Silhouette Body Measurement Benchmarks

Song Yan*, Johan Wirta†, Joni-Kristian Kämäräinen*
*Computing Sciences, Tampere University, Finland
†NOMO Technologies Ltd, Espoo Finland

***Abstract*—Anthropometric body measurements are important for industrial design, garment fitting, medical diagnosis and ergonomics. A number of methods have been proposed to estimate the body measurements from images, but progress has been slow due to the lack of realistic and publicly available datasets. The existing works train and test on silhouettes of 3D body meshes obtained by fitting a human body model to the commercial CAESAR scans. In this work, we introduce the BODY-fit dataset that contains fitted meshes of 2,675 female and 1,474 male 3D body scans. We unify evaluation on the CAESAR-fit and BODY-fit datasets by computing body measurements from geodesic surface paths as the ground truth and by generating two-view silhouette images. We also introduce BODY-rgb - a realistic dataset of 86 male and 108 female subjects captured with an RGB camera and manually tape measured ground truth. We propose a simple yet effective deep CNN architecture as a baseline method which obtains competitive accuracy on the three datasets.**

I. Introduction

Recovery of 3D body information from 2D images is an important yet challenging problem with applications in industrial design [1], garment fitting and online shopping [2], medical diagnosis [3] and ergonomics [4]. Recovery can be decomposed into 3D pose and 3D shape. Recently research on 3D pose estimation has been active [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], but shape recovery has received less attention.

Many applications do not need recovery of the full 3D shape, but only a set of suitable body variables, *anthropometric body measurements*, such as the head, waist and chest circumferences (Figure 1). Estimation of the body measurements is addressed in several works [15], [16], [17], [18], [19], [20], [21], [22] which process one or multiple silhouette images of persons in a fixed pose. The earliest works were based on engineered features and regression [15], [16], [17], [18], [19] and the recent works adopt various deep architectures [20], [21], [22]. For training and testing these methods use silhouettes rendered using 3D body meshes. The body meshes are either completely synthetic or generated by learning a parametric body model from real 3D body scans. A popular dataset is CAESAR [23] that contains 4,400 scans and tape measured ground truth. However, CEASAR is commercial and its license prevents its public use. The license allows derivative works such as the fitted meshes and measurement ground truth can be generated by defining geodesic paths on the mesh surfaces.

In this work, we provide a number of datasets that will be made publicly available to facilitate fair comparisons. Our main contributions are:

- A novel train/test dataset - *BODY-fit* - for benchmarking silhouette based body measurement methods. The dataset is obtained from the local clothing company who have 2,675 female and 1,474 male 3D scans of their customers.
- A novel testing dataset - *BODY-rgb* - of recently captured RGB images of 86 males and 108 females and tape measured ground truth.
- A strong *baseline* which achieves good accuracy on the new datasets and the existing data - *CAESAR-fit* - provided by Pishchulin *et al.* [24] and for which we define train/test splits files and generated silhouette images.

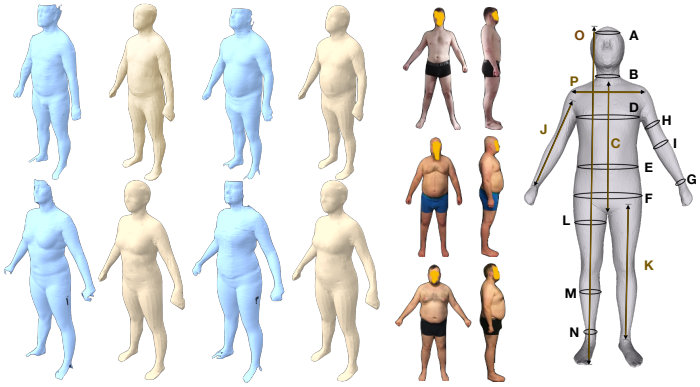

Fig. 1. Examples from the proposed benchmarks, "BODY-fit" and "BODY-rgb", and the 16 body measurements (A-P) used in the method comparison. The blue meshes represent the original BODY scans containing missing points and noise (mainly in the head, feet and hand regions). The yellow meshes result from non-rigid ICP fitting of the mean shape template from the CAESAR fits datasets [24] so that the both datasets now share the same topology. RGB images were captured using Apple iPad.

