Supplementary Material
HIT: Estimating Internal Human Implicit Tissues from the Body Surface

Introduction

In this document, we present additional information promised in the main manuscript. Please also see the companion **Supplemental Video**, which illustrates the method and results more visually, providing additional insight.

In Sec. 3 of the main manuscript, we describe how we build the HIT dataset containing the segmented MRI volumes, the extracted surface point clouds, and the corresponding SMPL registrations. Below, in Sec. 1 we detail how we segment the MRI images and in Sec. 2 we provide the two-step optimization process to fit SMPL body meshes to the MRI surface points.

In Sec. 4.2 of the main manuscript, we describe the HIT network architecture. Here, in Sec. 3, we provide the implementation details.

In Sec. 5.1 of the main manuscript, we extract per-tissue 3D meshes using a multi-label function and visualize the results. In Sec. 4 we explain how we extract these meshes.

And in Sec. 5 of the main manuscript, we provide results showing HIT predictions. In Sec. 5 we present further qualitative results complementing them.

1. Segmentation Process

To segment the data, we use the nnUNet [1] Auto-ML framework. It automatically configures a U-Net and adapts the training procedure to the input data. For segmentation tasks in the medical domain, it has been shown to be state-of-the-art in many benchmarks and it has been used by others in the creation of datasets [2, 4]. We use it for both $\mathbf{W_{bones}}$ and $\mathbf{W_{all}}$, which we present next, and for each model, we use the default settings configured by the Auto-ML framework.

To segment our dataset, we start by manually annotating the long bones (femur, tibia, fibula, humerus, ulna, radius and hips) in a small subset of the dataset (1105 slice images from 10 subjects). Fig. 1 presents examples of bone annotations. Then we train a segmentation model $\mathbf{W_{bones}}$, to segment the bones in the MRI images. The input to the network is a single-channel DICOM MRI that contains normalized MRI intensities and the output is a pixel-wise labeled mask with labels $\mathbf{L_{bones}} \in [0, 1]$. We empirically validate that

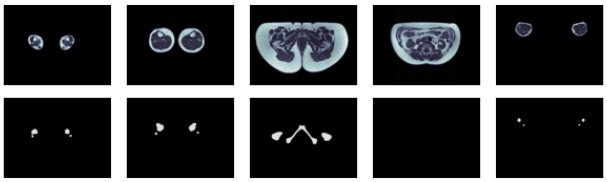

Figure 1. Examples of bone predictions from the model trained on manually annotated long bones.

1K images were enough to obtain a good generalization to left-out subjects (DICE score: mean 0.91/median 0.95).

In parallel, we use an automatic approach [5] that segments MRI images into adipose tissue (AT), empty (E), lean tissue (LT), as well as Visceral Adipose Tissue (VAT), i.e. fat around organs only in the abdominal region (see Fig. 2). The segmentations from this method show in general good results, but the method has empirically defined constants that do not generalize well across subjects. Most failures come from the sequential approach of the automatic method, first detecting anatomic landmarks and then segmenting the tissues. A landmark detection error often leads to some missing parts in the segmentation. Typical errors at this stage are shown in Fig. 3.

Screening the full dataset ($\sim 442 \times 110$ slices) is impractical, so we focus on a gender-balanced subset (80 subjects, $\sim$ 8900 images) and curate the generated segmentation artifacts. For the gender-balanced subset, we also infer the bone masks with $\mathbf{W_{bones}}$ and merge them with the curated segmentations to obtain one multi-tissue mask per image. We then post-process the merged segmentation masks in order to remove small artifacts and split erroneous adipose tissue (AT) segmentations from [5] into subcutaneous adipose tissue (SAT) and intra-muscular and visceral adipose tissue (IMVAT).

The obtained merged and post-processed segmentations were visually inspected and failure cases were corrected ($\sim$ 150 images from the total of $\sim$ 8900). Then, with the curated data, we train a new nnUNet model, $\mathbf{W_{all}}$, that takes an MRI image as input and predicts a label for each pixel corresponding to one of the 5 tissue types (BT, LT,

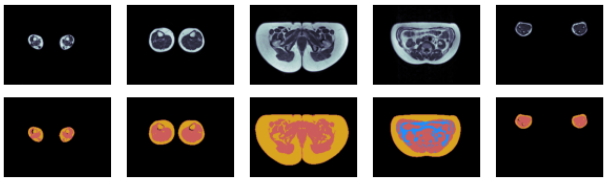

Figure 2. Examples of tissue predictions from the Würslin et al. [5] method.

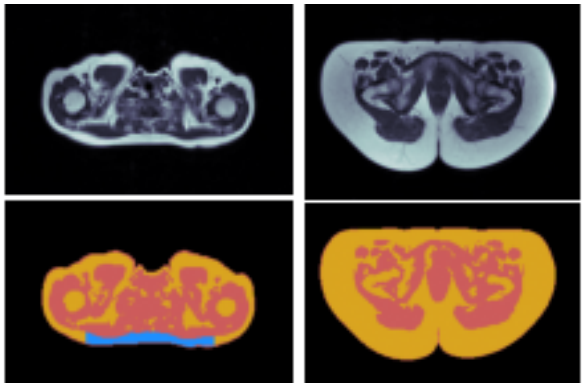

Figure 3. Typical errors from the Würslin et al. [5] method. Left: subcutaneous adipose tissue (SAT) here is inaccurately labeled as visceral adipose tissue (VAT). Right: visceral adipose tissue (VAT) is labeled as subcutaneous adipose tissue (SAT).

