Predicting Animation Skeletons for 3D Articulated Models via Volumetric Nets -Supplementary Material-

Ablation Study. Here we present evaluation of alternative choices for our method. All the variants are trained in the same split and tuned in the same hold-out validation set in the same manner as our original method. Table 1 reports the same evaluation measures described in Section 6 of our paper for different number of hourglass modules in our architecture. We observed that the performance saturates when we reach 4 hourglass modules.

We also evaluated the geometric features used as input to our architecture. Table 2 reports the evaluation measures when using the Signed Distance Function only (SDF), the SDF plus each of the other geometric features, and altogether. We can see that each geometric feature individually improves the performance, and integrating all of them achieves the best result.

Table 3 reports the performance when (a) we remove the granularity control parameter from the architecture, (b) use Euclidean distances as cost for Prim's algorithm instead of the predicted log probabilities for bones. Both degraded variants drop the performance especially in terms of precision and recall.

Dataset Statistics. Our de-duplicated dataset contained 3193 rigged characters from Models Resource. The average joint number per character is 26.4. Figure 1 shows a histogram over the number of joints across the models of our dataset.

Architecture details. Table 4 lists each layer used in our architecture along with the size of its output map. We refer the reader to our source code for more details.

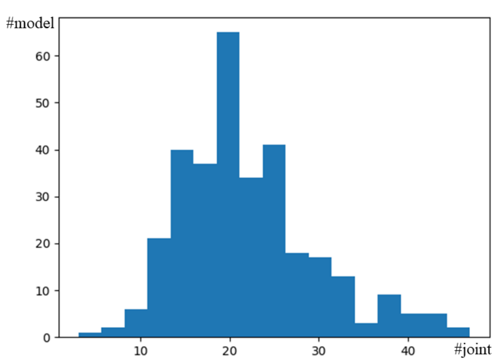

Figure 1. Histogram over the number of joints across the models of our dataset

Table 1. Evaluation of varying number of hourglass modules

(#) Modules	CD-joint	CD-joint2bone	MR-pred	MR-ref
1	5.2%	3.4%	55.7%	60.9%
2	4.9%	3.3%	60.0%	65.5%
3	4.7%	3.3%	61.4%	67.0%
4	**4.6%**	**3.2%**	**62.1%**	**68.3%**

Table 2. Evaluation of different input feature combinations

Input features	CD-joint	CD-joint2bone	MR-pred	MR-ref
SDF only	5.2%	3.5%	60.6%	56.0%
SDF+diam.	4.9%	3.3%	53.5%	61.8%
SDF+curv.	4.7%	**3.2%**	51.2%	66.4%
SDF+density	4.7%	**3.2%**	57.5%	63.2%
all features	**4.6%**	**3.2%**	**62.1%**	**68.3%**

Table 3. Evaluation of skipping the granularity control parameter (no control) and using Euclidean distances instead of log bone probabilities for Prim's algorithm (no bone prob)

variant	CD-joint	CD-joint2bone	MR-pred	MR-pref
no control	4.6%	3.2%	54.5%	67.9%
no bone prob.	4.6%	3.2%	57.8%	67.0%
full method	**4.6%**	**3.2%**	**62.1%**	**68.3%**

Table 4. **Architecture details.** ResBlock: The residual block is made of two volumetric convolutional layers with filters 3×3×3. Both produce the same number of feature maps. When the number of input/output feature maps differ, the skip path within any residual block contains an additional volumetric convolutional layer with 3×3×3 filters. Dropout: dropout layer with 0.2 probability.

	Layers	Output
	Input volume	88×88×88×5
	ReLU(BN(Conv(5x5x5, 5→8)))	88×88×88×8
	ResBlock	88×88×88×8
Encoder	ReLU(BN(Conv(2x2x2, stride=2)))	44×44×44×8 for 1st module, 44×44×44×10 for the rest
	ResBlock	44×44×44×16
	ReLU(BN(Conv(2x2x2, stride=2)))	22×22×22×16
	ResBlock	22×22×22×24
	ReLU(BN(Conv(2x2x2, stride=2)))	11×11×11×24
	ResBlock	11×11×11×36
	Concat with control param.	11×11×11×40
	ResBlock	11×11×11×40
Decoder	ResBlock	11×11×11×36
	ReLU(BN(ConvTrans(2x2x2, stride=2)))	22×22×22×24
	ResBlock	22×22×22×24
	ReLU(BN(ConvTrans(2x2x2, stride=2)))	44×44×44×16
	ResBlock	44×44×44×16
	ReLU(BN(ConvTrans(2x2x2, stride=2)))	88×88×88×8
Prediction	ResBlock	88×88×88×4
	Dropout(ReLU(BN(Conv(1x1x1, 4→4))))	88×88×88×4
	Conv(1x1x1, 4→1)	88×88×88×1