II. Related work

a) Body measurements.: Anthropometric body measurements provide detailed information about the body shape. The measurements are conventionally measured manually using a tape measure, but there is wide interest toward computerized tools. The first commercial products for 3D scan based anthropometric measurements were unsatisfactory. For example, Paquette et al. [25] reported that automatic point-cloud based measurements differ largely w.r.t manual tape measurements. They reported systematic errors of up-to $30-40$ mm despite the fact that standard measurement procedures were implemented in softwares used for their comparison (ISO-8559 and U.S. Army). Gordon *et al.* [26] defined accuracy thresholds for a number of anthropometric measurements and these vary from ± 4.0 mm (ankle, elbow and knee circumference) to ± 15.0 mm (chest).

b) Datasets.: Since the introduction of 3D scanners there have been several campaigns. For example, the UMTRI dataset was collected to find the safest sitting posture of young children in cars [27]. ANSUR 88 (1988) and ANSUR 2012 datasets contain 3D scans and 93 tape measured body measurements of US Army Force soldiers. ANSUR 2012 contains 4,082 male and 1,986 female subjects of varying age. UMTRI and ANSUR datasets are not publicly available. CAESAR dataset [23] is a commercial dataset that contains 3D scans of 2,400 U.S. and Canadian and 2,000 European civilians with tape measured ground truth. CAESAR license and expensive price prevent its wide adoption for research purposes. Several authors have performed mesh registration on the CAESAR scans to bring them in correspondence and use pre-defined geodesic distances as the ground truth [28], [20], [24], [22]. Pishchulin *et al.* [24] and Yang *et al.* [28] have made their data available. Pischulin *et al.* data covers nearly 98% of the original CAESAR scans and is thus the best for method comparisons.

c) Methods.: Human 3D pose recovery "in the wild" has recently gained momentum [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. A number of these methods also estimate body volume [9], [10], [11], [12], [13], [14], but only a few provide quantitative results [11], [14]. In this work we assume that the pose is approximately fixed which makes the problem substantially easier, but which is a fair assumption for many applications.

Earlier works use engineered features and regression [15], [16], [17], [18], [19], [29], [30], but recently deep architectures have become more popular. Dibra et al. [20], [21], [22] have proposed multiple architectures. In [21] the hand-crafted features are extracted from silhouettes and mapped to the shape PCA (Principal Component Analysis) sub-space via the Random Forest regressor, then the body measurements are obtained from the reconstructed meshes. HS-NET [20] learns a global mapping from silhouettes to shape parameters by training CNNs. In the most recent work [22] Dibra et al. firstly construct a rich body shape representation space from the pose invariant Heat Kernel Signature (HKS) descriptors, then learn a mapping from silhouettes to this embedded space.

III. Benchmark Datasets

The data used in the existing works can be divided to *generated body shapes* [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] and *fitted body shapes* [19], [20], [21], [22]. The generated shapes are not from real subjects, but are generated synthetically by varying shape parameters of a 3D body model such as SCAPE [31], BlendSCAPE [32] or SMPL [33]. Here we focus on fitted body shapes captured from real subjects.

A. CAESAR fits

The license of the CAESAR dataset prevents public use of the original body scans and the tape measured ground truth. However, a 3D body mesh can be fitted to the scans and geodesic surface paths can be defined as alternative ground truth measurements. For example, Dibra et al. [20], [21], [22] learn a statistical model from CAESAR dataset to synthesize training data. Yang *et al.* [28] provide 1,517 male and 1,531 female CAESAR fits and Pishchulin *et al.* [24] provide another set of 2,211 and 2,095 CAESAR fits. For our benchmark, we selected the fitted meshes of Pischulin *et al.* [24] since it covers 98% of the original samples

We rendered the silhouette images by a weak-perspective camera model with the focal length $f = 4.15\,\mathrm{mm}$ and physical pixel size $1.5\,\mu\mathrm{m}$. These settings correspond approximately the settings of *iPhone 5S rear camera*. The virtual camera was positioned to the height of 1.6 m from the ground and distance of 2.4 m from the body. 2240×2240 pix images were generated. In our rendering functionality these parameters can be easily adapted for specific purposes.

The anthropometric body measurements were defined as 16 circumference paths that match the definitions in [20], [21], [22], [19] and which are illustrated in Figure 1.