SAT, IMVAT, E). This network effectively replaces the previous network $\mathbf{W_{bones}}$ and the automatic method [5] which are not used anymore. We quantitatively evaluate the new segmentation predictions from $\mathbf{W_{all}}$ on a held-out test set, obtaining a mean/median DICE score of $0.92/0.98$;

To obtain the final segmentations, we use $\mathbf{W_{all}}$ to infer the segmentation masks for all 442 subjects. Fig. 14 and Fig. 15 show examples of final segmentation masks for a single female and male subject respectively. One final visual inspection of the obtained segmentations was performed to validate the full dataset of segmentations.

2. Two-step SMPL fit

The subjects of our dataset are lying down during the MRI scan, which causes the body shape to flatten. This skin compression is highly subject-specific, depending on their body composition, and SMPL is not able to model it. This section details the two-step process that we use to obtain SMPL fits that faithfully capture the shape of the subject as well as the compression.

For each subject i, we first extract the outer body contour from the segmented MRI images and, using the metric units of the volumetric MRI, we create a 3D point-cloud that we denote $\mathcal{S}_i$ for *skin* (see Fig. 5 left). Then, we compute a first approximation of the subject's shape and pose parameters $(\boldsymbol{\beta}_i^1, \boldsymbol{\theta}_i^1)$ that minimize the distance between the point cloud and body surface. As the subject's MRI poses are similar, we define a reference pose, $\boldsymbol{\theta}_{MRI}$, which we use to initialize fitting and we regularize the estimated pose so that it does not differ too much from the reference pose. We use the female or male version of SMPL according to the subject's gender and we denote these fitted SMPL meshes $\mathbf{S}_i^1$ (see Fig. 5 middle).

Next, we optimize the SMPL mesh vertices to deform and match the point-cloud $\mathcal{S}_i$. Inspired by the literature in the context of clothing capture [3, 6] we compute meshes in SMPL parametrization that tightly fit the segmented skin point clouds, $\mathcal{S}_i$. We optimize the new vertices locations, bound with Laplacian regularisation [6], and denote the resulting *free form meshes* $\mathbf{F}_i^1$. In Fig. 5 we show an example of an input MRI point-cloud $\mathcal{S}_i$ and the obtained meshes $\mathbf{S}_i^1$ and $\mathbf{F}_i^1$. It is worth noting, that while visually similar, the volumes of the meshes $\mathbf{S}_i^1$ and $\mathbf{F}_i^1$ can be very different. Fig. 4 shows the volume disagreement between $\mathbf{S}_i^1$ and $\mathbf{F}_i^1$ for the 442 subjects. This difference reveals that their current relative deformation includes other variations than the one solely created by the MRI table compression.

To overcome this problem, we start by computing the volume of the mesh $\mathbf{F}_i^1$, which we denote V_i^0. As this mesh is a tight fit to the point cloud, it approximates accurately the actual volume of the subject. Next, from the point-cloud $\mathcal{S}_i$, we now only consider the subset of points that are less affected by the table compression, i.e. those for which the normal vector is pointing in the same direction as the normal vector of the table, $\mathbf{n_T}$. Effectively, we weight the contribution of each point $\mathbf{s_p} \in \mathcal{S}_\mathbf{i}$ with $w(\mathbf{s_p}, \mathbf{n_p}) = \sigma(\mathbf{n_p} \cdot \mathbf{n_T})$ where σ is the sigmoid function. In Fig. 7 we show the effect of this weight on a SMPL mesh. Then, we compute a new SMPL mesh $\mathbf{S}_i$ and its parameters $(\boldsymbol{\beta}_i, \boldsymbol{\theta}_i)$ that match the weighted vertices with the additional volume constraint $||V_i^0 - V_i||_{L2}$, where V_i is the volume of $\mathbf{S}_i$. This enforces $\mathbf{S}_i$ to have a consistent volume with the MRI observation. In Fig. 6 we show the difference between the computed parameters $\boldsymbol{\beta}_i^1$ and $\boldsymbol{\beta}_i$, showing that the volume-preserving constraint effectively affects the computed body shape. The last step is to compute a deformed mesh $\mathbf{F}_i$ that is consistent with the new SMPL mesh $\mathbf{S}_i$. To this end, we recompute the free-vertex optimization starting from $\mathbf{S}_i$ to obtain a new tight fit $\mathbf{F}_i$. In Fig. 8 we show further results of the obtained SMPL meshes. These meshes allow us to compute the compression displacements $\mathbf{d}_{\text{comp}}$ between the $\mathbf{S}_i$ and $\mathbf{F}_i$ vertices. An animated illustration of this displacement can be seen in the accompanying video.

As SMPL can not model the stomach compression observed in the dataset, this goal two-step approach goal is crutial to get SMPL $\boldsymbol{\beta}$ values for each subject that actually match their body volume.

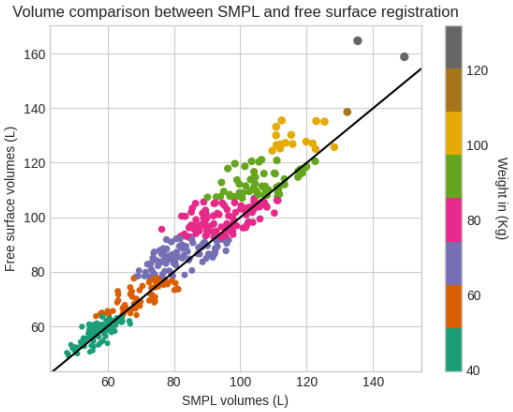

Figure 4. Volume differences between the naive SMPL body fit and the free-vertices version.

