

Human Body Measurement Estimation with Adversarial Augmentation

[Nataniel Ruiz](#) [Miriam Bellver](#) [Timo Bolkart](#) [Ambuj Arora](#) [Ming C. Lin](#) [Javier Romero](#) [Raja Bala](#)

Amazon

[Paper](#) | [Poster](#) | [Video](#) | [Dataset](#)

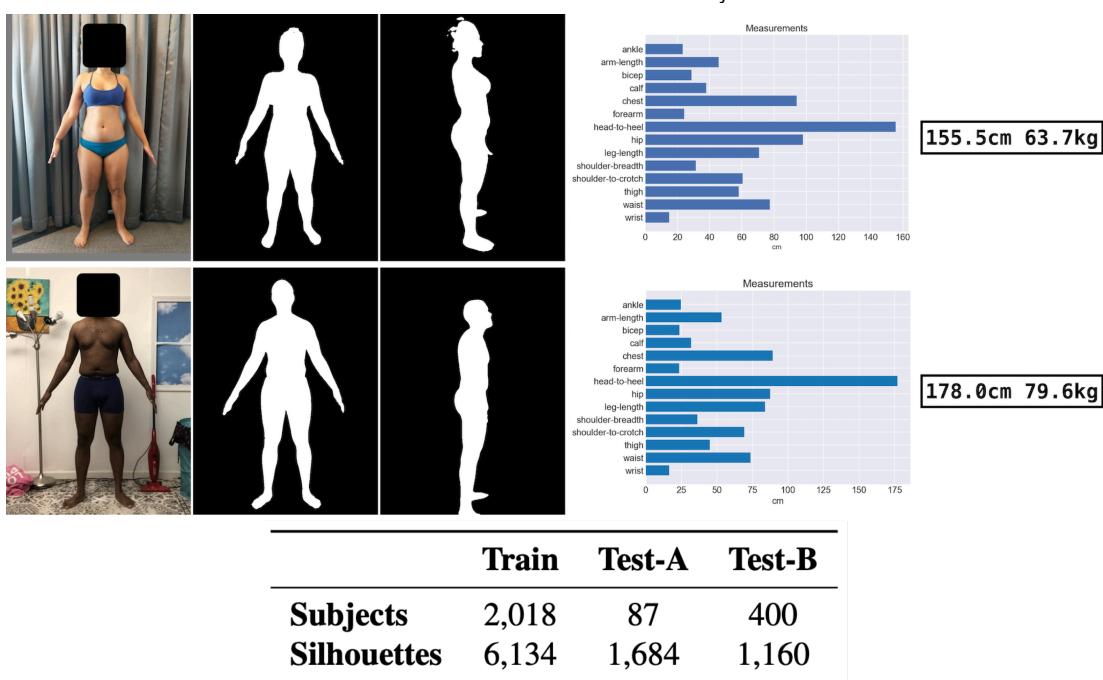
Abstract

We present a Body Measurement network (BMnet) for estimating 3D anthropomorphic measurements of the human body shape from silhouette images. Training of BMnet is performed on data from real human subjects, and augmented with a novel adversarial body simulator (ABS) that finds and synthesizes challenging body shapes. ABS is based on the skinned multiperson linear (SMPL) body model, and aims to maximize BMnet measurement prediction error with respect to latent SMPL shape parameters. ABS is fully differentiable with respect to these parameters, and trained end-to-end via backpropagation with BMnet in the loop. Experiments show that ABS effectively discovers adversarial examples, such as bodies with extreme body mass indices (BMI), consistent with the rarity of extreme-BMI bodies in BMnet's training set. Thus ABS is able to reveal gaps in training data and potential failures in predicting under-represented body shapes. Results show that training BMnet with ABS improves measurement prediction accuracy on real bodies by up to 10%, when compared to no augmentation or random body shape sampling. Furthermore, our method significantly outperforms SOTA measurement estimation methods by as much as 3x. Finally, we release BodyM, the first challenging, large-scale dataset of photo silhouettes and body measurements of real human subjects, to further promote research in this area.