B. BODY fits

A local clothing company provided us a dataset of real 3D body scans of people wearing only tight underwear. The scans were captured using a commercial TC2 device and software[1]. Subjects were instructed to step on the rotating platform and take a standing pose with the feet at around their shoulder width apart and the arms slightly raised to create a gap between the arms and torso, i.e. "A"-pose. The platform then rotates around once, during which three depth sensors produce a raw 3D scan of the customer and the process takes a few seconds. The scanner outputs a triangulated mesh structure in the regular OBJ file format. Each mesh contains on average 57,000 vertices and around 113,000 faces.

Similar to the original CAESAR scans also our scans are of various qualities and contain holes, in particular, near the feet, hand and head regions. To compensate the missing regions, the scans were converted to watertight meshes by applying the non-rigid ICP algorithm of Amberg et al. [34] and a 3D body template. The fitting process is explained in more details in [30]. As the body template we selected the the mean shape of the CAESAR fits that brings the additional benefit that our BODY fits and CAESAR fits from Pishchulin *et al.* [24] share the same topology. Then silhouette images were generated similar to the CAESAR fits using the same weak-perspective camera.

C. RGB Body

In addition to the 3D scan datasets we collected a small dataset of real RGB images of people in underwear (see examples in Figure 1). 8-20 body measurements were measured using a tape measure (multiple persons collected the data with varying expertise in tailoring). The dataset consists of 86 male and 108 female subjects. The approximate capturing distance was 2.4 m and the camera height 1.6 m. Images of front and side views and their manually segmented silhouettes are included. Silhouettes regions were extracted manually using a semi-automatic segmentation tool.

[1] https://www.tc2.com

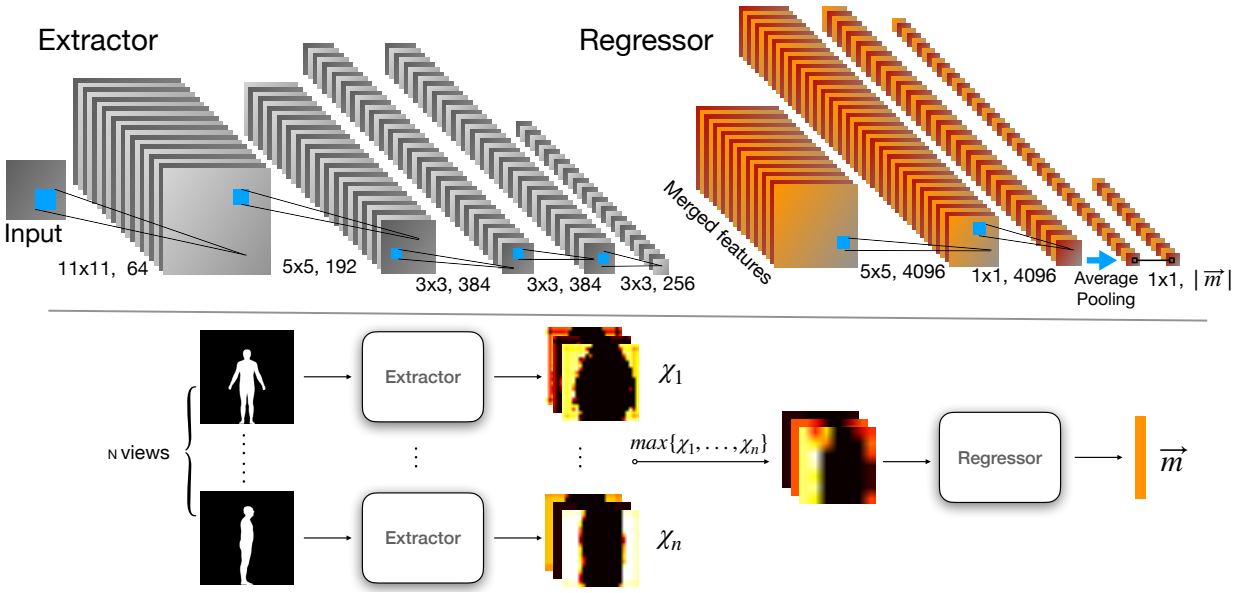

Fig. 2. The overall deep architecture for the baseline method. Blue blocks denote conv kernels, and grey and yellow blocks denote the feature maps. Kernel sizes and the number of output feature maps are shown as $< k \times k, C >$.