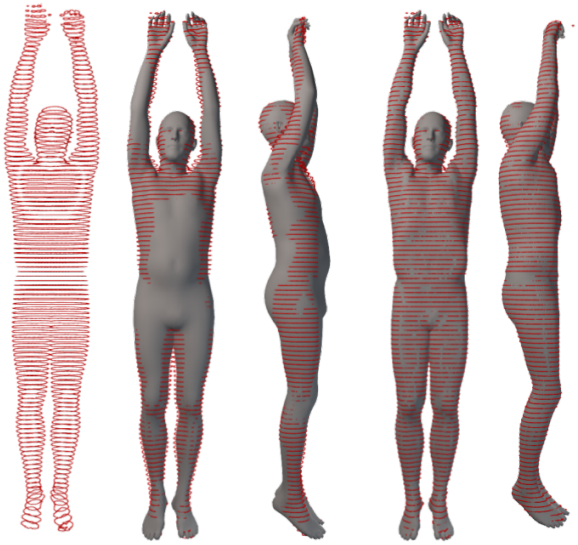

Figure 5. Initial fit to the MRI skin point-cloud. Left: point-cloud $\mathcal{S}_i$ extracted from the MRI. Middle: SMPL model fit $\mathbf{S}_i^1$. Right: Free-vertex fit $\mathbf{F}_i^1$.

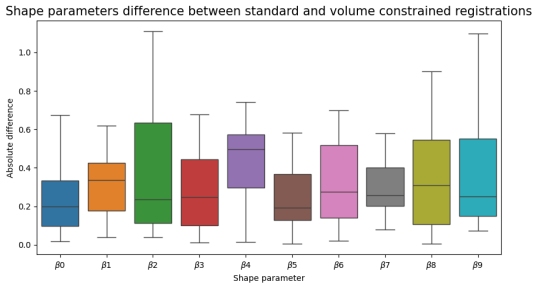

Figure 6. Boxplot of the shape coefficients difference between the $\mathbf{S}_i^1$ and the volume preserving $\mathbf{S}_i$ for the 442 subjects.

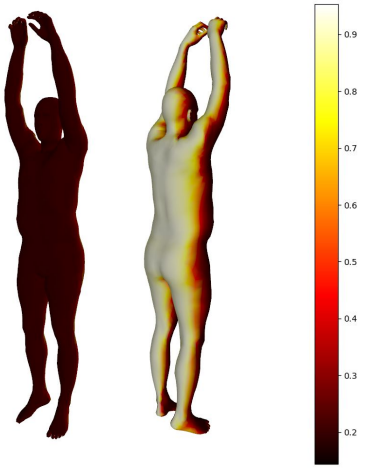

Figure 7. SMPL mesh vertices color-coded with the computed weights $w(\mathbf{s_p}, \mathbf{n_p})$. Vertices affected by the compression have a low weight, whereas vertices far from the MRI table have a high weight. These will be used to compute $\mathbf{S}_i$.

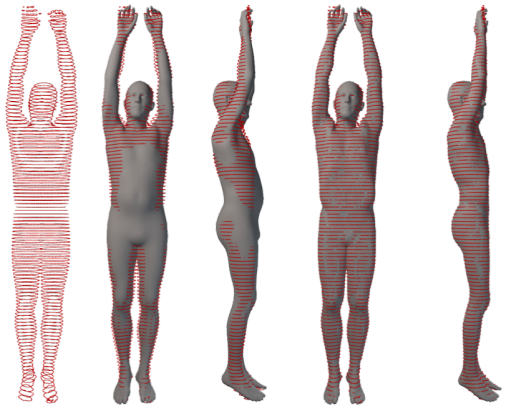

Figure 8. Examples of the obtained fit results. Left: point-cloud $\mathcal{S}_i$ extracted from the MRI. Middle: Volume preserving SMPL model fit $\mathbf{S}_i$. Right: Free-vertex fit $\mathbf{F}_i$.

3. Network Implementation Details

The HIT modules, described in Fig. 5 of the main document, define three networks: namely $\mathcal{B}$, $\mathcal{W}$ and $\mathcal{C}$. In addition, to predict the tissues inside the body in the canonical space, $\mathcal{T}$ is defined. All four networks are Multi Layer Perceptrons (MLP) with *softplus* activation functions. Next, we detail their architectures.

3.1. $\mathcal{B}$ MLP

The architecture of the network $\mathcal{B}$ is shown in Fig. 9. This network is used for converting a point from theshaped space into the canonical space. This is critical to enable learning of the implicit tissues in a single canonical representation given many training subjects of different shapes. $\mathcal{B}$ takes as input the shaped points $\mathbf{x}^{\beta}$ and the shape parameters $\boldsymbol{\beta}$ and it regresses a 3D offset $d_{\boldsymbol{\beta}}$. By applying this offset to the input point, the corresponding canonical point is obtained.

