Urban1960SatSeg: Unsupervised Semantic Segmentation of Mid-20th century Urban Landscapes with Satellite Imageries

Tianxiang Hao*
Fudan University
Shanghai, China

Lixian Zhang*†
National supercomputing center in Shenzhen
Shenzhen, China

Yingjia Zhang*
New York University Shanghai
Shanghai, China

Mengxuan Chen, Jinxiao Zhang
Tsinghua University
Beijing, China

Haohuan Fu†
Tsinghua University
Beijing, China

A Supplementary Experimental Results

Table 1: Performance comparison on the Urban1960SatISP dataset between Urban1960SatUSM(Ours) and existing unsupervised semantic segmentation methods, measured by Accuracy and mean IoU (in %) for both unsupervised and supervised probing.

Method	Backbone	Unsupervised		Supervised	
		mIoU	Acc	mIoU	Acc
Dino[1]	DINO ViT - S/8	43.8	71.4	80.1	92.9
+ STEGO [3]	DINO ViT - S/8	39.9	79.7	40.1	79.7
+HP [6]	DINO ViT - S/8	22.4	60.4	25.8	68.3
+ EAGLE [4]	DINO ViT - S/8	57.4	76.1	82.2	93.8
+ PriMaPs - EM [2]	DINO ViT - S/8	43.2	81.8	81.2	93.2
+ Urban1960SatUSM	DINO ViT - S/8	52.7	81.2	80.1	92.9
Dino[1]	DINO ViT - B/8	57.8	77.0	83.1	94.1
+ STEGO [3]	DINO ViT - B/8	74.0	88.9	78.4	92.3
+HP [6]	DINO ViT - B/8	19.0	53.1	35.6	73.3
+ EAGLE [4]	DINO ViT - B/8	39.0	62.9	78.6	93.0
+ PriMaPs - EM [2]	DINO ViT - B/8	60.4	79.0	83.1	94.1
+ Urban1960SatUSM	DINO ViT - B/8	61.5	80.1	83.1	94.0
Dinov2[5]	DINOv2 ViT - S/14	62.2	81.0	82.0	93.7
+ PriMaPs - EM [2]	DINOv2 ViT - S/14	57.3	77.6	82.7	93.9
+ Urban1960SatUSM	DINOv2 ViT - S/14	75.3	91.0	81.8	93.7
Dinov2[5]	DINOv2 ViT - B/14	70.7	87.2	79.0	92.0
+ PriMaPs - EM [2]	DINOv2 ViT - B/14	66.8	85.4	82.3	93.5
+ Urban1960SatUSM	DINOv2 ViT - B/14	75.0	91.1	80.6	93.6

References

[1] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 2021. Emerging Properties in Self-Supervised Vision Transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*.

[2] Oliver Hahn, Nikita Araslanov, Simone Schaub-Meyer, and Stefan Roth. 2024. Boosting Unsupervised Semantic Segmentation with Principal Mask Proposals. *Transactions on Machine Learning Research (TMLR)* (2024).

[3] Mark Hamilton, Zhoutong Zhang, Bharath Hariharan, Noah Snavely, and William T. Freeman. 2022. Unsupervised Semantic Segmentation by Distilling Feature Correspondences. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SaKO6z6Hl0c

[4] Chanyoung Kim, Woojung Han, Dayun Ju, and Seong Jae Hwang. 2024. EAGLE: Eigen Aggregation Learning for Object-Centric Unsupervised Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3523–3533.

[5] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research* (2024). https://openreview.net/forum?id=a68SUt6zFt Featured Certification.

[6] Hyun Seok Seong, WonJun Moon, SuBeen Lee, and Jae-Pil Heo. 2023. Leveraging hidden positives for unsupervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 19540–19549.

*All authors contributed equally to this research.
†The corresponding author.