IV. Baseline

The network architecture is depicted in Figure 2. The inputs are 224×224 binary silhouette images and the output is a body measurement vector $\boldsymbol{m} \in \mathcal{M}$. We adopt an AlexNet-like architecture as the feature extractor for the input silhouettes. The extractor consists of five conv layers followed by ReLU layers except the last layer and 3×3 max-pooling layers except the third and fourth ones. Feature maps extracted from each silhouette are merged via the element-wised max operation, then the merged feature maps are feed into the regressor for body measurement estimation. The regressor consists of three conv layers followed by the ReLU layers except the last layer. We adopt 1×1 conv layers instead of the fully-connected layers and an average pooling layer lies between the second and third conv layers. Kernel sizes and the number of feature maps in each layers are shown at the top of Figure 2. In principal the network can be extended to N silhouettes, but we found that more than two silhouettes provide only marginal improvement and therefore fixed $N = 2$. Moreover, small improvements can be achieved by optimizing the camera angles, but we fixed the angles to $\theta_1 = 0°$ and $\theta_2 = 90°$ corresponding to the frontal and side views used in the previous works.

To solve the task of body measurements estimation we adopt the weighted Mean Square Error loss function:

$$\mathcal{L}(\boldsymbol{m}, \hat{\boldsymbol{m}}) = \sum_{i=1}^{|\boldsymbol{m}|} w_i (m_i - \hat{m}_i)^2 \ , \qquad (1)$$

where $\boldsymbol{w}$ is the weight vector. There is an important finding related to setting the weight vector weights. In our preliminary experiments we found that selection of the weights has substantial effect to the accuracy of each measurement. Therefore, instead of using a single monolithic network for all measurements, a measurement specific network with optimized weights should be trained for each measurement (1 vs. 16 networks in our experiments). These findings are experimentally demonstrated in our ablation study in Section V-D and also discussed in several recent works [35], [36]. However, since exhaustive search of optimal weights using cross-validation is slow, this optimization was omitted in the method comparisons and a single monolithic network with equal weights $\boldsymbol{w} = \mathbf{1}_{16 \times 1}$ was used.

V. Experiments

A. Settings

a) Data and settings.: The training process of deep neural networks requires a huge amount of training samples and the corresponding ground truths. We synthesize training samples similar to the previous works [20], [21], [22], [14] using a statistical shape model. Since all subjects in the CAESAR-fit and BODY-fit are almost in the same "A" pose, we ignored the pose variation and learned a statistical shape model via performing PCA over the training set meshes. The first 20 principle components were selected and sampled from a multivariate normal distribution to synthesize samples from which geodesic distances were computed as the body measurements and silhouette images rendered. The CAESAR-fit and BODY-fit datasets were randomly split to the training set (80%) and test set (20%). For the both datasets we created 100k synthesized meshes for training.

The Dibra et al. [22] method used in the experiments was trained using the parameters from the original work and the same train/test splits as our baseline network. The baseline network was implemented in TensorFlow and used mini-batch size of 16, learning rate 1e-4 and the Adam optimizer.

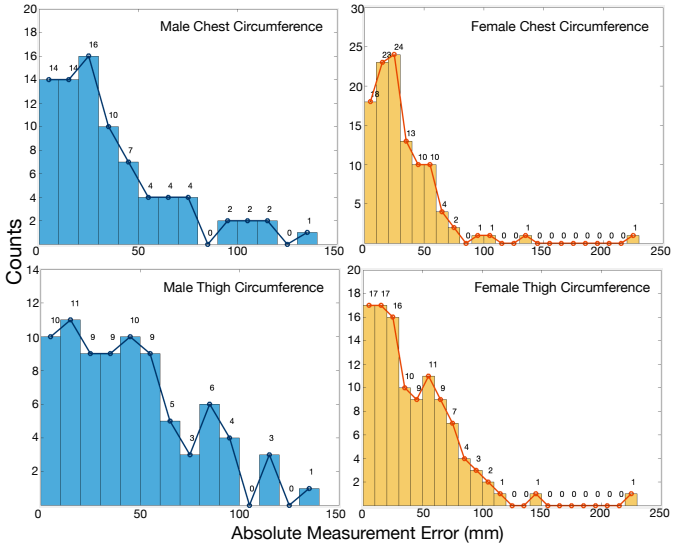

Fig. 3. The error histograms of male/female chest and thigh measurements in our realistic BODY-rgb dataset.

b) Performance Metrics.: The main performance metric is the Mean Absolute Error (MAE) of the anthropometric measurement estimates $\hat{m}_i$ against the ground truth m_i. For the i-th measurement, the MAE ϵ_i over j subjects is obtained from

$$\epsilon_i = \frac{1}{|j|} \sum_{j=1}^{|j|} |m_i^{(j)} - \hat{m}_i^{(j)}| \ . \tag{2}$$

It is important to notice that the minimum and maximum values of each measurement are different due to the human anatomy (e.g. wrist circumference vs. chest circumference) and therefore the mean over MAEs is a poor overall performance measure and therefore not reported.