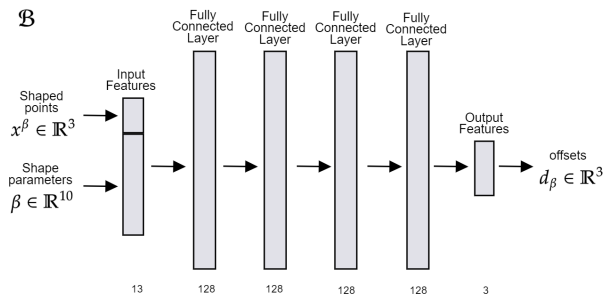

Figure 9. The network $\mathcal{B}$.

3.2. $\mathcal{W}$ MLP

The architecture of the network $\mathcal{W}$ is shown in Fig. 10. $\mathcal{W}$ takes as input a point in the canonical space $\mathbf{x}^{c}$ and regresses its skinning weights. The skinning weights are defined with respect to the 24 parts of the SMPL body model.

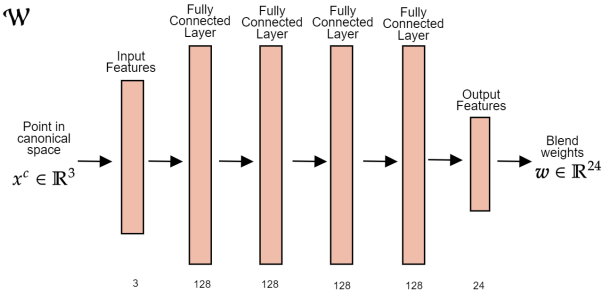

Figure 10. The network $\mathcal{W}$.

3.3. $\mathcal{C}$ MLP

The architecture of the network $\mathcal{C}$ is shown in Fig. 11. This network is important for undoing the effects of the table compression on the body. $\mathcal{C}$ takes as input shaped point $\mathbf{x}^{\beta}$, a shape parameter $\boldsymbol{\beta}$ and regresses a 3D offset $\mathbf{d}_{\text{comp}}$. By applying this offset to the corresponding point $\mathbf{x}^{m}$ in the compressed space, a point in the posed space is obtained.

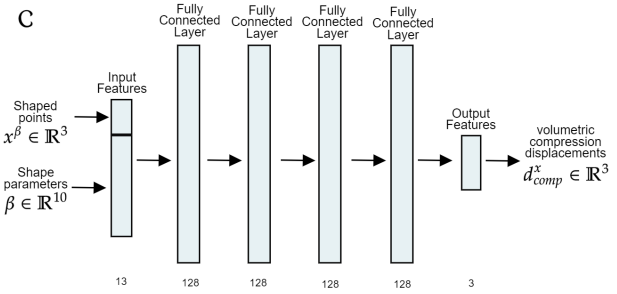

Figure 11. The network $\mathcal{C}$.

3.4. $\mathcal{T}$ MLP

The architecture of the network $\mathcal{T}$ is shown in Fig. 12. This network defines the implicit tissue classification at the heart of HIT. $\mathcal{T}$ takes as input a point in the canonical space $\mathbf{x}^{c}$, it encodes it using positional encoding, then regresses its 4-tissues probabilities. From these probabilities, the predicted tissue is obtained.

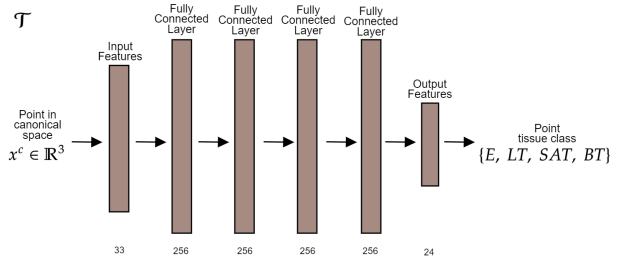

Figure 12. The network $\mathcal{T}$.

4. Multi-Tissue mesh extraction

A classical approach to visualize the occupancy of an implicit shape is to extract the mesh at a given level set of the implicit surface. Modeling occupancy as a multi-class classification problem has the advantage that we can extract meshes for each class that avoid overlap or interpenetrations of predictions. However, for a given class, class scores do not directly yield the continuous occupancy function that is required for level-set extraction.

Thus, given one tissue, our goal is to define a function to apply to the per-class probabilities with the following properties:

1. If class k has the highest probability, the function should yield an occupancy value > 0.5.
2. If class k does not have the highest probability, the function should yield an occupancy value < 0.5.
3. If class k has the highest probability, but is equal to another class, the function should yield an occupancy value $= 0.5$. This case defines the boundary between two or more tissues, which will be extracted by the level-set method.

For a point in the canonical space $\mathbf{x}^c$, let $\{p_i\}_{i\in[1,C]}$ with $p_i \in [0, 1]$ be the probabilities of each of the $C = 4$ classes ($\sum_{i=0}^{C} p_i = 1$). For a tissue k, we define the function l_k as:

$$l_k(\{p_i\}) = \frac{p_k}{p_k + \underset{\forall j!=k}{max}(p_j)} \tag{1}$$

which fulfills the desired properties. We can then pass $l_k(\{p_i\})$ to a marching cube algorithm to extract the k-th tissue mesh surface, and get tissue volumes that match the predicted occupancy.

The volumes displayed in Fig. 1, 7, 8, and 9 of the main manuscript, as well as the volumes in Fig. 18 were extracted using this technique.