BodyM Dataset

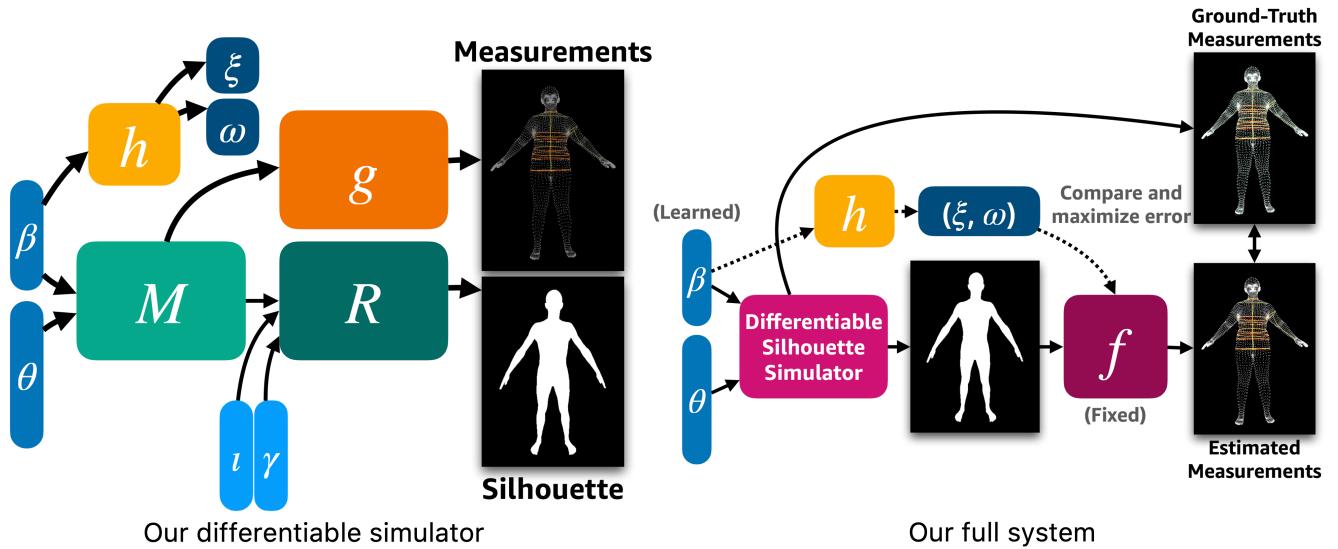
We present the first large public body measurement dataset including more than 8,000 frontal and lateral silhouettes from more than 2,500 real subjects, paired with height, weight and 14 body measurements. The figure above shows samples of silhouettes for a pair of subjects, including measurement bar charts with all 14 measurements and their respective height and weight. BodyM has been captured in a well-lit, indoor setup. Individuals are photographed in tight-fitting clothing and standing in A-Pose. 3D scans of each subject are acquired with a Treedy photogrammetric scanner, and registered to the SMPL mesh topology. The 14 body measurements are computed from these meshes. We break the dataset into 3 groups: Training, Test Set A, Test Set B. For the training and Test-A sets, subjects are photographed and 3D-scanned by in a lab by technicians. For the Test-B set, subjects are scanned in the lab, but photographed in a less-controlled environment with diverse camera orientations and lighting conditions, to simulate in-the-wild image capture. Some subjects have been photographed more than once with different clothing in order to test robustness of the dataset. The statistics of the 3 data subsets are shown in the table below, and breakdown of gender and body shape in terms of body-mass-index (BMI) is reported in Table 2 of the paper.

License: Creative Commons Attribution-Non Commercial 4.0 International Public License: <https://creativecommons.org/licenses/by/4.0/legalcode>



Simulated Adversarial Training

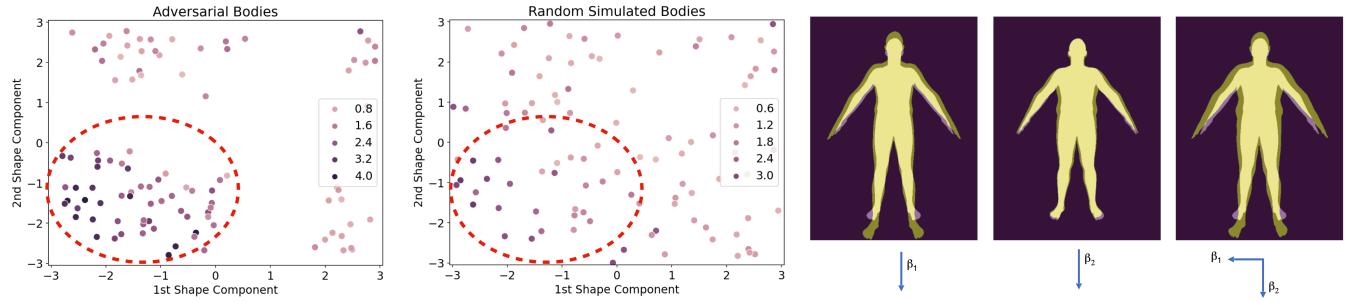
Our differentiable simulator is comprised of the SMPL model M that takes in shape β and pose θ . We generate height ϵ and weight ω using a trained regressor h , generate measurements using a deterministic measuring function g operating on vertices, and we render the silhouettes using a renderer R . In order to simulate adversarial bodies, we optimize the shape β by maximizing a loss between the ground-truth measurements and the measurements estimated by our model f . The resulting shape parameters give rise to an adversarial body, which are then used to train the model f .