B. Method Comparison

A number of different methods are discussed in Section II, but the authors of these methods do not provide code or pre-trained models. The two most recent works are HS-Net and UF-US by Dibra et al. [20], [22]. For these experiments, we obtained the code for UF-US [22] from the original authors. The two-view version of their method, UF-US-2, was trained using our train/test splits of the both datasets, CAESAR-fit and BODY-fit, and using their default parameters. The Heat Kernel Signature (HKS) step was not used as it is very slow to compute and provides only marginal improvement (see SFUS-HKS-1 vs. SFUS-1 in [22]). We also selected the geodesic distances of the same 16 anthropometric measurements in [22] as the metrics for comparison. The results for the proposed baseline ("Our") and UF-US-2 are in Table I. With the both datasets and for the most of the anthropometric measurements our model achieves better results than UF-US-2 using the same train/test splits.

C. RGB Dataset

5-fold cross-validation was run on the BODY-rgb dataset. Since some of the tape measurements were missing for some of the examples the missing values were replaced with the training set mean value on each fold and in testing the missing values were omitted. The BODY-fit trained model was used as the basis model that was fine-tuned using the RGB training data. The results are shown in Table II. The accuracy is clearly worse than for the 3D fit datasets in Table I, but also clearly better than the statistical baseline "1st-stat" indicating that the network learns essential features for anthropometric measurement estimation. The error histograms of the male and female chest (C.) and thigh (L.) circumference are shown in Figure 3. The histograms are better for female subjects which is partly due to the small number of male samples.

D. Ablation study

a) Amount of generated data:: An important factor for the proposed model is how many generated training samples are needed to reach good accuracy. This was experimented with CAESAR-fit male and the results for different number of synthetic samples in training are shown in Table III. We see that accuracies gradually improve with substantial variation between the runs, but converge at 100k.

b) Measurement specific networks:: One important finding in our preliminary experiments was that that measurement specific models can achieve better results than a single monolithic network for all measurements. The problem of specific networks is that the results strongly depend on the weights for each measurement during training. To experiment with this approach we tested the following three approaches: 1) Single network and the same weight value for all measurements ($w_i = 1.0$), 2) Specific networks with a larger weight for the main measurement ($w_i = 10.0$) and 3) Specific networks with single outputs ($w_i = \infty$). For the case 1) only one network needed to be trained, but for the cases 2) and 3) we needed to train 16 networks. The results are summarized in Table IV.

The results indicate that substantial improvement can be achieved by optimizing the cost weights. For example, the inside leg length measurement accuracy is 26.8 for a single output network, 13.5 for the network with equal weights and 9.6 for the network emphasizing the specific measurement. On the other hand, the male chest circumference performs best using the equal weights for all measurements.

VI. Conclusion

We introduced new benchmark datasets to boost research on methods that can estimate anthropometric body measurements from image data. The first dataset, BODY-fit, includes 2,675 female and 1,474 male 3D meshes constructed from the scans of real subjects. Similar to previous works, a number of geodesic distance paths on the meshes were measured to provide body measurement ground truth and silhouette images were generated. We provide the same measurements, similarly generated silhouettes and train/test splits for the existing 1,531 female and 1,517 male CAESAR fitted meshes

TABLE I

COMPARISON OF THE PROPOSED METHOD (OUR) TO THE PRIOR ART (UF-US-2 [22]). UF-US-2 CODE WAS OBTAINED FROM THE ORIGINAL AUTHORS. METHODS WERE TESTED USING THE SAME TRAIN/TEST SPLITS AND ALL UNITS ARE MILLIMETERS (MM).

