Joint Viewpoint and Keypoint Estimation with Real and Synthetic Data (Supplemental Material)

Pau Panareda Busto[0000−0002−2556−2531] and Juergen Gall[0000−0002−9447−3399]

Computer Vision Group, University of Bonn, Germany

1 Evaluation on the Pascasl3D+ Dataset

For the sake of completeness, we also evaluate our method on the popular *Pascal3D+* dataset [4], which also contains the same 12 classes evaluated on the *ObjectNet3D* dataset [5]. Compared to *ObjectNet3D*, this dataset provides object instances that are not centred in the middle of the image and can thus be found in any image location with different resolutions, producing more challenging scenarios. We follow the standard protocol when the keypoint and viewpoint accuracies are jointly evaluated. That is, for training our network we make use of all 5790 samples included in the training subset together with 28769 additional samples from the ImageNet dataset [2]. We only report our results on the fully visible objects of the test dataset, for a total of 2136 samples, with roughly 200-300 samples per class.

We report our trained network JVK with only real images (Re) and with real images together with the synthetic examples from the ShapeNet dataset [1] (Re-Sh). For both keypoint and viewpoint estimation results, we compare our method with the popular work by Tulsiani et al. (VpKp) and the State-of-the-Art method by Zhou et al. [6] (StarMap).

1.1 Keypoint Estimation

Pascal3D+ [4] (12 classes)		aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Avg.
PCK $\alpha = 0.1$	VpKp [3]	66.0	77.8	52.1	83.8	88.7	81.3	65.0	47.3	68.3	58.8	72.0	65.1	68.8
	StarMap [6]	75.2	83.2	54.8	87.0	94.4	90.0	**75.4**	58.0	68.8	**79.8**	54.0	**85.8**	78.6
	JVK (Re)	77.3	87.0	**68.1**	90.1	97.9	95.4	63.6	75.8	84.9	75.9	57.7.	70.7	78.7
	JVK (Re-Sh)	**80.4**	**89.9**	63.7	**90.5**	**98.0**	**96.8**	70.2	**78.9**	**86.7**	77.4	**63.6**	68.9	**80.5**

Table 1: Keypoint estimation on the Pascal3D+ dataset [4] for 12 object classes.

In Table 1, we show that JVK (Re) outperforms VpKP by a large margin and obtains comparable results to StarMap, but in a single joint pass and using a worse performing base CNN (a comparison between VGG-16 and ResNet-152

is shown in [6]). JVK (Re-Sh) obtains the best overall results, where additional synthetic data with only viewpoint annotations improves the keypoint accuracy (+1.8% with respect to JVK (Re)).

1.2 Viewpoint Estimation

Pascal3D+ [4] (12 classes)		aero	bike	boat	bottle	bus	car	chair	dtable	mbike	sofa	train	tv	Avg.
$Acc_{\frac{\pi}{6}}$	VpKp [3]	0.81	0.77	0.59	0.93	0.98	0.89	0.80	0.62	0.88	0.82	0.80	0.80	0.808
	StarMap [6]	0.82	0.86	0.50	0.92	0.97	0.92	0.79	0.62	0.88	0.92	0.77	0.83	0.823
	JVK (Re)	0.82	0.79	0.61	0.97	0.98	**0.95**	0.83	**0.65**	**0.89**	**0.97**	**0.83**	0.78	0.836
	JVK (Re-Sh)	**0.83**	**0.89**	**0.63**	**0.97**	**0.99**	0.94	**0.89**	**0.65**	0.86	0.90	0.82	**0.84**	**0.855**
MedError	VpKp [3] (pLike)	13.8	17.7	21.3	12.9	5.8	9.1	14.8	15.2	14.7	13.7	8.7	15.4	13.6
	StarMap [6]	**10.1**	14.5	30.0	9.1	**3.1**	6.5	11.0	23.7	14.1	11.1	7.4	13.0	10.4
	JVK (Re)	12.2	12.5	19.0	6.8	4.6	5.0	11.1	**9.4**	12.6	**10.6**	**5.9**	12.3	10.1
	JVK (Re-Sh)	11.3	**12.2**	**18.1**	**6.7**	3.8	**4.7**	**9.6**	11.6	**11.9**	11.1	6.3	**10.3**	**9.8**

Table 2: Viewpoint estimation on the Pascal3D+ dataset [4] for 12 object classes.

In Table 2, we show that JVK (Re) outperforms StarMap in both viewpoint accuracy and median error by +1.3% and -0.3, respectively. The best results are again obtained by JVK (Re-Sh), which improves by +1.9% the viewpoint accuracy and by -0.3 its median error compared to JVK (Re).

References

1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3D model repository. CoRR **abs/1512.3012** (2015)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
3. Tulsiani, S., Malik, J.: Viewpoints and keypoints. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1510–1519 (2015)
4. Xiang, Y., Mottaghi, R., Savarese, S.: Beyond Pascal: A benchmark for 3D object detection in the wild. In: IEEE Winter Conference on Applications of Computer Vision. pp. 75–82 (2014)
5. Xiang, Y., Kim, W., Chen, W., Ji, J., Choy, C., Su, H., Mottaghi, R., Guibas, L., Savarese, S.: Objectnet3d: A large scale database for 3D object recognition. In: IEEE European Conference on Computer Vision. pp. 160–176 (2016)
6. Zhou, X., Karpur, A., Luo, L., Huang, Q.: Starmap for category-agnostic keypoint and viewpoint estimation. In: IEEE European Conference on Computer Vision. pp. 318–334 (2018)