Accurate 3D Body Shape Regression using Metric and Semantic Attributes
Supplemental Material

Vasileios Choutas*[1], Lea Müller*[1], Chun-Hao P. Huang[1], Siyu Tang[2], Dimitrios Tzionas[1], Michael J. Black[1]
[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany [2]ETH Zürich
{vchoutas, lea.mueller, paul.huang, stang, dtzionas, black}@tuebingen.mpg.de
* Equal contribution, alphabetical order

1. Data Collection

1.1. Model-Agency Identity Filtering

We collect internet data consisting of images and height/chest/waist/hips measurements, from model agency websites. A "fashion model" can work for many agencies and their pictures can appear on multiple websites. To create non-overlapping training, validation and test sets, we match model identities across websites. To that end, we use ArcFace [2] for face detection and RetinaNet [3] to compute identity embeddings $E_i \in \mathbb{R}^{512}$ for each image. For every pair of models (q, t) with the same gender label, let Q, T be the number of query and target model images and $\boldsymbol{E_Q} \in \mathbb{R}^{Q\times 512}$ and $\boldsymbol{E_T} \in \mathbb{R}^{T\times 512}$ the query and target embedding feature matrices. We then compute the pairwise cosine similarity matrix $\mathcal{S} \in \mathbb{R}^{Q\times T}$ between all images in $\boldsymbol{E_Q}$ and $\boldsymbol{E_T}$, and the aggregate and average similarity:

$$\mathcal{S}_T(t) = \frac{1}{Q}\sum_q \mathcal{S}(q,t), \tag{1}$$

$$\mathcal{S}_{TQ} = \frac{1}{QT}\sum_q\sum_t \mathcal{S}(q,t). \tag{2}$$

Each pair with $\mathcal{S}$ and $\mathcal{S}_T$ that has no element larger than the similarity threshold $\tau = 0.3$ is ignored, as it contains dissimilar models. Finally, we check if $\mathcal{S}_{TQ}$ is larger than τ, and we keep a list of all pairs for which this holds true.

1.2. Crowd-Sourced Linguistic Shape-Attributes

To collect human ratings of how much a word describes a body shape, we conduct a human intelligence task (HIT) on Amazon Mechanical Turk (AMT). In this task, we show an image of a person along with 15 different gender-specific attributes. We then ask participants to indicate how strongly they agree or disagree that the provided words describe the shape of this person's body. We arrange the rating buttons from strong disagreement to strong agreement with equal distances to create a 5-point Likert scale. The rating choices are "strongly disagree" (score 1), "rather disagree" (score 2), "average" (score 3), "rather agree" (score 4), "strongly agree" (score 5).

We ask multiple persons to rate each body and image, to "average out" the subjectivity of individual ratings [15]. Additionally, we compute the Pearson correlation between averaged attribute ratings and ground-truth measurements. Examples of highly correlated pairs are "Big / Weight", and "Short / Height".

The layout of our CAESAR annotation task is visualized in Fig. R.1. To ensure good rating quality, we have several qualification requirements per participant: submitting a minimum of 5000 tasks on AMT and an AMT acceptance rate of 95%, as well as having a US residency and passing a language qualification test to ensure similar language skills and cultures across raters.

2. Mapping Shape Representations

2.1. Shape to Anatomical Measurements (S2M)

An important part of our project is the computation of body measurements. Following "Virtual Caliper" [11], we present a method to compute anatomical measurements from a 3D mesh in the canonical T-pose, i.e. after "undoing" the effect of pose. Specifically, we measure the height, $H(\boldsymbol{\beta})$, weight, $W(\boldsymbol{\beta})$, and the chest, waist and hip circumferences, $C_\text{c}(\boldsymbol{\beta})$, $C_\text{w}(\boldsymbol{\beta})$, and $C_\text{h}(\boldsymbol{\beta})$, respectively. Let $v_\text{head}(\boldsymbol{\beta}), v_\text{left heel}(\boldsymbol{\beta}), v_\text{chest}(\boldsymbol{\beta}), v_\text{waist}(\boldsymbol{\beta}), v_\text{hip}(\boldsymbol{\beta})$ be the head, left heel, chest, waist and hip vertices. $H(\boldsymbol{\beta})$ is computed as the difference in the vertical-axis "Y" coordinates between the top of the head and the left heel: $H(\boldsymbol{\beta}) = |v^y_\text{head}(\boldsymbol{\beta}) - v^y_\text{left heel}(\boldsymbol{\beta})|$. To obtain $W(\boldsymbol{\beta})$ we multiply the mesh volume by 985 kg/m^3, which is the average human body density. We compute circumference measurements using the method of Wuhrer et al. [17].

