ATLAS: Decoupling Skeletal and Shape Parameters for Expressive Parametric Human Modeling

Jinhyung Park[1,2] Javier Romero[1] Shunsuke Saito[1] Fabian Prada[1] Takaaki Shiratori[1]
Yichen Xu[1] Federica Bogo[1] Shoou-I Yu[1] Kris Kitani[1,2] Rawal Khirodkar[1]
[1]Meta [2]Carnegie Mellon University
https://jindapark.github.io/projects/atlas

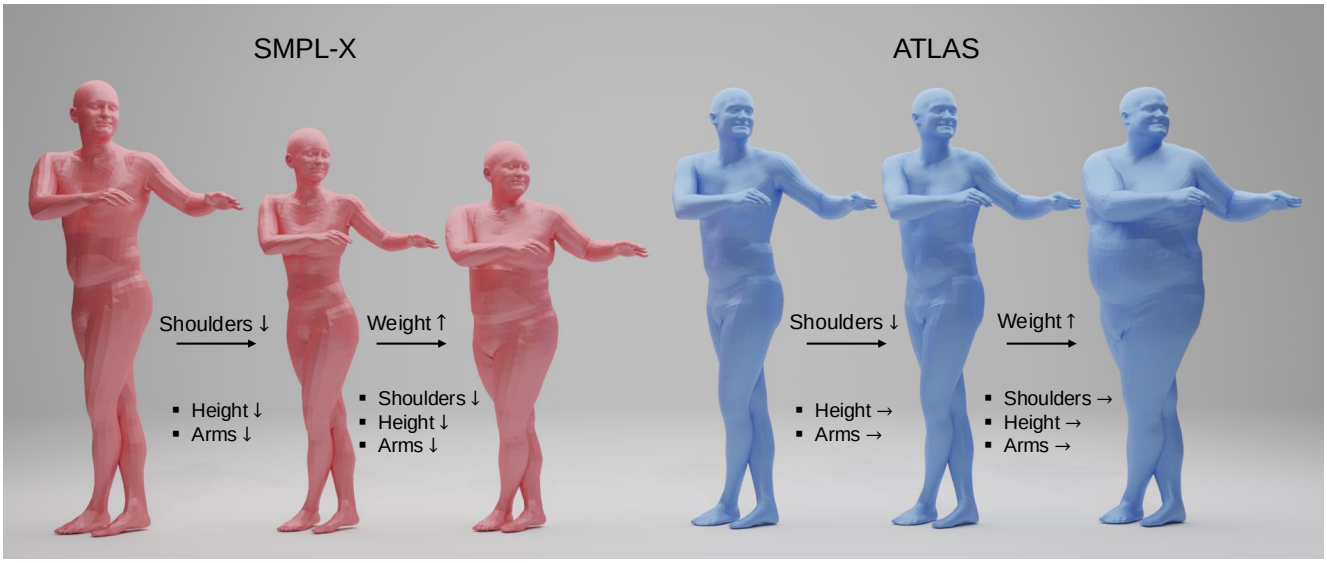

Figure 1. **ATLAS enables precise, decoupled control of skeletal and surface attributes.** Here, we customize a mesh to reduce shoulder width and increase body weight. This level of control is difficult to accomplish in prior work [39] due to undesirable correlations between joints and vertices, *e.g.* adjusting shoulder width affects the entire body and increasing weight reverses the shoulder adjustment. With ATLAS's decoupled skeleton and shape, this customization is a simple two-step deterministic edit.

Abstract

Parametric body models offer expressive 3D representation of humans across a wide range of poses, shapes, and facial expressions, typically derived by learning a basis over registered 3D meshes. However, existing human mesh modeling approaches struggle to capture detailed variations across diverse body poses and shapes, largely due to limited training data diversity and restrictive modeling assumptions. Moreover, the common paradigm first optimizes the external body surface using a linear basis, then regresses internal skeletal joints from surface vertices. This approach introduces problematic dependencies between internal skeleton and outer soft tissue, limiting direct control over body height and bone lengths. To address these issues, we present ATLAS, a high-fidelity body model learned from 600k high-resolution scans captured using 240 synchronized cameras. Unlike previous methods, we explicitly decouple the shape and skeleton bases by grounding our mesh representation in the human skeleton. This decoupling enables enhanced shape expressivity, fine-grained customization of body attributes, and keypoint fitting independent of external soft-tissue characteristics. ATLAS outperforms existing methods by fitting unseen subjects in diverse poses more accurately, and quantitative evaluations show that our non-linear pose correctives more effectively capture complex poses compared to linear models.

"*ATLAS*—a structure bearing human form."

1. Introduction

Recent years have seen significant advancements in human-centric applications, including 3D digitization of avatars for virtual reality [32, 51, 56, 58], efficient and performant motion capture [11, 41, 61, 62], physically plausible human-object interaction [6, 9, 55], and generative human character generation [31, 40, 45]. Supporting these methods are parametric models of the human body [5, 33, 39, 54, 59, 60] —

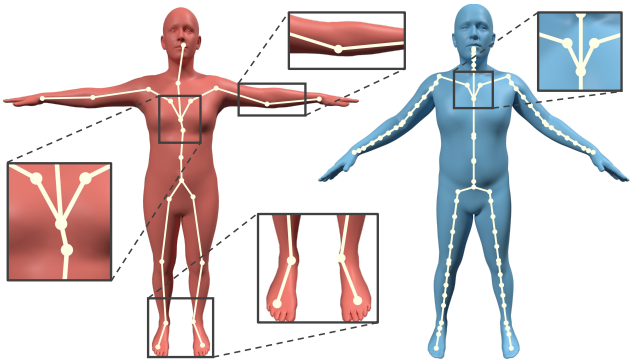

Figure 2. **Comparison of Skeleton Symmetries.** (*Left*) The mean SMPL-X mesh reveals significant skeletal asymmetries (elbows, spine, and feet) due to joint centers being derived from vertices. (*Right*) ATLAS mesh demonstrates a symmetric and consistent skeleton through decoupling of skeletal and surface parameters.

methods that derive diverse, articulated human meshes from low-dimensional shape and pose parameters. Expressive and controllable parametric human models are thus critical for advancing the broad field of human understanding.

The dominant approaches for parametric modeling of the human body follow a vertex-centric framework [33, 37–39, 59] where surface vertices are personalized through a linear basis, internal skeletal joints are derived from the surface through weighted sum, and the mesh is driven with linear blend skinning (LBS) [21] and pose dependent corrections. While achieving plausible 3D reconstruction, this paradigm presents several inherent limitations. First, deriving internal skeletal joints from surface vertices introduces incorrect correlations. Shown in Figure 2, the skeletal joints in SMPL-X [39] are asymmetrical, and the spine shifts left-to-right with changes in the second shape component which is associated with soft tissue variation. Second, skeletal attributes can only be modified by altering shape components, which inevitably affects other surface vertex attributes. For instance, shoulder width is intertwined with several components in SMPL-X [39] that affect soft tissue, inhibiting precise customization of internal attributes (refer Figure 1). Third, this correlation causes keypoint fitting to produce meshes with unwarranted soft-tissue deviations, although keypoints provide no information about these attributes.

To address these issues, we propose **ATLAS**, an expressive parametric model of the human body which explicitly decouples external shape and internal skeleton. Our model is natively trained at high resolution (115k vertices) and has an anatomically motivated skeleton with 77 joints. To drive ATLAS, we start with a template, unposed mesh and customize soft-tissue attributes (*e.g.* torso and leg volume *etc.*) with a linear basis over shape. At this stage, the skeletal joints remain unchanged. We use a skeletal basis to customize the internal skeleton, and then we scale and pose the mesh together with LBS [21]. By explicitly decoupling external shape and internal skeleton, ATLAS eliminates spurious vertex-joint correlations and enables more precise controllability of the human mesh as shown in Figure 1. Further, to improve the realism of the skinned mesh, we introduce sparse, non-linear pose corrective deformations prior to the LBS phase. The sparsity of this mapping prevents the pose correctives from fitting to spurious correlations in the data, such as actuation of one elbow affecting vertices of the other, and the non-linearity enables more accurate deformations around difficult joints (shoulders, elbow tips, *etc.*).

ATLAS is trained on a large-scale dataset of 600k high-resolution scans of minimally clothed subjects in diverse poses. Compared to prior work as shown in Table 1, our model is developed from a more diverse set of shapes, identities, and poses, resulting in a more expressive human body model. Additionally, for in-the-wild usage, we develop a single image model fitting pipeline. Our framework leverages ATLAS's decoupling of the shape and skeleton, as well as recent advancements in high-fidelity human-centric models [24], to first fit the skeleton and pose to keypoints, then personalize body shape to fit the human silhouette. Supporting the fitting is a VAE pose prior [25, 39] trained on 600k frames as well as a PCA prior over hand poses. Our evaluations show that the resulting pipeline better fits poses and derives more plausible shape compared to existing methods. To evaluate our model, we provide quantitative results on fitting a diverse dataset of body shapes and poses [49] and demonstrate that ATLAS is more expressive than existing state-of-the-art parametric body models.

