Supplementary Material
OSSO: Obtaining Skeletal Shape from Outside

Marilyn Keller[1] Silvia Zuffi[2] Michael J. Black[1] Sergi Pujades[3]

[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]IMATI-CNR, Milan, Italy
[3]Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

Introduction

In this supplementary material, we provide further details of our method and elaborate on the results presented in the main paper. Specifically:

In Sec. 1 we detail how we train a 2D landmark predictor from DXA silhouettes and quantitatively evaluate the accuracy of the 2D predicted landmarks on the synthetic data. This section extends Sec. 3.2 of the main document.

In Sec. 2, we provide further details about the skin and skeleton registrations to the DXA images. This section provides further details of Sec.s 3 and 4 of the main paper.

In Sec. 3 we present an evaluation of the skeleton shape space obtained in Sec. 4.1 of the main paper.

In Sec. 4 we provide quantitative and qualitative results to complement the Sec. 5 from the main document.

In Tab. 1 we summarize the notation used in the paper for an easy reference.

1. Predicting 2D landmarks on DXA scans

In order to register the skin and skeleton models to the DXA scans, we need 2D landmarks on the scans. In this section we explain how we generate the synthetic dataset (Sec. 1.1, Sec. 1.2) to train a 2D landmark predictor from DXA skeleton silhouettes (Sec. 1.3) and evaluate the prediction (Sec. 1.4). The 2D landmark prediction from DXA silhouette is illustrated in Fig. 1.

1.1. Initial model creation

To generate synthetic skeleton silhouettes that look similar to real DXA bone masks M_B, we create an articulated skeleton model K, rigged with the STAR body model [3] parameters.

We first generate 21 STAR bodies by sampling the STAR shape space $\mathcal{B}_S$. We consider the mean body, and then, for the $n_\beta = 10$ first components of the STAR shape space, we sample two new body shapes with the shape parameters $\boldsymbol{\beta} = \{-2, 2\}$. Using Anatomy Transfer (AT) [1], we register a template skeleton mesh to each of these body shapes. Effectively we enforce the skin of the AT mesh to match the STAR mesh.

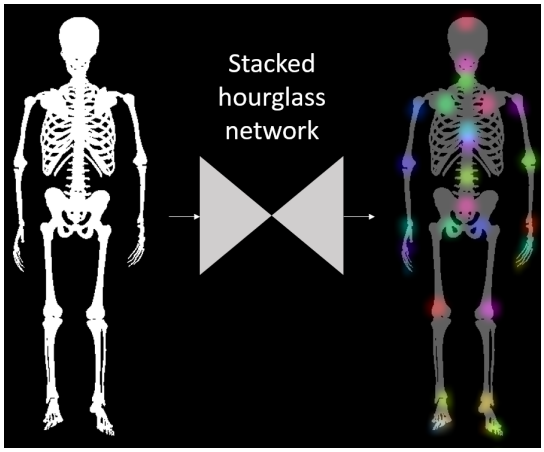

Figure 1. From a skeleton mask, a stacked hourglass network predicts the 2D locations of the landmarks $\tilde{\mathcal{L}}_I$.

With the obtained registrations, we define the mean skeleton shape $K(\beta = 0, \theta = 0)$, as the obtained AT skeleton on STAR's mean shape. Then, for each shape space component, we compute the skeleton offsets to the mean skeleton and use these offsets to define an initial skeleton shape space. From these, we compute the shape vectors of K as $\mathcal{B}_i = \mathbf{T}_{\boldsymbol{\beta}_i=2} - \mathbf{T}_{\boldsymbol{\beta}_i=-2}$ for i in $[0, n_\beta]$, else $\mathcal{B}_i = \mathbf{0}$.

To pose the skeleton, we rig it with the same kinematic tree as STAR. For each skeleton bone we manually define to which body part it belongs. This is straightforward as the initial template skeleton has the individual bones identified. It is important to note that the created skeleton model $K(\beta, \theta)$ can change its shape and pose using the same shape

Table of notation in OSSO

I_S	$\triangleq$	Dxa soft tissue image (skin)
I_B	$\triangleq$	Dxa bone image (skeleton)
M_S	$\triangleq$	Skin mask segmented from I_S
M_B	$\triangleq$	Skeleton mask segmented from I_B
$ST(\boldsymbol{\beta}_S, \boldsymbol{\theta}_S)$	$\triangleq$	STAR body model [3]
$\mathcal{B}_S$	$\triangleq$	STAR shape space
$K(\boldsymbol{\beta}_S, \boldsymbol{\theta}_S)$	$\triangleq$	The initial skeleton model rigged to the STAR shape and pose parameters
$SP(\mathbf{t}, \mathbf{r}, \boldsymbol{\beta}_B)$	$\triangleq$	Our *skeleton stitched puppet* model
$\hat{M}_B$	$\triangleq$	Synthetic skeleton mask generated with K
$\mathcal{L}_I$	$\triangleq$	29 3D landmarks whose 24 firsts correspond to STAR joints location and the closest skeleton vertices in K
$\tilde{\mathcal{L}}_I$	$\triangleq$	2D landmarks predicted from M_B.
$\mathbf{R}_S$	$\triangleq$	STAR body model registered to M_S
$\mathbf{R}_B$	$\triangleq$	Our *skeleton stitched puppet* model registered to M_S
$\mathbf{T}_B$	$\triangleq$	$\mathbf{R}_B$ unposed in T pose
$\mathcal{L}_B$	$\triangleq$	63 3D landmarks defined as vertices on the skeleton mesh template
$\mathcal{R}_B$	$\triangleq$	Regressor to predict skeleton landmarks $\mathcal{L}_B$ from a STAR body model registration $\mathbf{R}_S$
$\tilde{\mathcal{L}}_B$	$\triangleq$	$\mathcal{L}_B$ landmarks location inferred with the regressor $\mathcal{R}_B$
$\mathcal{B}_B$	$\triangleq$	PCA model of the skeleton learned from $\mathbf{T}_B$
$\mathcal{B}_S$	$\triangleq$	STAR PCA shape space
$\mathcal{R}_\beta$	$\triangleq$	Regressor to predict skeleton shape components $\beta_B \in \mathcal{B}_B$ from STAR shape components $\beta_S \in \mathcal{B}_S$
SI_{AT}	$\triangleq$	Skeleton mesh inferred with AT
SI_{OSSO}	$\triangleq$	Skeleton mesh inferred with OSSO

Table 1. Table of Notation

and pose parameters as STAR.

This initial model has an obvious drawback: the kinematic joint locations are not consistent with the anatomic skeleton articulations. Still, it is sufficient to easily generate plausible synthetic bone masks and the corresponding landmark annotations.