Here, $T \in \mathbb{R}^{F\times 3\times 3}$, where $F = 20,908$ is the number of triangles in the SMPL-X mesh, denotes "shaped" vertices of all triangles of the mesh $M(\boldsymbol{\beta}, \boldsymbol{\theta})$; we drop expressions, $\boldsymbol{\psi}$, which are not used in this work. Let us explain this using the chest circumference $C_\text{c}(\boldsymbol{\beta})$ as an example. We

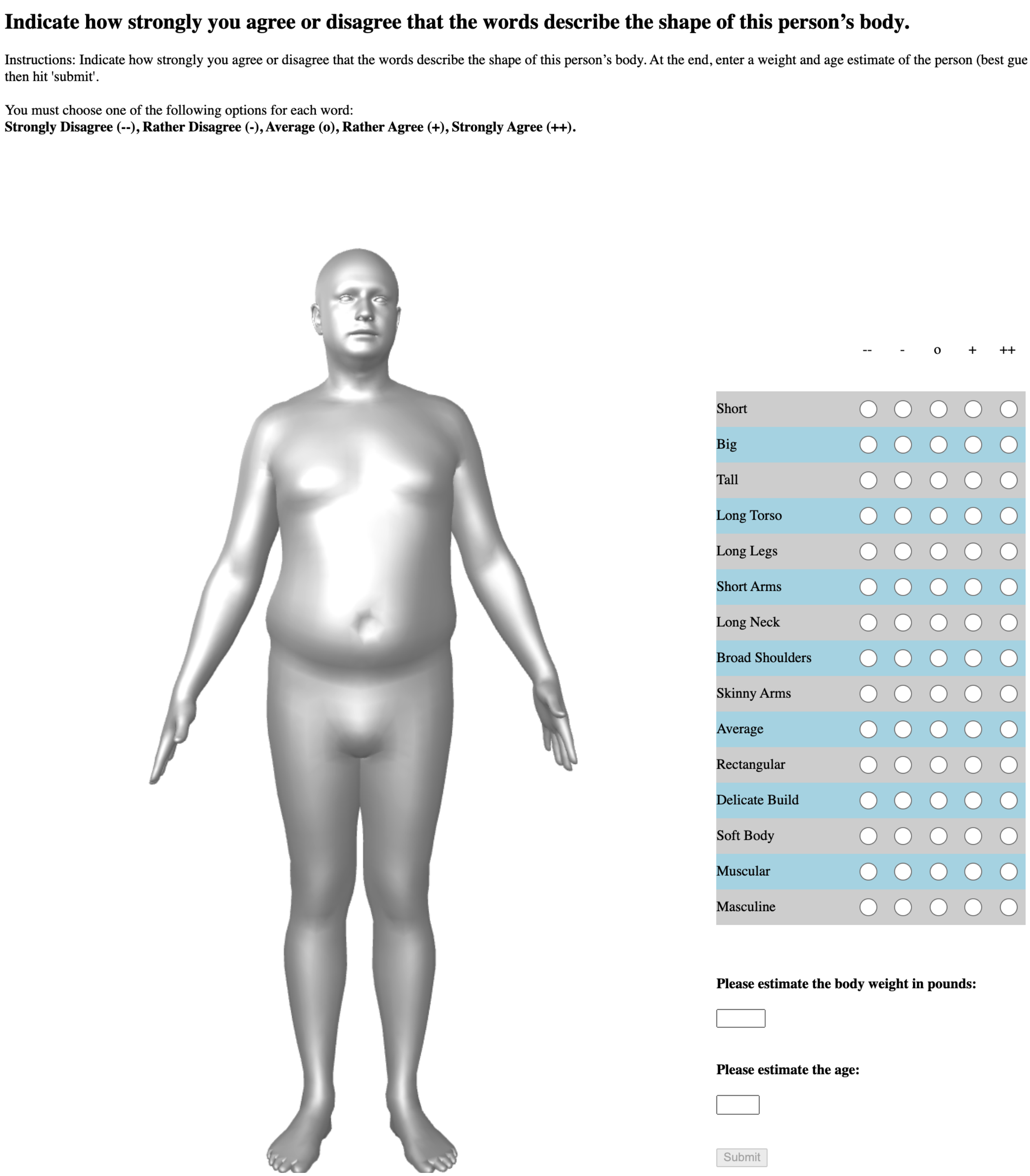

Figure R.1. Layout of the AMT task for a male subject. **Left:** the 3D body mesh in A-pose. **Right:** the attributes and ratings buttons.

form a plane P with normal $\boldsymbol{n} = (0, 1, 0)$ that crosses the point $v_{\text{chest}}(\boldsymbol{\beta})$. Then, let $\mathcal{I} = \{\boldsymbol{p}_i\}_{i=1}^{N}$ be the set of points of P that intersect the body mesh (red points in Fig. R.2). We store their barycentric coordinates $(\mathtt{u}_i, \mathtt{v}_i, \mathtt{w}_i)$ and the