5. Tissue prediction evaluation

5.1. Learned displacement fields visualization

HIT learns two volumetric displacement fields: $\mathbf{d}_\beta$ generalizing the SMPL shape space to $\mathbb{R}^3$, and $\mathbf{d}^x_{\text{comp}}$ accounting for the MRI table compression. Fig. 13 shows 2D slices of these fields at the hip level (tissue contours are shown). The field $\mathbf{d}_\beta$ computed for the shape component associated with weight (Fig. 13 left) shows a radial structure, which is consistent with the SMPL shape space. The field $\mathbf{d}^x_{\text{comp}}$ (Fig. 13 right) shows the displacements from compressed to uncompressed shape. Note how the central part experiences the most compression in the outwards direction, while the lateral parts have a milder, but lateral displacement. This is coherent on how the body shape is affected by the MRI table.

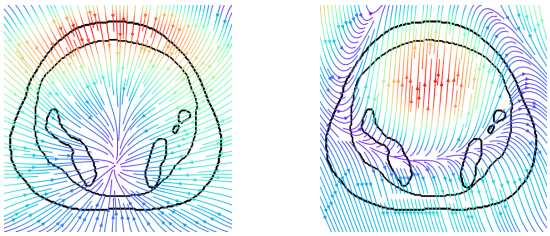

Figure 13. Slices of the learned volumetric displacement fields. Left: Shape field $\mathbf{d}_\beta$. Right: Compression field $\mathbf{d}^x_{\text{comp}}$. The arrows are colored from dark blue to red proportionally to the 3D field absolute value at each point.

5.2. Slice prediction

To complement Fig. 6 of the main manuscript, we present Fig. 16 and Fig. 17 with more examples of per-slice predictions on left-out subjects.

5.3. Volumetric prediction

To complement Fig. 7 of the main manuscript, in Fig. 18 we show more volumetric predictions on the left-out subjects. Each tissue's mesh is extracted in the canonical pose given the subject's shape $\boldsymbol{\beta}$, then posed to the target pose $\boldsymbol{\theta}$.

In Fig. 16 right column lines 9 to 11, we see that empty tissue is predicted inside the thigh. We conjecture this happens due to the root finding algorithm initialization, which rigs a query point to the closest SMPL skin vertex. In the cases where the SMPL fit mesh has self-penetration at the thigh level, which can happen when the legs are compressed together, the points in the intersection get rigged to the wrong leg. As a result, the occupancy is queried outside the body, leading to an empty prediction.

5.4. Comparison with OSSO

Finally, we compare the inferred bone volume with OSSO, we consider the GT point clouds of the bones in the test set and compute their distance to the predictions of OSSO and HIT. Fig. 19 reports the results in which the HIT predictions are systematically closer to GT than OSSO predictions.

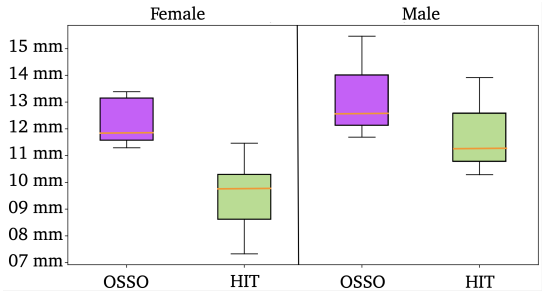

Figure 19. Comparison between OSSO and HIT bone predictions.