	CAESAR-fit				BODY-fit			
	Male		Female		Male		Female	
Measure	UF-US-2 [22]	Our	UF-US-2 [22]	Our	UF-US-2 [22]	Our	UF-US-2 [22]	Our
A. Head circ.	10.6	**8.6**	18.1	**15.9**	26.0	**17.2**	13.9	**9.2**
B. Neck circ.	11.6	**9.3**	**11.6**	15.5	13.4	**11.8**	**14.5**	14.6
C. Shoulder-b/c len.	9.9	**5.4**	**10.7**	16.3	12.3	**11.2**	9.4	**7.7**
D. Chest. circ.	27.4	**18.2**	32.3	**24.8**	32.1	**23.0**	26.2	**21.7**
E. Waist circ.	27.6	**17.0**	32.0	**22.9**	42.5	**16.5**	22.3	**17.1**
F. Pelvis circ.	**22.9**	30.6	29.0	**24.0**	24.8	**13.3**	20.6	**14.7**
G. Wrist circ.	**9.5**	10.7	**12.2**	13.3	4.2	**4.1**	**4.8**	5.2
H. Bicep circ.	14.9	**12.5**	16.6	**11.5**	13.8	**11.4**	11.9	**9.3**
I. Forearm circ.	12.4	**7.9**	13.5	**10.7**	8.7	**7.2**	8.6	**8.5**
J. Arm len.	8.9	**4.2**	**8.9**	13.1	9.2	**7.6**	7.4	**6.4**
K. Inside leg len.	**9.8**	13.5	**13.3**	14.8	11.9	**9.2**	10.0	**6.5**
L. Thigh circ.	21.9	**16.5**	28.2	**16.4**	**16.9**	17.8	14.8	**11.6**
M. Calf circ.	12.5	**7.2**	16.0	**10.3**	11.0	**8.8**	13.6	**9.2**
N. Ankle circ.	9.2	**4.6**	10.6	**6.1**	6.4	**5.4**	7.2	**6.1**
O. Overall height	**14.8**	15.1	**20.2**	34.7	25.8	**9.9**	17.1	**8.6**
P. Shoulder breadth	9.0	**5.6**	**9.8**	10.9	12.0	**9.2**	9.3	**7.6**

TABLE II

RESULTS FOR BODY-RGB (5-FOLD-CROSS-VALIDATION) WITH TAPE MEASURED GROUND TRUTH (A, C AND I WERE NOT AVAILABLE). "1ST-STAT" USES A TRAINING SET MEAN AS THE PREDICTION TO ALL TEST SAMPLES.

	BODY-rgb			
Measure	Male		Female	
	1st-stat	Our	1-stat	Our
A. Head circ.	-		-	
B. Neck circ.	19.8	14.3	20.3	13.8
C. Shoulder-b/c len.	-	-	-	
D. Chest. circ.	76.1	36.1	101.1	31.7
E. Waist circ.	97.6	35.3	121.9	42.7
F. Pelvis circ.	62.2	35.5	90.4	35.5
G. Wrist circ.	8.5	6.6	8.7	6.9
H. Bicep circ.	27.1	20.9	36.3	19.9
I. Forearm circ.	-	-	-	-
J. Arm len.	27.9	22.5	25.1	18.6
K. Inside leg len.	46.7	31.4	37.1	23.7
L. Thigh circ.	43.8	42.8	62.7	44.3
M. Calf circ.	23.1	12.8	29.6	16.7
N. Ankle circ.	12.3	8.5	17.1	13.8
O. Overall height	59.8	14.3	51.5	19.4
P. Shoulder breadth	21.8	15.8	22.0	19.6

TABLE III

TEST SET ACCURACY FOR DIFFERENT NUMBER OF GENERATED TRAINING SAMPLES.

	CAESAR-fit			
Measure	1k	5k	10k	100k
	Male			
A. Head circ.	17.5	14.3	19.4	8.6
B. Neck circ.	11.5	10.8	9.3	9.3
C. Shoulder-b/c len.	16.3	17.6	20.4	5.4
D. Chest. circ.	37.8	25.7	21.8	18.8
E. Waist circ.	41.2	24.2	17.1	17.0
F. Pelvis circ.	27.9	22.2	31.0	30.6
G. Wrist circ.	8.6	7.6	9.1	10.7
H. Bicep circ.	14.4	11.9	14.8	12.5
I. Forearm circ.	10.6	10.8	13.0	7.9
J. Arm len.	18.6	19.9	15.3	4.2
K. Inside leg len.	25.2	25.1	26.3	13.5
L. Thigh circ.	22.4	18.0	28.1	16.5
M. Calf circ.	12.3	10.2	15.7	7.2
N. Ankle circ.	8.8	7.4	12.0	4.6
O. Overall height	51.0	54.1	60.3	15.1
P. Shoulder breadth	10.5	13.4	11.5	5.6

