

Deterministic Measurement-Conditioned 3D Body Reconstruction for Precision Digital Fitting

Olga Barbina
0009-0007-8632-0847
DIVI Grupa Ltd
Fridriha Candra str. 1,
Riga, LV-1046, Latvia
Email: olga.barbina@di.lv

Zane Bicevska
0000-0002-5252-7336
University of Latvia,
Raina boulevard 19,
Riga, LV-1046, Latvia
Email: zane.bicevska@lu.lv

Abstract—Custom tailoring and digital fitting require geometrically precise avatars, while most existing methods prioritize visual realism. We present a method for reconstructing three-dimensional human bodies from 47 anthropometric measurements, with accuracy within one standard clothing-size step, ensuring suitability for pattern construction. We train a separate model per sex, each on 400 real scans, using a deterministic FiLM-conditioned regression network with per-dimension Fourier feature mapping, pose-robust geometric losses, and scale augmentation. We show that dense anthropometric measurements carry most of the shape information relevant to fitting. In our tests the deterministic model reaches lower measurement error than a conditional variational autoencoder and, unlike an unconstrained linear baseline, stays geometrically stable on out-of-distribution measurements.

Index Terms—anthropometric measurements, 3D body generation, deterministic regression, measurement-conditioned mesh generation, digital fitting.

I. INTRODUCTION

STANDARD clothing sizes are averaged and rarely fit individual bodies well. While made-to-measure tailoring solves the fit issue, developing and testing these physical prototypes is expensive and time-consuming. Finding real human models for all required measurement combinations is impractical, and producing physical prototypes for each variation requires substantial resources [1], [2].

Three-dimensional (3D) avatars make automated digital fitting possible. This requires translating simple 1D tape measurements into a complex 3D human body surface (~20,000 vertex coordinates). The main challenge is that one-dimensional measurements do not fully determine body shape as individuals with identical measurements may differ in muscle distribution, posture, or body composition. The model must resolve this ambiguity while maintaining anatomical plausibility. Our objective is to take a dense set

of measurements and output a 3D mesh that satisfies the target measurements without distorting the body.

Avatars used in retail for virtual try-on prioritize visual realism over geometric accuracy. These systems help visualize a garment's general appearance but lack the accuracy needed to predict whether it will physically fit a specific body.

Applications such as virtual apparel fitting and digital anthropometry require quantitatively accurate representations. For context, the European clothing standard [3] relies on 4 cm intervals for major measurements like the chest, waist, and hips. Errors of 20-40 mm in key measurements can lead to incorrect sizing and fit.

Existing methods for 3D human body modeling generally follow two main paths:

Parametric models rely on a small set of linear coefficients (typically 10-16 PCA components) derived from 3D scans to represent the body shape. While compact and suitable for animation, these models are linear approximations that affect all mesh vertices globally: adjusting one measurement (e.g., waist circumference) affects the entire mesh. This can limit their ability to satisfy multiple independent measurement constraints simultaneously [4].

Generative models, such as conditional variational autoencoders (CVAEs), predict geometry directly from measurements, but they include a latent variable that is sampled during generation. When it is sampled, the same measurements can produce slightly different bodies. For digital fitting this is undesirable as users expect repeatable, reliable results.

We address this with a dense set of 47 measurements. As this dense set describes most of body shape variance, we use a deterministic regression model that maps measurements directly to a repeatable 3D mesh, while a generative model produces variable output unless its latent is fixed.

The research leading to these results has received funding from the project “Competence Centre of Information and Communication Technologies for Digitalization” of the Recovery and Resilience Facility, contract No. 2.2.1.3.i.0/1/24/A/CFLA/006 signed between IT Competence Centre and Central Finance and Contracts Agency, Research No. 1.2 “Anthropometric Database for the Digitalization of the Apparel Industry”.

The model uses a Direct Regression Network (DRN) with Fourier feature mapping [5] to map each measurement into a multi-band sinusoidal representation. We use Feature-wise Linear Modulation (FiLM) [6] to inject the measurement signal into intermediate layers, preserving conditioning information throughout the network.

To improve accuracy on a limited dataset, we suppress unwanted variation (e.g., residual pose) using pose-robust losses that prioritize shape over pose. At the same time, we use custom scaling augmentations to synthetically expand the range of Fourier-mapped values the model encounters during training. We also apply per-dimension Fourier mapping [5], isolating each measurement into independent channels.

The resulting system achieves ~ 2 mm measurement error on in-distribution data and ~ 5 mm on out-of-distribution (OOD) inputs, within a single size gradation step, while training on a relatively small dataset of real scans.

II. RELATED WORK

Generating 3D body geometry from measurements requires mapping a low-dimensional vector to a high-dimensional mesh. Existing approaches vary mainly in how they handle this translation and represent the final shape.

A. Low-Dimensional Parametric Models

A common class of methods represents body shape in a low-dimensional parametric space derived from principal component analysis (PCA) learned from registered 3D scans. Models such as SCAPE [7], SMPL [8], STAR [4], and GHUM [9] build the body's geometry by blending global shape components with an underlying pose skeleton.

Measurements can be mapped to these parameters using regression or optimization-based techniques. For example, Virtual Caliper [10] optimizes SMPL parameters to match target measurements, while A2B [11] learns a direct mapping from measurements to shape coefficients.

These parametric methods are effective for generating visually plausible body shapes, but the output is mathematically constrained to a small set of PCA components that affect all vertices globally.

The disadvantage of these global parameters is that they are highly entangled. Adjusting one measurement can affect other parts of the body, sometimes even distant from the target measurement regions.

Recent work such as ATLAS [12] attempts to decouple skeletal and shape representations, improving control over proportions, but the surface geometry remains constrained by a low-dimensional shape space.

B. Probabilistic and Generative Models

To overcome limitations of parametric models, generative approaches predict geometry directly from measurements. AnthroNet [13], for example, trains a CVAE on synthetic meshes to predict high-resolution geometry directly from 36 specific body measurements.

While CVAEs increase expressivity, generating from measurements alone requires choosing a latent code: sampling it produces different outputs for identical measurements, while fixing it (e.g., to the prior mean) discards the latent the model was trained to use. Either way, the latent channel competes with the measurement conditioning, which we find degrades measurement fidelity (Section V.B). While suitable for applications emphasizing diversity or realism, this variability can be undesirable in settings requiring repeatable predictions, such as digital fitting, where the same customer measurements must produce identical avatars across sessions.

Optimization-based approaches can improve measurement accuracy, but the iterative process may be too slow for real-time inference. For example, Thach and Dat [14] use a genetic algorithm to iteratively search for SMPL shape parameters matching input measurements, but the approach requires ~ 500 generations of optimization per sample at inference.

C. Direct Regression of Geometry

Direct regression approaches learn a mapping from measurements to vertex positions without an intermediate shape space or latent variables. Earlier direct methods, like the work by Zeng et al. [15], relied on localized mappings driven by just 19 measurements.

