Scheinert, Sébastien Lefèvre, and Xiao Xiang Zhu. A Deep Active Contour Model for Delineating Glacier Calving Fronts. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 61, 2023. URL `https://ieeexplore.ieee.org/abstract/document/10195954`.

Wei Ji, Jingjing Li, Qi Bi, Tingwei Liu, Wenbo Li, and Li Cheng. Segment anything is not always perfect: An investigation of sam on different real-world applications, 2023.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

Christian Mattjie, Luis Vinicius de Moura, Rafaela Cappelari Ravazio, Lucas Silveira Kupssinskü, Otávio Parraga, Marcelo Mussi Delucis, and Rodrigo Coelho Barros. Zero-shot performance of the segment anything model (sam) in 2d medical imaging: A comprehensive evaluation and practical guidelines, 2023.

Lucas Prado Osco, Qiusheng Wu, Eduardo Lopes de Lemos, Wesley Nunes Gonçalves, Ana Paula Marques Ramos, Jonathan Li, and José Marcato Junior. The segment anything model (sam) for remote sensing applications: From zero to one shot, 2023.

Anirudh Potlapally, Potluri Sai Rohit Chowdary, S.S. Raja Shekhar, Nitin Mishra, Chella Sri Venkata Divya Madhuri, and A.V.V. Prasad. Instance segmentation in remote sensing imagery using deep convolutional neural networks. In *2019 International Conference on contemporary Computing and Informatics (IC3I)*, pages 117–120, 2019. doi: 10.1109/IC3I46837.2019.9055569.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

H Su, S Wei, M Yan, and et al. Object detection and instance segmentation in remote sensing imagery based on precise mask r-cnn. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1454–1457. IEEE, 2019.

Xian Sun, Bing Wang, Zhirui Wang, Hao Li, Hengchao Li, and Kun Fu. Research progress on few-shot learning for remote sensing image interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:2387–2402, 2021. doi: 10.1109/JSTARS.2021.3052869.

Ying Sun, Ziming Li, Huagui He, Liang Guo, Xinchang Zhang, and Qinchuan Xin. Counting trees in a subtropical mega city using the instance segmentation method. *International Journal of Applied Earth Observation and Geoinformation*, 106:102662, 2022. ISSN 1569-8432. doi: 10.1016/j.jag.2021.102662.

Justine Talpaert Daudon, Matteo Contini, Isabel Urbina-Barreto, Brianna Elliott, François Guilhaumon, Alexis Joly, Sylvain Bonhommeau, and Julien Barde. GeoAI for Marine Ecosystem Monitoring: a Complete Workflow to Generate Maps from AI Model Predictions. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W7-2023:223–230, June 2023. ISSN 2194-9034. doi: 10.5194/isprs-archives-XLVIII-4-W7-2023-223-2023. URL `https://archimer.ifremer.fr/doc/00844/95549/`. Publisher: Copernicus GmbH.

HUANG TAO, LI Hua, ZHOU Gui, LI Shaobo, and WANG Yang. Survey of research on instance segmentation methods. *Journal of Frontiers of Computer Science & Technology*, 17(4):810, 2023.

Shamsudeen Temitope Yekeen, Abdul-Lateef Balogun, and Khamaruzaman B. Wan Yusof. A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 167:190–200, September 2020. ISSN 0924-2716. doi: 10.1016/j.isprsjprs.2020.07.011. URL `https://www.sciencedirect.com/science/article/pii/S0924271620301982`.

Anastasios Tzepkenlis, Konstantinos Marthoglou, and Nikos Grammalidis. Efficient deep semantic segmentation for land cover classification using sentinel imagery. *Remote Sensing*, 15:2027, 2023. doi: 10.3390/rs15082027.

Di Wang, Jing Zhang, Bo Du, Minqiang Xu, Lin Liu, Dacheng Tao, and Liangpei Zhang. Samrs: Scaling-up remote sensing segmentation dataset with segment anything model, 2023.

Kejie Xu, Peifang Deng, and Hong Huang. Vision transformer: An excellent teacher for guiding small networks in remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15, 2022. doi: 10.1109/TGRS.2022.3152566.

Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot, 2023.