Segment Anything Model for Refugee-Dwelling Extraction with Few Samples from High-Resolution Satellite Imagery

Yunya Gao*, Hui Zhao**

* yunya.gao@plus.ac.at, Christian Doppler Laboratory for GEOHUM, Department of Geoinformatics – Z_GIS, Paris Lodron Universität Salzburg, Austria
** zhao.h@ufl.edu, School of Natural Resources and Environment, University of Florida, the United States

Abstract. Customizing Segment Anything Model (SAM) has recently attracted considerable attention in remote sensing domains. This study explores the performance of SAM-Adapter in refugee-dwelling extraction in three different refugee camps from high-resolution satellite images. The findings indicate that with scarce sample data, SAM-Adapter marginally outperforms other semantic segmentation models. This underscores SAM's promising potential for building extraction tasks when data is limited.

Keywords. Segment Anything Model, Dwelling Extraction

1. Introduction

The United Nations Statistical Commission approved multiple indicators for refugees under Sustainable Development Goals (SDGs) according to the commitment "Leave no one behind" in March 2020 (UNHCR, 2020). Providing adequate living resources such as clean water, nutritious food, medical services, and modern energy to refugees and their host communities is essential (UNHCR, 2020). Before delivering these resources, it is significant to know refugee population in need. Due to the difficulty of collecting such information on site, updated footprints of refugee dwellings from satellite imagery could be beneficial for the estimation purpose (Spröhnle et al., 2014), and thus can support urgent humanitarian operations for refugees.

LBS 2023
18th Conference on Location Based Services
November 20-22, Ghent, Belgium

Published in "Proceedings of the 18th International Conference on Location Based Services (LBS 2023)", edited by Haosheng Huang, Nico Van de Weghe and Georg Gartner, LBS 2023, 20-22 November 2023 Ghent, Belgium.

This contribution underwent single-blind peer review based on the paper.
https://doi.org/10.34726/5740

In recent years, deep learning approaches have shown high potential in extracting the footprints by learning essential features from images and annotated labels (Gao, Lang, Tiede, Gella, & Wendt, 2022). However, most research uses strong supervision that requires numerous high-quality labels beforehand.

The "Segment Anything Model (SAM)" introduced by Meta AI Research has recently garnered considerable interest. With training on an extensive segmentation dataset comprising over 1 billion masks, SAM excels at segmenting any entity within a given image (Kirillov et al., 2023). Lately, several studies have delved into evaluating SAM's efficacy in different applications for object recognition and segmentation.

SAM shows high potential in remote sensing domains yet is also challenging due to high variability in shape and size of objects (Ren et al., 2023). It is found that SAM does not outperform task-specific models on building extraction compared to trees or clouds. Hence, it's essential to augment the model's effectiveness by incorporating additional fine-tuning methods specifically for building extraction.

Currently, three primary adaptation methods are derived from recent studies (Zhang et al., 2023). They are 1) fine-tuning, 2) applying adapters, and 3) decoupling the mask decoder into two modules. Incorporating domain-specific adapters is a viable method to tailor SAM's architecture. These adapters aim to capture task-specific knowledge.

This research assesses the efficacy of SAM-Adapter (Chen et al., 2023) in extracting dwellings across three refugee camps, using two different training data sizes. The results prove that SAM-Adapter marginally outperforms other semantic segmentation models when sample data is limited. This study is currently ongoing. Associated codes can be accessible at https://github.com/YunyaGaoTree/SAM-Adapter-For-Refugee-Dwelling-Extraction.

2. Methodology

2.1. SAM-Adapter

SAM-Adapter integrates domain-specific data or visual prompts into the segmentation network through the use of efficient adapters. By melding task-specific insights with the broad knowledge acquired by the extensive model, SAM-Adapter notably boosts SAM's efficacy in intricate tasks like detecting camouflaged objects and identifying shadows (Chen et al., 2023). The codes for SAM-Adapter source from this GitHub. The pretrained model “sam_vit_h_4b8939.pth” are chosen as a foundation model in this study.

2.2. Models for ablation studies

We choose six advanced semantic segmentation models as for comparison against SAM-Adapter. These models include the Feature Pyramid Network (FPN) with 1) Mix Vision Transformer (MiT), 2) MobileNet-v2, 3) ResNet34 as their backbones, and Unet with 4) ResNet101, 5) ResNet34, and 6) MobileNet-v2. The codes for these models source from this GitHub.

2.3. Data processing

Three datasets are chosen for this study. They are from Kutupalong refugee camp in Kenya, Dagahley refugee camp in Tanzania and Minawao refugee camp in Cameroon. The dataset details about sensor, spatial resolution, retrieved date, spatial extent for training/validation/testing purposes, and number of patches in shape of 1024 by 1024 pixels are provided in *Table 1*.