Neural network approaches can handle non-linear mappings between measurements and 3D surfaces. Most existing methods condition on relatively sparse measurement sets (6–36 measurements), treating anthropometric inputs as partial shape descriptors. Smaller measurement sets do not capture the full picture of the body's geometry, requiring the model to infer missing information.

D. Our Work Positioning

Our work follows the direct regression approach but differs in the density of conditioning. We use a denser set of 47 measurements: circumferences, heights, widths, and lengths. This dense conditioning reduces ambiguity and enables the use of a deterministic model to provide repeatability required for digital fitting.

A linear least-squares map from the 47 measurements to vertex positions, fit on the training split, explains $\approx 75\%$ of vertex variance on held-out test meshes. The unexplained remainder concentrates in the most articulated regions - hands ($R^2 \approx 0.5$) and forearms (≈ 0.53). While the forearms are measured (circumference, length), they remain poorly reconstructed, indicating the gap is residual pose and articulation rather than missing measurements. Less pose-variant, fitting-relevant regions are well explained (torso 0.83, upper legs 0.72–0.78, neck 0.91, head 0.89). This mirrors our generative experiments, where the conditional latent encoded mostly residual pose, hand and feet position (Section V.B). A smaller part may be genuine shape (e.g., body composition) that distance measurements cannot capture, motivating non-geometric inputs (weight, age) in future work on an extended dataset.

Our non-linear model, trained with geometric losses, provides structural regularization that unconstrained linear meth-

ods lack. Instead of maximizing numerical precision at the expense of surface consistency, our non-linear approach explicitly balances accuracy with geometric plausibility and OOD stability.

III. METHOD

A. Problem Formulation

We formulate 3D body reconstruction as a supervised regression task, mapping a low-dimensional measurement vector to a high-dimensional surface.

Let $m \in \mathbb{R}^{\mathbb{D}_m}$ denote a vector of $D_m = 47$ anthropometric measurements, including circumferences, lengths, widths, and heights. Our objective is to predict a body mesh $V \in \mathbb{R}^{N \times 3}$ where $N = 6,890$.

Instead of predicting absolute coordinates, our model learns a mapping $f_\theta(m)$ that produces residual vertex offsets relative to a sex-specific mean template V_0 :

$$V = V_0 + f_\theta(m) \quad (1)$$

We train two independent models, one per sex, each conditioned on its own sex-specific template V_0 and trained only on meshes of that sex. All architecture and training details below apply identically to both.

B. Architecture

Our architecture (Fig. 1) is a deterministic model conditioned on measurements m . The mapped measurements serve a dual role: they are projected to a base shape code that seeds the main pathway and directly condition every residual block via FiLM [6] to continuously modulate intermediate features throughout the network.

1. per-dimension Fourier mapping,
2. projection to a base shape code (feature vector),
3. FiLM-conditioned residual blocks
4. offset prediction.

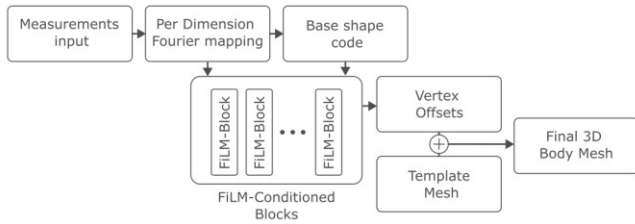


Fig. 1. Data flow of the Direct Regression Network. The base shape code serves as the initial input to the main pathway, while the mapped measurements independently condition every residual block via FiLM

We describe each stage below.

Per-dimension Fourier Feature Mapping

Anthropometric measurements vary widely in scale (e.g., total height vs. wrist circumference) and enter the network as single scalars - input from which an MLP struggles to resolve small differences between similar bodies. Following Tancik et al. [5], we expand each measurement into sine/cosine features across several frequency bands, giving a better-conditioned, than the raw scalar, multi-scale input.

While generative approaches like AnthroNet [13] have also used global Fourier mapping to process measurements, standard cross-dimensional projections make the network highly sensitive to OOD values as a single uncommon measurement can perturb the entire conditioning vector. To prevent this, we apply the mapping independently per dimension, so each measurement is isolated:

$$\phi_i(m_i) = [\sin(2\pi f_k m_i), \cos(2\pi f_k m_i)]_{k=1}^K \quad (2)$$

m_i - raw numerical value of the i -th measurement

f_0 - base frequency

f_k - frequency value for the current step k in the frequency loop, where, $f_k = f_0 \cdot 2^{k-1}$ defines log-spaced frequency bands

K - total number of frequency bands generated for a single measurement

This way, an OOD value affects only its own features. While this prevents anomalous inputs from corrupting the entire signal, a value outside the range the bands were tuned for can still drive the sinusoids past one period (phase wrapping), where distinct inputs share features and produce local surface defects.

We set the base frequency relative to a plausible measurement range, which was chosen to extend beyond the extremes seen in training to include a wider range produced by scale augmentation and cover physically realistic bodies. The lowest frequency band spans less than one period over the expected measurement range, so distinct values map to distinct features (the mapping stays one-to-one); this keeps it stable for unusual inputs, while higher bands add resolution. Each measurement is mapped with $K = 5$ log-spaced bands.

Tuning the base frequency to the plausible data range improves robustness, rather than expressiveness. Our ablations support this: standardized inputs perform comparably in-distribution (Section V.B), and the mapping's advantage appears mainly under OOD inputs. We map measurements in real scale rather than normalizing them per dimension, as per-dimension min-max normalization overfit in our experiments (Section V.B).

Base Shape Code Generation

The mapped measurements $m_{\text{fourier}} = \phi(m)$ are projected through a small MLP to produce a base shape code $c \in \mathbb{R}^{d_c}$, where d_c represents the bottleneck dimension. This compressed representation of the body shape serves as the initial input to the regression network, which subsequent layers refine.

FiLM-conditioned residual blocks

The base shape code is processed through residual blocks conditioned using FiLM. At each block, the Fourier-mapped measurement vector is passed through a small network to produce learned scale and shift parameters, which are applied feature-wise to the intermediate activations. This continuous re-injection of the measurement signal throughout the network allows each layer to dynamically adapt its representations to the specific body shape, capturing complex geometry better than conditioning only the first layer.

Offset Prediction

A final linear layer maps the modulated features to $N \times 3$ vertex offsets, initialized near zero so the untrained network starts from the mean template, computed as per-vertex mean of training meshes separately per-sex. This provides a geometric prior, faster convergence and slightly increased precision compared to predicting absolute vertex positions, which requires the network to learn the full spatial distribution of human body shapes. The final mesh is obtained by adding the predicted offsets to a fixed template mesh.

