Exploring Effective Data Augmentations for 3D Human Pose and Shape Estimation from a Single Egocentric Image

Bowen Dang

Xi'an Jiaotong University, Xi'an, China
dangbowen.bell@gmail.com

Abstract. In this paper, we present our solution to the EgoBody Challenge. The first phase of the challenge targets the task of 3D human pose and shape estimation from a single RGB image. Due to the existence of occlusion and motion blur, it is a challenging task to reconstruct the accurate 3D human mesh from an egocentric image. Based on SPIN, we explore some effective and general data augmentations to improve the model generalization on egocentric images. The model obtains 3rd place in the challenge after applying these data augmentations. Our code and model are publicly available at https://github.com/DangBowen-Bell/EgoBody-Challenge-Solution.

Keywords: 3D Human Pose and Shape Estimation

1 Introduction

3D human pose and shape estimation (3D HPS) from a single RGB image is a very challenging task and has a wide range of applications [2, 4–6, 8]. The EgoBody [9] dataset is shot in a real indoor scene and has high-quality 3D annotations which are obtained using the multi-view human reconstruction method. The difficulties of EgoBody dataset include two parts. The first is occlusion. Due to complex scenes and actions, the human may be occluded by objects in the scene, which makes it difficult to reconstruct the invisible parts correctly. The second is motion blur. Since the egocentric images are taken using HoloLens2 headset [1] worn by another person, it is difficult for the wearer to move smoothly, which leads to motion blur in some images.

In this paper, we explore some specially designed data augmentations to improve the model generalization on egocentric images. The experiments show that these tricks can effectively improve performance. In this challenge, our method achieves 3rd place.

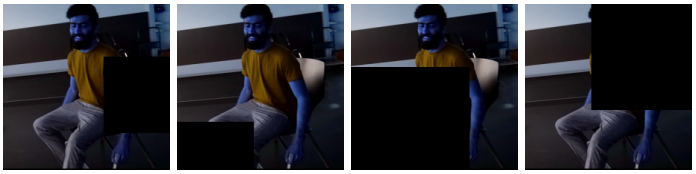
2 Method

We use SPIN [6] as our baseline and fine-tune the pre-trained model trained on EFT [4] dataset. We change the perspective projection model used in SPIN

back to the weak-perspective projection of HMR [5] the same as EFT. We remove the optimization module since EgoBody dataset provides high-quality 3D annotations.

2.1 Data Augmentation

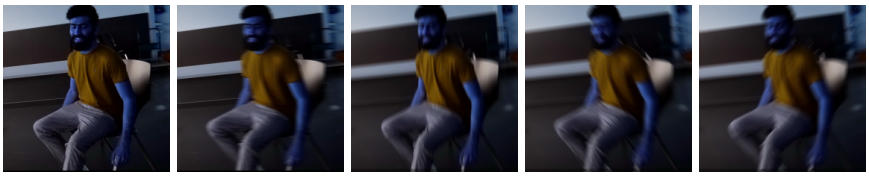
CutOut : We apply CutOut [3] data augmentation to the image to simulate possible occlusion during shooting. To be specific, we apply a square zero-mask to a random position of each image during each epoch of training. The square size is randomly selected from $[0, a \cdot R]$ where $R = 224$ is the image resolution. We find the performance is best when a is set to 0.8. The randomly generated sample is shown in Figure 1.



(a) (b) (c) (d)

Fig. 1: CutOut applied to images from the EgoBody dataset.

Motion Blur : We apply Blur data augmentation to the image to simulate possible motion blur while shooting. The usual Blur operation uses a fixed kernel for 2D filtering, which is too simple to simulate different situations [7]. We use a randomly generated kernel considering that the camera may move in different directions at different speeds. To be specific, we set up 4 directions (horizontal, vertical, and diagonal) and 3 sizes (3, 5, 7) to simulate various possible motion blur. The randomly generated sample is shown in Figure 2.



(a) Original (b) Horizontal (c) Vertical (d) Diagonal (e) Diagonal

Fig. 2: Motion Blur applied to images from the EgoBody dataset.

2.2 Results

Before the experiment, we need to modify the 3D annotations of the EgoBody dataset so that we can directly fit the pre-trained model on it. To be specific, we transform the global orientation of the SMPL model from world coordinate to camera coordinate using provided calibration data.

We add the data augmentations incrementally and get 4 configurations. We use the same evaluation metrics as EgoBody [9] and compare the results on the test set in table 1. As we can see from the table, the generalization of the model improves after adding the data augmentations.

Table 1: The results on EgoBody test set

Model	MPJPE	PA-MPJPE	V2V	PA-V2V
SPIN	130.1220	81.9946	140.0008	90.0855
SPIN-ft	93.1091	61.5779	104.8394	70.5731
SPIN-ft(CO)	89.7161	58.6065	100.4049	66.2072
SPIN-ft(CO+MB)	87.8209	58.0566	98.3177	65.8668

3 Conclusions

In this paper, we explore some effective and general data augmentations to improve the model generalization on egocentric images, which helps us win the 3rd place in the challenge.

References

1. Microsoft Hololens2. https://www.microsoft.com/en-us/hololens
2. Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M.J.: Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In: European conference on computer vision. pp. 561–578. Springer (2016)
3. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
4. Joo, H., Neverova, N., Vedaldi, A.: Exemplar fine-tuning for 3d human model fitting towards in-the-wild 3d human pose estimation. In: 2021 International Conference on 3D Vision (3DV). pp. 42–52. IEEE (2021)
5. Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J.: End-to-end recovery of human shape and pose. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7122–7131 (2018)
6. Kolotouros, N., Pavlakos, G., Black, M.J., Daniilidis, K.: Learning to reconstruct 3d human pose and shape via model-fitting in the loop. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2252–2261 (2019)
7. Rong, Y., Shiratori, T., Joo, H.: Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1749–1759 (2021)
8. Tian, Y., Zhang, H., Liu, Y., Wang, L.: Recovering 3d human mesh from monocular images: A survey. arXiv preprint arXiv:2203.01923 (2022)
9. Zhang, S., Ma, Q., Zhang, Y., Qian, Z., Kwon, T., Pollefeys, M., Bogo, F., Tang, S.: Egobody: Human body shape and motion of interacting people from head-mounted devices. In: European conference on computer vision (ECCV) (Oct 2022)