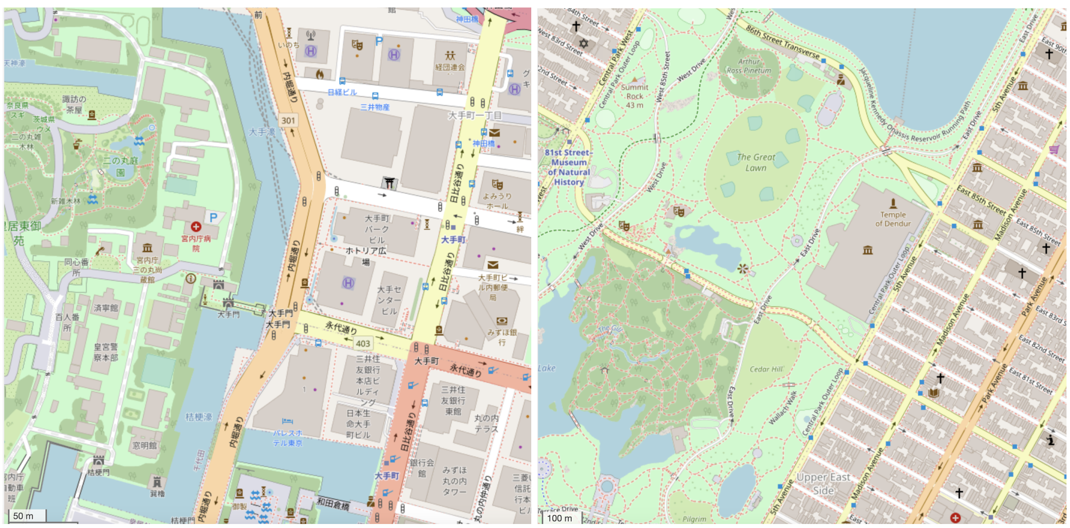
Proceedings of Machine Learning Research 222, 2023 ACML 2023

Supplementary Material "Lost and Found: How Self-Supervised Learning Helps GPS Coordinates Find Their Way"

0.1. Map Images



(*a*) Tokyo (*b*) New York

Figure 1: Two examples of map images from OpenStreeMap (OSM) OpenStreetMap contributors (2017).

As mentioned in the paper, we employ a map-based approach that converts GPS coordinates into map images to leverage the rich spatial and contextual information provided by maps. An example is provided in Figure 1. As can be observed, maps are rich sources of information. They contain various elements such as shop names, road structures, landmarks, points of interest, and other geographic features. This wealth of information allows us to capture a comprehensive understanding of the environment and enables our model to leverage the spatial context and relationships present in the maps.

0.2. Additional Results: Population Density Prediction

In the supplementary material, we also report results on another downstream task, population density prediction. We randomly selected 50,000 GPS coordinates from the dataset, and extracted the corresponding feature embeddings from their map images using the trained student network. We then trained a classification model to predict the population density based on the feature

Pretrain	1 epoch	25 epochs	50 epochs
Random Init	18.16/25.42	39.85/50.28	51.51/61.02
raw GPS	2.59/6.63	2.57/5.60	1.09/4.22
MoCo-V2 Chen et al. (2020)	34.15/50.13	48.42/61.21	54.11/67.43
DINO Caron et al. (2021)	37.83/54.28	52.67/66.40	66.24/78.32
MoCo-V2+Geo+TP Ayush et al. (2021)	49.30/66.31	61.29/76.79	71.76/85.34
Tile2Vec Jean et al. (2019)	38.49/54.25	55.09/67.20	62.22/73.71
GPS-SELM (geo)	56.12/71.47	64.23/80.11	72.34/85.17
GPS-SELM (geo+rec)	**58.20/72.10**	**65.19/81.29**	**73.40/86.28**
GPS-SELM (geo+rec+int)	54.98/70.23	63.41/80.09	72.95/86.21

Table 1: Top-1 accuracy/top-5 accuracy on the population density prediction. All results are averaged over ten evaluation trials.

embeddings. The cluster that the model aims to correctly predict is one cluster among 30 population density clusters. Similarly to the lance price experiment, the map images used for the population density task were not used for pretraining GPS-SELM. Table 1 presents the results of our experiments, including top-1 and top-5 accuracy of each method. Our method achieved the highest top-1 and top-5 accuracy, surpassing the previous baselines. These results highlight the effectiveness of enhancing the DINO framework with geo-predictive tasks and a high-level reconstruction task. Moreover, it is worth noting that GPS-SELM achieves a relatively good level of accuracy in estimating population density even after just a single fine-tuning epoch.

References

Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Geography-aware self-supervised learning. In *International Conference on Computer Vision*, pages 10181–10190, 2021.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *International Conference on Computer Vision*, pages 9650–9660, 2021.

Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. Tile2vec: Unsupervised representation learning for spatially distributed data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):3967–3974, Jul. 2019.

OpenStreetMap contributors. OpenStreetMap . https://www.openstreetmap.org, 2017.