Our contributions are summarized as follows.

- We propose ATLAS, a controllable and expressive parametric human body model with separate bases for external shape and internal skeleton.
- Our model, with sparse, non-linear pose correctives, demonstrates more expressivity in representing shaped, articulated human scans.
- We leverage our model to build a high-fidelity single RGB image to model parameter pipeline that captures diverse poses, body shapes, and expressions.

Method	Shape IDs	Pose IDs	Scans	Shape Basis	Skeletal Basis
SMPL [33]	3.8k	40	1.8k	✓	✗
SMPL-X [39]	3.8k	40	1.8k	✓	✗
STAR [38]	13k	40	1.8k	✓	✗
BLSM [54]	3.8k	10	41k	✓	✓
GHUM [59]	4.3k	48	60k	✓	✗
OSSO [22]	2k	-	2k	✓	✓
SUPR [37]	15k	-	1.2m†	✓	✗
BOSS [53]	300	-	300	✓	✓
SKEL [23]	3.9k	113	1m†	✓	✓
ATLAS (Ours)	15k	157	600k	✓	✓

Table 1. **Comparison of training data across state-of-the-art parametric models.** ATLAS leverages diverse registrations across shape and pose at large scale. †SUPR [37] and SKEL [23] use 60Hz and 30Hz scans while ATLAS uses 5Hz scans.

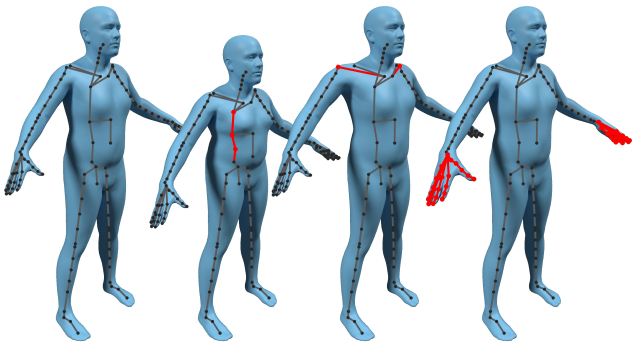

Figure 3. **Controllable Skeletal Attributes.** From left to right, we visualize the template mesh, decreasing spine length, increasing shoulder width, and increasing scale of both hands.

2. Related Work

3D Human Mesh Modeling. The early work SCAPE [4] separately models pose and shape changes with triangle deformations. Follow-up works refine or constrain the deformations, improve registrations, or apply it to soft-tissue dynamics [10, 13, 16, 18, 43]. SMPL [33] proposes a vertex-based model [3] with shape and pose corrective blendshapes and uses LBS to pose the mesh around joints regressed from vertices. STAR [38] compacts SMPL by using quaternions and sparsifies the corrective matrix. Frank [19] adds the FaceWarehouse [8] face model and an artist-designed hand model to the SMPL rig. SMPL-H [48] adds MANO, a hand model learned from scans, and SMPL-X [39] merges MANO, FLAME, and SMPL to learn a shape space for the entire body. SUPR [37] improves on SMPL-X with using a federated dataset of body parts and a better foot model. GHUM [59] proposes a non-linear shape space and pose correctives. Following the SMPL framework of regressing joints from surface vertices, these works introduce suboptimal correlations between external shape and internal skeleton. In contrast, ATLAS explicitly learns a separate skeleton space for better controllability and shape expressivity.

Skeleton Models. Many works in biomechanics define precise, anatomically accurate skeleton models. Some works [36, 44, 52] derive musculoskeletal models for modeling anatomically plausible movement. Other methods [12, 15, 20, 50, 64] optimize for underlying fat, muscle, and bone that drive surface deformations. While anatomically constrained, these methods rely on specialized simulators [28] or use models for bone or fat growth, making their usage in commercial graphics packages difficult. OSSO [22] and BOSS [53] derive the anatomical skeleton from SMPL meshes from medical segmentation masks, but driving them to new poses requires additional optimization.

Most related to our work are methods that introduce decoupled skeleton-driven human mesh models for graphics and commercial applications. An early work [17] optimizes internal bones for both pose changes and shape variation. In contrast, we maintain a separate space over vertices to represent shape, better modeling soft tissue deformations. BLSM [54] proposes a decoupled shape and bone-length space. While promising, the model is not open and lacks pose corrective deformations which limits realism. SKEL [23] places bony and soft markers on AMASS [34] using OSSO [22] and uses AddBiomechanics [57] to optimize for an internal skeleton. SKEL then learns a mapping from vertices joints and re-rigs the SMPL mesh. While effective in deriving biomechanical skeletons from SMPL meshes, SKEL inherits SMPL's surface vertex-based shape space to synthesize new human meshes. Skin generation for a specified skeleton requires optimization to find matching SMPL shape parameters, and it requires that the desired skeleton is represented in the SMPL shape space. Further, SKEL lacks finger control and inherits a limited set of pose correctives from SMPL. In contrast, our ATLAS enables direct controllability of decoupled shape and scale parameters, includes finger actuation, and provides expressive pose correctives.

Pose Corrective Deformations. To model pose-dependent deformations, Lewis [29] applies vertex offsets around joints to alleviate "collapsing joint" defects. Other works [2, 27, 46] interpolate between saved deformations for key poses and EigenSkin [26] adds a PCA space for each joint. SMPL [33] learns a mapping from joint rotations to vertex deformations, and STAR [38] sparsifies this mapping through geodesic initialization and regularization. GHUM [59] improves the expressiveness of this mapping through a non-linear network. However, by mapping the full body pose to a compressed 32-dim intermediate latent vector, GHUM's pose correctives remain dense. Our ATLAS seeks to achieve the best of both worlds through a sparse and non-linear mapping, avoiding spurious correlations while maintaining the expressiveness of non-linear correctives.

3. Method

In this section, we present the ATLAS model, detail its controllable skeletal attributes, describe our sparse non-linear pose-correctives, and present a single-image fitting method.

3.1. Decoupled Skeletal and Shape Body Model

Overview. The core strength of ATLAS lies in its explicit separation of the external surface from the internal skeleton. This characteristic enables precise, independent customization of surface and skeletal attributes. To support this capability, ATLAS derives a shaped and posed human mesh in two steps. **First**, surface vertices are customized while aligned to a fixed template skeleton. **Second**, this mesh is simultaneously scaled and posed using LBS [21], modifying the underlying skeleton using 76 individually controllable skeletal attributes that each control bone lengths and body part sizes (Section 3.2). The resulting mesh accurately captures subtle variations of the human shape.

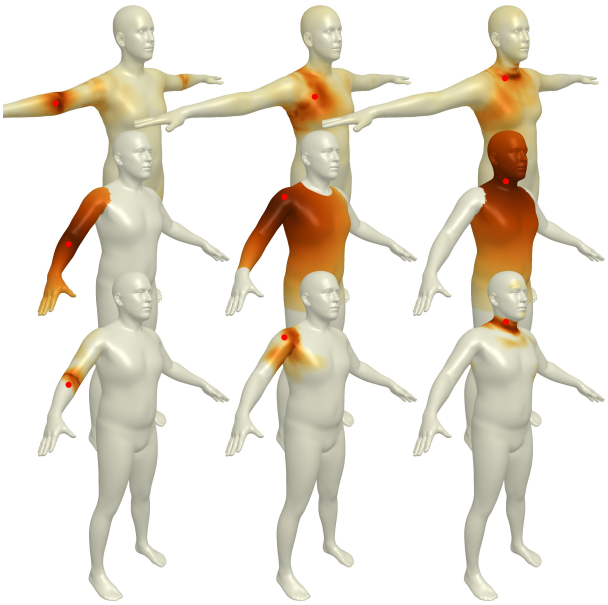

Figure 4. **Sparse Pose Correctives.** The first row displays pose correctives from SMPL-X. The second row shows the inverse geodesic initialization for our pose corrective activations, and the third row demonstrates their sparsity after convergence.

Surface Customization. To shape and pose an ATLAS mesh, we first obtain customized surface vertices $\tilde{X}$ aligned to the template skeleton in the default A-pose:

$$\tilde{X}(\beta^s, \beta^f, \theta) = \bar{X} + \mathcal{B}^s(\beta^s, \mathcal{S}) + \mathcal{B}^f(\beta^f, \mathcal{F}) + \mathcal{B}^p(\theta, \mathcal{P}) \tag{1}$$

where $\bar{X} \in \mathbb{R}^{3V}$ is the template A-pose shape, $\mathcal{B}^s(\beta^s, \mathcal{S}) = \sum_{n=1}^{|\beta^s|} \beta_n^s \mathcal{S}_n$ is the surface vertices' blend shape function, and $\mathcal{B}^f(\beta^f, \mathcal{F}) = \sum_{n=1}^{|\beta^f|} \beta_n^f \mathcal{F}_n$ is the facial expressions blend shape function. To correct artifacts caused by LBS [21], we add pose deformations $\mathcal{B}^p(\theta, \mathcal{P})$. Unlike prior works [33, 39] that derive joint centers from this customized identity shape, our mesh at this stage remains unposed, *unscaled*, and aligned to a fixed internal skeleton.