of Pishchulin et al. [24] (CAESAR-fit). Our meshes share the same topology to CAESAR-fit and therefore allows further 3D and 2D cross-dataset comparisons between them. We introduce another realistic dataset of 86 male and 108 female RGB images and corresponding manually made tape measured ground truth (BODY-rgb). As a baseline for these datasets we propose a simple yet effective deep CNN architecture that obtains competitive accuracy on all three datasets.

TABLE IV
TEST SET ACCURACY FOR SPECIFIC NETWORKS ($w = 1.0$: SINGLE NETWORK FOR ALL MEASUREMENTS; $w = 10.0$ TARGET MEASUREMENT WEIGHT IS 10.0 AND OTHER MEASUREMENTS 1.0; $w = \infty$: ONLY THE TARGET MEASUREMENT USED).

Measure	CAESAR-fit $w_i = 1.0$	CAESAR-fit $w_i = 10.0$	CAESAR-fit $w_i = \infty$
		Male	
A. Head circ.	8.6	**8.2**	11.9
B. Neck circ.	9.3	**9.2**	10.1
C. Shoulder-b/c len.	**5.4**	8.1	19.9
D. Chest. circ.	**18.2**	35.9	34.6
E. Waist circ.	**17.0**	23.5	24.6
F. Pelvis circ.	30.6	30.2	**23.0**
G. Wrist circ.	10.7	**6.8**	8.6
H. Bicep circ.	12.5	14.3	**10.6**
I. Forearm circ.	7.9	**6.9**	9.5
J. Arm len.	**4.2**	15.7	13.1
K. Inside leg len.	13.5	**9.6**	26.8
L. Thigh circ.	16.5	23.4	**15.5**
M. Calf circ.	**7.2**	8.5	10.9
N. Ankle circ.	4.6	**4.5**	6.5
O. Overall height	15.1	**8.3**	22.1
P. Shoulder breadth	**5.6**	5.7	7.6

REFERENCES

[1] B.-K. D. Park, S. Ebert, and M. Reed, "A parametric model of child body shape in seated postures," *Traffic Injury Prevention*, vol. 18, no. 5, 2017.

[2] H. Daanen and S.-A. Hong, "Made-to-measure pattern development based on 3D whole body scans," *International Journal of Clothing Science and Technology*, vol. 20, no. 1, 2008.

[3] C. Ogden, C. Fryar, M. Carroll, and K. Flegal, "Mean body weight, height, and body mass index, united states 1960–2002," Division of Health and Nutrition, Examination Surveys 347, 2004.

[4] S. Pheasant and C. Haslegrave, *Bodyspace: Anthropometry, Ergonomics and the Design of Work*, 3rd ed. Taylor & Francis, 2005.

[5] X. Zhou, M. Zhu, K. Derpanis, and K. Daniilidis, "Sparseness meets deepness: 3d human pose estimation from monocular video," in *CVPR*, 2016.

[6] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "VNect: Real-time 3d human pose estimation with a single rgb camera," 2017.

[7] G. Rogez, P. Weinzaepfel, and C. Schmid, "LCR-Net: Localization-Classification-Regression for Human Pose," in *CVPR*, 2017.

[8] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, "Learning from synthetic humans," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. IEEE, 2017, pp. 4627–4635.

[9] J. Tan, I. Budvytis, and R. Cipolla, "Indirect deep structured learning for 3d human body shape and pose prediction," in *BMVC*, vol. 3, no. 5, 2017, p. 6.

[10] I. K. R. Alp Güler, N. Neverova, "DensePose: Dense human pose estimation in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[11] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, "Learning to estimate 3d human pose and shape from a single color image," in *CVPR*, 2018.

[12] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-end recovery of human shape and pose," in *CVPR*, 2018.

[13] A. Zanfir, E. Marinoiu, and C. Sminchisescu, "Monocular 3d pose and shape estimation of multiple people in natural scenes–the importance of multiple scene constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2148–2157.

