DH-AUG: DH Forward Kinematics Model Driven Augmentation for 3D Human Pose Estimation (Supplementary Material)

Linzhi Huang, Jiahao Liang, and Weihong Deng*

Beijing University of Posts and Telecommunications
{huanglinzhi, jiahao.liang, whdeng}@bupt.edu.cn

1 Supplement of DH Parameter Model

We use DH parameters [2] to design a human kinematics model (DH parameter model). We provide our complete DH parameter table, as shown in Fig. 1. In Fig. 1, those marked with red triangles are variable parameters, and others are preset fixed parameters. There are 33 degrees of freedom (DOF) and 48 changeable DH parameters. In addition, we provide our complete DH parameter constraint table in Fig. 2. By constraining the joint angle, we limit the abnormal rotation of the joint. DH parameter constraint table is set according to personal experience. Better results can be obtained by adjusting these parameters.

2 Implementation Details

Generator. The specific structure of the DH-generator is shown in Fig. 3 (a). The input is the 128-dimensional vector sampled from the normal distribution. The output are changeable DH parameters, global rotation parameters and global translation parameters. The main module of the generator is the fully connected layer (FC) with residual block. Its activation function is $ReLU$. However, the last activation function is $tanh$. The feature dimensions in each FC layer are selected from 256 to 1000. After that, the parameters output by the FC network are transferred to the DH parameter model, and a new pose is obtained through a series of iterative calculations.

Discriminator. The specific structure of the multi-stream motion discriminator is shown in Fig. 3 (b). The input of the discriminator is the 3D and 2D coordinates of 16 key points, 15 joint angles, and their motion trajectories. Each branch in the discriminator has about 7 FC layers, and the merging part has about 4 FC layers. Its activation function is $ReLU$.

DH Parameter Table																	
	Right Leg					Left Leg					Trunk						
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a	0.25	0	0	0.60	0.50	-0.25	0	0	0.60	0.50	0	0	0	0	0	0	0
d	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0.2
α	0	-90	-90	0	0	0	90	90	0	0	0	-90	-90	-90	-90	-90	-90
θ	0	-90	180	0	0	180	-90	0	0	0	90	-90	-90	-90	-90	-90	-90
	Trunk					Right Hand					Left Hand						
ID	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	-
a	0	0	0	0	0	0.15	-0.30	0	0	0.40	0.35	0.30	0	0	0.40	0.35	
d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
α	-90	-90	-90	-90	-90	-90	-90	-90	-90	0	0	-90	90	90	0	0	
θ	-90	-90	-90	-90	0	0	-180	-90	-180	0	0	0	-90	0	0	0	

Fig. 1: **Complete DH parameter table.** There are 33 degrees of freedom (DOF) and 48 changeable DH parameters. a is the link length, d is the link offset, α is the twist angle, θ is the joint angle.

3 Training Details

Both generator and discriminator use Adam optimizer, and the learning rate remains 1e-4 unchanged. The training is carried out on one 1080ti GPU. Training about 100 to 140 epochs.

Single-frame. The batch size is 1024. The pose estimator uses the Adam optimizer with the learning rate of 1e-4, 1e-3. We first train GAN for about 2 to 5 epochs. Then both GAN and estimator are trained. The first 50 epochs use linear attenuation, and the subsequent epochs attenuate each epoch by 5% to 10%. See Fig. 4 for W-Distance during training.

Video. The batch size is 512. The pose estimator uses the Adam optimizer with the learning rate of 1e-4, 1e-3, or 2e-3. We first train the single-frame discriminator for 4 epochs and then turn on the multi-stream motion discriminator for training. After the 2 discriminators gradually stabilize, turn on the 3D pose estimator for training together. The first 50 epochs use linear attenuation, and the subsequent epochs attenuate each epoch by 5% to 10%. Fig. 5 show the W-Distance of the single-frame discriminator and the multi-stream motion discriminator respectively. Better results can be obtained by adjusting the parameters

4 More Analysis of The Data Distribution.

Here we provide the distribution of all joint angles. Before adding constraints (Fig. 6 (a)), the data distribution produced by the DH-generator is asymmetric and unreasonable. For example, DOF 3 and DOF 8 are the right knee and left knee joints of the human body respectively (the normal rotation range is about -180° to 0°). In Fig. 6 (a), the DOF 3 and DOF 8 have a lot of outward rotation values (between 0° and 180°), which indicates that the generator has learned the wrong human kinematics information, resulting in the ambiguity. In addition,