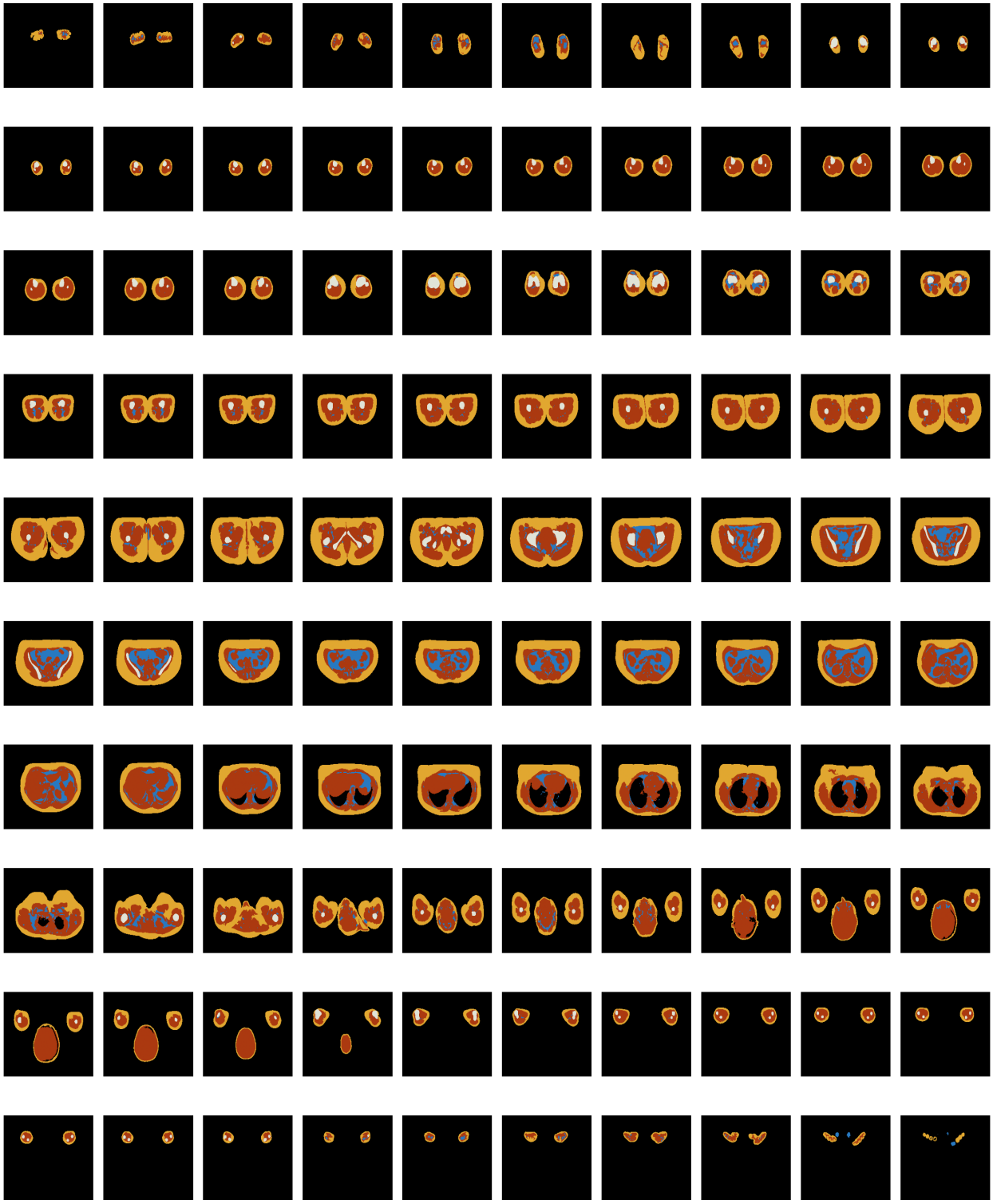

Figure 14. Example segmentation masks of a female subject

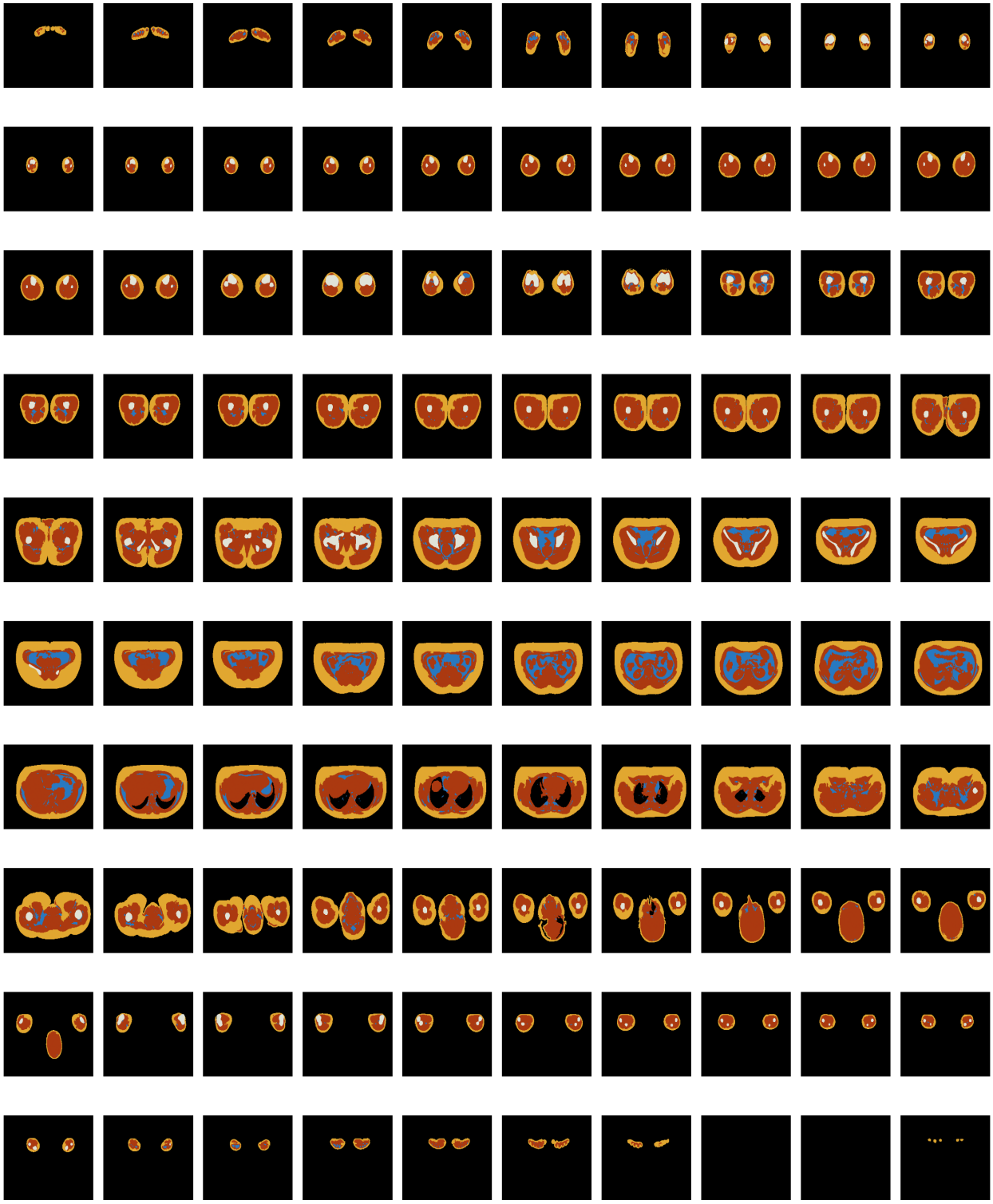

Figure 15. Example segmentation masks of a male subject

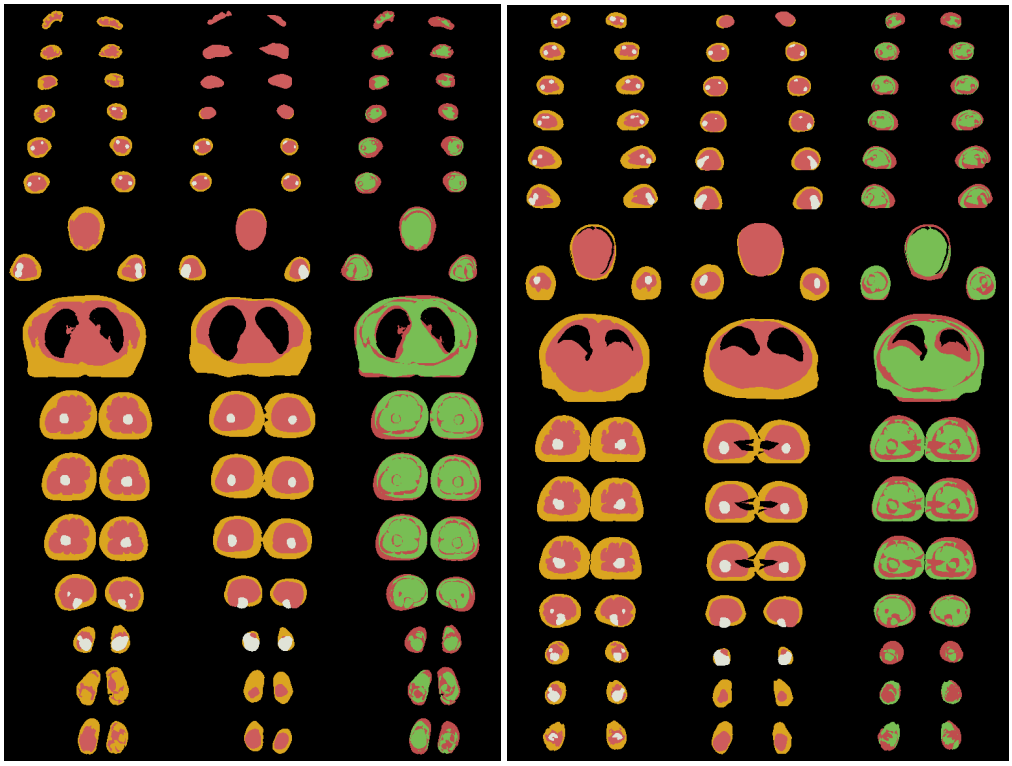