It's crucial to acknowledge that there are multiple methods for sampling image data. For instance, instead of maintaining a similar spatial extent for each study site, it may be worthwhile to keep the number of generated patches consistent. It may be valuable to explore different sampling strategies such as different proportions of data. However, in this initial research, we choose this straightforward data collection approach.

Refugee camp	Retrieved date	Sensor	Resolution (m)	Data type	Extent / pixel	Nr. of patches
Kutupalong	13/02/2018	UAV	0.1	Train_Large	13283, 12489	1848
				Train_Small	6828, 5346	420
				Validation	4334, 4078	226
				Test	9844, 9420	
Dagahaley	08/04/2017	WV3	0.3	Train_Large	5754, 5074	350
				Train_Small	2010, 1944	56
				Validation	1389, 1373	7
				Test	4783, 3101	
Minawao	12/02/2017	WV2	0.5	Train_Large	3832, 4625	224
				Train_Small	1682, 1744	56
				Validation	1188, 1187	7
				Test	1817, 3165	

Table 1. Dataset details utilized in this study.

2.4. Implementation details

We evaluate the performance of the proposed approach with Precision, Recall, F1-score and Intersection over Union (IoU) of refugee dwellings. All tests were conducted on a machine equipped with an RTX3090 GPU, utilizing the PyTorch 2.0 environment. Data augmentation by random rotation and flip are applied for model training. Additional details will be disclosed in an upcoming manuscript currently in the works.

3. Results and Discussion

The accuracy assessment results for the three refugee camps are shown in *Table 2 and Table 3. Figure 1* provides one example of predicted results for each refugee camp from FPN-MiT and SAM-Adapter trained on "small" dataset, and from SAM without any adjustments.

From these results, it's evident that SAM, without any modifications, falls short of yielding promising outcomes. The IoU values for the three datasets are considerably lower compared to SAM-Adapter and the other six segmentation models. Nonetheless, it's also discernible that as image resolution increases, the accuracy of the predicted results improves. This discovery aligns with the observation in (Osco et al., 2023), which recommends integrating SAM with the SR (super-resolution) model to enhance its effectiveness for low-resolution images.

In summary, SAM-Adapter significantly excels over other segmentation models, notably with small datasets. For instance, while the highest IoU value among the six segmentation models for the Minawao dataset stands at 0.195, it leaps to 0.571 with SAM-Adapter. Moreover, SAM-Adapter consistently produces comparable accuracy across both dataset sizes for all three refugee camps.

Model	Data size	Kutupalong				Dagahaley			
		IoU	F1	Precision	Recall	IoU	F1	Precision	Recall
FPN-MiT	Large	**0.666**	0.800	0.762	0.841	**0.523**	0.687	0.784	0.612
	Small	**0.656**	0.792	0.825	0.763	**0.297**	0.458	0.429	0.490
FPN-MobileV3L	Large	0.602	0.751	0.818	0.695	0.465	0.635	0.835	0.513
	Small	0.423	0.594	0.909	0.441	0.251	0.402	0.564	0.312
FPN-ResNet34	Large	0.594	0.745	0.808	0.692	0.351	0.519	0.803	0.384
	Small	0.587	0.740	0.805	0.684	0.138	0.239	0.721	0.143
Unet-ResNet101	Large	0.574	0.729	0.669	0.802	0.505	0.671	0.670	0.672
	Small	0.519	0.684	0.764	0.618	0.140	0.245	0.146	0.758
Unet-MobileV3L	Large	0.608	0.757	0.815	0.706	0.557	0.715	0.657	0.785
	Small	0.593	0.745	0.820	0.682	0.159	0.274	0.265	0.284
Unet-ResNet34	Large	0.606	0.755	0.695	0.826	0.432	0.604	0.793	0.488
	Small	0.601	0.750	0.790	0.715	0.129	0.229	0.141	0.612
SAM-Adapter	Large	**0.733**	0.846	0.879	0.815	**0.619**	0.765	0.793	0.738
	Small	**0.710**	0.831	0.810	0.852	**0.560**	0.718	0.626	0.842

Table 2. Accuracy assessment results for Kutupalong and Dadahaley refugee camp.

Model	Data size	Minawao			
		IoU	F1	Precision	Recall
FPN-MiT	Large	**0.515**	0.680	0.648	0.716
	Small	0.194	0.326	0.621	0.221
FPN-MobileV3L	Large	0.351	0.519	0.821	0.380
	Small	**0.195**	0.326	0.668	0.216
FPN-ResNet34	Large	0.158	0.273	0.769	0.166
	Small	0.139	0.180	0.114	0.429
Unet-ResNet101	Large	0.261	0.414	0.413	0.415
	Small	0.121	0.245	0.153	0.622
Unet-MobileV3L	Large	0.278	0.435	0.857	0.291
	Small	0.129	0.228	0.424	0.156
Unet-ResNet34	Large	0.300	0.461	0.631	0.364
	Small	0.144	0.252	0.164	0.541
SAM-Adapter	Large	**0.583**	0.736	0.779	0.698
	Small	**0.571**	0.727	0.699	0.757

Table 3. Accuracy assessment results for Minawa refugee camp.