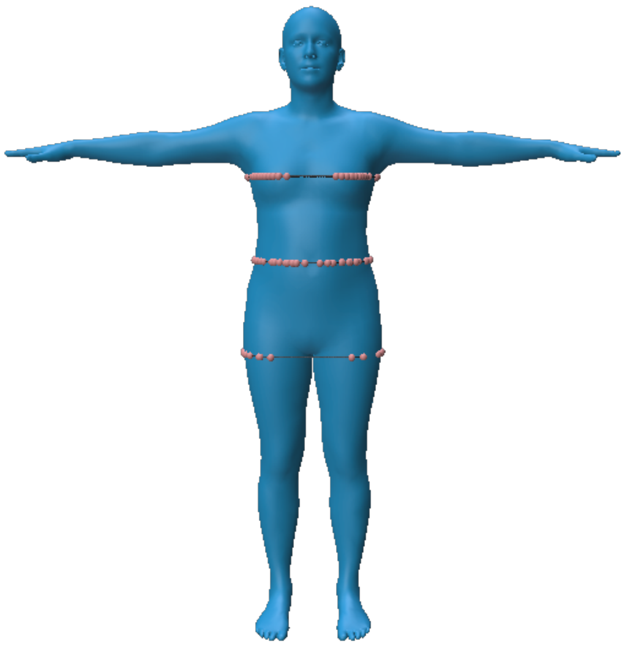

Figure R.2. Automatic anatomical measurements on a 3D mesh. The red points lie on the intersection of planes at chest/waist/hip height with the mesh, while their convex hull is shown with black lines.

corresponding body-triangle index t_i. Let $\mathcal{H}$ be the convex hull of $\mathcal{I}$ (black lines in Fig. R.2), and $\mathcal{E}$ the set of edge indices of $\mathcal{H}$. $C_{\text{c}}(\boldsymbol{\beta})$ is equal to the length of the convex hull:

$$C_{\text{c}}(\boldsymbol{\beta}) = \sum_{(i,j)\in\mathcal{E}} \left\| \begin{pmatrix} \mathtt{u}_i \\ \mathtt{v}_i \\ \mathtt{w}_i \end{pmatrix}^{\top} T_{t_i} - \begin{pmatrix} \mathtt{u}_j \\ \mathtt{v}_j \\ \mathtt{w}_j \end{pmatrix}^{\top} T_{t_j} \right\|_2 , \qquad (3)$$

where i, j are point indices for line segments of $\mathcal{E}$. The process is the same for the waist and hips, but the intersection plane is computed using $v_{\text{waist}}, v_{\text{hip}}$. All of $H(\boldsymbol{\beta}), W(\boldsymbol{\beta}), C_{\text{c}}(\boldsymbol{\beta}), C_{\text{w}}(\boldsymbol{\beta}), C_{\text{h}}(\boldsymbol{\beta})$ are differentiable functions of body shape parameters, $\boldsymbol{\beta}$.

Note that SMPL-X knows the height distribution of humans and acts as a strong prior in shape estimation. Given the ground-truth height of a person (in meter), $H(\boldsymbol{\beta})$ can be used to directly supervise height and overcome scale ambiguity.

2.2. Mapping Attributes to Shape (A2S)

We introduce A2S, a model that maps the input attribute ratings to shape components $\boldsymbol{\beta}$ as output. We compare a 2$^{\text{nd}}$ degree polynomial model with a linear regression model and a multi-layer perceptron (MLP), using the Vertex-to-Vertex (V2V) error metric between predicted and ground-truth SMPL-X meshes, and report results in Tab. R.1.

Model	Input	V2V mean $\pm$ std	
		Females	Males
Mean Shape		18.01 ± 8.73	19.24 ± 10.36
Linear Regression	A	10.83 ± 4.77	10.43 ± 4.63
Polynomial (d=2)	A	10.58 ± 4.67	10.25 ± 4.48
MLP	A	10.73 ± 4.62	10.33 ± 4.57
Linear Regression	A+H+W	7.00 ± 2.59	6.56 ± 2.21
Polynomial (d=2)	A+H+W	7.31 ± 2.56	6.71 ± 2.21
MLP	A+H+W	7.03 ± 2.6	6.68 ± 2.24
Linear Regression	A+H+ $\sqrt[3]{W}$	6.97 ± 2.58	6.54 ± 2.22
Polynomial (d=2)	A+H+ $\sqrt[3]{W}$	$\mathbf{6.88 \pm 2.55}$	$\mathbf{6.49 \pm 2.20}$

Table R.1. Comparison of models for A2S and AHW2S regression.

When using only attributes as input (A2S), the polynomial model of degree $d = 2$ achieves the best performance. Adding height and weight to the input vector requires a small modification, namely using the cubic root of the weight and converting the height from (m) to (cm). We. With these additions, the 2$^{\text{nd}}$ degree polynomial achieves the best performance.

2.3. Images to Attributes (I2A)

We briefly experimented with models that learn to predict attribute scores from images (I2A). This attribute predictor is implemented using a ResNet50 for feature extraction from the input images, followed by one MLP per gender for attribute score prediction. To quantify the model's performance, we use the attribute classification metric described in the main paper. I2A achieves 60.7 / 69.3% (fe-/male) of correctly predicted attributes, while our S2A achieves 68.8 / 76% on CAESAR. Our explanation for this result is that it is hard for the I2A model to learn to correctly predict attributes independent of subject pose. Our approach works better, because it decomposes 3D human estimation into predicting pose and shape. Networks are good at estimating pose even without GT shape [8]. "SHAPY 's losses" affect only the shape branch. To minimize these losses, the network has to learn to correctly predict shape irrespective of pose variations.