Diagnosing Model Weaknesses

We show that adversarially simulated bodies are concentrated in certain regions of body shape space. Here we can see that adversarial bodies are concentrated in negative regions of the 1st and 2nd shape component and have higher loss than random bodies. We also observe that the negative directions

of these components correspond to taller and wider bodies.



State-of-the-Art Results

SotA compared to off-the-shelf body reconstruction methods

We compare our method to SotA off-the-shelf body shape estimation methods such as SPIN, STRAPS and Sengupta et al. (ICCV 2021) and we find that our method outperforms them on body measurement estimation. Our method uses only one silhouette and no height and weight data for fairness.

	Overall			Chest	Hip	Leg Length	Waist
	TP90	TP75	TP50	MAE	MAE	MAE	MAE
SPIN [34]	81.10	57.33	33.96	74.45	65.41	35.81	77.39
STRAPS [72]	103.61	75.74	45.67	82.30	63.96	48.71	108.00
Sengupta et al. [69]	68.81	47.64	28.71	53.07	47.43	42.11	53.20
Ours (Single-View, No Metadata)	41.91	29.13	17.09	33.95	31.03	25.80	31.93

Comparison to SotA body shape estimation methods (errors in mm)

SotA compared to other body measurement estimation methods

We compare against the SotA methods for body measurement estimation and find that we outperform them overall, and over a variety of specific measurements on two different datasets.

	Ankle	Arm-L	Bicep	Calf	Chest	Forearm	H2H	Hip	Leg-L	S-B	S-to-C	Thigh	Waist	Wrist	Overall
Dibra et al. [19]	2.0	2.7	3.3	3.3	7.2	2.3	4.0	6.0	2.8	2.9	2.9	4.9	8.1	2.0	3.78
Smith et al. [78]	2.1	1.7	2.7	2.3	4.7	1.9	2.3	3.0	1.5	1.9	1.5	2.4	4.8	2.5	2.72
Ours	0.8	1.9	1.7	0.8	4.6	1.3	3.6	1.8	2.1	0.9	1.9	1.7	3.8	0.7	1.97
	Neck	Chest	Waist	Pelvis	Wrist	Bicep	Forearm	Arm	Leg	Thigh	Calf	Ankle	Height	Shoulder	Overall
Yan et al. [90]	11.8	23.0	16.5	13.3	4.1	11.4	7.2	7.6	9.2	17.8	8.8	5.4	9.0	9.2	11.0
Ours	11.0	15.2	15.7	17.3	3.8	4.7	3.9	7.7	10.0	7.5	8.6	10.0	13.4	7.1	9.7
Yan et al. [90]	14.6	21.7	17.1	14.7	5.2	9.3	8.5	6.4	6.5	11.6	9.2	6.1	8.6	7.6	10.5
Ours	4.4	9.1	10.8	7.7	5.2	3.9	5.3	6.4	10.2	13.2	9.8	12.2	20.7	6.5	9.0

Comparison to SotA body measurement estimation methods (errors in mm)

Simulated adversarial training improves body measurement estimation

We perform an ablation study over several setups (respectively: single-view, multi-view, and multi-view w/ 10x less real training data). We find that simulated adversarial body augmentation outperforms no augmentation and augmentation using randomly sampled simulated bodies.

	Overall			Chest	Hip	Waist
	TP90	TP75	TP50	MAE	MAE	MAE
Single-View (No Aug.)	19.10	13.00	7.64	19.18	11.53	16.12
Single-View (Random Aug.)	18.98	12.84	7.50	19.13	11.43	15.76
Single-View (Adv. Aug.)	18.90	12.82	7.44	18.84	11.14	15.78
Multi-View (No Aug.)	16.45	11.06	6.51	14.40	10.88	13.40
Multi-View (Random Aug.)	16.43	11.06	6.48	14.66	10.60	13.17
Multi-View (Adv. Aug.)	16.00	10.00	6.53	14.52	10.00	13.00
Multi-View (No Aug.)	26.52	17.64	10.04	24.60	19.55	21.75
Multi-View (Random Aug.)	26.13	17.35	9.90	23.09	18.87	22.44
Multi-View (Adv. Aug.)	25.00	16.28	9.50	22.98	18.09	21.10

Ablation study w.r.t. no augmentation, random aug. and adversarial aug. (errors in mm)