Skeleton Customization. Next, $\tilde{X}$ is both posed and scaled through LBS [21]. During this, the skeleton is customized using $N_k = 76$ controllable attributes: 15 modify body part sizes and 61 adjust bone lengths (Sec. 3.2 and Fig. 3). We denote these as $\ell = \sigma \oplus t$ where $\ell \in \mathbb{R}^{N_k}$ consists of scale $\sigma \in \mathbb{R}^{15}$ and $t \in \mathbb{R}^{61}$ bone length modifications. While these attributes can be set individually, we also learn a blend shape function over them $\mathcal{B}^k(\beta^k) = \sum_{n=1}^{|\beta^k|} \beta_n^k \mathcal{K}_n$ with $\mathcal{K}_n \in \mathbb{R}^{N_k}$ to capture common variations.

ATLAS is driven by Euler 3DoF poses $\theta \in \mathcal{R}^{3(J+1)}$ and skin weights $\omega \in \mathbb{R}^{V \times I}$, where each vertex is affected by up to $I = 8$ joints. Altogether, the surface vertices $\tilde{X}$ are then driven by the modified skeleton and the pose using the LBS function M:

$$X(\beta, \theta) = M(\tilde{X}(\beta^s, \beta^f, \theta), \mathcal{B}^k(\beta^k), \theta, \omega) \tag{2}$$

We emphasize that in ATLAS, the joint locations used for posing are *independent* of the vertex components β^s, which are only used for specifying the A-pose shape $\tilde{X}$. Rather, only the skeletal components β^k and the pose θ specify the joint locations, enabling precise decoupling of the surface and the skeleton. We refer to Section C of the supplement for a detailed mathematical formulation of M.

3.2. Controllable Skeletal Attributes

ATLAS incorporates $N_k = 76$ skeletal attributes, including 15 scale modifications that directly alter the overall size of body parts (body size, head, hands, feet, and individual fingers) and 61 bone length parameters that adjust joint translations relative to their kinematic parents. These bone length attributes encompass major body bones including spine, neck, upper and lower arms, upper and lower legs, and fingers. Figure 3 demonstrates some examples of individual scale attribute control, and visualizations of their effects are in Figure 12 of the supplementary.

3.3. Sparse, Non-Linear Pose Correctives

Overview. Pose-dependent deformations before LBS [21] are crucial for realistic meshes. Prior work highlight the benefits of both sparse-linear correctives [37, 38] – which restrict joint influences to local vertices to avoid spurious correlations – and dense non-linear correctives [59], where each joint non-linearly contributes to all vertices. We reconcile these methods by introducing sparse non-linear correctives that leverage the locality of sparse-linear approaches while preserving the expressivity of non-linear correctives.

Pose Correctives Formulation. Our correctives function $\mathcal{B}^p(\theta, \mathcal{P}) \in \mathbb{R}^{6J} \rightarrow \mathbb{R}^{3V}$ takes joint angles in 6D [63] and outputs vertex offsets. We decompose $\mathcal{B}^p$ into a local, non-linear operation and a sparse, geodesic-initialized linear operation. The former encodes local joint groups together, effectively enabling non-linear expressivity, while the latter constrains the extent of vertices each joint group can affect, avoiding spurious joint-vertex correlations.

First, the local, non-linear operation processes pose angles of joint j and those of its immediate immediate kinematic neighbors $n(j)$ using a lightweight MLP:

$$\text{Non-Linear}_j(\theta) = \text{MLP}\left(\{R_{6d}(\theta_a) - R_{6d}(\vec{0}) \mid a \in n(j)\}\right), \tag{3}$$

yielding a c-dimensional feature that encodes their poses.

This local feature is then transformed into pose corrective vertex offsets around j using a learned mapping:

$$\mathcal{B}_j^p = \phi(A_j) \odot (P_j \times \text{Non-Linear}_j(\theta)) \tag{4}$$

Here, $P_j \in \mathbb{R}^{3V \times c}$ is the pose corrective weight matrix, and the multiplication $P_j \times \text{Non-Linear}_j(\theta)$ produces the raw pose-dependent vertex offsets. Following STAR [38], the function ϕ is a ReLU applied to the joint mask $A_j \in \mathbb{R}^V$

to enforce vertex sparsity per joint. For vertex i, we initialize the i-th element of A_j as $(1 - d(i,j))\mathbf{1}_{i \in \text{seg}(j)}$, where $d(i,j)$ is the normalized geodesic distance from vertex i to the vertex ring around j, and $\mathbf{1}_{i \in \text{seg}(j)}$ indicates if vertex i belongs to joint j's corresponding or adjacent body part. This initialization, coupled with L1 regularization on $\phi(A)$, encourages sparsity in activation. Figure 4 shows the activation mask pre- and post-training, showing pose correctives concentrated around the actuated joint. Altogether, our pose correctives integrate the expressivity of non-linear functions with the sparsity of regularized linear correctives.

3.4. Single-Image Mesh Fitting Application

We demonstrate the applicability of ATLAS to real-world images by developing a single-image mesh fitting pipeline.

Objective. Our framework improves upon previous approaches [7, 39] by explicitly decoupling skeleton and shape fitting while leveraging predictions from high-fidelity human-centric models [24]. We optimize shape, skeleton, and pose parameters using the objective [7, 39]:

$$E(\beta^s, \beta^f, \beta^k, \theta) = E_{data} + E_{\theta_{body}} + E_{\theta_{hand}} + E_{\beta^s} + E_{\beta f} + E_{\beta^k} \quad (5)$$

E_{data} comprises of three components: $E_{kps2d} + E_{depth} + E_{mask}$. E_{kps2d} minimizes the distance between projected 3D ATLAS keypoints and 2D detector keypoints using a robust loss [14]. E_{depth} minimizes differences between rendered mesh depth and Sapiens [24] relative depth predictions. E_{mask} uses efficient Edge Gradients [42] to minimize differences between rendered and predicted foreground masks. For pose regularization, we train a VAE prior on full-body poses (excluding hands) and optimize in the latent space with L2 prior $E_{\theta_{body}}$. Hand poses are optimized in PCA 6D [63] space with L2 prior $E_{\theta_{hand}}$. Similarly, we apply L2 regularization to shape, expression, and skeleton attribute latents via E_{β^s}, $E_{\beta f}$, and E_{β^k}.

Optimization. We use multi-stage optimization, first fitting major body keypoints followed by hands and expressions. We explicitly decouple skeleton and shape parameter optimization: skeleton latents β^k are optimized using only pose and skeletal structure of the subject, $E_{kps2d} + E_{depth}$, while surface shape latents β^s are optimized using the mask term E_{mask}. This separation enables clean optimization of pose and body structure through keypoint and depth terms, while accurately capturing soft tissue variations through mask fitting. Unlike prior work [39] that entangles skeleton and shape, our approach prevents keypoint-induced soft tissue hallucinations while fitting subject silhouettes through shape variation. For facial expression fitting, we introduce an improved approach that first aligns projected 3D expression keypoints with target 2D keypoints using optimal rotation and translation, then minimizes their difference using a robust loss term. This enables realistic expression capture even with head misalignment.

4. Experiments

In this section, we first describe ATLAS training, followed by an extensive comparison with state-of-the-art body models across multiple scenarios. Lastly, we provide insights into the importance of pose correctives in ATLAS and demonstrate its fitting to in-the-wild images.

4.1. Training ATLAS

Goliath Dataset. We collect a dataset of $600K$ high-resolution scans from 130 subjects in dynamic poses, named *Goliath* for its scale. These scans are captured with a calibrated and synchronized multi-view camera system with 240 cameras at 4K resolution. Figure 7 shows images of participating subjects in our capture setup. Notably, this dataset is substantially larger than existing datasets; *e.g.* SMPL [33] consists of $1.2K$ scans from 27 subjects. Our dataset's scale enables learning a more generalizable human body model. Additionally, following [39], we also use existing datasets to train ATLAS, including CAESAR [47] and SizeUSA [1] processed by Meshcapade, consisting of 4391 and 10123 scans respectively. This dataset captures diverse body shapes, represent a broader section of the population (aged 18 to 65+), and complement our captured data.

Body Model. We design ATLAS to support multiple mesh resolutions. At the highest resolution, the ATLAS mesh consists of 115,834 vertices, approximately 16 times more than the 6,890 vertices in an SMPL [33] mesh. Our lowest resolution defaults to the SMPL resolution. We train 128 and 16 components for the surface vertex space and the skeletal space, respectively and our pose-corrective features are 24 dimensional. Additionally, we extract a hand pose PCA space, and re-target expressions from FLAME [30].