DH Parameter Constraint Table										
	Right Leg					Left Leg				
ID	1	2	3	4	5	6	7	8	9	10
Δa	(-0.2, 0.2)	(0, 0)	(0, 0)	(-0.2, 0.2)	(-0.2, 0.2)	(-0.2, 0.2)	(0, 0)	(0, 0)	(-0.2, 0.2)	(-0.2, 0.2)
Δd	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
Δα	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
Δθ	(-110, 65)	(-110, 65)	(-110, 180)	(-180, 0)	(0, 0)	(-65, 110)	(-65, 110)	(-110, 180)	(-180, 0)	(0, 0)
	Right Hand					Left Hand				
ID	24	25	26	27	28	29	30	31	32	33
Δa	(-0.2, 0.2)	(0, 0)	(0, 0)	(-0.2, 0.2)	(-0.2, 0.2)	(-0.2, 0.2)	(0, 0)	(0, 0)	(-0.2, 0.2)	(-0.2, 0.2)
Δd	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
Δα	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 0)
Δθ	(-155, 65)	(-155, 65)	(-100, 180)	(0, 180)	(0, 0)	(-65, 155)	(-65, 155)	(-100, 180)	(0, 180)	(0, 0)

Fig. 2: **Complete DH parameter constraint table.** Δa, Δd, $\Delta \alpha$, $\Delta \theta$ is the change in the DH parameter. Better results can be obtained by adjusting these parameters.

DOF 3 and DOF 8 are asymmetric. However, by observing Fig. 6 (b), it will be found that the distribution is symmetrical and reasonable.

5 More Qualitative Results

LSP [4], MPII [1] and MSCOCO [7] are three large 2D pose datasets containing a large number of outdoor scenes (excluding video). We use the single-frame version of VPose [8] as the baseline to compare the qualitative results before and after using our method (DH-AUG). DH-AUG achieves more accurate estimations as shown in Fig. 7.

6 Skeleton Video Generated by DH-AUG

We visualize the generated skeleton video (Fig. 8). It is observed that the action and trajectory are continuous and smooth. In addition, no wrong rotation occurred.

7 DH-3DP Dataset

We use a tool for making 2D-3D datasets offline. In this tool, we can manually set DH parameters to get 3D poses. In addition, we can also use this tool and pretrained DH-AUG to generate new datasets, as shown in Fig. 9.

We synthesized a new dataset (DH-3DP) with more than 1 million 2D-3D data pairs. The specific synthesis method of this dataset is: 1) We set DH parameter constraint table according to personal experience. 2) S15678 of H36M is used as the training set to train DH-AUG, with a total of 110 epochs. 3) We use the

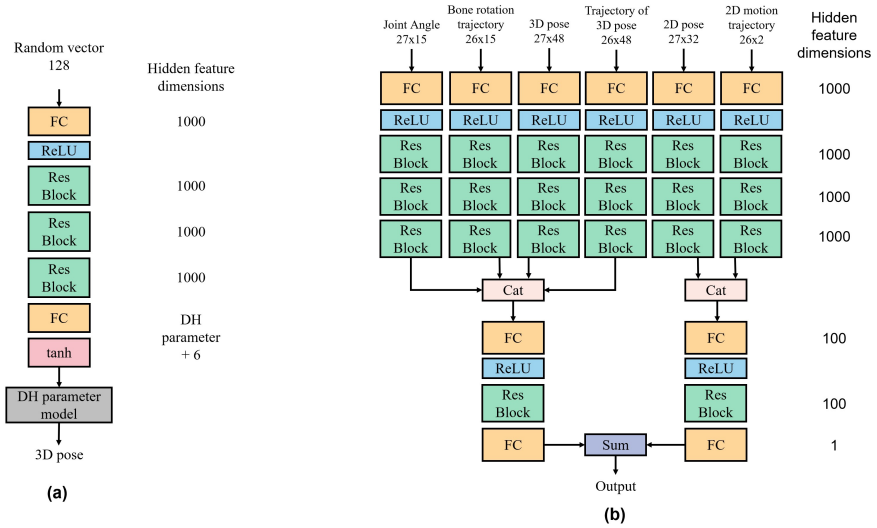

Fig. 3: **(a) DH-generator.** 128-dimensional vectors are sampled from the normal distribution and input into the fully connected network to obtain DH parameters, global rotation and translation parameters. **(b) Multi-stream motion discriminator (MSMD).** It has 3 two-stream branches. Better results can be obtained by adjusting these parameters.