We define 29 landmarks on the skeleton mesh (Fig. 2). The first 24 correspond to the closest vertex to the STAR joint locations. Additionally we select the tip of the head,

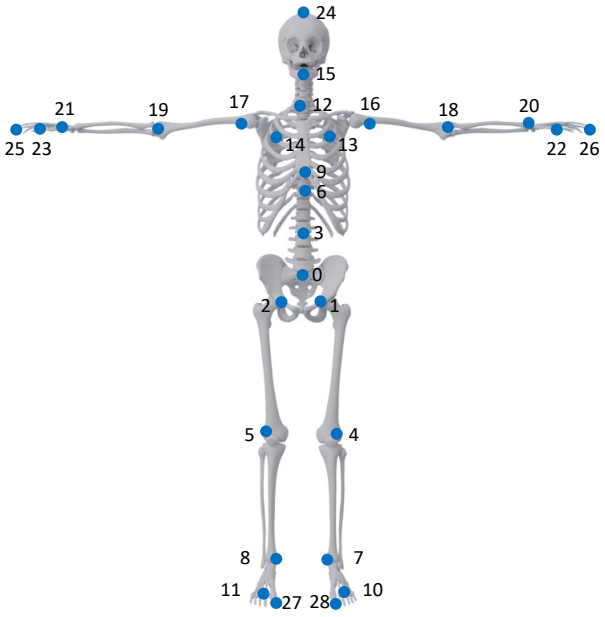

Figure 2. Position of the 3D landmarks $\mathcal{L}_I$ on the Stitched Puppet skeleton model P_B. These markers correspond to the location of the STAR 3D joints plus 5 additional landmarks.

fingers and feet. We denote these initial landmarks $\mathcal{L}_I$ or $\mathcal{L}_I(\mathbf{M})$ if we make explicit the mesh $\mathbf{M}$ on which the landmarks are defined.

1.2. Generating synthetic DXA masks

We use the skeleton model K to generate synthetic skeleton binary masks $\hat{M}_B$ with their corresponding 2D landmarks, that we denote $\tilde{\mathcal{L}}_I$ to explicitly distinguish them from the 3D landmarks $\mathcal{L}_I$.

We generate synthetic skeleton shapes by uniformly sampling the STAR shape space $\boldsymbol{\beta}$ in the range $[-2.5, 2.5]^{10}$. As the poses in DXA scans are relatively constrained, i.e. lying down with arms at the side, we manually define a *lying pose* $\boldsymbol{\theta}_L$ and sample new angles from a uniform distribution centered at $\boldsymbol{\theta}_L$ within a small range.

With the sampled shape and pose parameters, we render the silhouette of the skeleton and the corresponding landmark image. The virtual camera is orthographic to match the DXA scanner camera, and the field of view is set depending on the sample body height to leave a specific margin on the top and bottom of the image. This margin is sampled to match the margin distribution observed on the DXA dataset. A sample of the generated paired data is presented in Fig. 3.

To bridge the domain gap between the synthetic silhouettes and the DXA ones, we augment the data by eroding, and partially masking the rendered skeleton silhouettes, while keeping the landmarks fixed.

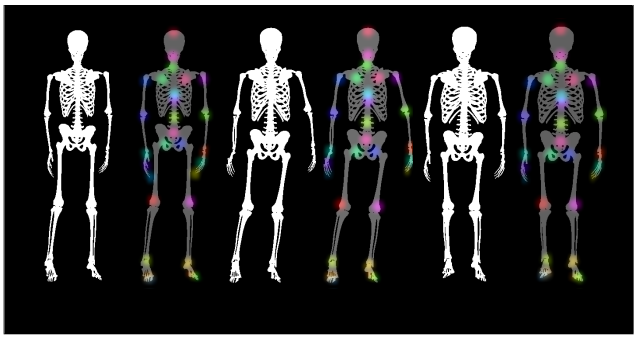

Figure 3. Pairs of synthetic skeleton masks (in white) and 2D landmarks $\tilde{\mathcal{L}}_I$ (color-coded) overlayed on the mask (in gray).

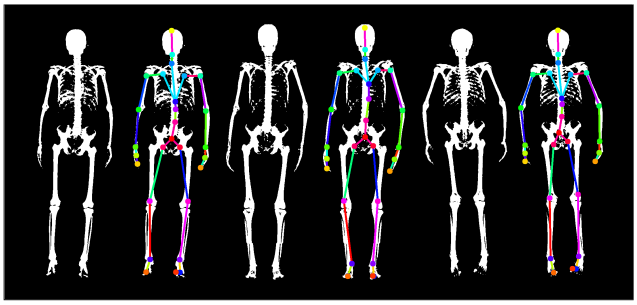

Figure 4. Pairs of input and predicted 2D landmarks $\tilde{\mathcal{L}}_I$ on real DXAs. The network learned on synthetic data generalizes well to real data.

1.3. Training a 2D landmarks predictor

From the synthetic silhouettes of the skeleton $\hat{M}_B$, we train the landmark detector using a stacked hourglass network [2] with 8 stacks. The network takes a 256x256 binary silhouette as input and outputs a 29x64x64 tensor, where each channel contains the position for one of the 29 landmarks $\tilde{\mathcal{L}}_I$.

In Fig. 4, we show qualitative results of the predicted landmarks on binary masks from real DXA images. We visually inspected the predicted 2D landmarks and observe that the silhouette simplification strategy combined with our data augmentation technique allows to obtain very good qualitative results on real DXA images.

1.4. 2D landmarks prediction evaluation

As the original DXA images do not have annotations, we only evaluate quantitatively on the synthetic dataset. We evaluate the landmarks predicted by the stacked hourglass network on 100 unseen synthetic skeleton silhouettes. The prediction error is measured in pixels on an image of size 256x256 pixels. The per landmark errors are reported in Table 2.