Figure 16. Transverse slices (female): (left) GT tissues, (middle) HIT predictions, (right) accuracy (green correct, red otherwise).

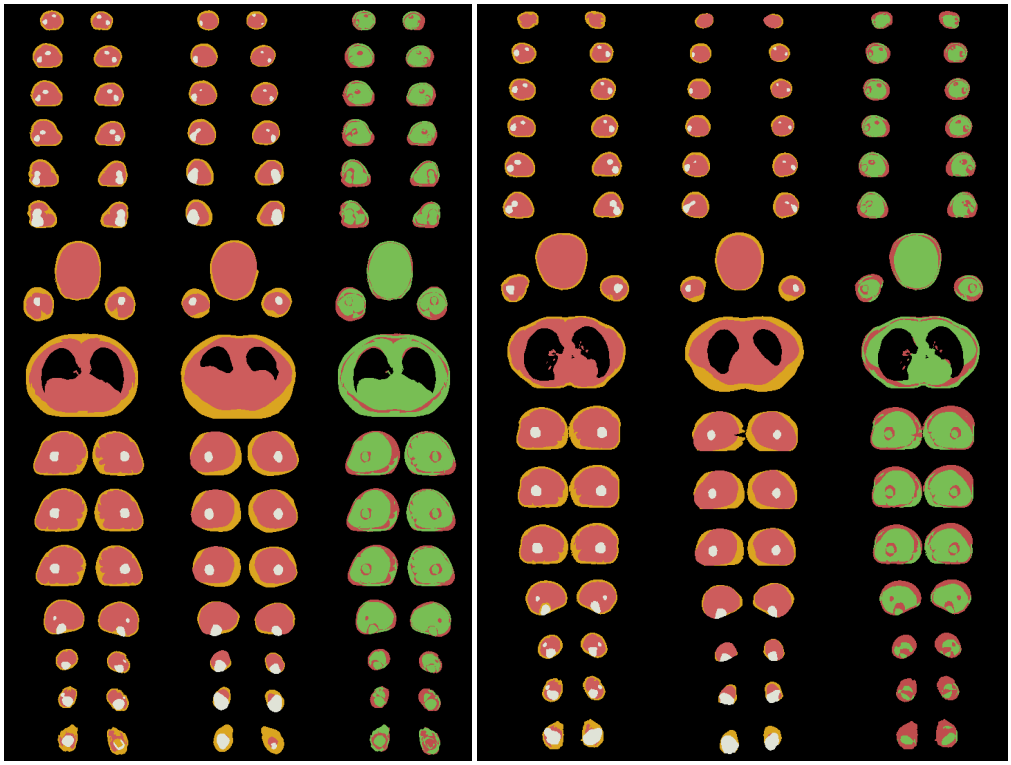

Figure 17. Transverse slices (male): (left) GT tissues, (middle) HIT predictions, (right) accuracy (green correct, red otherwise).