Table 1: Quantitative comparison.

Method	train data	3DHP (↓)	H36M (↓)
VPose [8]	H36M[14]	86.6	41.8
VPose [8]	DH-3DP	79.2	42.6
VPose [8]	DH-3DP + H36M	**77.4**	**41.4**

pretrained DH-AUG and this tool to generate more than 1 million 2D-3D data pairs. Several cases in DH-3DP dataset are shown in Fig. 10. We release source code and new dataset (DH-3DP) at https://github.com/hlz0606/DH-AUG-DH-Forward-Kinematics-Model-Driven-Augmentation-for-3D-Human-Pose-Estimation. It is worth mentioning that our DH-AUG can be used in other datasets, which is helpful to relevant workers. We also suggest using other data to train DH-AUG and generate new data online or offline. We did a simple quantitative experiment, as shown in table1. It is worth noting that the generated dataset DH-3DP and the dataset H36M used for training DH-AUG can be mixed and trained in turn to get better results. We strongly recommend that **each epoch** generate new data **online** to get the best results.

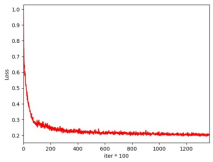

Fig. 4: The W-Distance of single -frame discriminator.

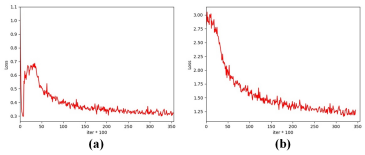

Fig. 5: **Video. (a)**: The W-Distance of single-frame discriminator. **(b)**: The W-Distance of multi-stream motion discriminator.

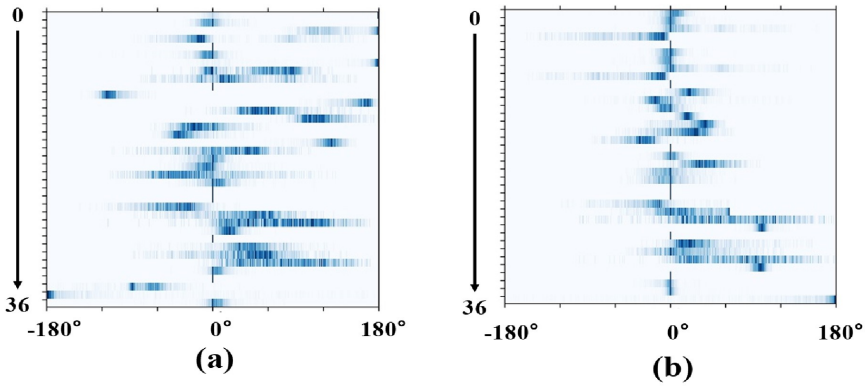

Fig. 6: **Data distribution of joint angle. (a)**: No constraints. **(b)**: Add constraints. **Y axis**: Joint angle index. **X axis**: Angle value. The darker the color, the greater the probability of generating the value. There are 33 joint angles and 3 global rotation angles. DOF 0,1,2,3 and 5,6,7,8 are symmetrical leg joints, DOF 24,25,26,27 and 29,30,31,32 are symmetrical hand joints.

8 More Quantitative Experiments

We conducted experiments on the CMU Panoptic[5] dataset. The training data used by DH-AUG and pose estimator are consistent. In addition, we also mixed H36M [3] dataset and CMU Panoptic[5] dataset to build a larger dataset for training. The experimental results show that using DH-AUG in large datasets can still greatly improve the performance of the model, as shown in table2. In addition, the discriminator used by VIBE[6] is used for SMPL. We modify its input modality and conduct quantitative experiments, as shown in the table3.

