Supplementary material for More Photos are All You Need: Semi-Supervised Learning for Fine-Grained Sketch Based Image Retrieval

Ayan Kumar Bhunia[1] Pinaki Nath Chowdhury[1,2] Aneeshan Sain[1,2] Yongxin Yang[1,2]
Tao Xiang[1,2] Yi-Zhe Song[1,2]
[1]SketchX, CVSSP, University of Surrey, United Kingdom
[2]iFlyTek-Surrey Joint Research Centre on Artificial Intelligence.
{a.bhunia, p.chowdhury, a.sain, yongxin.yang, t.xiang, y.song}@surrey.ac.uk

A. Necessity for rasterization

This is common practice in the FG-SBIR literature. Rasterized sketch-images tend to have better spatial encoding than coordinate-based alternatives [4, 1]. This is also verified in our work – removing rasterization and using sketch-coordinate retrieval reduces acc@1 to only 7.6% on Shoe-V2.

B. Motivation behind using discriminator for certainty score:

We are mostly inspired by recent image generation works [8, 3] that use the discriminator scores to iteratively improve generation quality. We also did an ablative study to investigate further (see L768-785 and Fig. 4(a)). We found that synthetic sketch-photo pairs having higher discriminator score, tend to have much better quality, and vice-versa. We will add some qualitative examples to further illustrate this correlation in supplementary materials. Defining a hard threshold (optimised) to eliminate bad generated sketches is an option – new experiments show acc@1 lags by around 2% compared to ours on Shoe-V2.

C. More details on experimental setup and analysis:

(i) Our self-designed baselines use the same backbone network, while joint-training is employed for Ours-F-Pix2Pix, Ours-F-L2S and Ours-F-Full.

(ii) SOTA data-augmentation strategies are already adopted by existing FG-SBIR works [4, 6]. However, they fail to capture the significant style variations that exist in real human sketches. In fact we already compare with [9] which employed such sketch specific augmentation strategies, and it is found to be much inferior to our semi-supervised framework (see Table. 2).

(iii) Optimising the final layer (using Eq. 9 in our case) is a very common practice during fine-tuning with RL, and is heavily adopted by the image-captioning literature [2], and very recently by on-the-fly FG-SBIR [1].

(iv) Edge-map hardly resembles the highly abstracted and subjective nature of amateur human sketches. For example, sketches do not follow the perfect edge boundary unlike edge-maps, thus model trained on pseudo-sketches via edge2sketch [5] falls short to generalise to real human sketches.

(v) Acc@1 without using RL scheme for Ours-F-Pix2Pix is 34.14%.

(vi) In future, our photo-to-sketch generation model could further be evaluated on Sketchy [7], however, it seems to be comparatively difficult than that of QMUL-ShoeV2 due to more noisy background.

References

[1] Ayan Kumar Bhunia, Yongxin Yang, Timothy M Hospedales, Tao Xiang, and Yi-Zhe Song. Sketch less for more: On-the-fly fine-grained sketch based image retrieval. In *CVPR*, 2020. 1

[2] Junlong Gao, Shiqi Wang, Shanshe Wang, Siwei Ma, and Wen Gao. Self-critical n-step training for image captioning. In *CVPR*, 2019. 1

[3] Minyoung Huh, Shao-Hua Sun, and Ning Zhang. Feedback adversarial learning: Spatial feedback for improving generative adversarial networks. In *CVPR*, 2019. 1

[4] Kaiyue Pang, Ke Li, Yongxin Yang, Honggang Zhang, Timothy M Hospedales, Tao Xiang, and Yi-Zhe Song. Generalising fine-grained sketch-based image retrieval. In *CVPR*, 2019. 1

[5] Umar Riaz Muhammad, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Learning deep sketch abstraction. In *CVPR*, 2018. 1

[6] Aneeshan Sain, Ayan Kumar Bhunia, Yongxin Yang, Tao Xiang, and Yi-Zhe Song. Cross-modal hierarchical modelling for fine-grained sketch based image retrieval. In *BMVC*, 2020. 1

[7] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database: learning to retrieve badly drawn bunnies. *ACM TOG*, 2016. 1

[8] Firas Shama, Roey Mechrez, Alon Shoshan, and Lihi Zelnik-Manor. Adversarial feedback loop. In *ICCV*, 2019. 1

[9] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy M Hospedales, and Chen-Change Loy. Sketch me that shoe. In *CVPR*, 2016. 1