Most errors are on the order of one pixel. The highest prediction errors are for the tip of the middle fingers (L25 and L26) and the toes (L27 and L28). We observe that due to the resizing of the skeleton mask from the original image size (approx 800x800) to the size of the network (256x256), fine structures such as fingers and toes are degraded or lost. This is numerically visible with the standard deviations of the finger markers which are over 1 pixel.

	err. (mean $\pm$ std)		err. (mean $\pm$ std)
L0	0.73 $\pm$ 0.35	L15	0.78 $\pm$ 0.37
L1	0.95 $\pm$ 0.40	L16	1.01 $\pm$ 0.47
L2	0.81 $\pm$ 0.38	L17	0.87 $\pm$ 0.50
L3	0.90 $\pm$ 0.46	L18	1.22 $\pm$ 0.62
L4	1.14 $\pm$ 0.54	L19	1.01 $\pm$ 0.54
L5	1.12 $\pm$ 0.60	L20	1.22 $\pm$ 0.69
L6	0.78 $\pm$ 0.46	L21	1.21 $\pm$ 0.56
L7	1.16 $\pm$ 0.63	L22	1.08 $\pm$ 0.75
L8	1.24 $\pm$ 0.68	L23	1.04 $\pm$ 0.69
L9	1.07 $\pm$ 0.37	L24	0.75 $\pm$ 0.43
L10	1.18 $\pm$ 0.52	L25	1.87 $\pm$ 1.39
L11	1.18 $\pm$ 0.62	L26	1.53 $\pm$ 1.02
L12	0.87 $\pm$ 0.41	L27	1.23 $\pm$ 0.61
L13	0.87 $\pm$ 0.41	L28	1.23 $\pm$ 0.67
L14	1.01 $\pm$ 0.43		

Table 2. Prediction error in pixels of the predicted 2D landmark $\tilde{\mathcal{L}}_I$ on synthetic skeleton silhouettes. Landmark numbers are visually shown on the mesh in Fig. 2.

2. Skin and skeleton registrations to DXA

This section provides further details to complement the sections 3.3, 3.4 and 3.5 of the main paper.

2.1. Skeleton model based on Stitched Puppet

We create a parametric skeleton model to align to the DXA skeleton silhouettes based on the *stitched puppet* [6].

The *stitched puppet* model, as the name implies, represents an articulated deformable structure, the human body, as a collection of parts that are stitched together at the part interfaces. The model has per-part shape spaces and a pose parametrization in terms of location of each part center and its global rotation. The *stitched puppet* can be seen as a graphical model, where part parameters are defined at each node, and edge potentials represent stitching costs, that favor the parts to be connected and have smooth skin connections. The original model [6] is fit to 3D scans of people with non-parametric particle belief propagation. In order to define a stitched puppet model given an existing mesh, one needs to define a segmentation of the faces into parts, duplicate the vertices that belong to different adjacent parts, and define stitching potentials that act as springs between the corresponding duplicated vertices.

In our skeleton model, we manually define 21 groups of bones that belong to the same anatomic part, and define

the interfaces between these parts. In Fig. 5 we show the different parts with color codes, their interfaces, as well as the 3D landmarks $\mathcal{L}_B$ defined on the bones.

2.2. Registration costs

In this section, we detail the costs used for the skeleton registration (Sec 3.5 of the main paper) and the final reposing (Sec 4.3 of the main paper). In this section, we denote the vertices of SP as v_{sp}, the vertices of ST as v_{st} and z the anterior-posterior axis. v^z denotes the z component of vertex v and v^n the mesh normal at this vertex.

Skeleton to DXA registration. In Sec. 3.5 of the main paper, we introduce the cost E_i to constrain the skeleton inside the body. We decompose E_i as $E_i = E_{in} + E_p + E_{ct}$ and illustrate the intuition of each cost in Fig. 6.

The energy term E_{in} forces the skeleton to be inside the body along the front-back axis.

$$E_{in} = \max(0, D_z(SP(\boldsymbol{\beta}_B, \mathbf{t}, \mathbf{r}), \mathbf{R}_S)) \quad (1)$$

where D_z is the distance along z between a SP vertex and the closest skin vertex.

The term E_p forces vertices of the skeleton to be close to specific areas of the skin along the front-back axis. For several manually defined pairs of skeleton vertices and skin area A, we define

$$E_p = v_{sp}^z - \sum_{v_{st} \in A} (v_{st}^z). \quad (2)$$

The energy E_{ct} forces the *contact* between some specific vertices of the skeleton and the skin, like the elbow or the finger tips.

We define pairs of skin and skeleton vertices (v_{sp}, v_{st}) and want them to be at a fixed small distance $e = 5mm$. Effectively, E_{ct} is the per vertex distance:

$$E_{ct} = v_{sp} - (v_{st} - e \cdot v_{st}^n) \quad (3)$$

Skeleton unposing. In Sec. 3.5 of the main paper, we introduce E_d, a cost that enforces the conservation of the skeleton to skin distance when changing the pose. In Fig. 7 we illustrate the pairs of skin and skeleton vertices that are used for this cost. Our heuristic is that each of these pairs has a fixed distance d_0 that should be constant independent of the 3D pose.

Skeleton reposing. In Sec. 4.3 of the main paper, we use the costs E_j and E_l in the skeleton inference optimization.

The term E_j models ball joints in the shoulders, elbows and hips. It forces the bone heads to stay in their sockets. For each articulation, we define vertices s_i, s_j on the skeleton template that define a joint socket of a bone head. At each optimization step, we fit spheres with centers S_i, S_j to each groups of vertex and force each of spheres to stay at a similar distance during the optimization:

$$E_j(\mathbf{t}, \mathbf{r}; SP_0) = ||S_i(\mathbf{t}, \mathbf{r}) - S_j(\mathbf{t}, \mathbf{r})|| - d_{s0} \quad (4)$$

This cost is not sufficient to model the knee movement, so we define stitching costs approximating the human knee ligaments. We create pairs of vertices (l_i, l_j) at the bone locations where the ligaments are attached, and define the per-vertex cost $E_l = ||l_i - l_j|| - d_{l_0}$.

The distances d_{l0} and d_{s0} are defined such that $E_j(\mathbf{t_0}, \mathbf{r_0}; SP_0) = 0$ and $E_l(\mathbf{t_0}, \mathbf{r_0}; SP_0) = 0$.

3. Skeleton shape space evaluation

In section 3.6 of the main paper, we detail how we learn a skeleton shape space from the unposed skeleton meshes. In this section, we present an evaluation of the compactness of the shape space as well as its generalization ability.

3.1. Variance

To evaluate the compactness of our skeleton shape space, we compute the variance explained by each component of the PCA space. The cumulative variance plot is shown Fig. 8. With 3, 5 and 10 components, the male PCA model respectively captures 91.1%, 94.8% and 97.8% of the skeleton's variance. The female model respectively 92.7%, 95.6% and 98.1%.

3.2. Shape space generalisation

We next evaluate how the skeleton shape space generalises to unseen subjects. We compute the skeleton shape space from the training dataset and we evaluate how accurately it can reconstruct 200 left out unposed skeletons. We project each of the test set meshes onto the first N basis vectors of the shape space and we reconstruct the bones using only these coefficients.

We then measure how much information is lost in this projection by computing the per-vertex distance between the original mesh and the projected and reconstructed mesh. We aggregate this per-vertex error for each mesh and obtain the errors reported in Table 3.

