HIT: Estimating Internal Human Implicit Tissues from the Body Surface

Marilyn Keller[1], Vaibhav Arora[2], Abdelmouttaleb Dakri[2], Shivam Chandhok[2], Jürgen Machann[3], Andreas Fritsche[3], Michael J. Black[1], and Sergi Pujades[2]

[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France
[3]IDM, Helmholtz Center Munich at the University of Tübingen

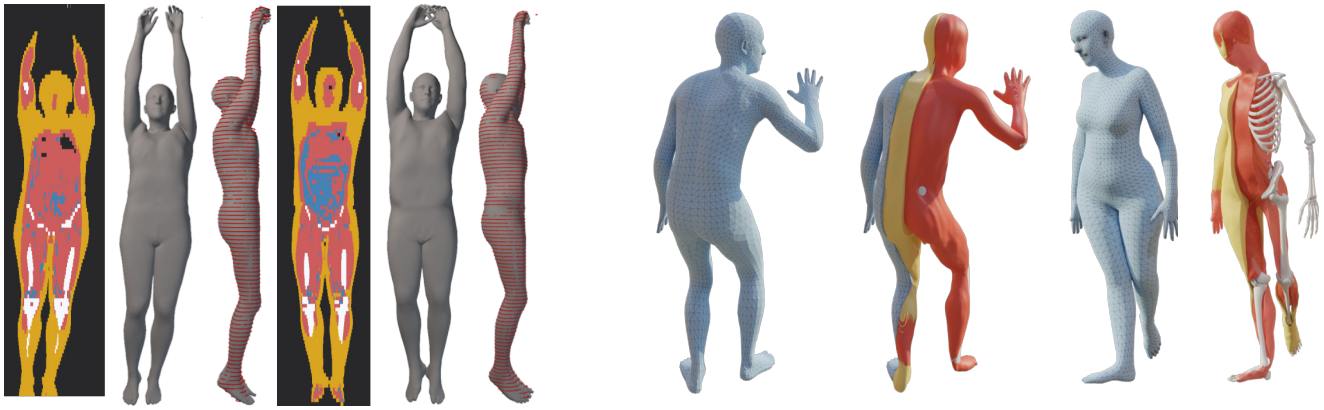

Figure 1. Left half: From volumetric human MRI scans, we learn to segment human internal tissues: subcutaneous adipose tissue (yellow), intra-muscular and visceral adipose tissue (blue), lean tissue (red), and long bones (white). We segment the MRI to extract a point cloud of the human body surface (red rings) to which we fit a human body model (SMPL, gray mesh). From this internal and external paired data, we learn Human Implicit Tissues (HIT), an implicit volumetric model that predicts the type and location of internal tissue. Right half: input body (blue mesh) and predicted tissues: subcutaneous adipose tissue (yellow) and lean tissue (red). We use OSSO [32] to infer the bones.

Abstract

The creation of personalized anatomical digital twins is important in the fields of medicine, computer graphics, sports science, and biomechanics. To observe a subject's anatomy, expensive medical devices (MRI or CT) are required and the creation of the digital model is often time-consuming and involves manual effort. Instead, we leverage the fact that the shape of the body surface is correlated with the internal anatomy; e.g. from surface observations alone, one can predict body composition and skeletal structure. In this work, we go further and learn to infer the 3D location of three important anatomic tissues: subcutaneous adipose tissue (fat), lean tissue (muscles and organs), and long bones. To learn to infer these tissues, we tackle several key challenges. We first create a dataset of human tissues by segmenting full-body MRI scans and registering the SMPL body mesh to the body surface. With this dataset, we train HIT (Human Implicit Tissues), an implicit function that, given a point inside a body, predicts its tissue class. HIT leverages the SMPL body model shape and pose parameters to canonicalize the medical data. Unlike SMPL, which is trained from upright 3D scans, MRI scans are acquired with subjects lying on a table, resulting in significant soft-tissue deformation. Consequently, HIT uses a learned volumetric deformation field that undoes these deformations. Since HIT is parameterized by SMPL, we can repose bodies or change the shape of subjects and the internal structures deform appropriately. We perform extensive experiments to validate HIT's ability to predict a plausible internal structure for novel subjects. The dataset and HIT model are available at https://hit.is.tue.mpg.de *to foster future research in this direction.*

1. Introduction

Creating personalized anatomical digital twins of humans is key in fields such as medicine, sports science, biomechanics, and computer graphics. They play an important role in early diagnosis of diseases, performance evaluation, and gait analysis among others. Yet, creating accurate and detailed patient-specific avatars requires expensive medical imaging devices like Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scanners. Each subject needs to be scanned and the images segmented, often with manual intervention, making the creation of personalized anatomical avatars tedious [21, 28].