Pose Robust-Loss Design

Registered meshes carry residual pose variation. A global vertex loss would force the model to reproduce that pose, spending capacity on irrelevant information instead of the measurement-to-shape mapping. We avoid global vertex loss and rely primarily on local geometric losses. We weight the rotation-sensitive differential reconstruction loss more lightly, while the edge length loss, invariant to both rotation and translation, carries the highest weight.

Edge length loss enforces strict structural consistency between the predicted and target meshes by penalizing differences in mesh edge lengths.

Differential reconstruction loss compares Laplacian coordinates [16] of predicted and target meshes, encouraging consistent local surface structure.

Segmented Procrustes loss aligns the predicted mesh to the target using rigid transformation estimated from a stable body region (torso), reducing sensitivity to residual pose differences.

A distortion term penalizes the Laplacian of the deformation (prediction – template), suppressing creases; it references the template, so it is pose-neutral.

We initially experimented with direct measurement supervision, but it induced surface creases on extreme OOD samples. Fixing this required heavy regularization, such as soft smoothing mask along measurement paths and stronger Laplacian displacement penalty to suppress the defect. In practice, pose-robust geometric losses provided comparable measurement accuracy in our experiments with more stable surfaces, so we adopted this approach.

Scale Augmentation for Generalization

A model trained on a small dataset will encounter unseen measurement combinations. To improve generalization in these cases, we apply two complementary scaling-based augmentation strategies applied stochastically during training:

Uniform scaling: For each training sample, with probability 25% we scale the mesh uniformly. All distance-based measurements are multiplied by the same scaling factor maintaining geometric consistency.

Anisotropic scaling: With probability of 30%, we apply independent scale factors along the vertical s_y for heights and horizontal s_{xz} for circumferences and widths. This approach helps to increase exposure to proportion diversity (tall-thin, short-wide) beyond what uniform scaling provides.

We use a small anisotropic range to limit distortion and set different proportions of horizontal and vertical axis influence

on different diagonally oriented measurements. For example, shoulder slope is primarily influenced by the horizontal axis, while arms lengths mostly by the vertical axis.

Both augmentations scale the mesh in real space, ensuring geometrically consistent training examples, while measurements are updated analytically: exactly for uniform scaling, approximately for anisotropic scaling.

Implementation details.

Each sex-specific model maps a 470-dim per-dimension Fourier input ($47 \text{ measurements} \times 5 \text{ bands} \times 2$) to a 48-dim base shape code, refined by 4 FiLM-conditioned residual blocks of width 1024. We train with AdamW (lr 5×10^{-4} , weight decay 1×10^{-2} , 10-epoch warmup then ReduceLROnPlateau (factor 0.5), batch size 100, early stopping on validation loss; best-validation epoch (mean) ≈ 475 (range 355–613 across seeds).

The edge length (weight (w.) 50), differential (w. 10) and segmented-Procrustes (w. 1.5) losses use L1 error; the distortion term (w. 10) penalizes the squared Laplacian of the deformation (L2).

Augmentation uses uniform scaling ($p = 0.25, \pm 30\%$) and anisotropic scaling ($p = 0.30, \pm 15\%$). Inference takes ≈ 0.3 ms/mesh (≈ 0.014 ms/mesh batched) on an NVIDIA GeForce RTX 5090, excluding one-time model loading.

IV. EXPERIMENTS

A. Evaluation Objectives and Metrics

We evaluate our method for:

1. Measurement accuracy: how well the generated mesh matches the input measurements.
2. Geometric consistency: whether the predicted surface preserves local shape properties independent of pose.
3. Generalization: whether the model produces stable and plausible outputs for OOD measurement combinations.

We quantify these objectives with four metrics.

1. Measurement MAE: mean absolute error between digitally computed measurements on the predicted mesh and the GT mesh.
2. Edge length MAE: mean absolute difference in mesh edge lengths between predicted and GT meshes.
3. Per-part volume error: mean relative volume difference between corresponding anatomical segments (torso, upper/lower arms, upper/lower legs, head) of the predicted and ground-truth meshes.
4. Per-part convex hull error: mean relative difference between convex-hull volumes of the reconstructed and GT anatomical segments, averaged over segment; used in OOD evaluation.

B. Dataset and Preprocessing

We use the BODY-fit [17] dataset, which provides SMPL-registered meshes (6890 vertices, 13776 faces) fitted via non-rigid ICP to real 3D body scans. After removing duplicates and meshes with visible registration defects (collapsed limbs,

squeezed body parts), we retain 906 male and 1975 female meshes. Because we train a separate model per sex, each model uses 400 randomly sampled meshes of its own sex for training and 100 for validation; the remaining meshes form per-sex test sets (406 male, 1475 female; 1881 total). This training size reflects realistic constraints of high-quality scanned data.

The dataset provides 16 geodesic circumference measurements, which are insufficient for our application. Instead, we compute a set of 47 measurements directly on the mesh surface using our measurement module. Those include geodesic paths, Euclidean widths, vertical heights, and landmark-based distances. Defining custom measurement protocols on 3D body representations is common in this setting [11], [13], [18].

Our training set (400 samples per sex) is smaller than those used in several prior methods (AnthroNet [13]: 100K synthetic; Zeng et al. [15]: 3,048; CAESAR [19]: ~4,000) but is comparable to the ~500 profiles estimated by Bicevskis et al. [1] as the target for representative size-group coverage.

The DRN operates in real space: measurements are provided in real-world units (meters) inside the per-dimension Fourier mapping (Eq. 2), and the network predicts real-space vertex offsets added to a centered, sex-specific template (Eq. 1). We precompute two training-set statistics used in the pipeline: per-vertex positional variance, for the variance analysis and inverse-variance weighting (Section IV.C); and per-measurement mean/variance, for the standardized-input ablation (Section V.B).

Measurement computation

Ground-truth (GT) measurements are computed on the mesh surface using three methods: geodesic path lengths (for circumferences and surface-traced lengths/widths), Euclidean distances between anatomical landmarks (for straight-line measurements), and vertical distances from a per-mesh ground plane defined by the lowest vertex (for heights). The full measurement set is presented in Table I.

TABLE I.
MEASUREMENT TYPES

Measurement	Method
Ankle, bicep, calf, elbow, forearm, knee, thigh, wrist circumference (for left and right side (L&R)). Chest, head, hip, neck-base, neck-middle, waist-low, waist-natural circumference	Geodesic circumferences: closed-loop path between predefined 2-4 control points
Upper arm, lower arm, shoulder length (L & R); Crotch arc length; Shoulder and torso (back, front) lengths. Neck, underbust heights.	Geodesic lengths: open path between predefined 2-3 control points
Bust-point width, shoulder width	Euclidean distance between 2 points
Ankle, calf, knee height (L & R); crotch, natural-waist height, total height	Euclidean distance, difference (on vertical axis only) between control point and individual ground plane

The measurement module is being extended to support additional landmark-based measurements. The final set of measurements for training is selected based on variance coverage of underrepresented body regions, inter-measurement correlation structure [1], and relevance to garment fitting.