As we can see, with a small number of components, such as 5, mean errors are below 6 mm. When using 10 components, the reconstruction mean errors are below 4 mm. The created bones shape space can capture the shape of left out subjects with errors below 4 millimeters.

4. Extended results

This section complements the presented results in Sec. 5 of the main document.

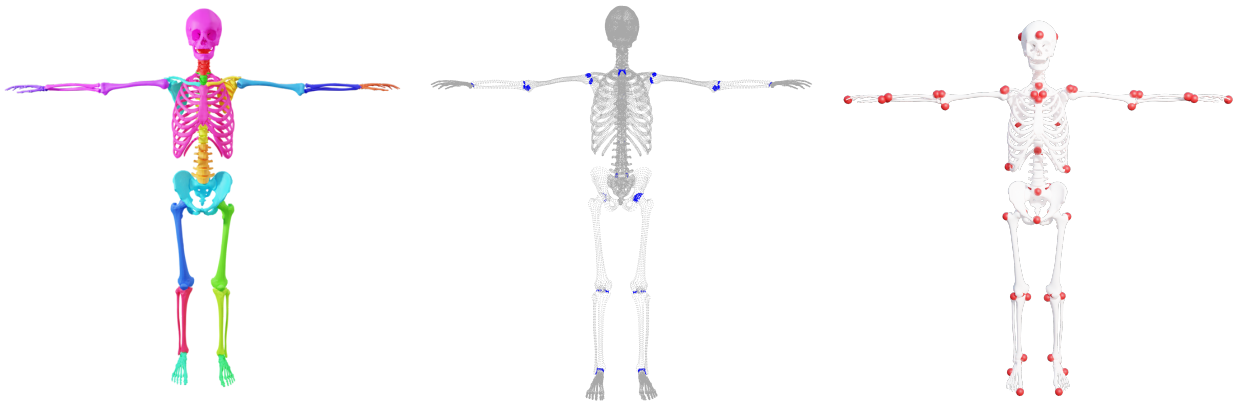

Figure 5. Our *stitched puppet* skeleton model, with the different bone groups (left), the interface point between the groups (center) and the 3D landmarks $\mathcal{L}_B$ (right).

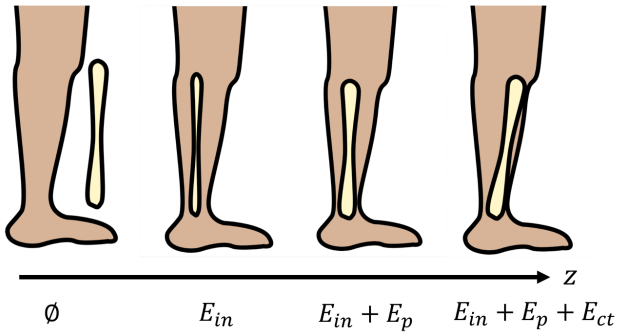

Figure 6. We illustrate the intuition behind the costs on a profile view of the tibia in the leg. From the frontal projected silhouette, there is no constraint for the bone to be inside the body along the z axis. We use E_{in} to force it to be inside. Forcing it inside is not enough as it could squeeze and collapse; thus, we enforce the bone to be close to the skin surface with E_p. In addition, there are regions where the bones are not covered by muscle and fat and should, therefore, lie close to the skin surface. We use E_{ct} to enforce these manually defined areas of contact.

	error (mm) (mean $\pm$ std)	
Nb components	Male	Female
3	7.59 ± 4.79	7.79 ± 4.86
5	5.55 ± 3.49	5.14 ± 3.27
10	3.14 ± 2.19	3.02 ± 2.14

Table 3. Skeleton reconstruction error given the number of principal components used. The errors are in millimeters.

4.1. Skin alignment qualitative evaluation

In this section we illustrate the alignment results of the STAR model on the DXA images. Those alignments were obtained with the optimization presented in Sec. 3.3 of the main paper. These results complement the quantitative evaluation reported in Sec. 5.1 of the main manuscript, where the intersection over union coefficient between the DXA mask M_S and the computed skin silhouette is 94% for females and 95% for males. In Figure 9, we show the qualitative results. The color-coded images show that the skin registrations faithfully capture the DXA skin silhouettes.

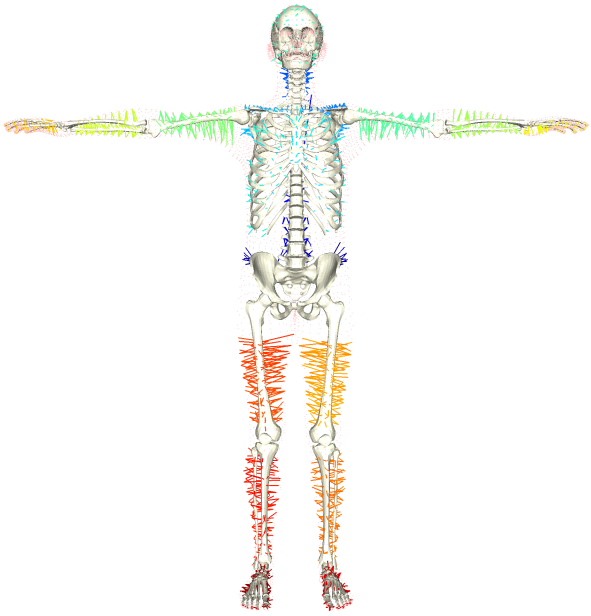

Figure 7. Skin to skeleton pairs used in the cost E_d. We color the links in each part with a different color for visualization purposes.

As mentioned in the last paragraph of Sec. 3.3, we use the quality of the fit to detect and remove failure cases from our datasets, i.e. subjects whose body shape can not be explained with STAR. In Fig. 10, we show some failure cases with low intersection over union values. These examples include subjects with atrophied or swollen limbs, severe sco-

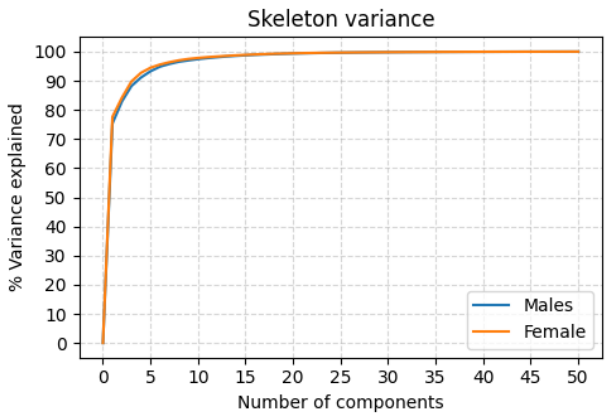

Figure 8. Cumulative variance of the skeleton shape space.

liosis or very low BMI. In practice, we used the alignment score to remove outliers of the available DXA scans (about 1%) to constitute a curated dataset containing a training set of 1000 subjects and a test set of 200 subjects for each gender.