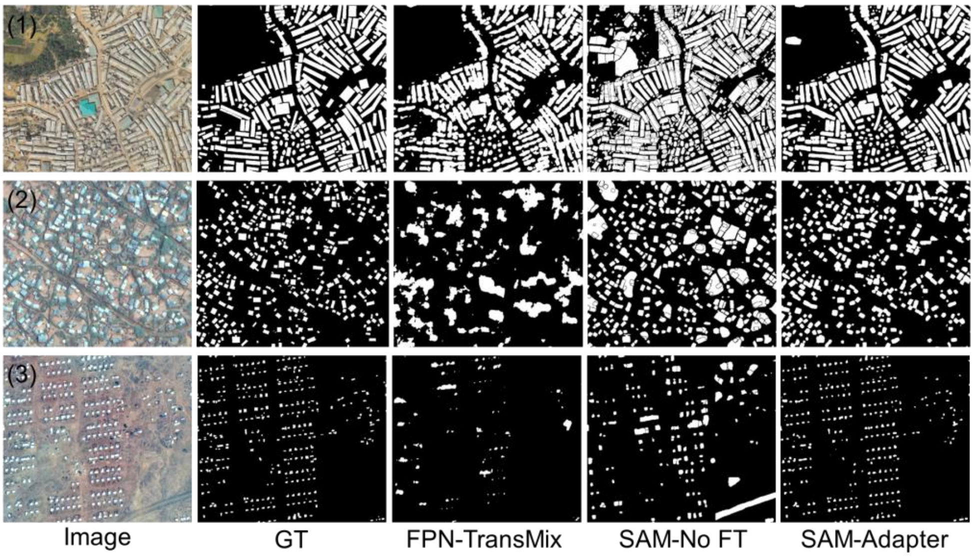

Figure 1. Predicted results from FPN-TransMix and SAM-Adapter trained on “small” dataset, and from SAM without any adjustments. From top to bottom, the images represent: (1) Kutupalong refugee camp; (2) Dagahaley refugee camp; (3) Minawao refugee camp.

4. Potential future improvements

As a pioneering study on SAM for refugee dwelling extraction tasks, there's significant room for enhancement. Firstly, enhancing the fine-tuning speed of SAM is essential. Secondly, while the SAM-Adapter's architecture is uncomplicated and easy to implement, there's some room for modifications. Thirdly, there's a potential in devising more efficient sampling strategies. Last but not least, creating SR models tailored for satellite imagery that can surpass a spatial resolution of 0.1 m could be significant for building extraction tasks.

References

Checchi, F., Stewart, B. T., Palmer, J. J., & Grundy, C. (2013). Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *International Journal of Health Geographics*, *12*. https://doi.org/10.1186/1476-072X-12-4

Chen, T., Zhu, L., Ding, C., Cao, R., Wang, Y., Li, Z., ... Zang, Y. (2023). SAM Fails to Segment Anything? -- SAM-Adapter: Adapting SAM in Underperformed Scenes: Camouflage, Shadow, Medical Image Segmentation, and More. Retrieved from http://arxiv.org/abs/2304.09148

Gao, Y., Lang, S., Tiede, D., Gella, G. W., & Wendt, L. (2022). Comparing the robustness of U-Net, LinkNet, and FPN towards label noise for refugee dwelling extraction from satellite imagery. *2022 IEEE Global Humanitarian Technology Conference, GHTC 2022*, 88–94. https://doi.org/10.1109/GHTC55712.2022.9911036

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... Girshick, R. (2023). Segment Anything. Retrieved from http://arxiv.org/abs/2304.02643

Osco, L. P., Wu, Q., de Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). The Segment Anything Model (SAM) for Remote Sensing Applications: From Zero to One Shot. Retrieved from http://arxiv.org/abs/2306.16623

Ren, S., Luzi, F., Lahrichi, S., Kassaw, K., Collins, L. M., Bradbury, K., & Malof, J. M. (2023). Segment anything, from space?, 1–10. Retrieved from http://arxiv.org/abs/2304.13000

Spröhnle, K., Tiede, D., Schoepfer, E., Füreder, P., Svanberg, A., & Rost, T. (2014). Earth observation-based dwelling detection approaches in a highly complex refugee camp environment - A comparative study. *Remote Sensing*, *6*(10), 9277–9297. https://doi.org/10.3390/rs6109277

UNHCR. (2020). *The Sustainable Development Goals and the Global Compact on Refugees*.

Zhang, C., Puspitasari, F. D., Zheng, S., Li, C., Qiao, Y., Kang, T., ... Hong, C. S. (2023). A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering. Retrieved from http://arxiv.org/abs/2306.06211