Learning Semantic Correspondence with Sparse Annotations
- Supplementary Material -

Shuaiyi Huang[1], Luyu Yang[1], Bo He[1], Songyang Zhang[2],
Xuming He[3,4], and Abhinav Shrivastava[1]

[1] University of Maryland, College Park
[2] Shanghai AI Laboratory
[3] ShanghaiTech University
{huangshy,loyo,bohe}@umd.edu, zhangsongyang@pjlab.org.cn,
hexm@shanghaitech.edu.cn, abhinav@cs.umd.edu

In this supplementary material, we provide more detailed quantitative analysis and qualitative results of our method as follows: i) Apart from the PCK results by category reported in the main paper, we additionally provide the PCK results by variation factors in Sec. A; ii) We further analyse model complexity by computing FLOPS in Sec. B; iii) Finally, we provide more qualitative results on PF-PASCAL [2], PF-WILLOW [3], and SPair-71k [1] in Sec. C.

A. More Quantitative Results on SPair-71k

To have a better understanding of our method in different challenging scenarios, we report quantitative performance with respect to different levels of four variation factors (viewpoint, scale, truncation, and occlusion) on SPair-71k benchmark [1], as summarized in Table S1. Large PCK gains for all levels of image pairs indicate the robustness and effectiveness of our method.

Table S1. PCK analysis by variation factors on SPair-71k [1] ($\alpha_{bbox} = 0.1$). The variation factors include view-point, scale, truncation, and occlusion with various difficulty levels. Numbers in bold indicate the best performance and underlined ones are the second best.

Methods	View Point			Scale			Truncation				Occlusion				**All**
	easy	medi	hard	easy	medi	hard	none	src	tgt	both	none	src	tgt	both	
CNNGeo$_{\text{ResNet-101}}$ [4]	28.8	12.0	6.4	24.8	18.7	10.6	23.7	15.5	17.9	15.3	22.9	16.1	16.4	14.4	20.6
A2Net$_{\text{ResNet-101}}$ [5]	30.9	13.3	7.4	26.1	21.1	12.4	25.0	17.4	20.5	17.6	24.6	18.6	17.2	16.4	22.3
WeakAlign$_{\text{ResNet-101}}$ [6]	29.3	11.9	7.0	25.1	19.1	11.0	24.0	15.8	18.4	15.6	23.3	16.1	16.4	15.7	20.9
NC-Net$_{\text{ResNet-101}}$[7]	26.1	13.5	10.1	24.7	17.5	9.9	22.2	17.1	17.5	16.8	22.0	16.3	16.3	15.2	20.1
HPF$_{\text{ResNet-101}}$[8]	35.6	20.3	15.5	33.0	26.1	15.8	31.0	24.6	24.0	23.7	30.8	23.5	22.8	21.8	28.2
SCOT$_{\text{ResNet-101}}$[9]	42.7	28.0	23.9	41.1	33.7	21.4	39.0	32.4	30.0	30.0	39.0	30.3	28.1	26.0	35.6
DHPF$_{\text{ResNet-101}}$[10]	43.1	31.0	27.3	42.0	35.6	25.0	40.3	34.7	32.5	30.9	40.4	32.5	30.3	28.1	37.3
CATs$_{\text{ResNet-101}}$[11]	54.0	45.5	43.1	54.7	49.3	35.3	48.1	53.7	42.3	42.4	44.0	53.2	42.9	41.7	49.9
PMNC$_{\text{ResNet-101}}$[12]	53.3	<u>47.4</u>	<u>45.9</u>	53.7	49.6	<u>41.5</u>	**54.3**	46.8	45.0	41.9	**54.2**	43.9	43.0	38.4	50.4
Ours(ST)$_{\text{ResNet-101}}$	<u>57.1</u>	47.1	44.8	<u>56.3</u>	<u>52.2</u>	39.6	48.7	<u>56.5</u>	<u>45.9</u>	<u>43.5</u>	46.4	<u>55.6</u>	<u>45.9</u>	<u>43.1</u>	<u>52.4</u>
Ours(MT)$_{\text{ResNet-101}}$	**59.6**	**50.7**	**48.3**	**59.0**	**55.3**	**43.4**	<u>52.5</u>	**59.3**	**48.8**	**46.0**	<u>50.3</u>	**58.3**	**49.0**	**46.1**	**55.3**

B. FLOPS Comparison

We compare the model complexity of our proposed network with existing work [11,7] by computing FLOPS with facebookresearch/fvcore library. We summarize the results in Table S2. Our proposed network has 1.54M and 310.27M lower total FLOPS compared with CATs [11] and NCNet [7], respectively, as we do not use any conv4d or self-attention layers for correlation refinement.

Table S2. FLOPS Comparison between baselines and ours.

Model	Corr Refine	Total FLOPS (M)	Conv Op. FLOPS (M)	Linear Op. FLOPS (M)
CATs [11]	Self-Attention	3.52	1.83	1.54
NCNet [7]	Conv4d	312.25	312.07	0.00
Ours	None	1.98	1.83	0.11

C. More Qualitative Results

More qualitative results from our method (MT) on SPair-71k [1], PF-PASCAL [2] and PF-WILLOW [3], are shown in Figure S1, S2 and S3, respectively.