4.2. Skeleton 3D landmarks regression evaluation

In Sec. 4.1 of the main paper, we explain how we train a regressor that, taking as input the vertices of the skin, predicts the 3D location of the landmarks $\mathcal{L}_B$ (presented in Fig. 5 right). This regression is learned in a normalized lying down pose as illustrated in Fig. 12.

To evaluate the $\mathcal{L}_B$ regressor accuracy, we learn the regressor from the 1000 train subjects and evaluate on the 200 left out subjects. We compute the 3D distance between the regressed landmarks position and its ground truth position. In Sec. 5.2 of the main paper we provide a general evaluation on the accuracy of the regressor as well as a discussion of the results. The detailed per landmark errors are listed in Table 4.

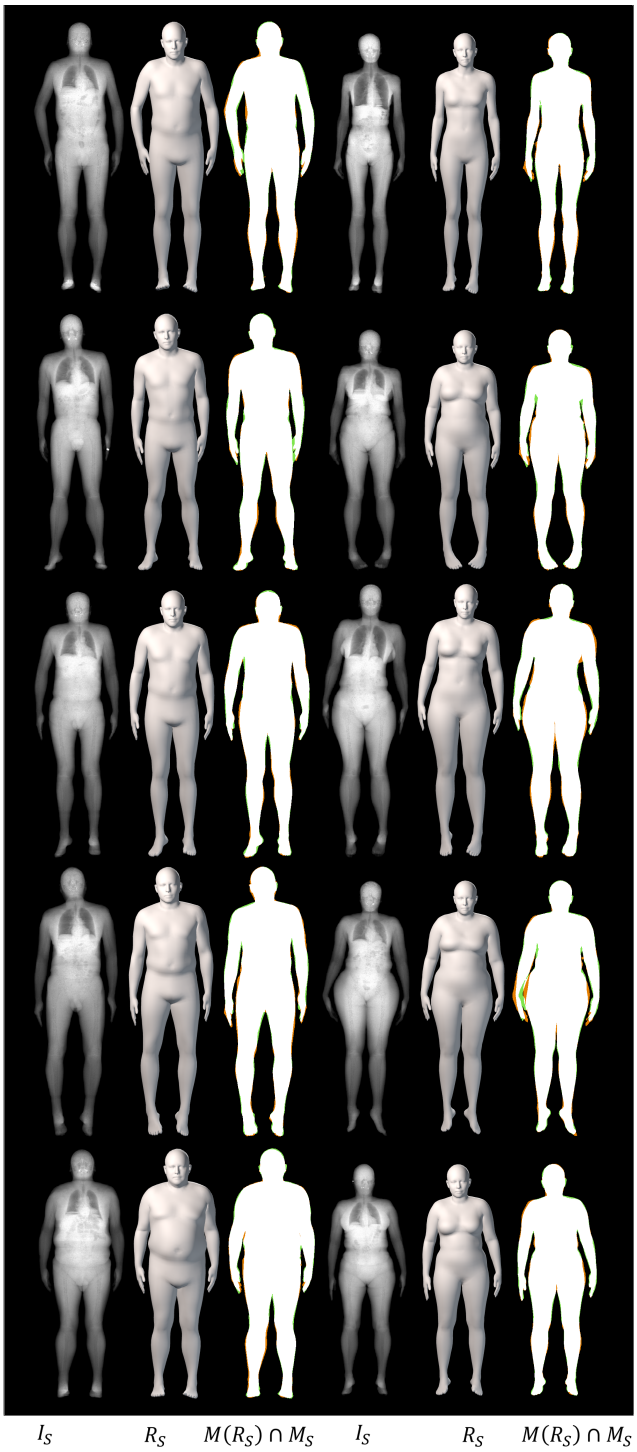

Figure 9. Comparison of the aligned STAR models $\mathbf{R}_S$ with the target DXA masks M_S for subjects sampled from the curated dataset. On the left we show males and on the right females. The masks intersection is color-coded as follow: green: $\mathbf{R}_S$ only, orange: M_S only, white: both.

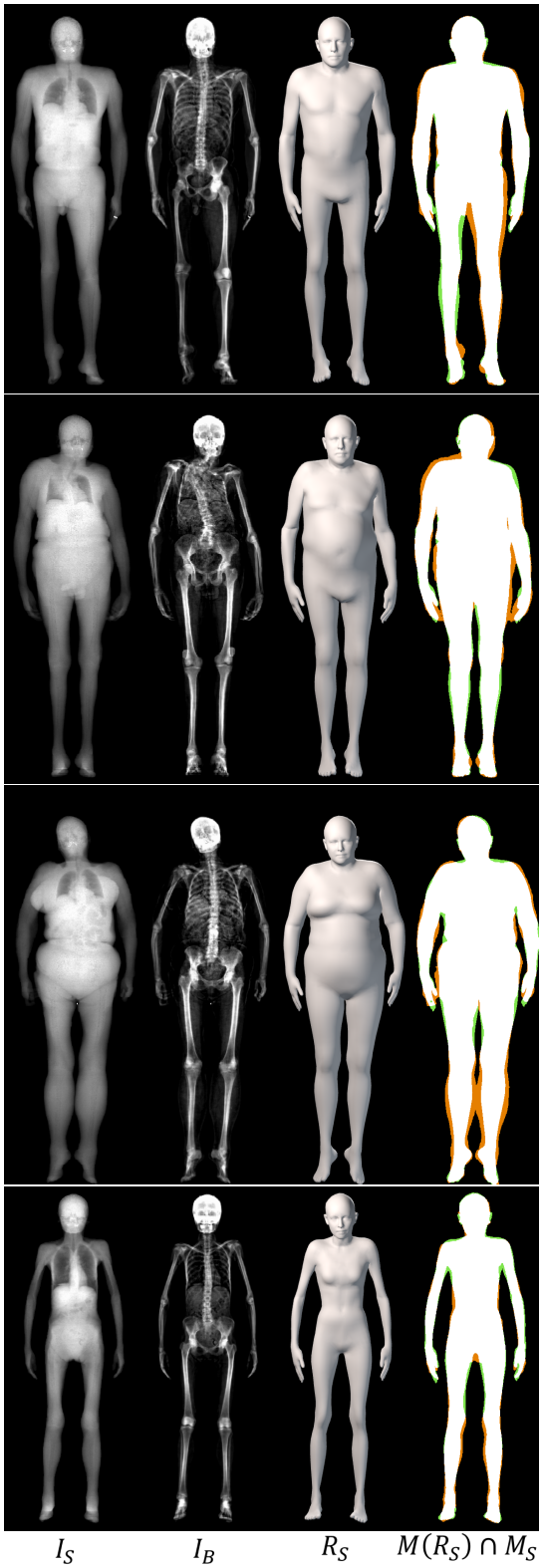

Figure 10. Failure cases. For each subject, we show I_S, I_B, the fitted skin mesh $\mathbf{R}_S$ and the intersection of both masks. The masks intersection is color-coded as follow: green: $\mathbf{R}_S$ only, orange: M_S only, white: both. The STAR model can not faithfully capture the shape of these subjects.