3. SHAPY- 3D Shape Regression from Images

Implementation details:: To train SHAPY, each batch of training images contains 50% images collected from model agency websites and 50% images from ExPose's [1] training set. Note that the overall number of images of males and females in our collected model data differs significantly; images of female models are many more. Therefore, we randomly sample a subset of female images so that, eventually, we get an equal number of male and female images. We also use the BMI of each subject, when available, as a sampling weight for images. In this way, subjects with

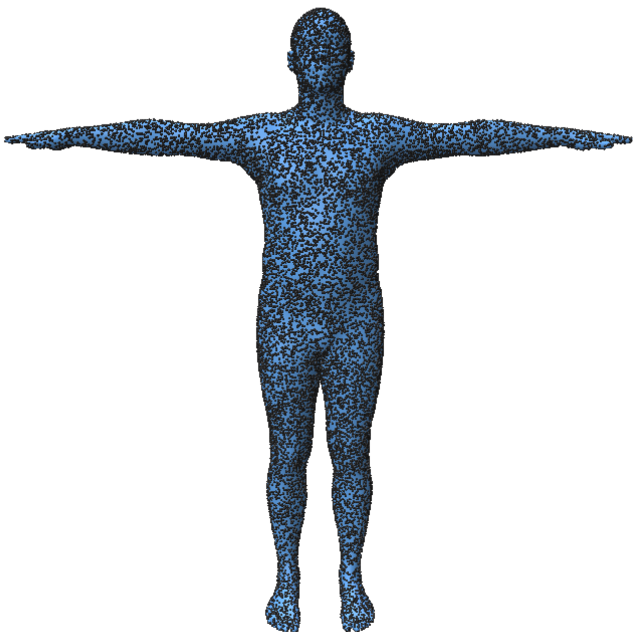

Figure R.3. The 20K body mesh surface points (in black) used to evaluated body shape estimation accuracy.

	Method -	P2P_{20K} (mm)	Height (mm)	Weight (kg)	Chest (mm)	Waist (mm)	Hips (mm)
female	A2S	10.9 ± 5.2	27 ± 21	5 ± 5	30 ± 26	32 ± 31	28 ± 22
	H2S	12.8 ± 7.0	5 ± 5	12 ± 11	93 ± 72	101 ± 88	60 ± 52
	AH2S	7.2 ± 2.8	4 ± 3	3 ± 4	27 ± 23	29 ± 28	23 ± 19
	HW2S	7.9 ± 3.2	5 ± 5	1 ± 1	25 ± 22	22 ± 18	26 ± 25
	AHW2S	6.4 ± 2.5	4 ± 3	1 ± 1	14 ± 12	14 ± 12	17 ± 14
	C2S	19.5 ± 10.8	58 ± 46	8 ± 6	54 ± 36	57 ± 42	47 ± 36
	AC2S	9.6 ± 4.3	24 ± 18	3 ± 2	18 ± 15	19 ± 16	19 ± 14
	HC2S	7.3 ± 2.8	5 ± 5	2 ± 2	19 ± 16	16 ± 14	15 ± 13
	AHC2S	6.3 ± 2.4	4 ± 3	1 ± 1	15 ± 12	14 ± 12	14 ± 12
	HWC2S	7.2 ± 2.9	5 ± 5	1 ± 1	14 ± 12	13 ± 11	14 ± 12
	AHWC2S	6.2 ± 2.4	4 ± 3	1 ± 1	11 ± 9	12 ± 10	13 ± 11
male	A2S	11.1 ± 5.2	29 ± 21	5 ± 4	30 ± 22	32 ± 24	28 ± 21
	H2S	12.1 ± 6.1	5 ± 4	11 ± 11	81 ± 66	102 ± 87	40 ± 33
	AH2S	6.8 ± 2.3	4 ± 3	3 ± 3	27 ± 21	29 ± 23	24 ± 18
	HW2S	8.1 ± 2.7	5 ± 4	1 ± 1	24 ± 17	26 ± 20	21 ± 18
	AHW2S	6.3 ± 2.1	4 ± 3	1 ± 1	19 ± 15	19 ± 14	20 ± 16
	C2S	19.7 ± 11.1	59 ± 47	9 ± 8	55 ± 41	63 ± 49	37 ± 28
	AC2S	9.6 ± 4.4	25 ± 19	3 ± 3	23 ± 19	21 ± 17	18 ± 14
	HC2S	7.7 ± 2.6	5 ± 4	2 ± 2	28 ± 23	18 ± 15	13 ± 11
	AHC2S	6.0 ± 2.0	4 ± 3	2 ± 2	21 ± 17	17 ± 14	13 ± 10
	HWC2S	7.3 ± 2.6	5 ± 4	1 ± 1	20 ± 15	14 ± 12	13 ± 11
	AHWC2S	5.8 ± 2.0	4 ± 3	1 ± 1	16 ± 13	13 ± 10	13 ± 10

Table R.2. Results of A2S and its variations on CMTS test set, in mm or kg. Trained with gender-specific SMPL-X model.

higher BMI are selected more often, due to their smaller number, to avoid biasing the model towards the average BMI of the dataset. Our pipeline is implemented in PyTorch [10] and we use the Adam [6] optimizer with a learning rate of $1e-4$. We tune the weights of each loss term with grid search on the MMTS and HBW validation sets. Using a batch size of 48, SHAPY achieves the best performance on the HBW validation set after 80k steps.