C. Per-Vertex Variance Analysis

Even with a shared template topology, the dataset contained residual rigid-body misalignments (minor variations in global position and orientation across meshes). We apply ICP alignment of all meshes to a canonical template via SVD-based rigid transformation, followed by centroid subtraction. This rigid step cannot remove articulated pose differences, which remain in the data.

To locate that residual pose variation, we measured per-vertex positional variance across the dataset (Fig. 2a, 2b). It is highest in pose-variant extremities (fingertips, toes), whereas measurement-dense regions (torso, proximal limbs) vary mainly with shape. A uniform per-vertex loss would be influenced by extremity pose noise, rather than focusing only on regions that drive measurement accuracy. This motivated our pose-robust losses, which spend most of capacity on shape instead of reproducing pose. We additionally down-weight high-variance vertices (inverse-variance weighting) inside Procrustes-aligned reconstruction loss to rebalance this; its effect is smaller and is reported in the ablations (Section V.B).

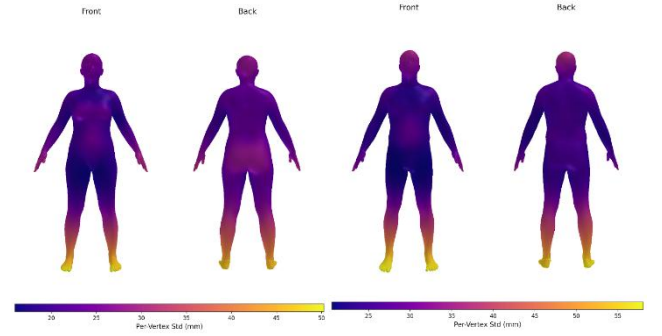


Fig. 2a (female). Fig. 2b (male). Per-vertex positional standard deviation across the dataset (separate for female, male), visualized on the template mesh. Yellow indicates high variance (pose-variable regions), blue indicates low variance. This distribution motivates pose-robust losses

V. RESULTS AND ANALYSIS

We evaluate geometric precision by comparing generated meshes with GT digital measurements. These metrics reflect how accurately the system reproduces the target geometry.

We discuss a separate challenge of calibrating digital measurements to physical tape readings in Section VI.E.

Reconstruction quality is evaluated using both global measurement accuracy and local geometric metrics. As standard vertex MAE is highly sensitive to pose noise, we rely on less pose-dependent metrics (edge length MAE and per-part volume error; Section IV.A) to evaluate surface geometry.

The dataset consists of real 3D scans, so it provides true human shape diversity. Our primary test set (406 male samples) is a random holdout split and is largely in-distribution, although it includes a long tail of rare body proportions. Approximately 7.4% of test samples contain at least one measurement outside the min–max range of that measurement type in the training set, serving as an initial built-in stress test for the model.

A. Baselines and Ablation Design

To isolate the contribution of each component, we compare the full model with several variants, including linear regression, a CVAE, and versions of our model with different input representations (standardized vs Fourier), output formulations (absolute vs offset), and training strategies (with/without scale augmentation, variance weighting, and alignment).

All models share identical preprocessing and mesh alignment (rigid ICP to a canonical template, then centroid subtraction; Section IV.C). The only exception is the unaligned baseline, which omits this alignment to assess its impact.

In Table II and Table IV all reported numbers use a single representative seed, for DRN we report results of the model nearest the 8-seed mean. Table III presents results averaged over 8 seeds (per sex).

B. Ablation Study Results Analysis

All ablations were run on the male split. As a baseline, we evaluate an Ordinary Least Squares (OLS) regression mapping measurements directly to vertex positions. Linear regression interpolates highly accurate shapes from familiar data, achieving a mean measurement error of 1.02 mm on our in-distribution test set. However, this precision comes without geometric regularization. As summarized in Table IV, OLS lacks structural constraints to prevent geometric collapse on anomalous inputs (which we detail in the OOD analysis section V.D).

Ridge regression ($\alpha = 1.0$) improves stability on OOD inputs by biasing predictions toward the mean shape, but at the cost of reduced accuracy (7.61 mm mean error and 89.19 mm worst-case error). This illustrates the primary challenge our pipeline addresses: while the measurement-to-shape mapping already captures about three-quarters, reliable deployment requires structural guides for inputs the model has not seen.

The CVAE baseline shares the DRN's FiLM decoder and conditions it on the same Fourier-mapped measurements; an encoder maps the mesh through residual blocks to the latent (dim 48). It differs from the DRN in its stochastic latent and a KL regularizer. Losses, optimizer, and training length are matched, and its latent dimension and KL weight were tuned on the validation set.

Measurement Accuracy

The proposed model achieves approximately 2 mm mean measurement error across all measurements, well below a single clothing size gradation step, indicating a level of precision suitable for digital fitting applications.

Effect of Deterministic vs Probabilistic Modeling

Even with this matched setup, the CVAE was less accurate (Table II): per-part volume error rose from ~1.8% to 2.8% and the worst single-measurement error more than doubled; 98.8% of its meshes had a measurement above 1 cm (59.6% above 2 cm), versus about a third of male meshes for the DRN (3.9% above 2 cm, none above 3 cm). In our runs the latent carried mostly residual pose regardless of loss settings, so fixing $z = 0$ at inference lost little and optimizing z per sample gave no precision gain. This gap traces to the probabilistic formulation: during training the decoder receives a KL-regularized sampled latent alongside the measurements, so it never learns to rely on measurements alone, which weakens measurement conditioning at inference.

Scale Augmentation

Removing scale augmentation significantly reduces robustness to extreme body proportions, increasing the worst individual measurement error from ~26 mm to ~190 mm. Visual inspection reveals localized surface distortions consistent with phase wrapping.

Measurement Representation

In our ablations replacing Fourier mapping with simple standardization slightly reduces reconstruction accuracy. While both standard and per-dimension Fourier mapping perform similarly in-distribution, the per-dimension variant is more stable when some measurements are strongly OOD or missing. Both standard Fourier mapping and standardized inputs underperform under such conditions and produce defective meshes even if scaling augmentation was active during the training. Per-dimension Fourier mapping improves stability for OOD inputs, although we must note that this approach is effective when combined with scaling augmentation only.

TABLE II.

ABLATION STUDY – COMPARISON OF MODEL ARCHITECTURES

Method	Meas. MAE (mm)	Worst ind. meas. (mm)	Edge length MAE (mm)	Per-part vol. error (%)	Notes
OLS	1.02	14.68	0.56	1.81	High in-distribution precision but lacks structural regularization. Leads to surface collapse and non-human topologies on OOD samples.
CVAE	4.87	88.09	0.66	2.82	Probabilistic latent raises measurement error and volume/worst-case error.
DRN (no scale aug.)	2.64	189.64	0.6	2.04	Limited robustness to OOD body proportions.
DRN	2.01	26.06	0.56	1.84	Provides good balance between numerical accuracy and anatomical plausibility. Maintains stable, human-like geometry under OOD stress.