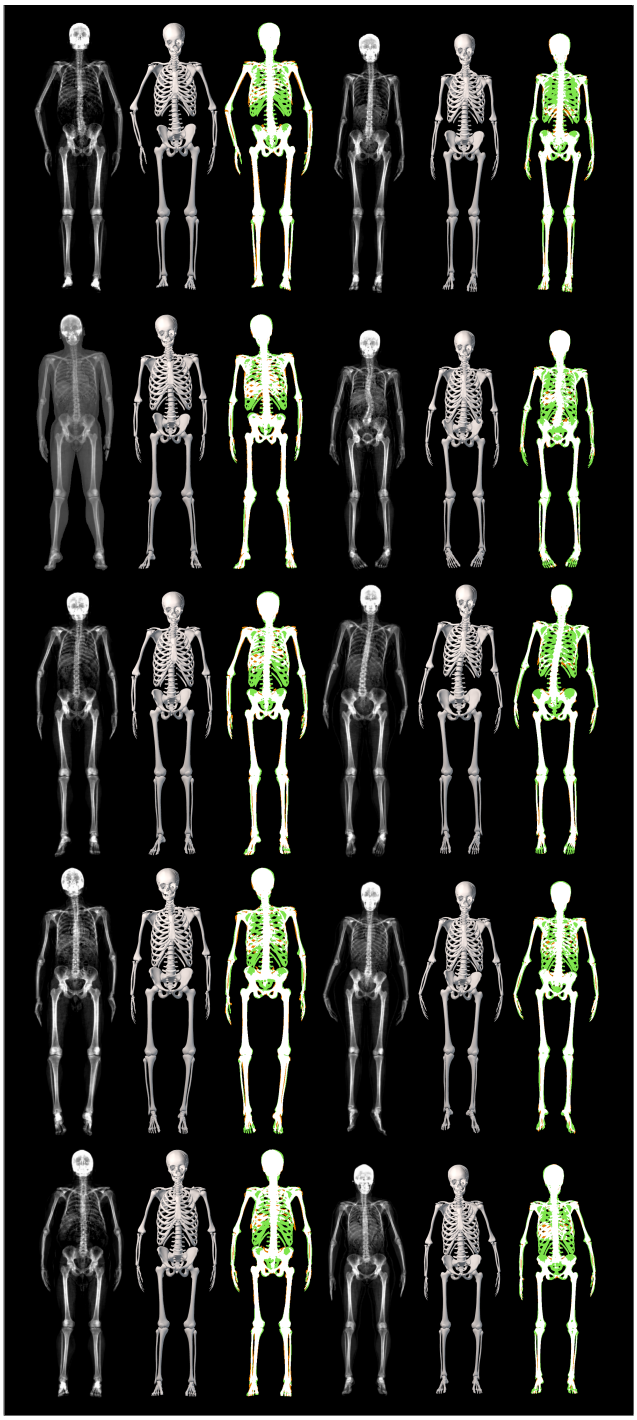

Figure 11. Comparison of the registered skeleton $\mathbf{R}_B$ with the target DXA masks M_B for subjects sampled from the training dataset. On the left we show males and on the right females. The masks difference is color-coded as follow: green: $\mathbf{R}_B$ only, orange: M_B only, white: both.

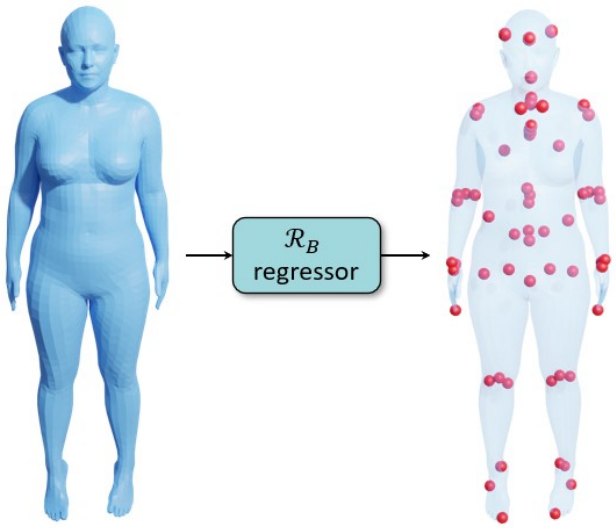

Figure 12. Given a skin mesh, the landmark regressor lets us compute the landmark 3D locations as a linear combination of the skin mesh vertices locations.

4.3. Skeleton registration qualitative evaluation

Next we show qualitative results of the skeleton registrations $\mathbf{R}_B$ in Fig. 11. The subjects are the same as in Fig. 9. These results complement the Sec. 5.3 of the main document, and precisely, the numeric value reported in the first row of Table 1 in the main document.

4.4. OSSO VS Anatomy Transfer comparison

In Figure 14, we present a qualitative comparison between our OSSO predictions and the ones from Anatomy Transfer. This results complement Sec. 5.3 of the main document.

From the DXA test set, we select 5 subjects spanning the dataset BMI distribution. From the skin alignment $\mathbf{R}_S$, we infer the skeleton and compare it to the subject's skeleton DXA image. We denote SI_{AT} the skeleton inferred with AT and SI_{OSSO} the skeleton inferred with OSSO. $M(SI)$ is the mask rendered from the mesh SI.

As can be seen from the images, our predictions do better capture the global shape of the skeletons. Particularly, Anatomy Transfer often estimates the location of the hips to be too low with respect to the actual hips location. Our method predicts a skeleton which is visually closer to the one observed in the DXA images.

4.5. Skeleton inference qualitative evaluation

Lateral view Fig. 15 shows side views of the inference result in T-pose. While there is no ground truth to evaluate this pose with, the results are plausible.