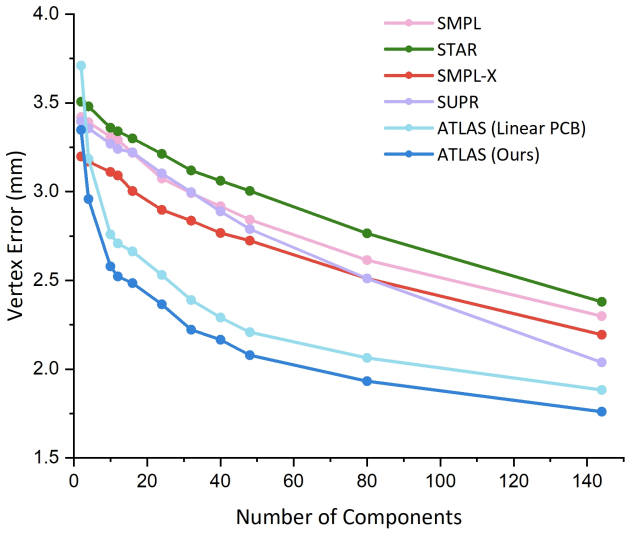

Figure 5. **Quantitative Evaluation on 3DBodyTex.** We report vertex-to-vertex error (mm) with different numbers of fitting components. For ATLAS, we report the combination of the number of shape and scale components used.

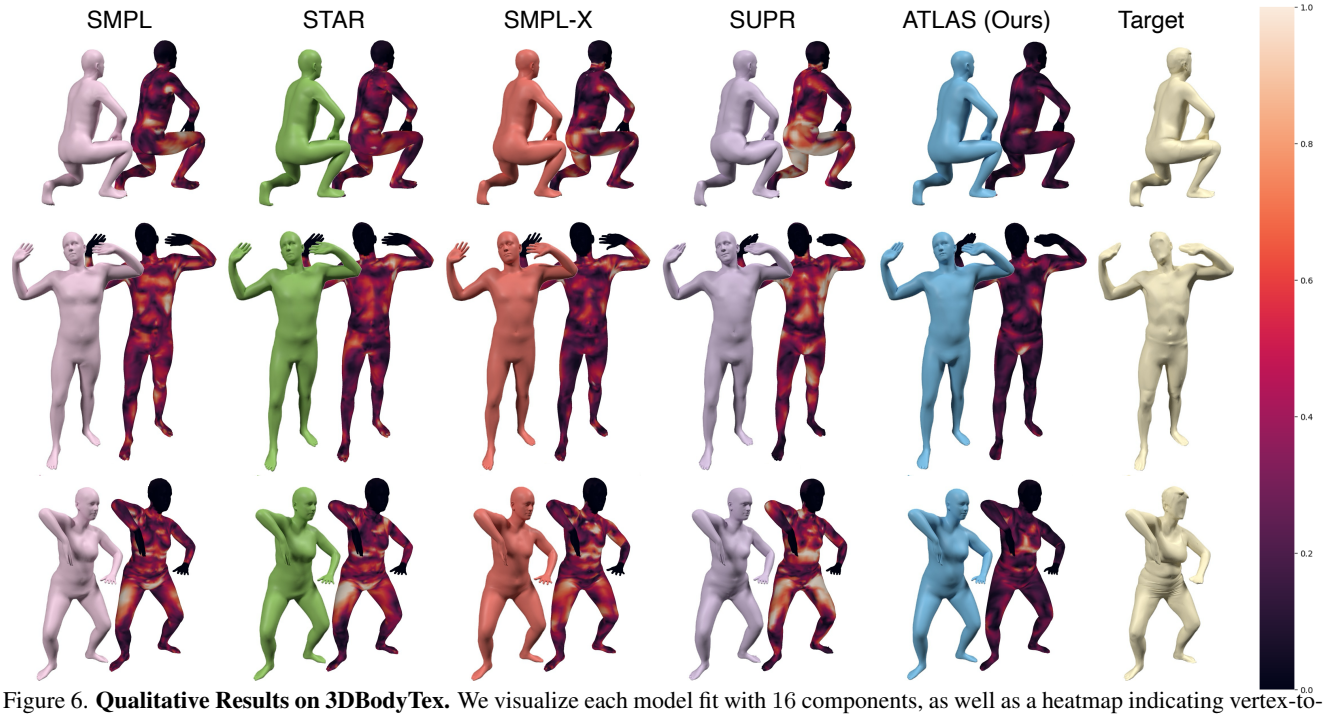

Figure 6. **Qualitative Results on 3DBodyTex.** We visualize each model fit with 16 components, as well as a heatmap indicating vertex-to-vertex error. Overall ATLAS exhibits tighter fits and has fewer blending artifacts at the elbows, knees, and shoulders compared to baselines.

Implementation Details. We decouple the surface and the skeleton in our data by first optimizing registrations with only skeletal parameters and poses. Using triangulated keypoints to regularize joints, these skeletal-only fits capture variations in height, arm length, finger size, etc. Then, we optimize surface shape to model soft tissue attributes like body weight and arm width. We then train ATLAS to capture these skeletal and surface spaces with autoencoders. Please refer to Section E of the supplement for more details.

4.2. Comparison with existing Body Models

We compare ATLAS with state-of-the-art body models, including SMPL [33], STAR [38], SMPL-X [39] and SUPR [37] on two datasets: 3DBodyTex [49] and Goliath-Test, a held-out test set of our captured dataset. In contrast to these baselines, only our proposed ATLAS body model decouples the skeletal and shape spaces.

3DBodyTex [49]. The dataset consists of 100 male and 100 female scans. For evaluation, we register each body model to the ground-truth 3D scans using the SMPL [33] topology to ensure a fair comparison. Due to missing or noisy data in the ground-truth face and hand regions, we mask these areas during evaluation.

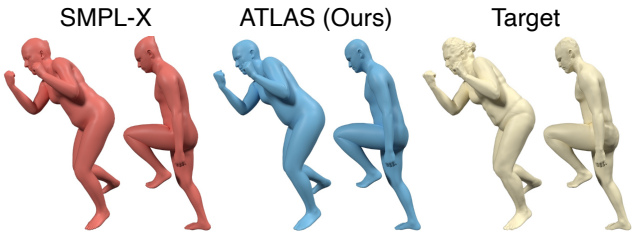

Figure 7. **Qualitative Results on the Goliath-Test set.** From the left, we compare SMPL-X's fits, ATLAS's fits, and registrations. ATLAS is noticeably better at capturing areas around joints, resulting in sharper knees and elbows.

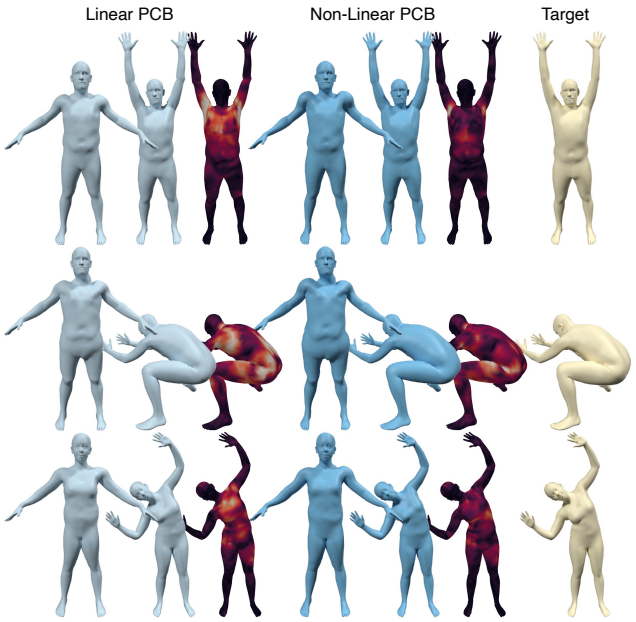

Figure 8. **Qualitative comparison of linear and non-linear pose correctives (PCB) on the SMPL pose dataset.** For each of linear and non-linear PCBs, we visualize the predicted vertex offsets in the rest pose, the posed mesh, and the fitting error heatmap.

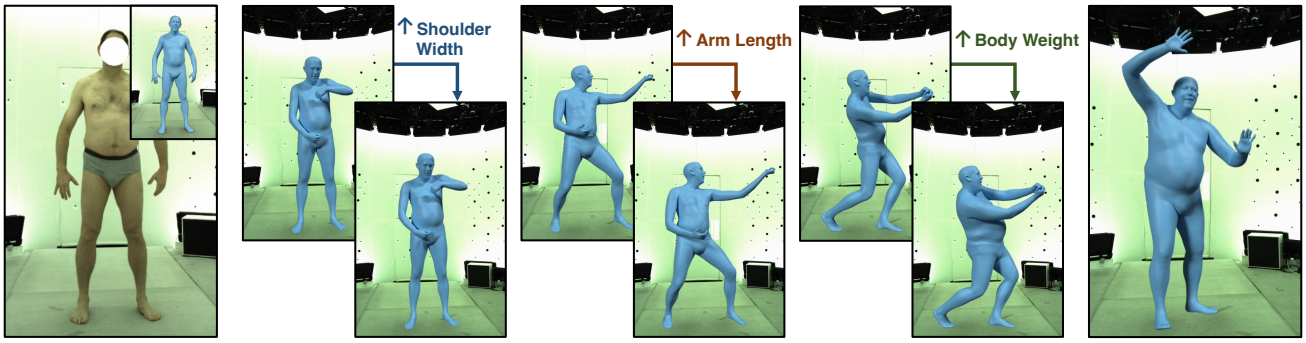

Figure 9. **Precise, decoupled control of skeletal and surface attributes.** Starting with a detailed ATLAS mesh of a subject, we sequentially increase shoulder width, change arm length, and adjust body weight. The resulting mesh is highly realistic, and it maintains details of the original subject shape and skeleton while naturally incorporating the added customizations.