We also tried normalizing each measurement to a fixed range (per-dimension min–max) before mapping; this markedly widened the train–validation gap and roughly doubled the measurements error, so we map measurements in real scale as does AnthroNet [13].

Minor Improvements

During our ablation tests several additional design choices provide smaller but consistent improvements. Predicting offsets instead of absolute coordinates slightly improves accuracy and reduces high-frequency noise in poorly constrained regions.

Variance-aware loss weighting and ICP alignment help to stabilize training by reducing the influence of pose variation and high-variance extremities. In our ablations without those stabilizations for both cases more than 45% of generated meshes exhibited at least one error exceeding 1 cm.

C. Results (in-distribution)

The male model was evaluated on 406 test samples, while the female model was evaluated on 1475. Both models show similar performance (Table III). The main in-distribution results are averaged over 8 random seeds.

Of the samples evaluated with the female-trained model, 22% ($\pm 4\%$) had at least one measurement error above 1 cm, and 2.8% ($\pm 3\%$) had an error above 2 cm. Worst average measurement MAE per sample was 5.05 mm for males and 6.15 mm for females.

TABLE III.

MODEL EVALUATION RESULTS ON IN-DISTRIBUTION INPUTS

Metric	Male	Female
Meas. MAE (mean, mm)	2.02 ± 0.10	1.89 ± 0.04
Meas. MAE (median, mm)	1.87 ± 0.10	1.76 ± 0.04
Edge length MAE (mm)	0.56 ± 0.00	0.5 ± 0.00
Per-part volume error: (%)	1.84 ± 0.02	1.8 ± 0.03

Worst single-measurement error for the male test set was ≈ 27 mm mean across 8 seeds; max 29 mm, performance is stable across seeds ($\sigma = 0.10$ mm on mean MAE). While the female test set exhibited a higher worst-case mean ≈ 33 mm and maximum error 37.22 mm (predominantly in waist circumferences and crotch arc length), analysis revealed that these outliers belonged to pregnant subjects. This minority group is present in both the training and test splits; its distinct abdominal geometry and high within-group variability likely make those the hardest cases to fit. However, even with these outliers, the maximum error across all test meshes remains within a single size gradation step. Maternity is a distinct sizing category that would benefit from a dedicated model, which we leave to future work.

The female model's slightly lower total error and slightly higher worst-case error are likely due to differences in test split size. Therefore, the error for the female test set may be more representative of overall pipeline performance, as this set has over three times as many samples. The qualitative results of model's evaluation are presented in Fig. 3.

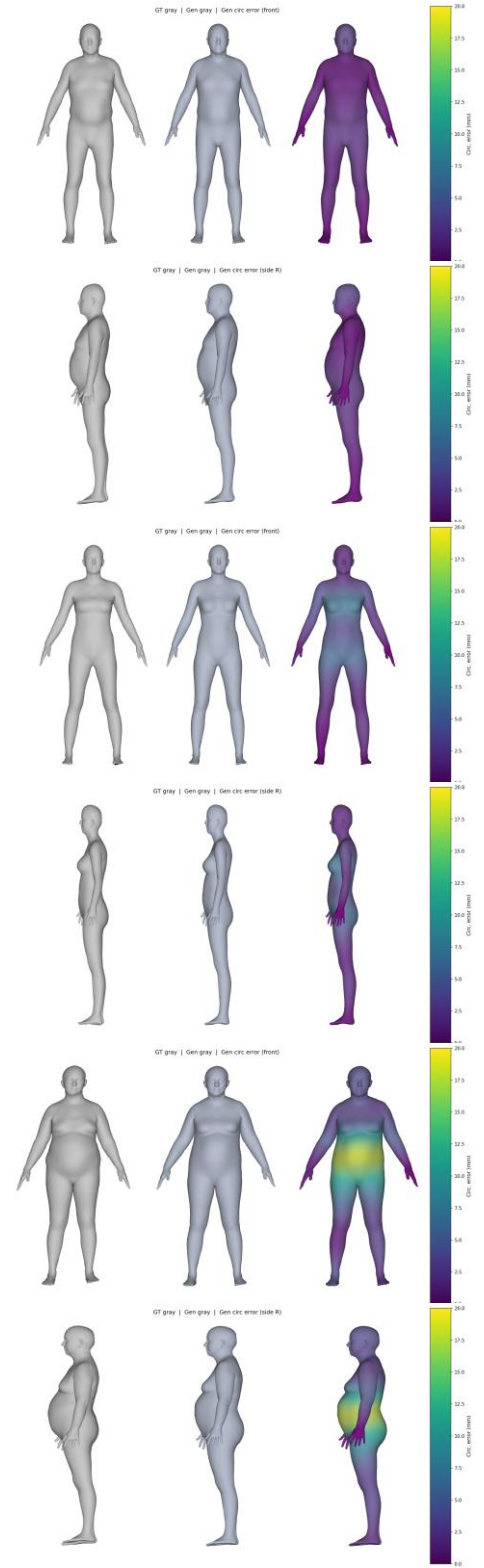


Fig. 3. Generated body meshes sorted by measurement MAE (mm): best – 1.05, median – 1.91, worst – 6.15. Localized circumference measurement errors overlaid as a heatmap. All error visualizations use a fixed 20 mm color scale ceiling, corresponding to our target precision threshold (half a clothing size gradation step); values above 20 mm saturate to the maximum color. The model outputs meshes in a canonical A-pose, creating minor visual discrepancies compared to the articulated pose variance of the original scans; the underlying geometric properties, such as volume and target measurements, remain consistent.

Error Distribution by Measurement Type

Mean absolute error varies by measurement type. The model is most accurate on Euclidean (1.28 ± 0.05 mm male / 1.13 ± 0.05 mm female); followed by geodesic lengths (1.57 ± 0.07 / 1.53 ± 0.02 mm) and circumferences (1.94 ± 0.12 / 1.92 ± 0.05 mm); vertical heights are highest (3.37 ± 0.19 / 2.65 ± 0.06 mm), with worst-case errors in total height and large circumferences.

While the model achieves high precision on complex surface circumferences, it shows slightly worse performance in vertical measurements (e.g., total height, ankle height). We attribute this primarily to articulated pose variations in the training data. Our heights are measured relative to an individual ground plane (e.g., the lowest vertical coordinate of the mesh) and are sensitive to stance, whereas our model predicts a canonical pose.

We suggest that for robust generation, height extraction should be independent of the surrounding space (even if the ground level is defined individually). Instead, heights should be redefined intrinsically as distances between anatomical keypoint chains, excluding stance variation, or extracted from a standing pose with feet placed close together.

