Domain Adaptive 3D Pose Augmentation (DAPA) for In-the-wild Human Mesh Recovery

Zhenzhen Weng [1], Kuan-Chieh Wang [1], Angjoo Kanazawa [2], Serena Yeung [1]

[1]Stanford University, [2] UC Berkeley

3DV 2022 International Conference on 3D Vision
Hybrid / Prague, Czechia
September 12-15, 2022

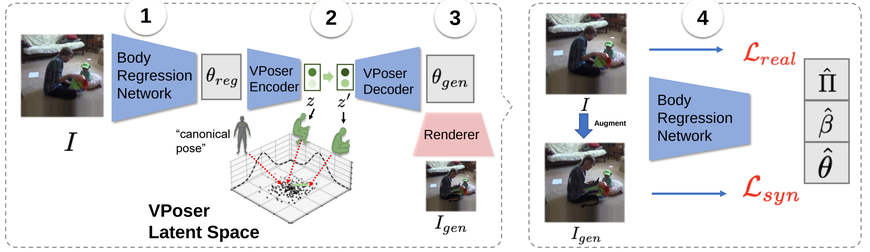
• Motivation

- Human mesh recovery (HMR) brings a lot of exciting opportunities in domains such as AR/VR. However, a fundamental challenge of HMR is in collecting the ground truth 3D mesh targets required for training
- Capturing such datasets usually require burdensome motion capturing systems in indoor scenes, and the captured poses often have limited diversity. As a result, models trained on those datasets tend not to generalize well to real world "in-the-wild" scenarios due to distribution shifts.
- We propose Domain Adaptive 3D Pose Augmentation (DAPA), a data augmentation method that **enhances the model's generalization ability** in **in-the-wild scenarios**.
- DAPA combines the strength of methods based on synthetic datasets by getting direct supervision from the synthesized meshes, and domain adaptation methods by using 2D keypoints from the target dataset.

Overview

- DAPA is **backbone-agnostic** - it takes the pose prediction of the backbone and generates synthetic data on-the-fly in an adaptive way. We showcase the effectiveness of DAPA using the backbone of SPIN [1].
- We first regress SMPL body parameters from input images using a pretrained body regression network. Next, we augment the estimated poses in the latent space of Vposer, a variational autoencoder trained on a massive dataset of real human poses. Then, we render synthetic training examples with augmented poses and learned textures. For real training examples, we minimize the 2D keypoint re-projection loss. For synthetic training examples, we have access to the ground truth information. Hence, we can directly supervise the predicted 2D/3D keypoints as well as the SMPL parameters.

• Method



1 Regress SMPL parameters

$$\{\beta_{reg}, \theta_{reg}, \Pi_{reg}\} = f(I)$$

2 Augment pose in VPoser latent space

$$\begin{aligned} \mu, \sigma &= \text{Encoder}_{\text{VPoser}}(\theta_{reg}) \\ z &\sim \mathcal{N}(\mu, \sigma I) \\ \tilde{z} &= z \odot (1 + s\epsilon), \quad \epsilon \sim \mathcal{U}[0,1] \\ \theta_{syn} &= \text{Decoder}_{\text{VPoser}}(\tilde{z}) \end{aligned}$$

3 Render augmented pose with learned textures

$$I_{gen} = \mathcal{R}(\mathcal{M}_{SMPL}(\beta_{reg}, \theta_{syn}, \Pi_{reg}); \mathcal{T}(I))$$

4 Supervise model on real and synthetic examples

$$\mathcal{L}_{real} = \lambda_{2D}||J_{reg} - J_{gt}||$$

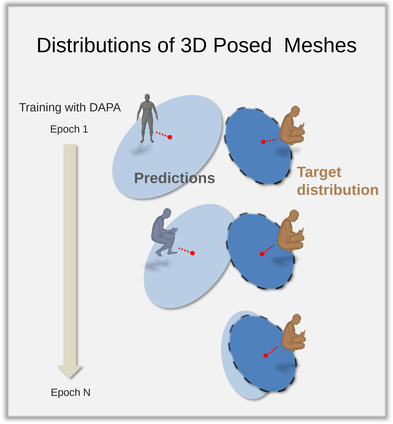
$$\mathcal{L}_{syn} = \lambda_{2D}\mathcal{L}_{syn,2D} + \lambda_{3D}\mathcal{L}_{syn,3D} + \lambda_{\theta}\mathcal{L}_{syn,\theta} + \lambda_{\beta}\mathcal{L}_{syn,\beta}$$

$$\mathcal{L}_{syn,2D} = ||J_{syn,reg} - J_{syn}||$$
$$\mathcal{L}_{syn,3D} = ||X_{syn,reg} - X_{syn}||$$
$$\mathcal{L}_{syn,\theta} = ||\theta_{syn,reg} - \theta_{syn}||$$
$$\mathcal{L}_{syn,\beta} = ||\beta_{syn,reg} - \beta_{syn}||$$

Overall Loss $\mathcal{L} = \mathcal{L}_{real} + \mathcal{L}_{syn}$

Notation

I : input image
f : body regression network
β_{reg} : body shape
θ_{reg} : body pose
Π_{reg} : camera translation
$\mathcal{R}$: Neural Mesh Renderer
$\mathcal{T}$: Texture model



References

[1] Kolotouros, et al. *CVPR* 2019.
[2] von Marcard, et al. *ECCV* 2018.
[3] Patel, et al. *CVPR* 2021.
[4] Bergelson, et al. *Developmental science 2019.*

• Experiments

- We consider the task of weakly-supervised domain adaptation where we finetune a pretrained model using 2D keypoints from the target dataset.
- We evaluate on 3D benchmarks 3DPW [2] and AGORA [3], and SEEDLingS [4], a dataset derived from real-world videos of parent-child interaction. We include quantitative results for SEEDLingS here. Please see paper for complete protocols and results.

Quantitative results on SEEDLingS (2D PCK)

Model	All joints	Eye	Shoulder	Elbow	Wrist	Hip	Knee	Ankle
SPIN-pt	52.8	83.3	51.1	71.2	65.0	50.1	43.2	25.2
SPIN-ft	53.1	**91.1**	62.2	73.9	**73.0**	35.6	36.1	22.6
DAPA (rand. Pose)	53.2	81.0	56.1	73.0	64.0	53.6	41.7	26.2
DAPA (static textures)	57.7	83.3	74.8	65.7	65.7	62.8	44.5	28.0
DAPA (full model)	**61.3**	85.9	**79.5**	**74.1**	67.5	**63.4**	**48.8**	**31.4**

Qualitative results (Top: SEEDLingS, bottom: Gymnastics video from YouTube)