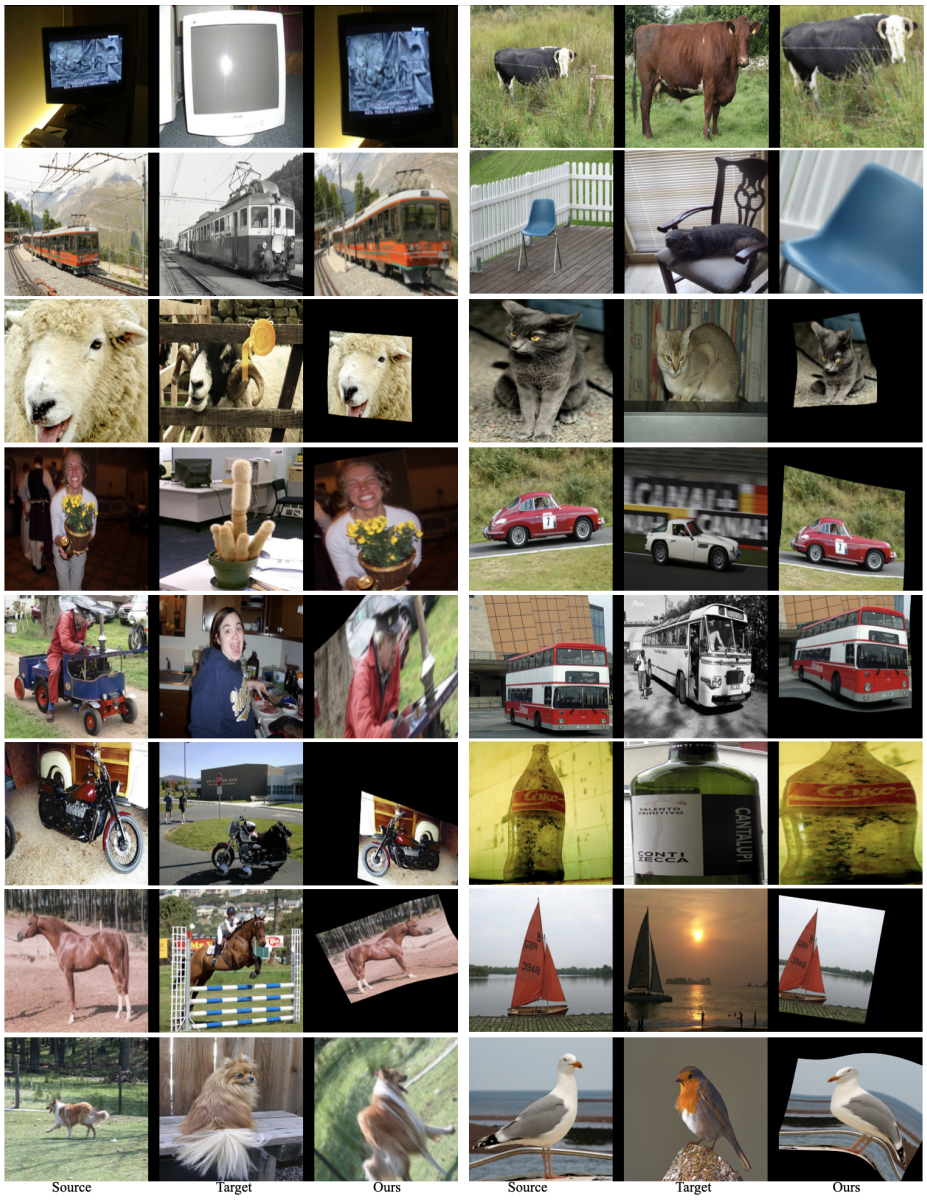

Fig. S1. Qualitative results on SPair-71k benchmark [1]. From left to right are source image, target image and results from our method, respectively.