Sensitivity to Conditional Input

To test how strongly the network responds to the input measurements, we perturbed them and measured the geometric impact. Masking (zeroing) the conditioning signal caused an average geometric deviation of 37.68 mm. Swapping it with another subject's measurements produced 18.24 mm deviation. These results indicate that the generation process is strongly driven by the measurement vector, with the network responding to input variations rather than relying on the initial base shape.

The smoothness of this mapping is further illustrated in Fig. 4, where linear interpolation between two measurement vectors produces geometrically coherent transitions.

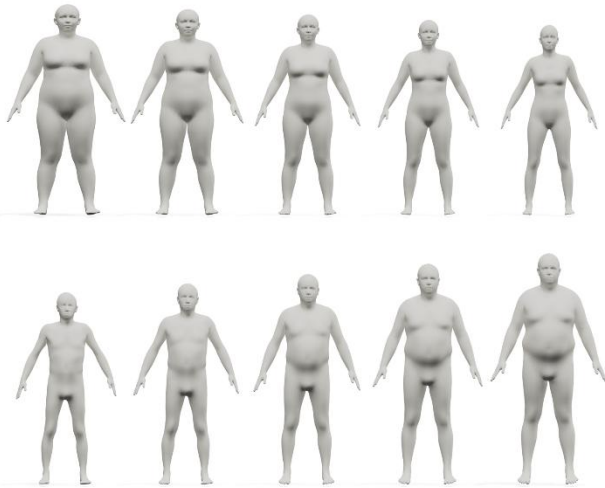


Fig. 4. Linear interpolation between measurement vectors produces smooth transitions, indicating a continuous mapping from measurements to shape.

D. OOD Analysis

To evaluate OOD robustness, we performed two primary stress tests:

1. Different subjects from AGORA dataset [20].
2. The same subject from AGORA dataset in different poses.

We note these AGORA samples differ from our training distribution along several confounded axes at once: registration process, pose, and measurement sliding, so this is a combined stress test rather than a controlled single factor shift.

Different subjects

Although the dataset contains posed meshes, we selected a small subset (152 meshes) of subjects as close to a neutral pose as possible, or at least standing on both feet, since most of our heights are measured from the individual ground level. We also excluded subjects crossing arms or holding them around the torso as this influences geodesic paths.

Dataset samples have the same SMPL topology (6890 vertices), so we extracted measurements from posed meshes using the same measurements definitions as we used for our training set. We acknowledge potential measurement sliding due to different template stretching and registration processes across datasets, as well as varying poses. Inconsistent poses introduce noise and more asymmetric measurements than our model observed during training.

When presented with measurement combinations outside the training distribution, such as these AGORA dataset samples, the linear baseline (OLS) produces geometrically degenerate meshes with squished limbs and non-human topologies. As the comparison below demonstrates (Fig. 5), while OLS minimizes local measurement error effectively, it lacks the structural priors for maintaining 3D surface integrity under OOD stress. OLS failures of this kind were consistent across dozens of OOD subjects; the inverse DRN failing while OLS did not - was not observed.

TABLE IV.

MODEL ROBUSTNESS ON OUT-OF-DISTRIBUTION INPUTS

Metric	OLS	DRN (ours)
Meas. MAE (mm)	4.16	5.1
Edge length MAE (mm)	1.17	0.99
Per-part volume error (%)	6.99	5.19
Per-part convex hull error (%)	8.14	6.53

To capture the gross shape distortions typical of OOD failures, we additionally report per-part convex hull error (defined in Section IV.A), as relying solely on measurement MAE is insufficient to evaluate quality. Table IV supports this: while OLS achieves lower measurement error, it does so by distorting the mesh (indicated by high convex hull and volume errors).

Under extrapolation, numerical accuracy alone is insufficient: in-distribution all methods including OLS preserve geometry (Table II), but on OOD inputs the unconstrained linear model distorts the surface (Table IV) while the structurally

regularized DRN maintains stable, human-like geometry (Fig. 5).

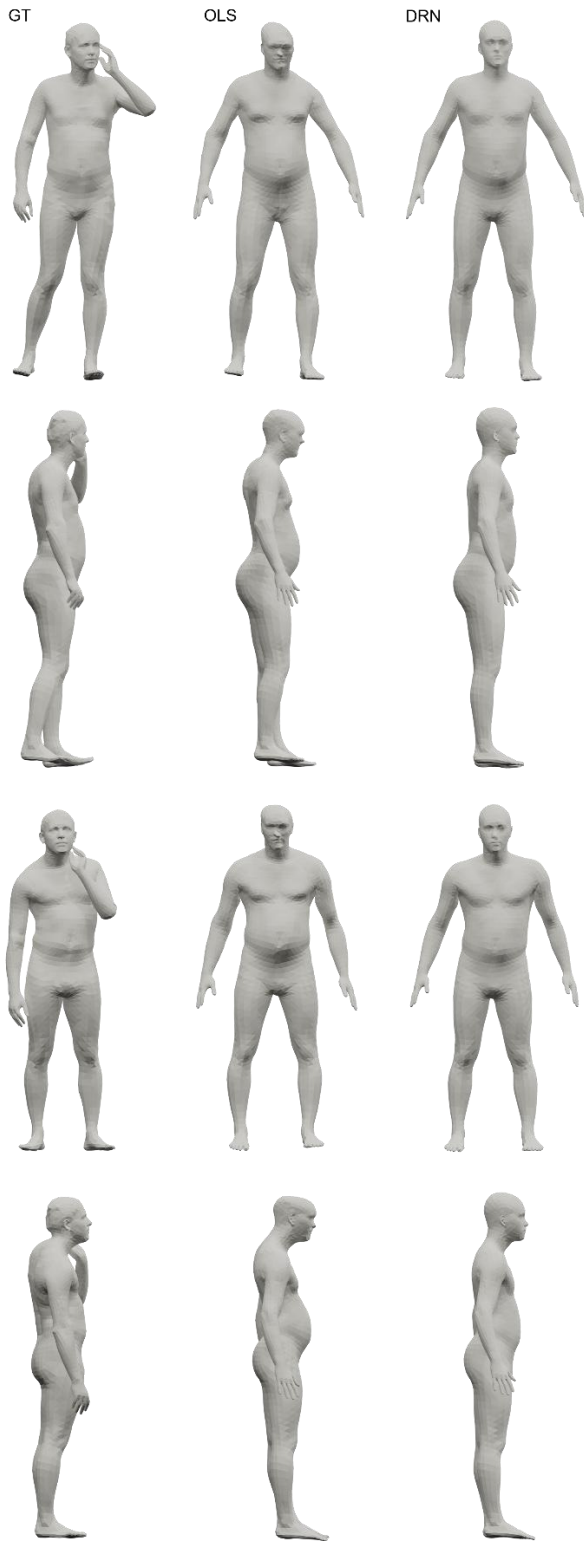


Fig. 5. Representative OOD failure cases from AGORA. Each row: ground truth (front or side), OLS output, DRN output. Cases were selected to show the qualitative gap: OLS produced geometric distortions of this kind across dozens of OOD subjects.