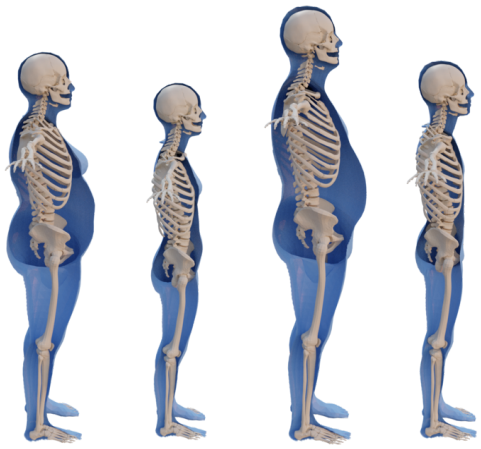

Figure 15. Lateral views of skeletons inferred with OSSO.

Inference on subjects from AGORA [4] Fig. 16 shows the inferred skeletons for subjects with different shapes and poses.

References

[1] Dicko Ali-Hamadi, Tiantian Liu, Benjamin Gilles, Ladislav Kavan, François Faure, Olivier Palombi, and Marie-Paule Cani. Anatomy transfer. *ACM Transactions on Graphics*, 32(6):1–8, Nov. 2013. 1

[2] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European Conf. on Computer Vision (ECCV)*, pages 483–499. Springer, 2016. 3

[3] Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. STAR: Sparse trained articulated human body regressor. In *European Conf. on Computer Vision (ECCV)*, volume LNCS 12355, pages 598–613, Aug. 2020. 1, 2

[4] Priyanka Patel, Chun-Hao P. Huang, Joachim Tesch, David T. Hoffmann, Shashank Tripathi, and Michael J. Black. AGORA: Avatars in geography optimized for regression analysis. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 13468–13478, June 2021. 8, 12

[5] RenderPeople. https://renderpeople.com, 2020. 12

[6] Silvia Zuffi and Michael J. Black. The stitched puppet: A graphical model of 3D human shape and pose. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3537–3546, June 2015. 3

	female	male		female	male
	err. (mm) (mean ± std)			err. (mm) (mean ± std)	
L0	9.03 ± 5.52	10.28 ± 10.28	L32	10.75 ± 5.10	11.65 ± 11.65
L1	14.41 ± 8.79	12.60 ± 12.60	L33	6.88 ± 3.37	6.40 ± 6.40
L2	15.74 ± 8.49	13.90 ± 13.90	L34	6.23 ± 2.58	6.42 ± 6.42
L3	9.99 ± 4.81	10.69 ± 10.69	L35	8.47 ± 4.79	7.96 ± 7.96
L4	4.23 ± 2.00	4.42 ± 4.42	L36	5.28 ± 2.53	5.21 ± 5.21
L5	8.38 ± 5.39	9.37 ± 9.37	L37	4.91 ± 2.63	4.24 ± 4.24
L6	9.72 ± 5.80	10.81 ± 10.81	L38	7.19 ± 3.00	6.95 ± 6.95
L7	14.76 ± 8.36	13.95 ± 13.95	L39	4.92 ± 2.52	4.28 ± 4.28
L8	15.93 ± 8.47	14.59 ± 14.59	L40	5.27 ± 2.66	4.47 ± 4.47
L9	4.06 ± 1.97	4.57 ± 4.57	L41	6.39 ± 3.76	4.65 ± 4.65
L10	10.76 ± 5.14	11.12 ± 11.12	L42	12.68 ± 7.17	10.93 ± 10.93
L11	9.46 ± 5.57	9.86 ± 9.86	L43	12.40 ± 7.77	11.08 ± 11.08
L12	2.03 ± 1.04	1.96 ± 1.96	L44	11.26 ± 6.14	10.44 ± 10.44
L13	2.89 ± 1.73	2.58 ± 2.58	L45	11.96 ± 5.93	9.85 ± 9.85
L14	3.34 ± 2.00	3.26 ± 3.26	L46	9.22 ± 4.40	9.37 ± 9.37
L15	3.67 ± 2.05	3.49 ± 3.49	L47	10.33 ± 5.51	10.13 ± 10.13
L16	2.42 ± 1.35	2.28 ± 2.28	L48	9.37 ± 4.21	9.78 ± 9.78
L17	3.33 ± 1.81	3.15 ± 3.15	L49	6.84 ± 3.29	7.69 ± 7.69
L18	11.20 ± 5.47	10.90 ± 10.90	L50	8.16 ± 3.93	7.62 ± 7.62
L19	9.91 ± 5.01	8.44 ± 8.44	L51	4.57 ± 2.21	4.53 ± 4.53
L20	11.50 ± 5.83	13.34 ± 13.34	L52	7.85 ± 3.95	6.68 ± 6.68
L21	9.96 ± 4.94	8.53 ± 8.53	L53	5.82 ± 2.89	5.13 ± 5.13
L22	6.76 ± 3.16	6.93 ± 6.93	L54	0.95 ± 0.52	0.98 ± 0.98
L23	7.17 ± 3.56	7.24 ± 7.24	L55	1.69 ± 0.89	1.90 ± 1.90
L24	5.29 ± 2.65	5.87 ± 5.87	L56	1.40 ± 0.74	1.47 ± 1.47
L25	5.31 ± 2.69	4.99 ± 4.99	L57	12.81 ± 7.43	11.38 ± 11.38
L26	7.74 ± 3.92	7.47 ± 7.47	L58	15.95 ± 9.94	13.96 ± 13.96
L27	5.72 ± 3.46	4.57 ± 4.57	L59	12.62 ± 6.91	11.32 ± 11.32
L28	5.44 ± 2.68	5.22 ± 5.22	L60	20.13 ± 10.65	17.36 ± 17.36
L29	6.66 ± 3.22	6.40 ± 6.40	L61	10.62 ± 4.44	8.80 ± 8.80
L30	10.83 ± 5.08	10.85 ± 10.85	L62	20.51 ± 11.31	16.47 ± 16.47
L31	8.94 ± 4.84	8.10 ± 8.10			

Table 4. Errors on the $\mathcal{L}_B$ landmarks regression in millimeters. In green the errors below 5 mm, in red the errors over 15 mm. The landmark numbers are visually shown in Fig. 13.