In recent years, researchers have shown that the shape of the human body surface is related to the internal body composition [36, 43, 48, 72], leading the way towards fast and non-invasive methods for early screening of body-composition-related pathologies. In addition, recent work shows that predicting internal anatomical structures from the outer surface is also possible [23, 32, 33], paving the way towards the automatic creation of digital twins solely from body surface observations. Body surfaces are now relatively easy to acquire using techniques that fit parametric body models like SMPL [41] to 3D scans or images.

In this work, we focus on three important body tissues: long bones, i.e. femur, tibia, fibula, humerus, ulna, radius, and hips; subcutaneous adipose tissue (SAT), i.e. fat under the skin; and lean tissue (LT), i.e. muscles and organs. From a medical perspective, these tissues are important: an excess of fat with respect to lean tissue is correlated with health risks such as the development of type-II diabetes and cardiovascular disease [22, 47]. From a biomechanics perspective, these tissues have different physical properties and dynamic behaviors, i.e. lean tissue is stiffer, adipose tissue is more elastic, whereas bones are rigid. These differences affect, for example, marker-based motion capture (mocap) systems [40], as markers on soft tissue exhibit artifacts [10]. Thus, having a good estimate of the tissue distribution could improve mocap accuracy and enable the simulation of soft-tissue compression in the apparel industry. In computer graphics, several methods assume [65] or optimize [34, 54] a *soft tissue layer* attached to a rigid structure to simulate physical interactions of the avatars in a virtual world. Also, artificial muscle systems [57] are widely used in character animation but these are complex to design by hand. Having a good estimate of the tissue distribution could improve the anatomic realism of these computational models.

To the best of our knowledge, the precise 3D prediction of these layers inside the body, *given only the outer body surface*, is a novel problem that has not been tackled in the literature. Specifically, our goal is to provide a prediction of the internal structures within a body model like SMPL for arbitrary body shapes and to be able to repose the predicted tissues. See Fig. 1 for a visualization of the resulting 3D representation.

Three main challenges must be overcome to learn a model that predicts the inside of the body from its surface. First, one needs paired observations of the inside and the outside of the body. While medical scanners can capture the raw data, datasets are scarce and usually need to be annotated (segmented). Another challenge is that scanners that can see inside a body, such as MRI, require the subject to be in a lying down position. This position introduces significant shape deformations due to the displacement of the soft tissues through contact with the scanning table. The last challenge is to design a neural network that can be effectively trained to extract the relevant information from the surface of the body to infer the inner tissues. Our approach, *Human Implicit Tissues (HIT)*, addresses these challenges.

To obtain paired *inside* and *outside* data, we acquired a dataset of full-body MRI scans (260 female and 182 male). We start with a small subset (40 female and 40 male) for which we compute initial segmentations of lean and adipose tissues [73]. We curate them and enrich them with manual segmentations of the long bones and then train a nnUNet [27] to segment all tissues in the full dataset. These segmented volumes provide the distributions of the tissues inside the body.

To represent the outer body surface, we use the SMPL body model [41], which lets us model the dependency of the tissue locations inside the body on the pose and shape of a subject. But unlike the surface of the body, an explicit mesh is not appropriate to represent the inner tissues, since their topology can significantly vary between subjects. Implicit functions are particularly well suited to model the occupancy in a given volume [14] and recent work has extensively explored their use in modeling the body surface, clothed bodies and clothing itself, but not for modeling internal body structures. In our approach, given a point inside a body, we predict its tissue class; that is, we formulate the problem as a multi-tissue classification problem. Inspired by recent work on modeling clothed humans and neural rendering [6, 12] we combine implicit and explicit models and learn to map a 3D point inside a SMPL body into a canonical space. This allows us to learn the multi-tissue classification function in the canonical space. The decomposition of the problem into canonicalization and tissue classification offers the advantage of allowing generalization to unseen poses and body shapes. Yet, one more problem remains. Since full-body MRI scans are performed in a prone pose, the bodies exhibit significant deformation which is not modeled by SMPL, as it was learned from upright scans of people. We capture these deformations by optimizing the SMPL mesh vertices to tightly fit the body surface extracted from the MRI. These tight fits allow us to quantify the geometric changes between the SMPL model mesh vertices and their deformed version. Our neural network can thus learn the 3D volumetric displacement of internal soft tissue caused by lying down. This allows us to uncompress the surface and internal structures from lying down to an upright position.

In summary, HIT provides a novel representation of the human body that connects the outer surface to the inner structure. It employs a hybrid of implicit and explicit shape representation and effectively extends the SMPL body model to

infer internal structures that can be reshaped and reposed. The key contributions of HIT are: (a) we formulate the new problem of estimating the 3D structure of human internal tissues from surface observations as a multi-tissue classification problem; (b) we contribute a new dataset, containing the volumetric tissue location inside the body, extracted from real MRI scans, as well as the corresponding SMPL meshes representing the body surface; (c) we learn a volumetric deformation field accounting for the compression between a standing body shape and its counterpart lying prone on an MRI table; (d) we propose a neural implicit formulation to represent the tissue locations inside the body and show that this generalizes to new subjects and new poses; (e) we evaluate and ablate the proposed model on the created dataset. The new dataset and learned models are made available for academic research at https://hit.is.tue.mpg.de.