4. Experiments

4.1. Metrics

P2P_{20K}: SMPL-X has more than half of its vertices on the head. Consequently, computing an error based on vertices overemphasizes the importance of the head. To remove this bias, we also report the mean distance between $P = 20k$ mesh surface points; see Fig. R.3 for a visualization on the ground-truth and estimated meshes. For this, we uniformly sample the SMPL-X template mesh and compute a sparse matrix $\mathbf{H}_{\text{SMPL-X}} \in \mathbb{R}^{P \times N}$ that regresses the mesh surface points from SMPL-X vertices V, as $\mathbf{P} = \mathbf{H}_{\text{SMPL-X}} V$.

To use this metric in a mesh with different topology, e.g. SMPL, we simply need to compute the corresponding $\mathbf{H}_{\text{SMPL}}$. For this, we align the SMPL model to the SMPL-X template mesh. For each point sampled from the SMPL-X mesh surface, we find the closest point on the aligned SMPL mesh surface. To obtain the SMPL mesh surface points from SMPL vertices, we again compute a sparse matrix, $\mathbf{H}_{\text{SMPL}} \in \mathbb{R}^{P \times 6{,}890}$. The distance between the SMPL-X and SMPL mesh surface points on the template meshes is 0.073 mm, which is negligible.

Given two meshes M_1 and M_2 of topology T_1 and T_2 we obtain the mesh surface points $P_1 = \mathbf{H}_{T_1} U_1$ and $P_2 = \mathbf{H}_{T_2} U_2$, where U_1 and U_2 denote the vertices of the shaped zero posed (t-pose) meshes. To compute the P2P_{20K} error we correct for translation $t = \bar{P}_2 - \bar{P}_1$ and define

$$\text{P2P}_{\text{20K}}(U_1, U_2) = ||\mathbf{H}_{T_1} U_1 + t - \mathbf{H}_{T_2} U_2||_2^2.$$

4.2. Shape Estimation

A2S and its variations: For completeness, Table R.2 shows the results of the female A2S models in addition to the male ones. The male results are also presented in the main manuscript. Note that attributes improve shape reconstruction across the board. For example, in terms of P2P_{20K}, AH2S is better than just H2S, AHW2S is better than just HW2S. It should be emphasized that even when many measurements are used as input features, i.e. height, weight, and chest/waist/hip circumference, adding attributes still improves the shape estimate, e.g. HWC2S vs. AHWC2S.

Attribute/Measurement ablation: To investigate the extent to which attributes can replace ground truth measurements in network training, we train SHAPY's variations in a leave-one-out manner: SHAPY-H uses only height and SHAPY-C only hip/waist/chest circumference. We compare these models with SHAPY-AH and SHAPY-AC, which use attributes in addition to height and circumference measurements, respectively. For completeness, we also evaluate SHAPY-HC and SHAPY-AHC, which use all measurements; the latter also uses attributes. The results are reported in Tab. R.3 (MMTS) and Tab. R.4 (HBW).

Mean absolute error (mm) ↓				
Method	Height	Chest	Waist	Hips
SHAPY-H	**52**	113	172	108
SHAPY-HA	60	**64**	**96**	**77**
SHAPY-C	119	66	**70**	70
SHAPY-CA	**74**	**60**	82	**69**
SHAPY-HC	**54**	62	**72**	**69**
SHAPY-HCA	57	**61**	85	73

Table R.3. Leave-one-out evaluation on MMTS.

Mean absolute error (mm) ↓					
Method	Height	Chest	Waist	Hips	$P2P_{20K}$
SHAPY-H	54	90	77	**54**	22
SHAPY-HA	**49**	**62**	**71**	58	**20**
SHAPY-C	72	65	**77**	60	26
SHAPY-CA	**54**	**69**	78	**58**	**22**
SHAPY-HC	53	**61**	77	55	23
SHAPY-HCA	**47**	66	**75**	**52**	**20**

Table R.4. Leave-one-out evaluation on the HBW test set.

The tables show that attributes are an adequate replacement for measurements. For example, in Tab. R.3, the height (SHAPY-C vs. SHAPY-CA) and circumference errors (SHAPY-H vs. SHAPY-AH) are reduced significantly when attributes are taken into account. On HBW, the $P2P_{20K}$ errors are equal or lower, when attribute information is used, see Tab. R.4. Surprisingly, seeing attributes improves the height error in all three variations. This suggests that training on model images introduces a bias that A2S antagonizes.

S2A: Table R.5 shows the results of S2A in detail. All attributes are classified correctly with an accuracy of at least 58.05% (females) and 68.14% (males). The probability of randomly guessing the correct class is 20%.

