Self-supervising Fine-grained Region Similarities for Large-scale Image Localization (Supplementary Materials)

Yixiao Ge[1], Haibo Wang[3], Feng Zhu[2], Rui Zhao[2], and Hongsheng Li[1]

[1] The Chinese University of Hong Kong
[2] SenseTime Research
[3] China University of Mining and Technology
{yxge@link,hsli@ee}.cuhk.edu.hk
{zhufeng,zhaorui}@sensetime.com haibo@cumt.edu.cn

A Related Work (Cont.)

Image-based Localization (IBL). Besides the previous weakly-supervised methods [2, 6, 7] and our work [5], there exists another stream of fully-supervised research which focused on datasets with 6DoF camera pose information [3, 4, 8, 9]. Such datasets are generally difficult to collect and scale up, since they require extra human and computational costs to capture dense images and obtain 6DoF information via post-processing, *e.g.* mapping, SfM [12], *etc.* In contrast, datasets [11, 10] studied in weakly-supervised methods are much easier to scale up, since the data and GPS information can be easily obtained for free from the internet without extra costs, *e.g.* Google Street View [1], Baidu Total View.

B Progressively Refined Supervisions

As shown in Tab. 1, the retrieval accuracies gradually increase as the network generation proceeds and saturate after the 4^{th} one, which indicates that the self-predicted soft supervisions in our work [5] are progressively refined by training in generations.

Table 1. Performance of our proposed method in different generations on Tokyo 24/7, in terms of Recall@1/5/10 (%)

Metric	1^{st}	2^{nd}	3^{rd}	$\mathbf{4^{th}}$
R@1	80.6	82.6	84.2	**85.4**
R@5	87.6	89.2	90.5	**91.1**
R@10	90.8	92.7	92.8	**93.3**

References

1. Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J.: Google street view: Capturing the world at street level. Computer **43**(6), 32–38 (2010)
2. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: Netvlad: Cnn architecture for weakly supervised place recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5297–5307 (2016)
3. Carlevaris-Bianco, N., Ushani, A.K., Eustice, R.M.: University of Michigan North Campus long-term vision and lidar dataset. International Journal of Robotics Research **35**(9), 1023–1035 (2015)
4. Chen, D.M., Baatz, G., Köser, K., Tsai, S.S., Vedantham, R., Pylvänäinen, T., Roimela, K., Chen, X., Bach, J., Pollefeys, M., et al.: City-scale landmark identification on mobile devices. In: CVPR 2011. pp. 737–744. IEEE (2011)
5. Ge, Y., Wang, H., Zhu, F., Zhao, R., Li, H.: Self-supervising fine-grained region similarities for large-scale image localization. In: European Conference on Computer Vision (2020)
6. Kim, H.J., Dunn, E., Frahm, J.M.: Learned contextual feature reweighting for image geo-localization. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3251–3260. IEEE (2017)
7. Liu, L., Li, H., Dai, Y.: Stochastic attraction-repulsion embedding for large scale image localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2570–2579 (2019)
8. Maddern, W., Pascoe, G., Linegar, C., Newman, P.: 1 Year, 1000km: The Oxford RobotCar Dataset. The International Journal of Robotics Research **36**(1), 3–15 (2017)
9. Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A., Fitzgibbon, A.: Scene coordinate regression forests for camera relocalization in rgb-d images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2930–2937 (2013)
10. Torii, A., Arandjelovic, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 place recognition by view synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1808–1817 (2015)
11. Torii, A., Sivic, J., Pajdla, T., Okutomi, M.: Visual place recognition with repetitive structures. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 883–890 (2013)
12. Ullman, S.: The interpretation of structure from motion. Proceedings of the Royal Society of London. Series B. Biological Sciences **203**(1153), 405–426 (1979)