The contour overlays (Fig. 6) show that OLS and DRN often produce nearly identical silhouettes, yet OLS attains lower measurement MAE (Fig. 7) while underperforming on local geometric metrics (edge length MAE, per-part convex hull error).

This reflects how each handles the constraints: OLS minimizes measurement error directly, which can distort the local surface; DRN model preserves surface structure at the cost of slightly higher measurement error. This trade-off achieves a smooth, structurally stable, anatomically plausible surface, which is relevant to downstream physical simulations like digital garment draping.

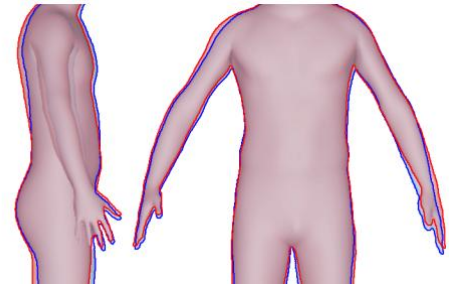


Fig. 6. Waist silhouette looks almost identical both for front and side views, while heatmap shows higher measurement MAE in waist region. Blue - OLS, Red - DRN.

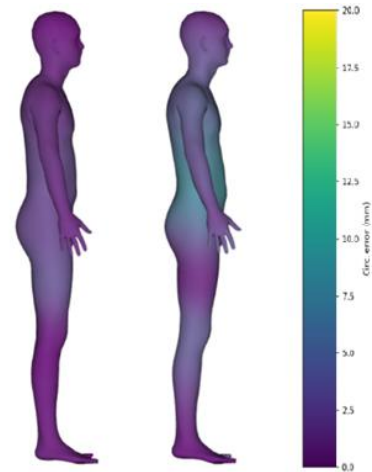


Fig. 7. On the left OLS generated mesh, on the right - DRN

Mesher generated from measurements of the same subject in different poses

We performed an evaluation test on the same subject, to compare the consistency of the output mesh (Fig. 8). As the results show, the OLS approach produces visibly distorted meshes, and inconsistencies in the input are amplified, while DRN produces visibly more consistent outputs across poses.

We note that variations in heights seen in picture are caused by inconsistent pose as height is measured between individual ground plane and specified points on the mesh (head top - for total height, waist point for waist level).

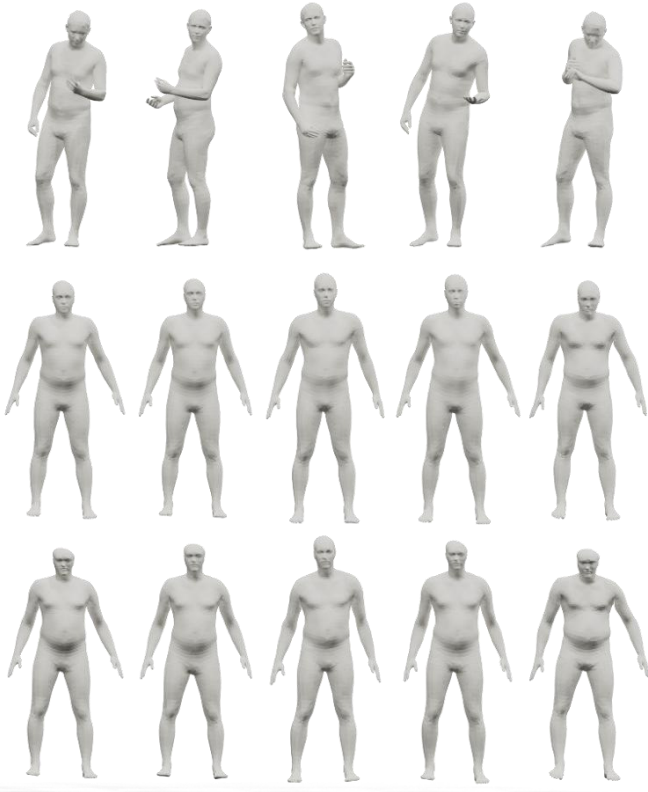


Fig. 8. Top row GT meshes, middle - DRN (ours) generated, bottom - OLS. Despite pose differences, DRN model better preserves overall geometry and measurement consistency.

In deployment, the system assumes validated measurement inputs. Nevertheless, user-provided measurements may still form unseen combinations. The proposed design improves robustness in such cases, even if in-distribution accuracy remains slightly below the linear baseline.

VI. DISCUSSION

A. Precision vs. dataset size

The results suggest that careful pipeline design (variance reduction, augmentation and architectural choices) can partially compensate for limited training data. With ~ 400 training meshes per sex, we achieve a mean measurement error of approximately 2 mm for males and 1.9 mm for females. The worst-case errors remain within a single size gradation step. This is achieved with substantially less training data than several prior methods.

B. Context from prior work

We cannot easily run a one-to-one comparison against previous methods due to differing goals, evaluation protocols, and measurement definitions, alongside the varying use of real versus synthetic datasets. Reported measurement-level errors vary widely:

For example, AnthroNet [13] uses a fully synthetic pipeline. While they achieve low overall vertex errors by mapping

digitally extracted measurements to synthetic meshes, they report large deviations in key circumferences on synthetic and real data. Working with real scan data introduces natural anatomical variance. Zeng et al. [15] train on real 3D scans paired with 19 digital surface measurements, reporting an 8.3 mm average error. Thach and Dat [14] map 10 physical tape measurements to real scans from the CAESAR [19] dataset, achieving 2.6-3.8 mm errors through iterative optimization. In this context, our approach achieves precision (~ 2 mm) by mapping a dense set of 47 digital surface measurements to real 3D scans, while maintaining the rapid processing required for real-time digital fitting.

C. On disentanglement under dense conditioning

Measurement-level disentanglement (the ability to change one measurement while holding all others fixed) is not well-defined under dense conditioning. When many measurements are highly interdependent, modifying a single value (e.g., waist circumference) without adjusting correlated measurements (e.g., hips, crotch arc length) often leads to implausible geometry.

For digital fitting, precision with correlated control is preferable to independent control with lower geometric fidelity.

D. Limitations

We use a dataset with SMPL topology for current evaluation; generalization to other topologies requires further validation.

Extreme OOD inputs may still produce lower quality meshes despite the proposed defect mitigation strategies.

Anisotropic augmentation relies on axis-aligned scaling, which is a simplification for diagonally oriented limbs.