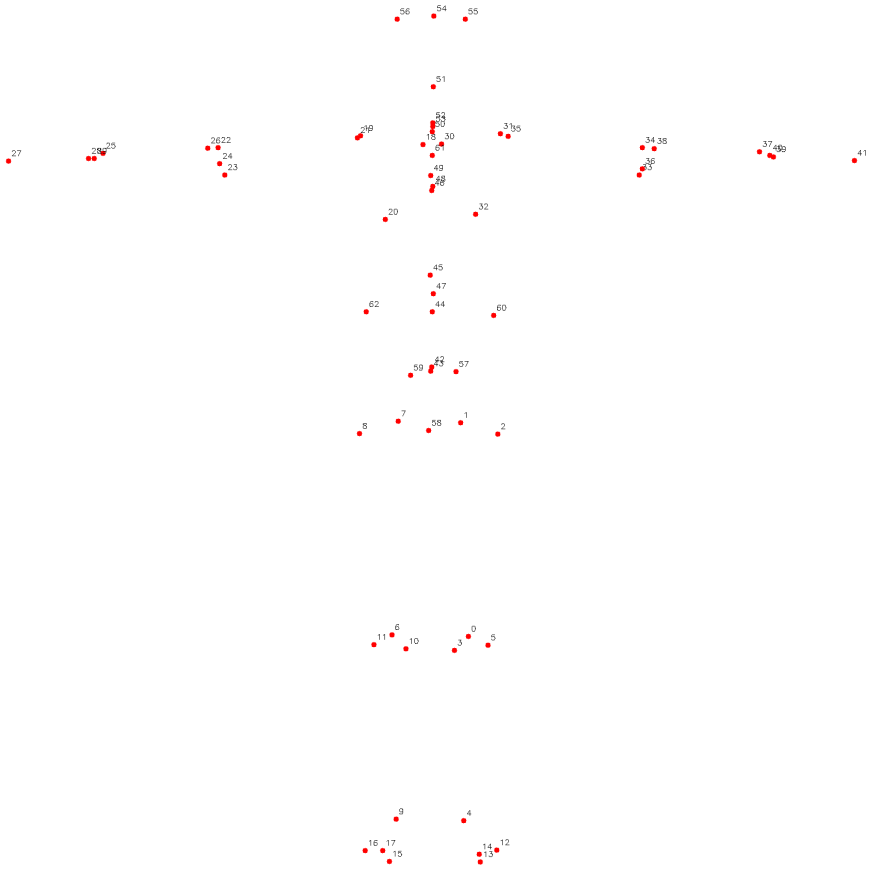

Figure 13. Landmarks $\mathcal{L}_B$ on the skeleton mesh with landmark number.

Figure 14. For each subject, we show in the order (1) $\mathbf{R}_S$, (2) SI_{AT} superimposed with the ground truth DXA I_B, (3) the overlap of $M(SI_{AT})$ and I_B, (4) SI_{OSSO} superimposed with the ground truth DXA I_B, (5) the difference between $M(SI_{OSSO})$ and M_B, (6) the ground truth DXA I_B

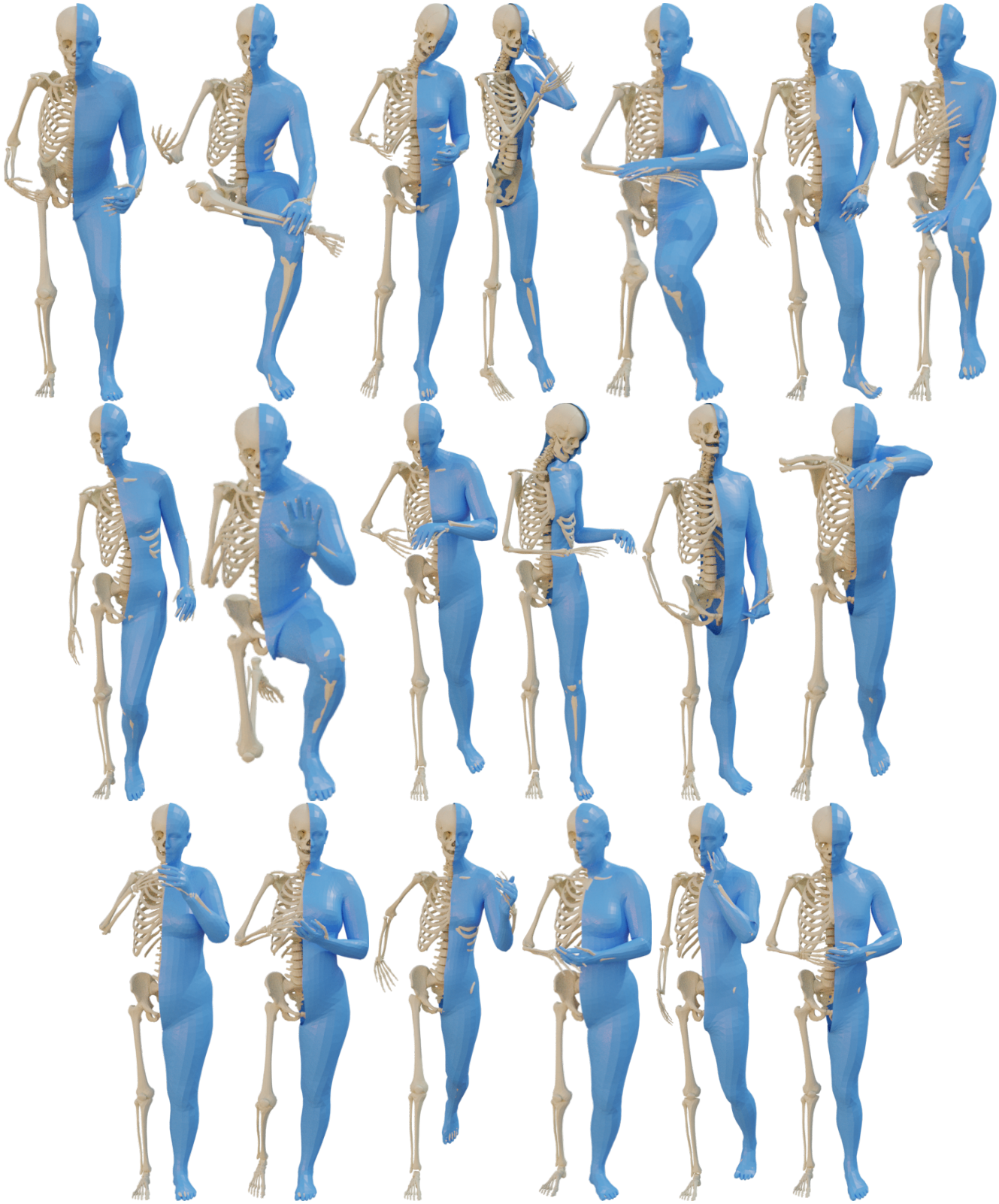

Figure 16. Given SMPL bodies aligned to RenderPeople subjects [4, 5], we use OSSO to infer the underlying skeleton.