[14] Z. Ji, X. Qi, Y. Wang, G. Xu, P. Du, and Q. Wu, "Shape-from-mask: A deep learning based human body shape reconstruction from binary mask images," *arXiv preprint arXiv:1806.08485*, 2018.

[15] L. Sigal, A. Balan, and M. J. Black, "Combined discriminative and generative articulated pose and non-rigid shape estimation," in *Advances in neural information processing systems*, 2008, pp. 1337–1344.

[16] Y. Chen, T.-K. Kim, and R. Cipolla, "Inferring 3d shapes and deformations from single views," in *European Conference on Computer Vision*. Springer, 2010, pp. 300–313.

[17] Y. Chen, D. P. Robertson, and R. Cipolla, "A practical system for modelling body shapes from single view measurements." in *BMVC*, 2011.

[18] Y. Chen, T.-K. Kim, and R. Cipolla, "Silhouette-based object phenotype recognition using 3d shape priors," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 25–32.

[19] J. Boisvert, C. Shu, S. Wuhrer, and P. Xi, "Three-dimensional human shape inference from silhouettes: reconstruction and validation," *Machine vision and applications*, vol. 24, no. 1, pp. 145–157, 2013.

[20] E. Dibra, H. Jain, C. Öztireli, R. Ziegler, and M. Gross, "HS-Nets: Estimating human body shape from silhouettes with convolutional neural networks," in *3D Vision (3DV)*. IEEE, 2016, pp. 108–117.

[21] E. Dibra, C. Öztireli, R. Ziegler, and M. Gross, "Shape from selfies: Human body shape estimation using cca regression forests," in *European Conference on Computer Vision*. Springer, 2016, pp. 88–104.

[22] E. Dibra, H. Jain, C. Öztireli, R. Ziegler, and M. Gross, "Human shape from silhouettes using generative HKS descriptors and cross-modal neural networks," in *CVPR*, 2017.

[23] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming, "Civilian american and european surface anthropometry resource (CAESAR) final report," US Air Force Research Laboratory, Tech. Rep. AFRL-HE-WP-TR-2002-0169, 2002.

[24] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele, "Building statistical shape spaces for 3d human modeling," *Pattern Recognition*, 2017.

[25] S. Paquette, J. D. Brantley, B. D. Corner, P. Li, and T. Oliver, "Automated extraction of anthropometric data from 3d images," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 38. SAGE Publications Sage CA: Los Angeles, CA, 2000, pp. 727–730.

[26] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, I. Tebbetts, and R. A. Walker, "Anthropometric survey of us army personnel: Summary statistics, interim report for 1988," ANTHROPOLOGY RESEARCH PROJECT INC YELLOW SPRINGS OH, Tech. Rep., 1989.

[27] K. Kim, M. Jones, S. Ebert, L. Malik, M. Manary, M. Reed, and K. Klinich, "Development of virtual toddler fit models for child safety restraint design," University of Michigan Transportation Research Institute, Tech. Rep. UMTRI-2015-38, 2015.

[28] Y. Yang, Y. Yu, Y. Zhou, S. Du, J. Davis, and R. Yang, "Semantic parametric reshaping of human body models," in *Int. Conference on 3D Vision (3DV)*, 2014.

[29] A. Tsoli, M. Loper, and M. Black, "Model-based anthropometry: Predicting measurements from 3d human scans in multiple poses," in *Winter Conference on Applications of Computer Vision (WACV)*, 2014.

[30] S. Yan, J. Wirta, and J.-K. Kämäräinen, "Anthropometric clothing measurements from 3d body scans," *Machine Vision and Applications*, vol. 31, no. 1, p. 7, 2020.

[31] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: shape completion and animation of people," in *SIGGRAPH*, 2005.

[32] D. A. Hirshberg, M. Loper, E. Rachlin, and M. J. Black, "Coregistration: Simultaneous alignment and modeling of articulated 3d shape," in *European Conference on Computer Vision*. Springer, 2012, pp. 242–255.

[33] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, vol. 34, no. 6, pp. 248:1–248:16, Oct. 2015.

[34] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid ICP algorithms for surface registration," in *CVPR*, 2007.

[35] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *CVPR*, 2018.

[36] O. Zener and V. Koltun, "Multi-task learning as multi-objective optimization," in *NeurIPS*, 2018.