E. Bridging Synthetic and Real-World Measurements

For real-world deployment, the digital measurements used to condition the model must correspond to physical tape measurements obtained from a real subject. Our measurements are defined on the mesh surface (geodesic paths and landmark distances) and may differ systematically from manual measurements.

We can bridge this gap using the validation workflows outlined by Bicevskis et al. [1] and Neimanis et al. [2]. It enables consistency checks and missing values prediction for input measurements. In practice, this can be used to validate user inputs before reconstruction and to verify generated measurements after mesh generation.

F. Future Work

We plan to:

1. extend to a custom mesh topology with additional landmark-based measurements and non-geometric inputs (weight, age) to capture body composition variation;
2. train on a higher-precision dataset currently being acquired;
3. conduct paired digital-vs-physical measurement validation studies;

4. evaluate cross-dataset transfer from SMPL-registered meshes to the target topology.

VII. CONCLUSIONS

Dense measurement conditioning provides enough geometric constraint for a deterministic regressor to reach within-size-gradation accuracy, the relevant bar for fitting. By relying on a comprehensive set of 47 measurements, this approach achieves precision suitable for tailoring without the variability introduced by generative latent spaces.

By combining per-dimension Fourier mapping with scale augmentations and pose-robust losses we improved stability and generalization. Even with a limited dataset, the system achieves an error of roughly 2 mm.

Despite low measurements error, we conclude that in the context of digital tailoring, for robust deployment optimizing for numerical precision alone is an insufficient objective. By enforcing structural constraints, our system provides balance between measurement accuracy, geometric plausibility, and OOD stability.

While dense conditioning limits independent control of individual measurements, it provides consistent and anatomically plausible outputs, which are more relevant for practical fitting applications.

GENERATIVE AI DISCLOSURE STATEMENT

During the preparation of this work the authors used Claude (Anthropic) to assist with language editing and style polishing of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

- [1] J. Bicevskis, Z. Bicevska, E. Diebelis, L. Purina, "Quality control of body measurement data using linear regression methods," in Proc. *19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, ACSIS, vol. 39, pp. 289–300, 2024, doi:10.15439/2024F6463.
- [2] A. Neimanis, L. Purina, Z. Bicevska, E. Diebelis, J. Bicevskis, "Application of statistical methods in the product development of customized clothing," in *Information Technology for Management: Intelligent Alignment of IT with Business and Society*, pp.105-125, Springer Nature, Switzerland, 2026, doi:10.1007/978-3-031-93580-0_5.
- [3] European Committee for Standardization, "Size designation of clothes - Part 3: Size labelling based on body measurements and intervals, ISO 8559-3:2018," *International Organization for Standardization*, 2018.
- [4] A. A. A. Osman, T. Bolkart, M. J. Black, "STAR: Sparse trained articulated human body regressor," in Proc. *Computer Vision - ECCV 2020*, Proceedings, Part VI, pp. 598-613, 2020, doi:10.1007/978-3-030-58539-6_36.
- [5] M. Tancik et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in Proc. *The 34th International Conference on Neural Information Processing Systems (NIPS '20)*, Red Hook, USA, art. 632, pp. 7537–7547, 2020, doi:10.48550/arXiv.2006.10739.
- [6] E. Perez, F. Strub, H. de Vries, V. Dumoulin, A. Courville, "FiLM: Visual reasoning with a general conditioning layer," in Proc. *The Thirty-Second AAAI conference on artificial intelligence (AAAI-18)*, art. no.: 483, pp. 3942-3951, 2018, doi:10.1609/aaai.v32i1.11671.
- [7] D. Anguelov et al., "SCAPE: Shape completion and animation of people," in *ACM Transactions on Graphics* vol. 24 no.3, pp. 408-416, 2005, doi:10.1145/1073204.1073207.
- [8] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. J. Black, "SMPL: A skinned multi-person linear model," in *ACM Transactions on Graphics*, vol. 34, no. 6, art.no. 248, pp. 1-16, 2015, doi:10.1145/2816795.2818013.
- [9] H. Xu et al., "GHUM&GHUML: Generative 3D human shape and articulated pose models," in Proc. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6184-6193, 2020, doi:10.1109/CVPR42600.2020.00622.
- [10] S. Pujades et al., "The Virtual Caliper: Rapid creation of metrically accurate avatars from 3D measurements," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1887–1897, 2019, doi:10.1109/TVCG.2019.2898748.
- [11] K. Ludwig, J. Lorenz, D. Kienzle, T. Bui, R. Lienhart, "Leveraging anthropometric measurements to improve human mesh estimation and ensure consistent body shapes," in Proc. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5862-5871, 2025, doi:10.1109/CVPRW67362.2025.00585.
- [12] S. Park et al., "ATLAS: Decoupling skeletal and shape parameters for expressive parametric human modeling," in Proc. *IEEE/CVF International Conference on Computer Vision*, pp. 6508-6518, 2025, doi:10.1109/ICCV51701.2025.00614.
- [13] F. Picetti et al., "AnthroNet: Conditional generation of humans via anthropometrics," arXiv preprint arXiv:2309.03812, 2023, doi:10.48550/arXiv.2309.03812.
- [14] H. N. Thach, N. T. Dat, "3D reconstruction human body from anthropometric measurements using DCGA," in *MENDEL*, vol. 27, no. 1, pp. 49–57, 2021, doi:10.13164/mendel.2021.1.049.
- [15] Y. Zeng, J. Fu, H. Chao, "3D human body reshaping with anthropometric modeling," in Proc. *Internet Multimedia Computing and Service, ICIMCS 2017, Communications in Computer and Information Science*, vol. 819, pp. 96-107, Springer, Singapore, 2018, doi:10.1007/978-981-10-8530-7_10.
- [16] O. Sorkine et al., "Laplacian surface editing," in Proc. *Eurographics Symposium on Geometry Processing*, pp. 175-184, 2004, doi:10.1145/1057432.1057456.
- [17] S. Yan, J. Wirta, J.-K. Kamarainen, "Silhouette body measurement benchmarks," in Proc. *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, pp. 7804-7809, 2021, doi:10.1109/ICPR48806.2021.9412708.
- [18] D. Škorvanková, A. Riečický and M. Madaras, "Automatic estimation of anthropometric human body measurements," arXiv preprint, arXiv:2112.11992, 2021, doi:10.48550/arXiv.2112.11992.
- [19] K. M. Robinette, H. Daanen and E. Paquet, "The CAESAR project: A 3-D surface anthropometry survey," in Proc. *Second International Conference on 3-D Digital Imaging and Modeling (Cat. no.PR00062)*, Ottawa, Canada, pp. 380-386, 1999, doi:10.1109/IM.1999.805368.
- [20] P. Patel et al., "AGORA: Avatars in geography optimized for regression analysis," in Proc. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13463-13473, 2021, doi:10.1109/CVPR46437.2021.01326.