Figure 5 shows the vertex error of all body models with respect to the number of fitting components. ATLAS achieves lower fitting error with fewer components due to its explicit decoupling of skeletal and shape spaces. For instance, at 32 components, ATLAS achieves 21.6% lower vertex-to-vertex error compared to SMPL-X. This validates ATLAS's ability to generalize to unseen identities. Figure 6 shows a qualitative comparison of ATLAS with existing baselines on the 3DBodyTex dataset. We observe that ATLAS especially performs well at the tip of the actuated joints (elbows and knees) and fits the shoulders of the target scan more closely compared to SMPL-X [39].

Goliath-Test. We evaluate on 100 unseen 3D scans in unique poses from 10 held-out subjects. Our evaluation protocol remains similar to 3DBodyTex but include the face and hands. The qualitative results are shown in Figure 7. In addition to having sharper joints, ATLAS better captures subtle deformations of the clenched hand and angled chin, and it achieves a lower fitting error of 2.34 mm compared to SMPL-X's 2.78 mm.

4.3. Discussion

Controllability. Compared to prior work, ATLAS's separation of surface and skeleton allows for precise control over the human mesh. In Figure 9, we demonstrate customization of the ATLAS mesh of a subject. We easily control shoulder width and arm length by adjusting a single skeletal attribute each, then adjust body weight by updating the first surface component. Changes in skeletal attributes precisely maintain the original surface details, and changes in surface attributes keep the internal skeleton constant. The resulting mesh is realistic and can readily be driven by the subject's own motion or by pose sequences from other sources. We encourage readers to refer to the supplemental video.

Linear vs Non-Linear Pose Correctives. Unlike existing methods, ATLAS uses non-linear pose correctives, which introduces more parameters but provides higher capacity to model pose-correlated vertex corrections. To evaluate, we compare ATLAS against a version with linear pose correctives on the SMPL [33] dataset, isolating the influence of pose-corrective blendshapes by fitting a single rest-pose mesh and internal skeleton across the entire sequence. A qualitative comparison is shown in Figure 8. The non-linear correctives achieve more realistic fitting, particularly around complex joints such as the shoulders, and better capture muscle bulging in extreme poses. Quantitatively, the fitting error decreases from 1.82 mm to 1.61 mm, with improvements concentrated around joint locations.

Computational Analysis. We compare the cost of generating a 3D mesh from the body model parameters in Table 2. ATLAS achieves significantly faster inference times than SMPL-X [39] for the same number of vertices, leveraging its optimized CUDA-based implementation. Moreover, our model supports higher resolutions (10× more vertices) with minimal latency increase.

Method	# Vertices	Runtime (ms)
SMPL-X	10475	3.74
ATLAS (SMPL topology)	6890	2.39
ATLAS (SMPL-X topology)	10475	2.47
ATLAS (High-resolution)	115834	5.37

Table 2. Runtime comparison of mesh skinning on an A100 GPU.

Method	**Vertex Error** (mm)	**Joint Error** (mm)
SMPLify-X	87.7	73.2
ATLAS (Ours)	**55.4**	**53.7**
no rel. depth	60.7	54.5
no rel. depth, no mask	61.8	55.7

Table 3. **Evaluating mesh prediction from a single image.** Our model better predicts the 3D human mesh from a single image, and each data term improves fitting.

Figure 10. **Qualitative results of fitting ATLAS to in-the-wild images.** Our multi-stage body fitting procedure robustly handles clothed subjects in varying poses along with detailed facial expressions.

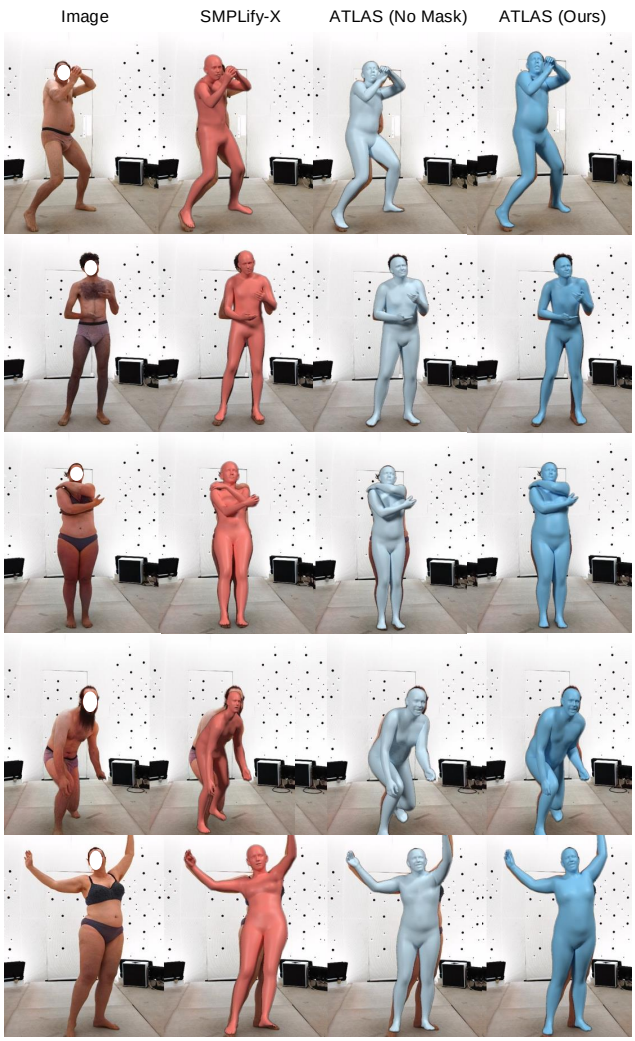

Figure 11. **3D pose, shape, and skeleton estimation from a single image.** Our fitting pipeline captures pose more accurately, and edge gradient optimization of shape captures soft tissue attributes.

4.4. Monocular Mesh Fitting

We evaluate our proposed single image mesh fitting approach (Sec. 3.4) on 200 scans from 10 unseen subjects in the Goliath-Test dataset. Table 3 reports mean vertex-to-vertex error (mm) and 3D joint error (mm) after Procrustes alignment. ATLAS achieves better vertex and joint fits compared to SMPLify-X [39], with further improvements from relative depth and mask optimization. Figure 11 shows that, unlike SMPLify-X, ATLAS fits pose and skeleton to keypoints without spuriously altering body shape. With edge gradient optimization, the body shape better aligns with subjects, particularly in the torso and legs. Our decoupled skeleton and shape body model, combined with decoupled keypoint and mask fitting, enables accurate pixel-aligned fitting across diverse images, as in Figure 10.

5. Conclusion

We propose ATLAS, an expressive body model that explicitly decouples surface shape from internal skeleton. Our body model enables direct controllability of the internal skeleton, avoids spurious and incorrect correlations between surface vertices and internal joint centers, and enables decoupled skeleton and shape fitting. We additionally propose a sparse and non-linear pose corrective function that demonstrates improved generalizability for 3D human mesh modeling. Additionally, we also present a keypoint fitting framework that achieves accurate, pixel-aligned fits using ATLAS on monocular images. ATLAS represents a step toward addressing limiting assumptions in body modeling, advancing the field toward realistic, accurate, and anatomically consistent 3D human mesh modeling.

Limitations. While ATLAS captures diverse body shapes, our 15,000 subjects do not span the full range of human variation. High-resolution human scan collection and processing remains time-consuming and costly, creating a bottleneck for scaling human modeling. However, ATLAS provides an accurate prior for human scan registration, enabling development of next-generation parametric models.