AHWC and AHWC2S noise: To evaluate AHWC's robustness to noise in the input, we fit AHWC using the per-rater scores instead of the average score. The $P2P_{20K} \downarrow$ error only increases by 1.0 mm to 6.8 when using the per-rater scores.

4.3. Pose evaluation

3D Poses in the Wild (3DPW) [16]: This dataset is mainly useful for evaluating body *pose* accuracy since it contains few subjects and limited body shape variation. The test set contains a limited set of 5 subjects in indoor/outdoor videos with everyday clothing. All subjects were scanned to obtain their ground-truth body shape. The body poses are pseudo ground-truth SMPL fits, recovered from images and IMUs. We convert pose and shape to SMPL-X for evaluation.

We evaluate SHAPY on 3DPW to report pose estimation accuracy (Tab. R.6). SHAPY's pose accuracy is slightly behind ExPose which also uses SMPL-X. SHAPY's performance is better than HMR [5] and STRAPS [12]. However, SHAPY does not outperform recent pose estimation methods, e.g. HybrIK [8]. We assume that SHAPY's pose estimation accuracy on 3DPW can be improved by (1) adding data from the 3DPW training set (similar to Sengupta et al. [13] who sample poses from 3DPW training set) and (2) creating pseudo ground-truth fits for the model data.

Attribute	Male		Female	
	MAE ± SD	CCP	MAE ± SD	CCP
Big	0.25 ± 0.18	71.68%	0.31 ± 0.23	70.00%
Broad Shoulders	0.26 ± 0.20	73.75%	0.33 ± 0.24	63.90%
Long Legs	0.23 ± 0.17	81.12%	0.43 ± 0.33	58.05%
Long Neck	0.27 ± 0.21	73.75%	0.29 ± 0.21	69.51%
Long Torso	0.27 ± 0.20	70.80%	0.36 ± 0.27	62.68%
Muscular	0.31 ± 0.24	69.03%	0.26 ± 0.21	73.17%
Short	0.28 ± 0.22	72.27%	0.27 ± 0.21	67.56%
Short Arms	0.20 ± 0.15	84.07%	0.27 ± 0.22	72.20%
Tall	0.27 ± 0.22	70.80%	0.30 ± 0.23	70.98%
Average	0.27 ± 0.19	78.76%	n / a	n / a
Delicate Build	0.21 ± 0.16	78.17%	n / a	n / a
Masculine	0.23 ± 0.18	78.17%	n / a	n / a
Rectangular	0.27 ± 0.20	80.24%	n / a	n / a
Skinny Arms	0.25 ± 0.19	76.40%	n / a	n / a
Soft Body	0.32 ± 0.23	68.14%	n / a	n / a
Large Breasts	n / a	n / a	0.31 ± 0.23	72.93%
Pear Shaped	n / a	n / a	0.32 ± 0.22	64.39%
Petite	n / a	n / a	0.40 ± 0.30	61.95%
Skinny Legs	n / a	n / a	0.25 ± 0.18	81.22%
Slim Waist	n / a	n / a	0.30 ± 0.23	71.71%
Feminine	n / a	n / a	0.26 ± 0.20	73.41%

Table R.5. S2A evaluation. We report mean, standard deviation and percentage of correctly predicted classes per attribute on CMTS test set.

	Model	MPJPE	PA-MPJPE
HMR [5]	SMPL	130	81.3
SPIN [7]	SMPL	96.9	59.2
TUCH [9]	SMPL	84.9	55.5
EFT [4]	SMPL	-	54.2
HybrIK [8]	SMPL	**80.0**	**48.8**
STRAPS [12]*	SMPL	-	66.8
Sengupta et al. [14]*	SMPL	-	61.0
Sengupta et al. [13]*	SMPL	84.9	53.6
ExPose [1]	SMPL-X	93.4	60.7
SHAPY (ours)	SMPL-X	95.2	62.6

Table R.6. Evaluation on 3DPW [16]. * uses body poses sampled from the 3DPW training set for training.

4.4. Qualitative Results

We show additional qualitative results in Fig. R.5 and Fig. R.7. Failure cases are shown in Fig. R.8. To deal with high-BMI bodies, we need to expand the set of training images and add additional shape attributes that are descriptive for high-BMI shapes. Muscle definition on highly muscular bodies is not well represented by SMPL-X, nor do our attributes capture this. The SHAPY approach, however, could be used to capture this with a suitable body model and more appropriate attributes.

Figure R.4. Qualitative results of SHAPY predictions for female bodies.