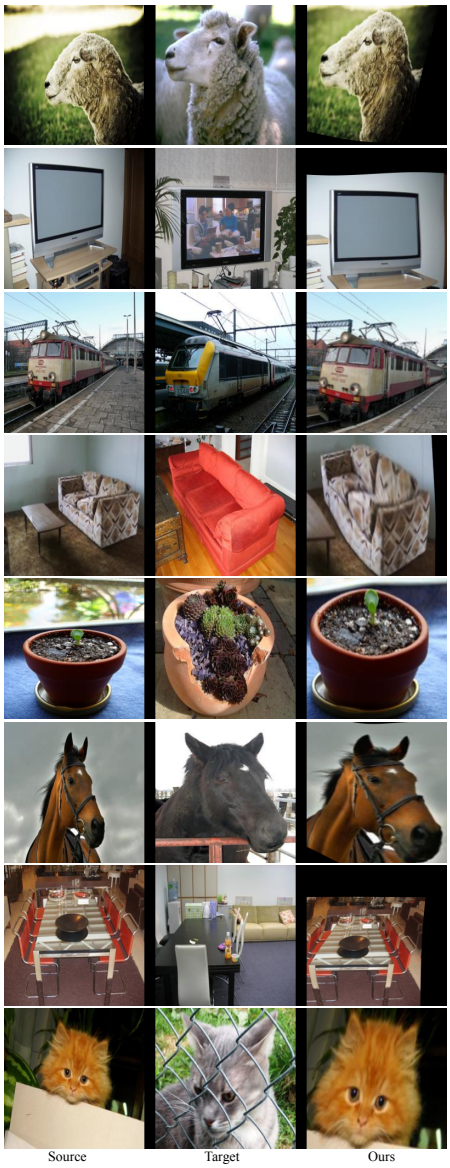

Fig. S2. Qualitative results on PF-PASCAL benchmark [2]. From left to right are source image, target image and result from our method, respectively.

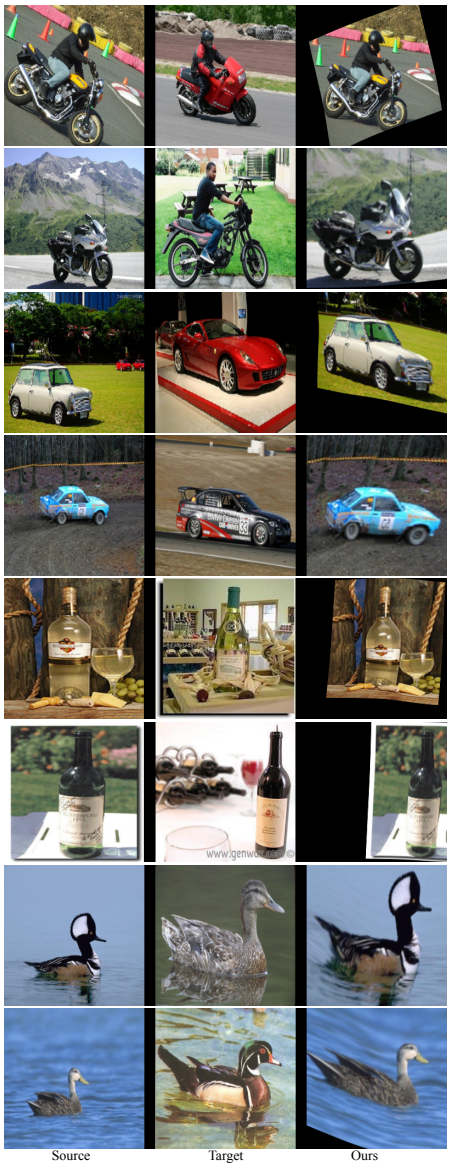

Fig. S3. Qualitative results on PF-WILLOW benchmark [3]. From left to right are source image, target image and result from our method, respectively.

References

1. Juhong Min, Jongmin Lee, Jean Ponce, and Minsu Cho. Spair-71k: A large-scale benchmark for semantic correspondence. *arXiv preprint arXiv:1908.10543*, 2019. 1, 2, 3
2. Bumsub Ham, Minsu Cho, Cordelia Schmid, and Jean Ponce. Proposal flow: Semantic correspondences from object proposals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 1, 2, 4
3. Bumsub Ham, Minsu Cho, Cordelia Schmid, and Jean Ponce. Proposal flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2016. 1, 2, 4
4. Ignacio Rocco, Relja Arandjelović, and Josef Sivic. Convolutional neural network architecture for geometric matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2017. 1
5. Paul Hongsuck Seo, Jongmin Lee, Deunsol Jung, Bohyung Han, and Minsu Cho. Attentive semantic alignment with offset-aware correlation kernels. In *Proceedings of the European Conference on Computer Vision(ECCV)*, 2018. 1
6. Ignacio Rocco, Relja Arandjelović, and Josef Sivic. End-to-end weakly-supervised semantic alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2018. 1
7. Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Neighbourhood consensus networks. In *Advances in Neural Information Processing Systems(NeurIPS)*, 2018. 1, 2
8. Juhong Min, Jongmin Lee, Jean Ponce, and Minsu Cho. Hyperpixel flow: Semantic correspondence with multi-layer neural features. In *Proceedings of the IEEE International Conference on Computer Vision(ICCV)*, 2019. 1
9. Yanbin Liu, Linchao Zhu, Makoto Yamada, and Yi Yang. Semantic correspondence as an optimal transport problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2020. 1
10. Juhong Min, Jongmin Lee, Jean Ponce, and Minsu Cho. Learning to compose hypercolumns for visual correspondence. In *Proceedings of the European Conference on Computer Vision(ECCV)*, 2020. 1
11. Seokju Cho, Sunghwan Hong, Sangryul Jeon, Yunsung Lee, Kwanghoon Sohn, and Seungryong Kim. Cats: Cost aggregation transformers for visual correspondence. In *Advances in Neural Information Processing Systems(NeurIPS)*, 2021. 1, 2
12. Jae Yong Lee, Joseph DeGol, Victor Fragoso, and Sudipta N Sinha. Patchmatch-based neighborhood consensus for semantic correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2021. 1