Supplementary: Towards Effective Synthetic Data Sampling for Domain Adaptive Pose Estimation

Isha Dua* **Arjun Sharma*** **Shuaib Ahmed** **Rahul Tallamraju**
Mercedes-Benz Research and Development India
{isha.dua, arjun.a.sharma, shuaib.ahmed, rahul.tallamraju}@mercedes-benz.com

1 Dataset

SURREAL[5] is a synthetically-generated dataset rendered from sequences of human motion capture data. The dataset has 6 million labeled frames of human body poses covering a wide variety of actions.

Leeds Sports Pose [3] (LSP) is a real-world outdoor human pose dataset capturing individuals in a wide range of poses, including challenging scenarios with occlusions and intricate body positions. Comprising 2000 images, it provides annotations for key human body joint locations, primarily gathered during sports activities.

Human3.6M [1] (H3.6M) is a real-world video dataset for human body pose estimation that includes data of diverse indoor activities. It has a total of 3.6 million frames.We follow the training and evaluation splits defined in [4]. The dataset has 5 subjects (S1, S5, S6, S7, S8) for training and the remaining 2 subjects (S9, S11) for testing. This split is typically adopted to train and evaluate models for human pose estimation.

Rendered Hand Pose Dataset [7] (RHD), is a synthetic dataset for the task of hand pose estimation. It encompasses a wide range of hand poses captured under varying lighting conditions and comprises $41.2k$ training images, $2.7k$ test images, and annotations for 21 hand keypoints.

Hand-3D-Studio dataset [6] (H3D), abbreviated as H3D, is a real-world dataset capturing multi-view indoor hand poses. It has a collection of $22k$ frames. Following a similar partitioning approach as used in the RegDA [2] framework, a subset of $3.2k$ frames is designated as the test set.

2 Additional Qualitative Results

References

[1] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36:1325–1339, 2014.

[2] Junguang Jiang, Yifei Ji, Ximei Wang, Yufeng Liu, Jianmin Wang, and Mingsheng Long. Regressive domain adaptation for unsupervised keypoint detection. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6776–6785, 2021.

[3] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *British Machine Vision Conference*, 2010.

[4] Sijin Li and Antoni B. Chan. 3d human pose estimation from monocular images with deep convolutional neural network. In *Asian Conference on Computer Vision*, 2014.

*Both authors contributed equally.

NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research.

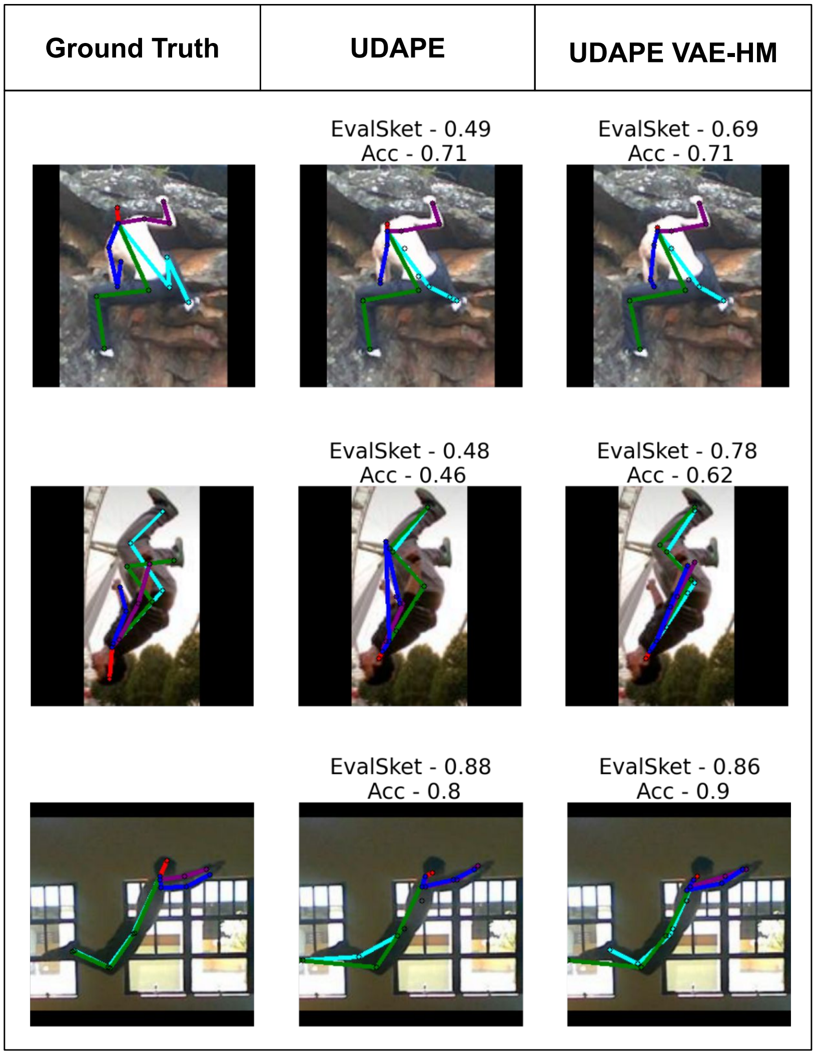

Figure 1: More qualitative results on SURREAL→ LSP

[5] Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4627–4635, 2017.

[6] Zheng Fa Zhao, Tianyao Wang, Siyu Xia, and Yangang Wang. Hand-3d-studio: A new multi-view system for 3d hand reconstruction. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2478–2482, 2020.