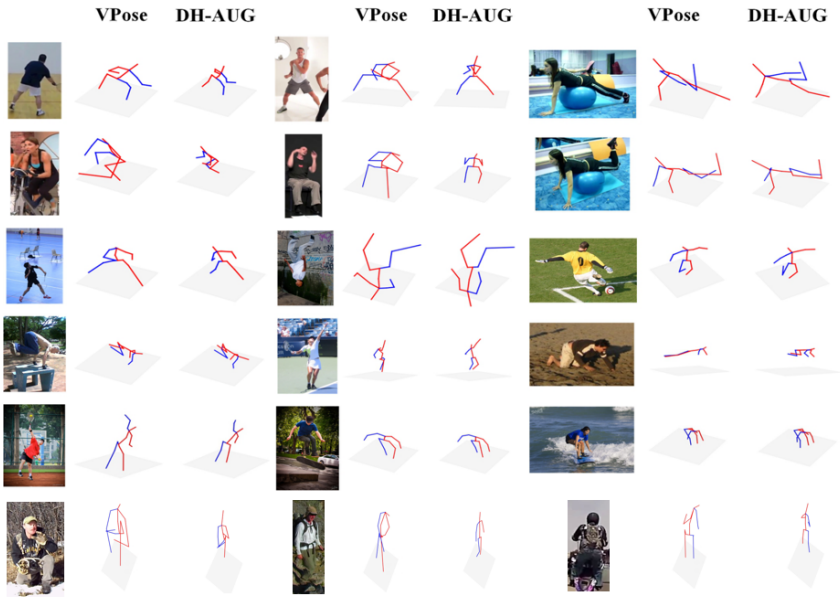

Fig. 7: **Qualitative results on MPII, LSP and MSCOCO data sets.** The first column is the result of VPose [8] before data augmentation, and the second column is the result of VPose after data augmentation (DH-AUG).

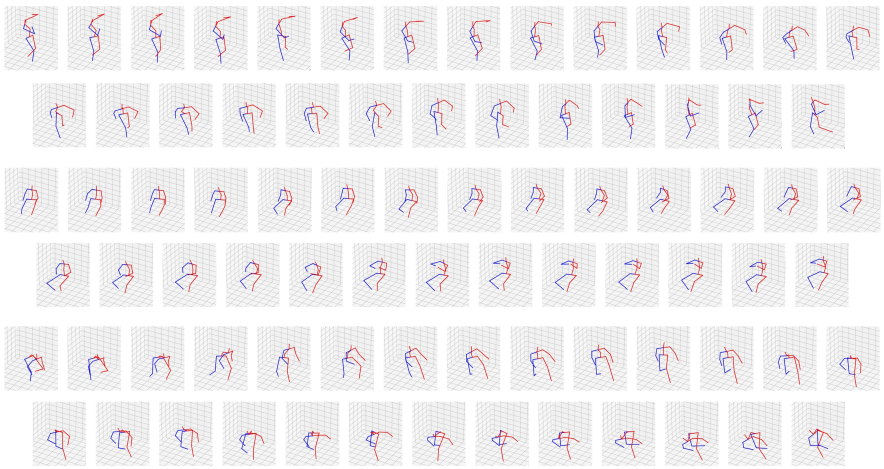

Fig. 8: **Skeleton videos generated by DH-AUG (27 frames).** There are 3 skeleton videos in total (standing, sitting and squatting).

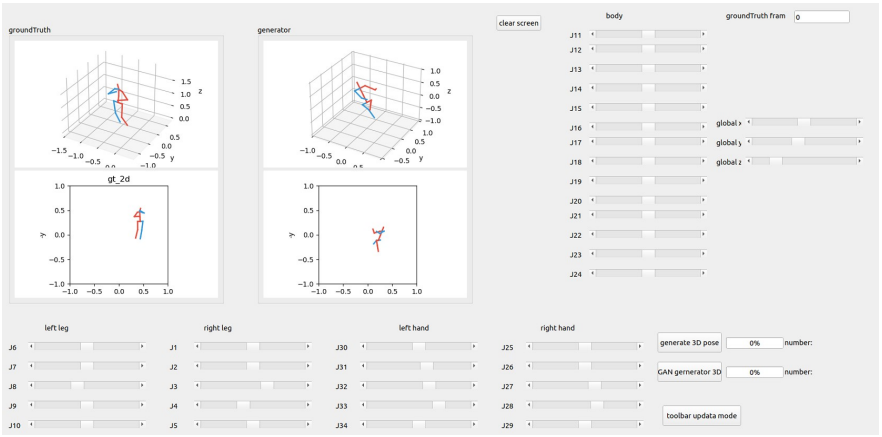

Fig. 9: **Tool for making datasets.**

Table 2: **Results on H36M, MPI and CMU.** Evaluation criteria: MPJPE. Best in bold.