References

[1] SizeUSA dataset. https://www.tc2.com/size-usa.html, 2017. 5

[2] Brett Allen, Brian Curless, and Zoran Popović. Articulated body deformation from range scan data. *ACM Transactions on Graphics, (Proc. SIGGRAPH)*, 21(3):612–619, 2002. 3

[3] Brett Allen, Brian Curless, Zoran Popović, and Aaron Hertzmann. Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 147–156. Citeseer, 2006. 3

[4] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: Shape Completion and Animation of PEople. *ACM TOG*, 24(3):408–416, 2005. 3

[5] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005. 1

[6] Bharat Lal Bhatnagar, Xianghui Xie, Ilya A Petrov, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. Behave: Dataset and method for tracking human object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15935–15946, 2022. 1

[7] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *Computer Vision – ECCV 2016*. Springer International Publishing, 2016. 5

[8] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, 2014. 3

[9] Yu-Wei Chao, Zhan Wang, Yugeng He, Jiaxuan Wang, and Jia Deng. Hico: A benchmark for recognizing human-object interactions in images. In *Proceedings of the IEEE international conference on computer vision*, pages 1017–1025, 2015. 1

[10] Yinpeng Chen, Zicheng Liu, and Zhengyou Zhang. Tensor-based human body modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 105–112, 2013. 3

[11] Wei Cheng, Ruixiang Chen, Siming Fan, Wanqi Yin, Keyu Chen, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, et al. Dna-rendering: A diverse neural actor repository for high-fidelity human-centric rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19982–19993, 2023. 1

[12] Ali-Hamadi Dicko, Tiantian Liu, Benjamin Gilles, Ladislav Kavan, François Faure, Olivier Palombi, and Marie-Paule Cani. Anatomy transfer. *ACM Transactions on Graphics (TOG)*, 32:1 – 8, 2013. 3

[13] Oren Freifeld and Michael J Black. Lie bodies: A manifold representation of 3d human shape. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I 12*, pages 1–14. Springer, 2012. 3

[14] Stuart Geman. Statistical methods for tomographic image restoration. *Bull. Internat. Statist. Inst.*, 52:5–21, 1987. 5

[15] Benjamin Gilles, Lionel Reveret, and Dinesh K Pai. Creating and animating subject-specific anatomical models. In *Computer Graphics Forum*, pages 2340–2351. Wiley Online Library, 2010. 3

[16] Nils Hasler, Carsten Stoll, Martin Sunkel, Bodo Rosenhahn, and H-P Seidel. A statistical model of human pose and body shape. In *Computer graphics forum*, pages 337–346. Wiley Online Library, 2009. 3

[17] Nils Hasler, Thorsten Thormählen, Bodo Rosenhahn, and Hans-Peter Seidel. Learning skeletons for shape and pose. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*, pages 23–30, 2010. 3

[18] David A Hirshberg, Matthew Loper, Eric Rachlin, and Michael J Black. Coregistration: Simultaneous alignment and modeling of articulated 3d shape. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12*, pages 242–255. Springer, 2012. 3

[19] Hanbyul Joo, Tomas Simon, and Yaser Sheikh. Total capture: A 3D deformation model for tracking faces, hands, and bodies. In *CVPR*, pages 8320–8329, 2018. 3

[20] Petr Kadleček, Alexandru-Eugen Ichim, Tiantian Liu, Jaroslav Křivánek, and Ladislav Kavan. Reconstructing personalized anatomical models for physics-based body animation. *ACM Transactions on Graphics (TOG)*, 35(6):1–13, 2016. 3

[21] Ladislav Kavan, Steven Collins, Jiří Žára, and Carol O'Sullivan. Skinning with dual quaternions. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games*, pages 39–46, 2007. 2, 3, 4

[22] Marilyn Keller, Silvia Zuffi, Michael J Black, and Sergi Pujades. Osso: Obtaining skeletal shape from outside. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20492–20501, 2022. 2, 3

[23] Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C Karen Liu, and Michael J Black. From skin to skeleton: Towards biomechanically accurate 3d digital humans. *ACM Transactions on Graphics (TOG)*, 42(6): 1–12, 2023. 2, 3

[24] Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. Sapiens: Foundation for human vision models. In *European Conference on Computer Vision*, pages 198–213. Springer, 2025. 2, 5

[25] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2

[26] Paul G Kry, Doug L James, and Dinesh K Pai. Eigenskin: real time large deformation character skinning in hardware. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 153–159. ACM, 2002. 3

[27] Tsuneya Kurihara and Natsuki Miyata. Modeling deformable human hands from medical images. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 355–363. Eurographics Association, 2004. 3

[28] Sung-Hee Lee, Eftychios Sifakis, and Demetri Terzopoulos. Comprehensive biomechanical modeling and simulation of the upper body. *ACM Transactions on Graphics (TOG)*, 28 (4):1–17, 2009. 3

[29] J. P. Lewis, Matt Cordner, and Nickson Fong. Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 165–172, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 3

[30] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), 2017. 5

[31] Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, et al. Rich human feedback for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19401–19411, 2024. 1

[32] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019. 1

[33] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, 2015. 1, 2, 3, 4, 5, 6, 7, 12

[34] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *ICCV*, pages 5442–5451, 2019. 3

[35] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011. 14

[36] Marlies Nitschke, Eva Dorschky, Dieter Heinrich, Heiko Schlarb, Bjoern M Eskofier, Anne D Koelewijn, and Antonie J van den Bogert. Efficient trajectory optimization for curved running using a 3d musculoskeletal model with implicit dynamics. *Scientific reports*, 10(1):17655, 2020. 3

[37] Ahmed AA Osman, Timo Bolkart, Dimitrios Tzionas, and Michael J Black. Supr: A sparse unified part-based human representation. In *European Conference on Computer Vision*, pages 568–585. Springer, 2022. 2, 3, 4, 6

[38] Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. STAR: Sparse trained articulated human body regressor. In *ECCV*, pages 598–613, 2020. 2, 3, 4, 6

[39] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 3, 4, 5, 6, 7, 8, 14

[40] Hao-Yang Peng, Jia-Peng Zhang, Meng-Hao Guo, Yan-Pei Cao, and Shi-Min Hu. Charactergen: Efficient 3d character generation from single images with multi-view pose canonicalization. *ACM Transactions on Graphics (TOG)*, 43(4): 1–13, 2024. 1

[41] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *CVPR*, 2021. 1

[42] Stanislav Pidhorskyi, Tomas Simon, Gabriel Schwartz, He Wen, Yaser Sheikh, and Jason Saragih. Rasterized edge gradients: Handling discontinuities differentiably. *ArXiv*, abs/2405.02508, 2024. 5

[43] Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J Black. Dyna: A model of dynamic human shape in motion. *ACM Transactions on Graphics (TOG)*, 34(4):1–14, 2015. 3

[44] Apoorva Rajagopal, Christopher L Dembia, Matthew S DeMers, Denny D Delp, Jennifer L Hicks, and Scott L Delp. Full-body musculoskeletal model for muscle-driven simulation of human gait. *IEEE transactions on biomedical engineering*, 63(10):2068–2079, 2016. 3

[45] Jianqiang Ren, Chao He, Lin Liu, Jiahao Chen, Yutong Wang, Yafei Song, Jianfang Li, Tangli Xue, Siqi Hu, Tao Chen, et al. Make-a-character: High quality text-to-3d character generation within minutes. *arXiv preprint arXiv:2312.15430*, 2023. 1

[46] Taehyun Rhee, John P Lewis, and Ulrich Neumann. Real-time weighted pose-space deformation on the gpu. In *Computer Graphics Forum*, pages 439–448. Wiley Online Library, 2006. 3

[47] Kathleen M. Robinette, Sherri Blackwell, Hein Daanen, Mark Boehmer, Scott Fleming, Tina Brill, David Hoeferlin, and Dennis Burnsides. Civilian American and European Surface Anthropometry Resource (CAESAR) final report. Technical Report AFRL-HE-WP-TR-2002-0169, US Air Force Research Laboratory, 2002. 5

[48] Javier Romero, Dimitrios Tzionas, and Michael J Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM TOG*, 36(6):245:1–245:17, 2017. 3

[49] Alexandre Saint, Eman Ahmed, Abd El Rahman Shabayek, Kseniya Cherenkova, Gleb Gusev, Djamila Aouada, and Bjorn Ottersten. 3dbodytex: Textured 3d body dataset. In *2018 International Conference on 3D Vision (3DV)*, pages 495–504, 2018. 2, 6

[50] Shunsuke Saito, Zi-Ye Zhou, and Ladislav Kavan. Computational bodybuilding: Anatomically-based modeling of human bodies. *ACM Transactions on Graphics (TOG)*, 34(4): 1–12, 2015. 3

[51] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 84–93, 2020. 1

[52] Ajay Seth, Ricardo Matias, António P Veloso, and Scott L Delp. A biomechanical model of the scapulothoracic joint

to accurately capture scapular kinematics during shoulder movements. *PloS one*, 11(1):e0141028, 2016. 3

[53] Karthik Shetty, Annette Birkhold, Srikrishna Jaganathan, Norbert Strobel, Bernhard Egger, Markus Kowarschik, and Andreas Maier. Boss: Bones, organs and skin shape model. *Computers in Biology and Medicine*, 165:107383, 2023. 2, 3

[54] Haoyang Wang, Riza Alp Güler, Iasonas Kokkinos, George Papandreou, and Stefanos Zafeiriou. Blsm: A bone-level skinned model of the human mesh. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 1–17. Springer, 2020. 1, 2, 3