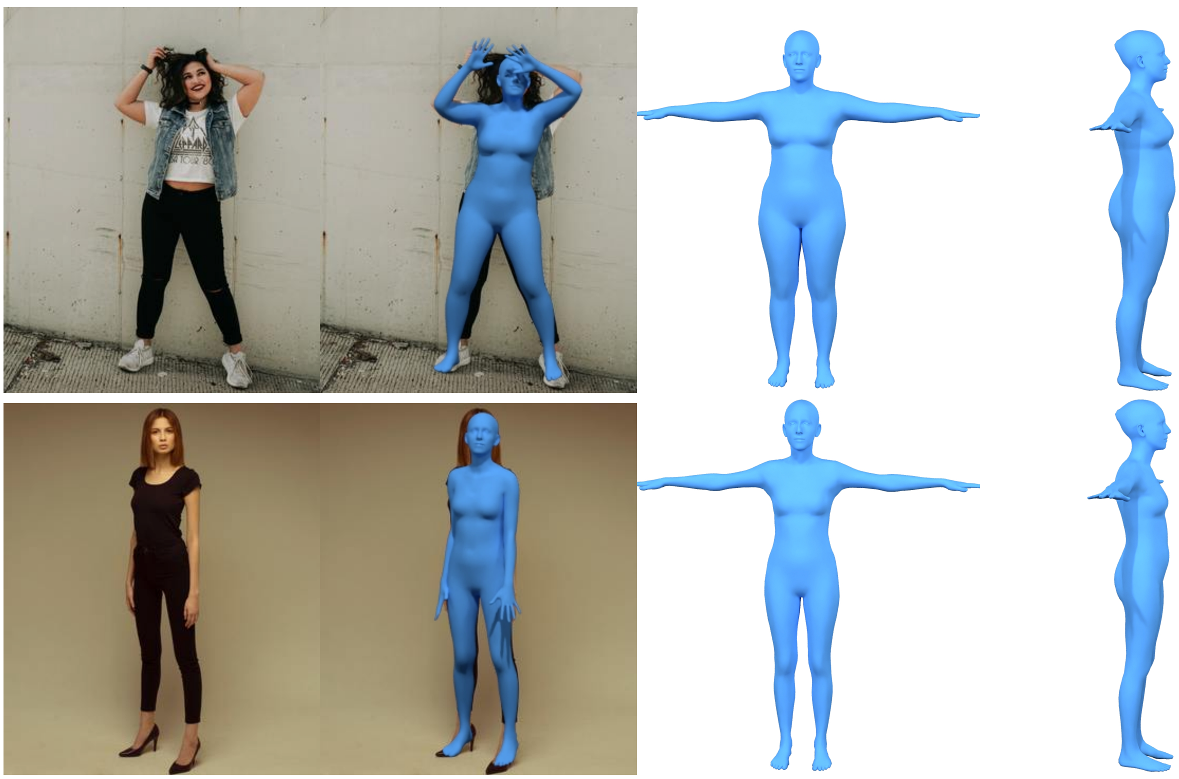

Figure R.5. Qualitative results of SHAPY predictions for female bodies. (Cont.)