Method	Train data	3DHP	H36M	CMU-Test
VPose [8]	CMU-Train	127.6	105.1	40.6
VPose + DH-AUG (Ours)	CMU-Train	**103.5**	**78.3**	**33.3**
VPose [8]	CMU-Train + H36M	78.4	45.2	35.2
VPose + DH-AUG (Ours)	CMU-Train + H36M	**70.1**	**37.0**	**32.2**

Table 3: Quantitative comparison between VIBE's[55] discriminator and our discriminator.

Method	3DHP (↓)	H36M (↓)
VPose (f=9)[32]	90.7	42.14
VPose (f=9) + DH-AUG (VIBE's discriminator[55])	83.4	41.63
VPose (f=9) + DH-AUG (Ours)	**80.39**	**41.21**

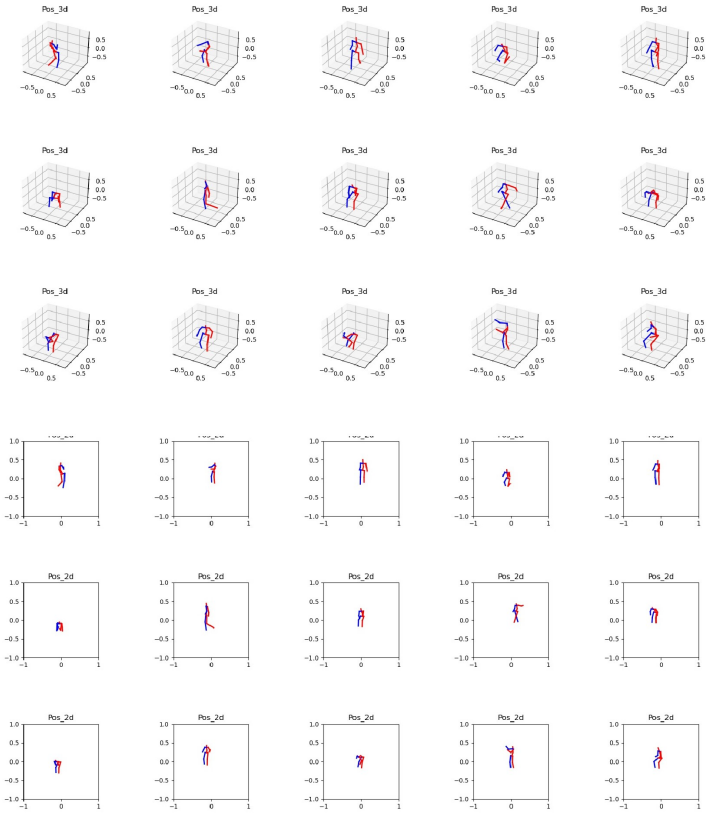

Fig. 10: **Several cases in our DH-3DP dataset.**

References

1. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2d human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. pp. 3686–3693 (2014)
2. Craig, J.J.: Introduction to robotics: mechanics and control, 3/E. Pearson Education India (2009)
3. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE transactions on pattern analysis and machine intelligence **36**(7), 1325–1339 (2014)
4. Johnson, S., Everingham, M.: Clustered pose and nonlinear appearance models for human pose estimation. In: bmvc. vol. 2, p. 5. Citeseer (2010)
5. Joo, H., Liu, H., Tan, L., Gui, L., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., Sheikh, Y.: Panoptic studio: A massively multiview system for social motion capture. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3334–3342 (2015)
6. Kocabas, M., Athanasiou, N., Black, M.J.: Vibe: Video inference for human body pose and shape estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5253–5263 (2020)
7. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
8. Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M.: 3d human pose estimation in video with temporal convolutions and semi-supervised training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7753–7762 (2019)