[7] Christiane Zimmermann and Thomas Brox. Learning to estimate 3d hand pose from single rgb images. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4913–4921, 2017.

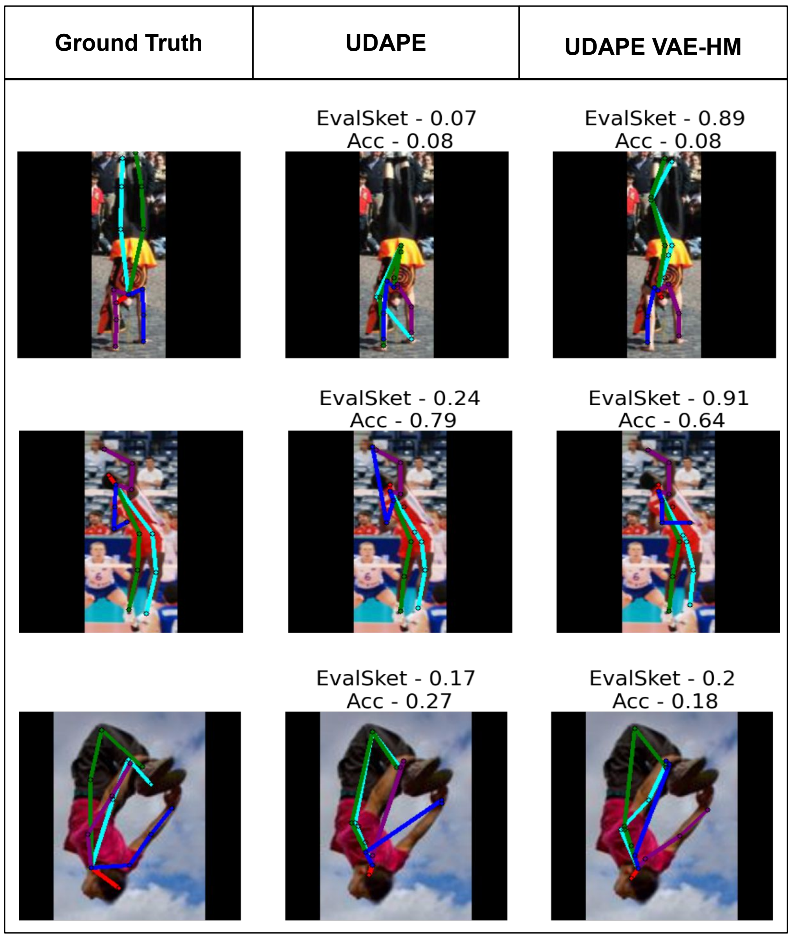

Figure 2: More qualitative results on SURREAL$\rightarrow$ LSP

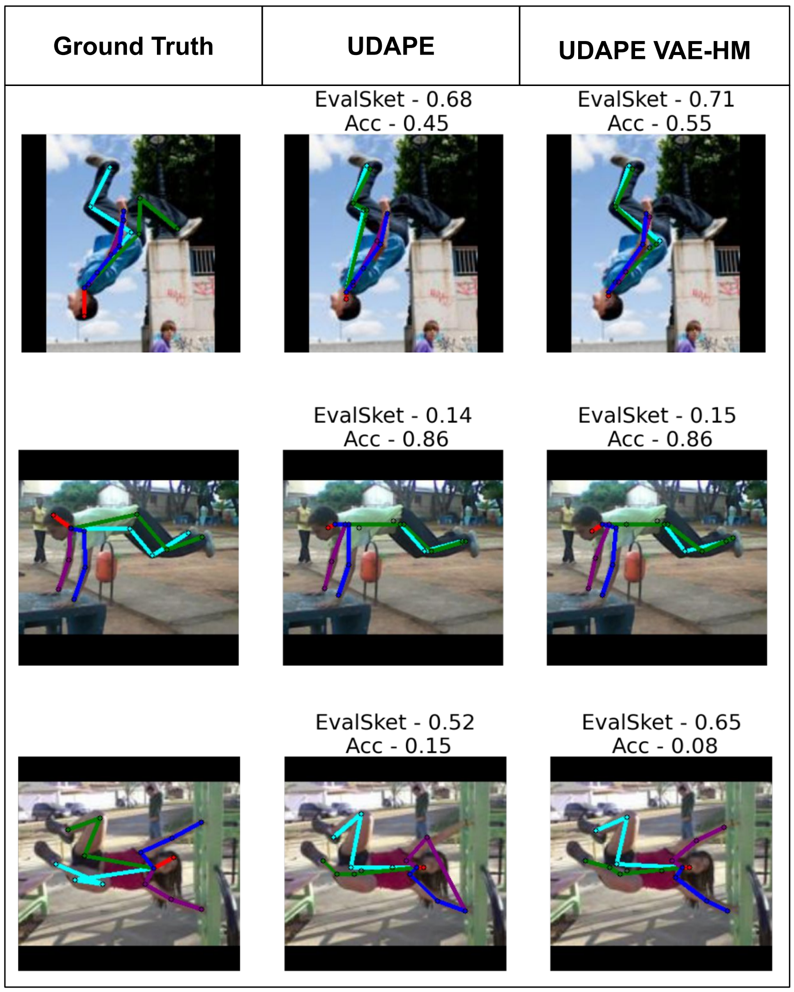

Figure 3: More qualitative results on SURREAL$\rightarrow$ LSP