[55] Yuxiao Wang, Qiwei Xiong, Yu Lei, Weiying Xue, Qi Liu, and Zhenao Wei. A review of human-object interaction detection. *arXiv preprint arXiv:2408.10641*, 2024. 1

[56] Chung-Yi Weng, Brian Curless, Pratul P Srinivasan, Jonathan T Barron, and Ira Kemelmacher-Shlizerman. Humannerf: Free-viewpoint rendering of moving people from monocular video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition*, pages 16210–16220, 2022. 1

[57] Keenon Werling, Michael Raitor, Jon Stingel, Jennifer L Hicks, Steve Collins, Scott L Delp, and C Karen Liu. Rapid bilevel optimization to concurrently solve musculoskeletal scaling, marker registration, and inverse kinematic problems for human motion reconstruction. *bioRxiv*, pages 2022–08, 2022. 3

[58] Yuliang Xiu, Jinlong Yang, Dimitrios Tzionas, and Michael J Black. Icon: Implicit clothed humans obtained from normals. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13286–13296. IEEE, 2022. 1

[59] Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, William T Freeman, Rahul Sukthankar, and Cristian Sminchisescu. GHUM & GHUML: Generative 3D human shape and articulated pose models. In *CVPR*, pages 6184–6193, 2020. 1, 2, 3, 4

[60] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction. In *CVPR*, pages 601–610, 2020. 1

[61] Yifei Yin, Chen Guo, Manuel Kaufmann, Juan Jose Zarate, Jie Song, and Otmar Hilliges. Hi4d: 4d instance segmentation of close human interaction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17016–17027, 2023. 1

[62] Zerong Zheng, Han Huang, Tao Yu, Hongwen Zhang, Yandong Guo, and Yebin Liu. Structured local radiance fields for human avatar modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15893–15903, 2022. 1

[63] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5745–5753, 2019. 4, 5

[64] Lifeng Zhu, Xiaoyan Hu, and Ladislav Kavan. Adaptable anatomical models for realistic bone motion reconstruction. In *Computer Graphics Forum*, pages 459–471. Wiley Online Library, 2015. 3

A. Supplementary Overview

In the supplementary video, we present video results of fitting ATLAS to high-fidelity 3D scans, demonstrate controllability of skeletal attributes for a dynamnic sequence, and show results of fitting ATLAS to RGB videos in the wild.

In this supplementary document, we provide additional details on skeletal attributes, visualizations of the training data, implementation details, and qualitative results of ATLAS. The sections are organized as follows:

- Section B provides additional details regarding the 76 individually controllable skeletal attributes of ATLAS.
- Section C outlines the specific formulation of the Linear Blend Skinning (LBS) function.
- Section D visualizes some sample registrations from our Goliath dataset.
- Section E contains additional details regarding the training of ATLAS and the pose prior.
- Section F shows the skin weights before and after training.
- Section G includes visualizations of the first few external shape & internal skeleton latent components.
- Section H demonstrates the full expressiveness of ATLAS by visualizing generated subjects through random sampling of external shape, internal skeleton, body poses, hand poses, and facial expressions.
- Section I provides additional results on our single image to mesh prediction pipeline on in-the-wild images.

B. Details on Controllable Skeletal Attributes

ATLAS defines 76 controllable skeletal attributes that modify different parts of the skeleton. As described in the main paper, 15 of these attributes directly scale a local joint space (and those of its kinematic children). These consist of scales that affect the full-body, head, hands, feet, and individual fingers. The remaining 61 are bone length parameters that directly adjust each joint's center location with respect to its kinematic parent. These include the spine, neck offset, neck length, shoulder width, upper arms, lower arms, hip location, upper legs, lower legs, and each bone in the finger for precise controllability. We visualize the skeletal attributes that affect major parts (excluding individual finger bone adjustments) in Figure 12.

Further, we demonstrate that the surface shape basis and the skeletal basis are both necessary and are complementary by evaluating mesh fitting with disentangled parameters in Table 4. Shape alone misses height and limb length variations, while skeleton alone overlooks soft tissue. Using both, like ATLAS, best captures diverse body shapes.

Shape	Skeleton	3DBodyTex	Goliath-Test
✓	✗	6.47	4.76
✗	✓	3.17	2.67
✓	✓	**2.48**	**2.34**

Table 4. Mesh fitting error (mm) with shape and skeleton params.

C. Linear Blend Skinning Formulation

In this section, we provide the precise formulation for the LBS skinning function M used in Section 3.1 of the main paper. This transformation M to yield a scaled and posed vertex x_i is written as:

$$x_i = \sum_{j=1}^{I} \omega_{ij} \mathcal{T}_j(\bar{\theta}, \bar{t}, \theta, \sigma, t) \mathcal{T}_j(\bar{\theta}, \bar{t}, \vec{0}, \vec{0}, \vec{0})^{-1} \tilde{x}_i \quad (6)$$

where $\bar{\theta}$ and $\bar{t}$ define the rest pose of the skeleton. These rest pose definitions of each joint's rotation and offset with respect to its parent are necessary because unlike SMPL [33] where each joint's coordinate system is root axis-aligned, our rotations are skeleton-aligned. The forward kinematic transformation $\mathcal{T}_j$ is then defined by:

$$\mathcal{T}_j(\bar{\theta}, \bar{t}, \theta, \sigma, t) = \Pi_{a \in \boldsymbol{K}(j)} \begin{bmatrix} 2^{\sigma(a)} R(\theta_a) R(\bar{\theta}_a) & t(a) t_e(a) + \bar{t}_l \\ 0 & 1 \end{bmatrix} \quad (7)$$

where $\boldsymbol{K}(j)$ are the kinematic tree parents of joint a in ascending order, $\sigma(a)$, $t(a)$, and $t_e(a)$ are zero if joint a lacks a corresponding skeleton modification. Thus, $\mathcal{T}_j(\bar{\theta}, \bar{t}, \vec{0}, \vec{0}, \vec{0})^{-1}$ transforms from global to joint-j's local coordinates through kinematic tree traversal of an unposed, unscaled skeleton, while $\mathcal{T}_j(\bar{\theta}, \bar{t}, \theta, \sigma, t)$ transforms from joint-j's local to global coordinates with skeleton posing and bone scale/length modifications.

D. Visualization of scans from the Goliath dataset

In Figure 14 we provide a sample of our Goliath dataset To assemble a large and diverse set of of scans to train our model, we capture 130 subjects in a diverse suite of poses including conversational settings, charades acting, and dynamic movements. The frames are captured using 240 high-resolution, synchronized cameras that yields meshes with approximately 1 million vertices. The scans are captured at 30-90 FPS, and we use furthest-point-sampling on pose to select an interesting and diverse set of 600k frames to train ATLAS.

E. Training Details

E.1. ATLAS Body Model Details

E.1.1. Vertex Resolutions

While ATLAS is natively trained at the highest resolution with 115,834 vertices, we define mappings to the 6890 and

(a) Body Scale (b) Spine Length (c) Neck Offset (d) Neck Length

(e) Head Scale (f) Shoulder Width (g) Upper Arm Length (h) Lower Arm Length

(i) Right Hand Scale (j) Left Hand Scale (k) Hip Width (l) Hip Height

(m) Hip Depth (n) Upper Leg Length (o) Lower Leg Length (p) Feet Scale

Figure 12. **Visualization of Body Skeletal Attributes.** For each skeletal attribute, we show three meshes - increasing the skeletal parameter, the base mesh, and decreasing the parameter. Bones affected by the changed parameter are colored red if they have increased in size, and blue if they have decreased. Each attribute either directly scales a local joint space, including those of its kinematic children, or adjusts joint translations relative to its own kinematic parent. For instance, Figure 12i shows an instance of the former, where the entirety of the right hand changes in size, while Figure 12f is an instance of the latter, where the shoulder joint center is moved, driving an increase or decrease in shoulder width.

10475 vertices of SMPL and SMPL-X. This enables transformations between ATLAS and SMPL/SMPL-X and allows ATLAS to operate with fewer vertices for improved efficiency.

E.1.2. Body Model Design

ATLAS leverages a joint structure designed by expert sculpting artists to ensure anatomical consistency. The joint locations adhere to the human bone structure, and in place of a standard single 3DoF rotation for major joints, ATLAS decomposes them into anatomically accurate subjoints. For example, the shoulder includes a scapular joint,

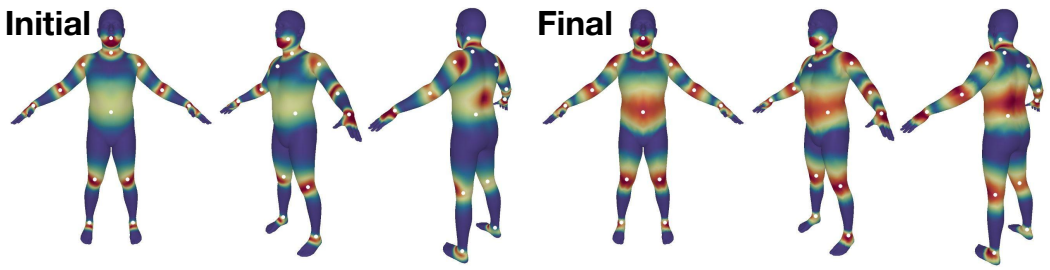

Figure 13. Skinning weights for the jaw, neck, upper arm, elbow, wrist, lower spine, knee, and ankle before and after optimization.

and the ankle is divided into subtalar and talocrural joints.