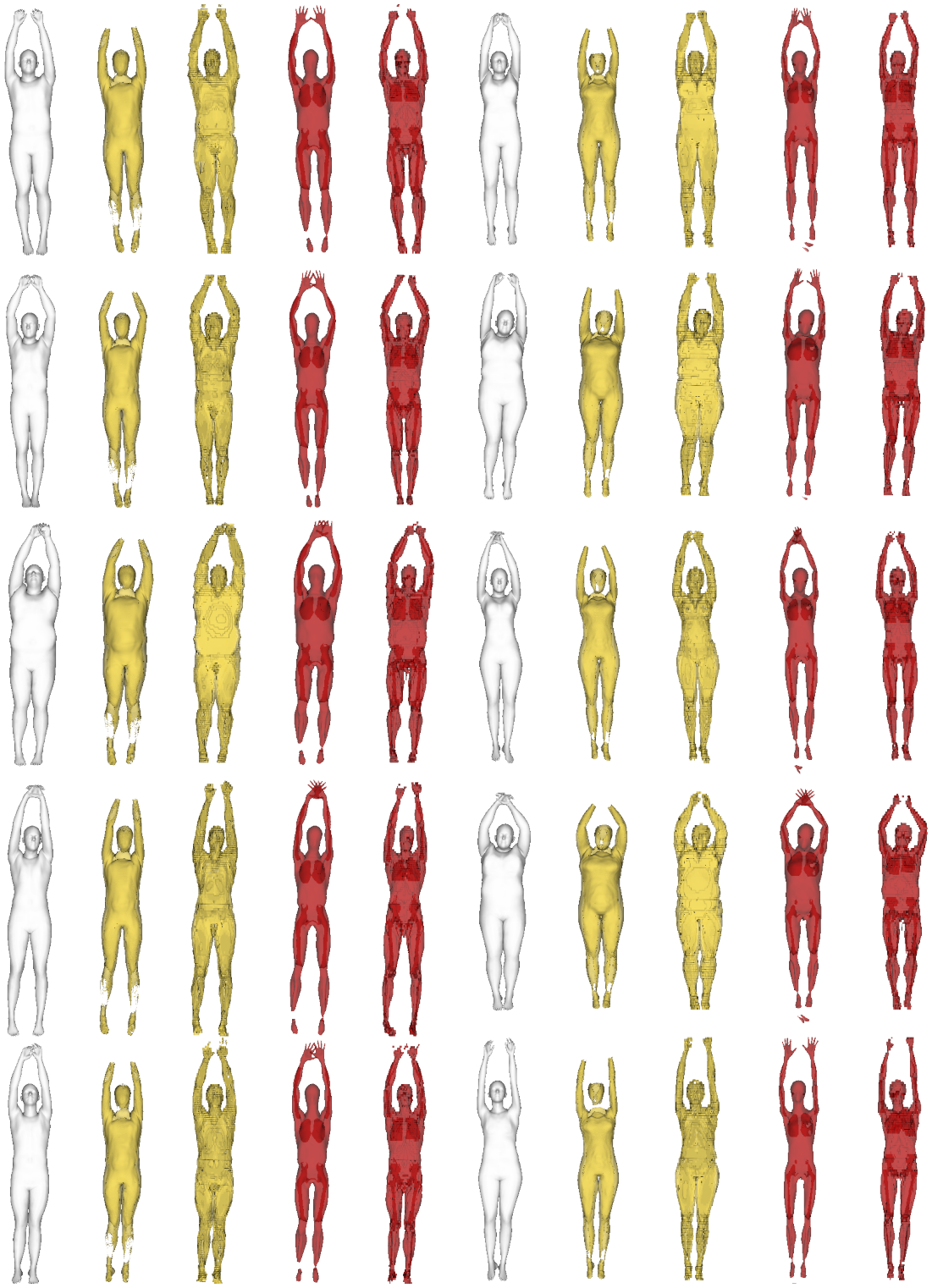

Figure 18. Volumetric tissue predictions for males (left column) and females (right column). From left to right: SMPL fit $\mathbf{S}_i$ (gray), HIT LT prediction, GT LT, HIT SAT prediction, GT SAT.

References

[1] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmenta-

tion. *Nature Methods*, 18:203 – 211, 2020. 1

[2] Alexander Jaus, Constantin Seibold, Kelsey Hermann, Alexandra Walter, Kristina Giske, Johannes Haubold, Jens Kleesiek, and Rainer Stiefelhagen. Towards unifying anatomy segmentation: Automated generation of a full-body ct dataset via knowledge aggregation and anatomical guidelines, 2023. 1

[3] Gerard Pons-Moll, Sergi Pujades, Sonny Hu, and Michael J. Black. ClothCap: Seamless 4D clothing capture and retargeting. *ACM TOG*, 36(4):1–15, 2017. 2

[4] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5), 2023. 1

[5] Christian Würslin, Jürgen Machann, Hansjörg Rempp, Claus Claussen, Bin Yang, and Fritz Schick. Topography mapping of whole body adipose tissue using a fully automated and standardized procedure. *Journal of Magnetic Resonance Imaging*, 31(2):430–439, 2010. 1, 2

[6] Chao Zhang, Sergi Pujades, Michael J. Black, and Gerard Pons-Moll. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In *CVPR*, pages 4191–4200, 2017. 2