Supplementary Material
Perceptual Loss for Robust Unsupervised Homography Estimation

Daniel Koguciuk, Elahe Arani, Bahram Zonooz
Advanced Research Lab, NavInfo Europe, Eindhoven, The Netherlands
{daniel.koguciuk, elahe.arani}@navinfo.eu, bahram.zonooz@gmail.com

1. Photometric or Perceptual

In this Section, we want to additionally explore the possible reason behind the effectiveness of perceptual loss functions. Photometric loss is known to be sensitive to illumination conditions, while high-level features extracted from pretrained networks care more about perceptual similarity [1, 2]. To better understand the behavior of both losses we prepare a simple experiment, where the target image is shifted by a given number of pixels in the X and Y axis w.r.t the source image. Then we report L_1 distance in pixel space and L_1 distance in feature space produced by the *Loss Network* g between both images. To bring both distances in the same range we normalize them by the maximum observed distance. The average of one hundred images for both S-COCO and PDS-COCO is presented in Figure 1.

The distance curve of photometric loss and perceptual loss on S-COCO is similar, so we expect the comparable performance of both loss functions. However, when photometric distortion is introduced, the perceptual loss function is smoother and will likely produce better results. Indeed, both conclusions are supported by the illumination robustness experiments shown in Section 5.3 of the main paper.

2. biHomE Performance on Real-World Dataset

The dataset proposed by Zhang *et al*. [3] is composed of $80k$ image pairs extracted from short video clips containing small camera movements and dynamic objects on the scene. The image pairs are divided into 5 categories: regular (RE), low-texture (LT), low-light (LL), small-foregrounds (SF), and large-foreground (LF) scenes. We reproduced the original Zhang *et al*. [3] method and using the same learning setting we also learned our *biHomE* loss.

As one can observe in Table 1 the performance of the original Zhang *et al*. [3] method is only slightly better than using our *biHomE* loss. A similar effect could be observed also in Section 5.3, where for small viewpoint change ρ and small photometric distortion δ original Zhang method is also better than with our *biHomE* loss.

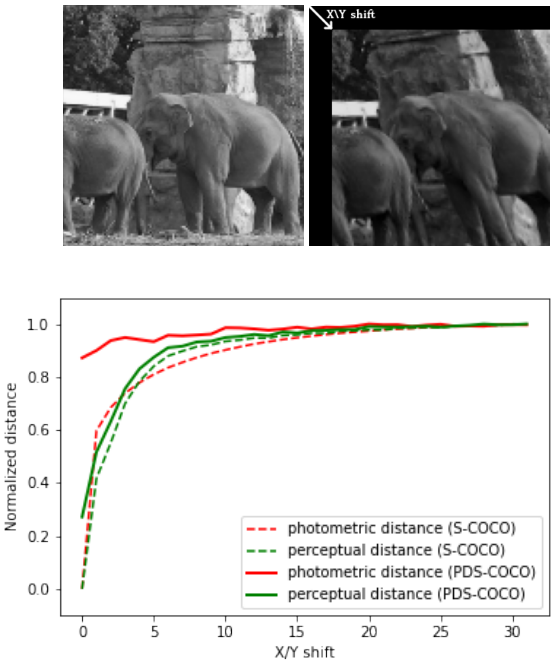

Figure 1: Normalized distance as a function of image shift. The Figure is prepared for one hundred random images from S-COCO (dashed) PDS-COCO dataset (solid), where the target image was shifted by a given number of pixels in both X and Y axes. We used pretrained ResNet34 up to the first residual block as perceptual *Loss Network* and MSE as photometric distance. For S-COCO distance statistics is similar for both distances, but for PDS-COCO perceptual distance is smoother.

References

[1] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016. 1

[2] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *Proceedings of the European Conference on Computer Vi-*

	RE	LT	LL	SF	LF	Avg
Zhang (reported)	1.81	1.90	1.94	1.75	1.72	1.82
Zhang (reproduced)	1.813 $\pm$0.013	2.139 $\pm$0.013	1.906 $\pm$0.013	1.837 $\pm$0.010	1.894 $\pm$0.006	1.918 $\pm$0.006
Zhang + biHomE	1.822 $\pm$0.006	2.178 $\pm$0.031	1.924 $\pm$0.011	1.842 $\pm$0.006	1.994 $\pm$0.009	1.941 $\pm$0.008

Table 1: The performance of the original Zhang *et al*. [3] method (reported in the paper and reproduced by us) and trained with our *biHomE* loss on their dataset [3]. The performance of the original Zhang *et al*. [3] method is only slightly better than using our biHomE loss. We hypothesize this is because this dataset consists mostly of image pairs with small viewpoint and illumination changes.

sion (ECCV) Workshops, pages 0–0, 2018. 1

[3] Jirong Zhang, Chuan Wang, Shuaicheng Liu, Lanpeng Jia, Nianjin Ye, Jue Wang, Ji Zhou, and Jian Sun. Content-Aware Unsupervised Deep Homography Estimation. *arXiv preprint arXiv:1909.05983*, 2019. 1, 2