E.2. ATLAS Training Details

ATLAS is trained end-to-end by sampling registrations with their corresponding rest-pose surface vertices, internal skeletal parameters, and full body pose. The surface vertices and skeletal parameters are input into their respective linear autoencoders [35], the pose is input into our sparse, non-linear pose correctives function, and the mesh is rigged with the reconstructed vertices, reconstructed skeletal parameters, pose correctives, and trainable skin weights.

We initialize autoencoders [35] using PCA of surface vertices and skeletal parameters from our multi-shape dataset. For each training iteration, we sample the number of components $n \in [1, \max]$ and preserve only the first n features in the autoencoder latent bottleneck, zeroing out the remainder. This ordered dropout strategy maintains the component importance hierarchy throughout optimization. We use 128 components for the shape and 16 components for the skeleton, as we found fitting error plateaued beyond these components.

ATLAS is trained by minimizing the loss:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{shape\_reg}} + \mathcal{L}_{\text{skele\_reg}} + \mathcal{L}_{\text{skin\_lapl}} + \mathcal{L}_{\text{pc\_lapl}} + \mathcal{L}_{\text{skin\_init}} + \mathcal{L}_{\text{pc\_act\_reg}}$$

where $\mathcal{L}_{\text{data}}$ is the main data term minimizing vertex-to-vertex distance between the registration and the predicted mesh. $\mathcal{L}_{\text{shape\_reg}}$ and $\mathcal{L}_{\text{skele\_reg}}$ are L2 losses that regularize the intermediate latents of the surface vertex and skeleton attribute autoencoders. $\mathcal{L}_{\text{skin\_lapl}}$ and $\mathcal{L}_{\text{pc\_lapl}}$ regularize the skin weights and pose corrective blendshapes with a cotangent laplacian loss. $\mathcal{L}_{\text{skin\_init}}$ regularizes the skin weights towards their artist-defined initialization through L2. $\mathcal{L}_{\text{pc\_act\_reg}}$ imposes an L1 regularization loss on the pose corrective activation matrix, which is geodesic initialized, to encourage sparsity in vertex-joint correlations.

E.3. Pose Prior Implementation Details

For our pose prior, we adopt a lightweight VAE architecture similar to that of SMPL-X [39]. The VAE has a 32 latent dimension, takes as input 6D continuous rotation vectors for the full body excluding hands, and is trained to reconstruct samples from our 600k multi-pose dataset. The model is trained for 40 epochs with a batch size of 512 and a learning rate of 5e-3. We minimize three losses - the KL divergence loss, a reconstruction loss, and the angle difference loss between the input and output.

F. Optimized Skin Weights

We initialize skin weights Ω with artist-defined values and optimize them end-to-end during training. The weights before and after training are shown in Figure 13.

G. Skeleton and Shape Latent Spaces

Our skeletal attribute definitions allow for direct controllability of individual aspects of the internal skeleton. Furthermore, for lower-dimensional keypoint fitting, scan registration, and skeleton modification, our skeleton latent space provides data-driven correlations between different aspects of the body. We visualize the first four components of the skeletal components in Figure 15. We find that the skeletal attributes themselves capture most of the variation in the human body, such as overall body size, shoulder width, arm length, etc. While the skeletal components focus on the internal structure of humans, the surface components, shown in Figure 16 instead focus on the external soft tissue changes. Our surface components are more subtle than the shape components of prior work, as previous methods entangle skeletal and surface attributes, forcing the same components to capture variations in both soft tissue attributes and internal skeleton.

H. Latent Sampling of Shape, Skeleton, Pose, and Expressions

In this section, we further demonstrate the expressiveness of ATLAS by randomly sampling shaped, articulated human subjects in Figure 17. More specifically, we sample from our surface and skeletal latent spaces to model a random identity, then sample from our pose prior and hand PCA space for full-body pose, and finally sample facial expressions. The resulting meshes are realistic, and they span a

Figure 14. **Sampled Visualizations of Our Multi-Pose Dataset.** We train ATLAS on a diverse set of 600k scans captured by a high-resolution scanner with 240 synchronized cameras.

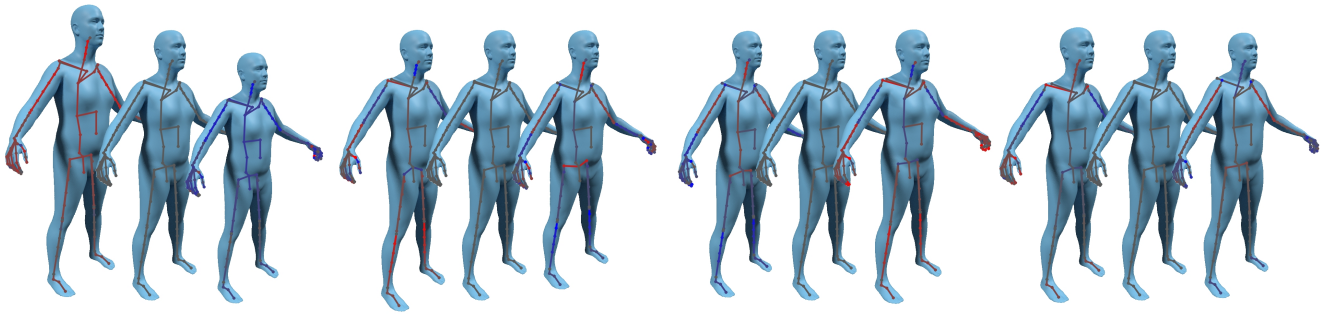

Figure 15. **Visualizations of the first four internal skeleton components.** For each component, we visualize changes in the mesh from decreasing and increasing the component. The skeleton is colored such that red indicates an increase in bone length while blue indicates a decrease. The skeletal components alone are sufficient to capture most human body variation. The first component is correlated with overall size of the subject, the second captures the neck and the hips, the third focuses on the shoulders and arms (decoupling upper and lower arm lengths), while the fourth captures length of the full arm.

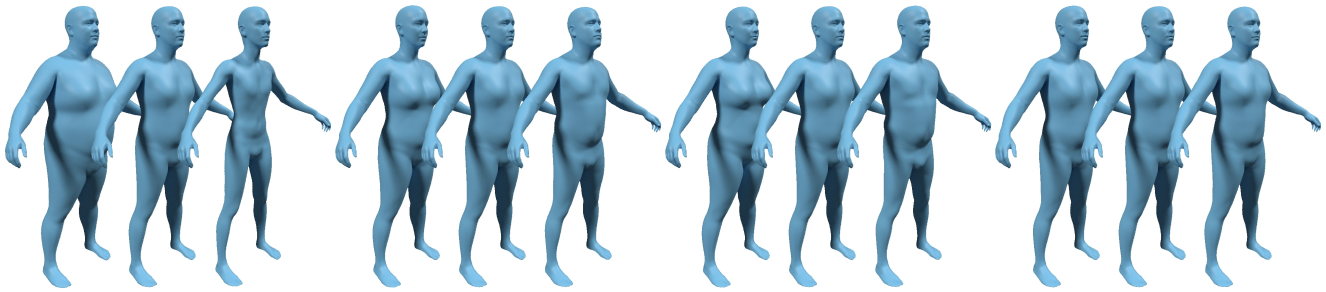

Figure 16. **Visualizations of the first four external surface components.** As most of the body variation (height, arm length, hand size, etc) are already captured by the skeleton, the surface components focus on soft tissue changes such as weight, neck width, arm thickness, and facial attributes. Note that we do not display the skeleton as it remains unchanged with variations in the surface vertices.

wide range of diverse human subjects in a variety of poses.

I. Additional Results on Mesh Prediction in the Wild

We extend the results in Figure 11 of the main paper by providing additional results on in-the-wild images in Figure 18. Our fitting procedure complements ATLAS by yielding shape, scale, pose, and expression parameters from 2D RGB images in the wild. Of particular note is ATLAS's ease at capturing undersized subjects such as children. By explicitly modeling the size of each skeletal part, ATLAS naturally predicts realistic shapes for children, accounting for their relatively larger heads compared to the rest of their body.

Figure 17. **Visualization of Random Latent Samples from ATLAS.** We randomly sample subject surface vertices and internal skeleton from the latent spaces, sample pose from our VAE pose prior and hand PCA space, and facial expressions from the FLAME space. ATLAS captures a wide breadth of realistic human shapes and articulates them into realistic poses.

Figure 18. **Additional Visualizations of Fitting ATLAS to Single Images.** Our fitting pipeline can capture a wide range of poses and shapes in addition to facial expressions.