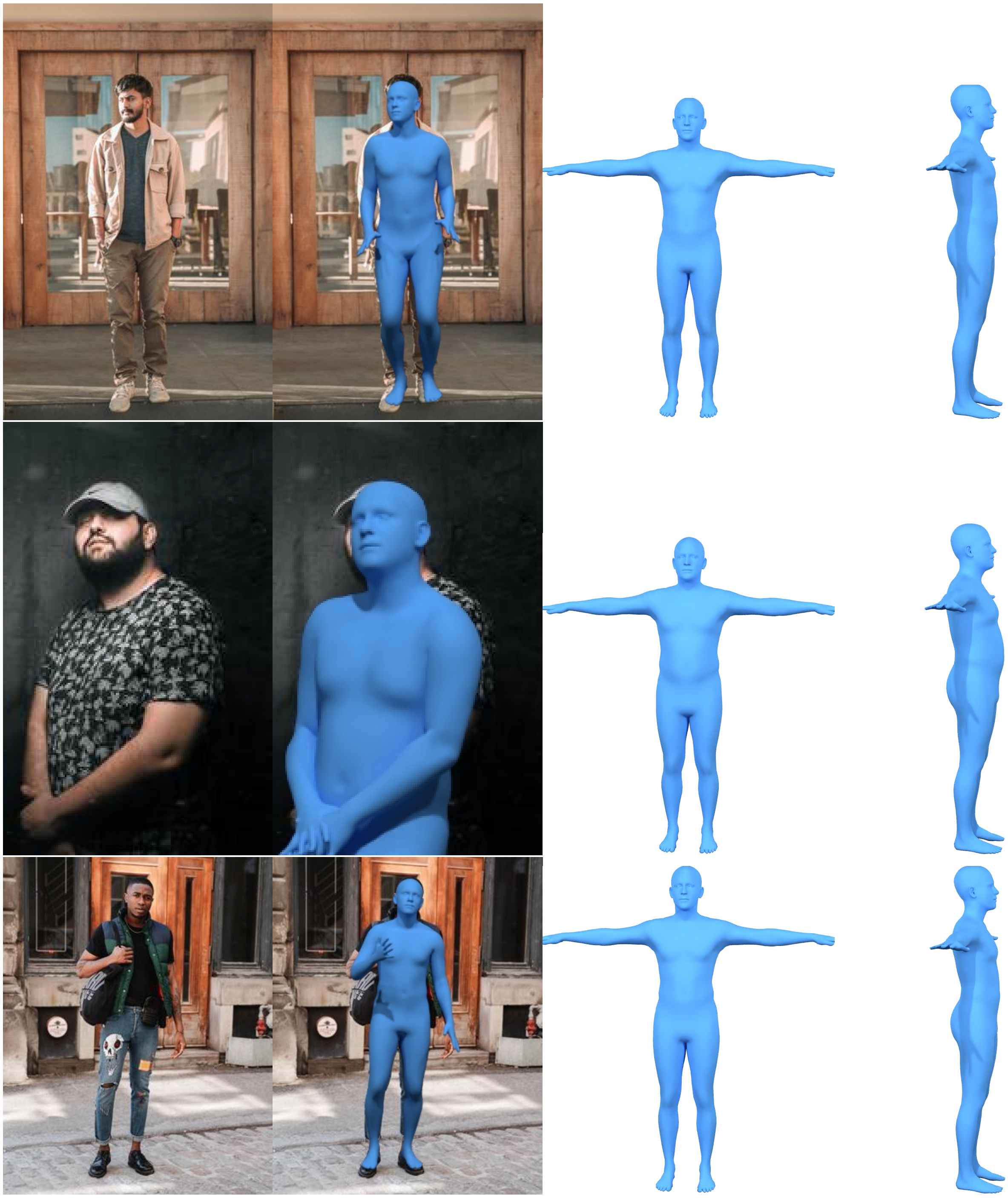

Figure R.6. Qualitative results of SHAPY predictions for male bodies.

Figure R.7. Qualitative results of SHAPY predictions for male bodies (Cont.) .

Figure R.8. Failure cases. In the first example (upper left) the weight is underestimated. Other failure cases of SHAPY are muscular bodies (upper right) and body shapes with high BMI (second row).

References

[1] Vasileios Choutas, Georgios Pavlakos, Timo Bolkart, Dimitrios Tzionas, and Michael J. Black. Monocular expressive body regression through body-driven attention. In *European Conference on Computer Vision (ECCV)*, volume 12355, pages 20–40, 2020. 3, 5

[2] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019. 1

[3] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. RetinaFace: Single-shot multi-level face localisation in the wild. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5202–5211, 2020. 1

[4] Hanbyul Joo, Natalia Neverova, and Andrea Vedaldi. Exemplar fine-tuning for 3D human pose fitting towards in-the-wild 3D human pose estimation. In *International Conference on 3D Vision (3DV)*, pages 42–52, 2020. 5

[5] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, pages 7122–7131, 2018. 5

[6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 4

[7] Nikos Kolotouros, Georgios Pavlakos, Michael J. Black, and Kostas Daniilidis. Learning to reconstruct 3D human pose and shape via model-fitting in the loop. In *International Conference on Computer Vision (ICCV)*, pages 2252–2261, 2019. 5

[8] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. HybrIK: A hybrid analytical-neural inverse kinematics solution for 3D human pose and shape estimation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3383–3393, 2021. 3, 5

[9] Lea Müller, Ahmed A. A. Osman, Siyu Tang, Chun-Hao P. Huang, and Michael J. Black. On self-contact and human pose. In *Computer Vision and Pattern Recognition (CVPR)*, pages 9990–9999, 2021. 5

[10] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Conference on Neural Information Processing Systems (NeurIPS)*, pages 8024–8035, 2019. 4

[11] Sergi Pujades, Betty Mohler, Anne Thaler, Joachim Tesch, Naureen Mahmood, Nikolas Hesse, Heinrich H Bülthoff, and Michael J. Black. The virtual caliper: Rapid creation of metrically accurate avatars from 3D measurements. *Transactions on Visualization and Computer Graphics (TVCG)*, 25(5):1887–1897, 2019. 1

[12] Akash Sengupta, Ignas Budvytis, and Roberto Cipolla. Synthetic training for accurate 3D human pose and shape estimation in the wild. In *British Machine Vision Conference (BMVC)*, 2020. 5

[13] Akash Sengupta, Ignas Budvytis, and Roberto Cipolla. Hierarchical kinematic probability distributions for 3D human shape and pose estimation from images in the wild. In *International Conference on Computer Vision (ICCV)*, pages 11219–11229, 2021. 5

[14] Akash Sengupta, Ignas Budvytis, and Roberto Cipolla. Probabilistic 3D human shape and pose estimation from multiple unconstrained images in the wild. In *Computer Vision and Pattern Recognition (CVPR)*, pages 16094–16104, 2021. 5

[15] Stephan Streuber, M. Alejandra Quiros-Ramirez, Matthew Q. Hill, Carina A. Hahn, Silvia Zuffi, Alice O'Toole, and Michael J. Black. Body Talk: Crowdshaping realistic 3D avatars with words. *Transactions on Graphics (TOG)*, 35(4):54:1–54:14, 2016. 1

[16] Timo von Marcard, Roberto Henschel, Michael Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3D human pose in the wild using IMUs and a moving camera. In *European Conference on Computer Vision (ECCV)*, volume 11214, pages 614–631, 2018. 5

[17] Stefanie Wuhrer and Chang Shu. Estimating 3D human shapes from measurements. *Machine Vision and Applications (MVA)*, 24(6):1133–1147, 2013. 1