2. Related work

Motivated by prior work on the prediction of body composition from 3D scans [48, 71, 72], silhouettes [36], or images [43], we go further to predict the location of subcutaneous adipose tissue, lean tissue, and the long bones, solely from the external body surface.

Anatomic models. Early models [20, 66] use the Visible Human data [1], consisting of high-quality images from a cadaver, to build an anatomic model that can be animated. Other works address the creation of detailed personalized anatomic models from data of the hand [2] or the combination of multiple scans of the body [55] into one full-body avatar. Many other personalized anatomic human models have been created, with a focus on physical simulation of the tissues [26, 29, 57, 68, 76], pedagogic purposes [3, 60], or biomechanics [53]. The recent statistical model BOSS [63] includes the skeleton and several organs, but, unlike HIT, it does not model lean and adipose tissues.

Several methods create avatars with **soft tissue** deformation, enabling physics simulation. These typically model the soft tissue as a continuum layer coupled to an articulated skeleton [34, 54, 56, 57, 65]. This layer can be manually defined [57], estimated [56], obtained with an actual scan [54, 65], inferred from skin motion observations [34], or estimated using contact sensors [51]. None of these are validated against clinical data. Some other works have also addressed the modeling of deformation of the hands [37] and feet [7, 50] due to contact with the world.

Anatomy inference. Most internal anatomical structures cannot be inferred from skin observations alone, but some can, such as estimating the skull or jaw from the face shape [26, 77]. Anatomy Transfer [3] deforms an anatomical template model to be consistent with a new body surface. Similarly, Bauer et al. [5] leverage [3] to infer the skeleton inside a body from an RGBD image. Guo et al. [23] estimate the deformation of organs as a patient moves, but the organs' initial shapes are obtained by a scan of the patient. Anatomy Completor [38] can complete the shape of missing organs from the shape of the neighboring ones and OSSO [32] can infer the skeletal bones from the skin surface. Only the last three works [23, 32, 38] evaluate on clinical data. Recently, SKEL [33] goes further and parameterizes the SMPL body model with a biomechanical skeleton that can be inferred from the body surface.

Datasets. Training and evaluation data, i.e. segmented full-body volumetric images, are key for solving this problem. While databases with medical scans [24, 64] exist, their per-pixel automatic segmentation into tissues is not straightforward. To create our paired dataset, we use an MRI protocol [42] and an automatic method [73] to obtain initial segmentations, that we manually curate and enrich.

Human implicit shape models. Implicit shape representations have a long history and have recently become more popular due to the use of neural networks to learn occupancy or signed distance fields. Here we focus on methods that model deformable volumes like the body surface [4, 17, 31, 39, 44, 46, 49, 58, 70], clothed bodies [6, 11, 12, 19, 25, 52, 58, 67, 69, 74], and clothing [16, 59]. Implicit shape representations enable efficient inside/outside tests, allowing the models to take into account the surrounding scene [46, 59], as well as supporting arbitrary topologies. Implicit functions for representing human bodies mainly use three approaches to encode the input query point: part-based, relative, and global. *Part-based* approaches [17, 44, 46] learn the occupancy in each part's canonical space whereas *relative encoding* approaches encode a point's occupancy with respect to joint locations [75], sparse skin vertices [16], or detected keypoints [45]. HIT uses a *global* approach, which learns occupancy in a canonical pose (namely a "star" pose). SCANimate [58], Meta Avatar [69] and ARAH [70] learn a subject-specific avatar in a canonical space and train a neural network to predict the skinning weights of any point in space. This learned inverse LBS (Linear Blend Skinning) lets them transform a point to the canonical pose space before querying the occupancy. In gDNA [12], a multi-subject occupancy model is learned in a canonical pose. A root-finding algorithm [13] enables unposing points, and a displacement field that maps shaped points to the canonical space is learned. We leverage a pretrained SMPL occupancy network to generalize to new shapes and poses, and add a new module to model the body compression and pose-dependent deformations.

A few works jointly model two surfaces, e.g. hand object interaction [30] or multiple clothes [59], by adding interpenetration losses. Our multi-tissue classification formulation naturally avoids reasoning about interpenetration.

3. Human Tissue Data

A crucial requirement for learning the relationship between the body's inner tissues and the body surface is a structured dataset of paired observations. We use MRI scans of human subjects that we segment into several tissues. Each pixel of the MRI volume is classified as Bone Tissue (BT), Lean Tissue (LT), Intra-Muscular and Visceral Adipose Tissue (IMVAT),

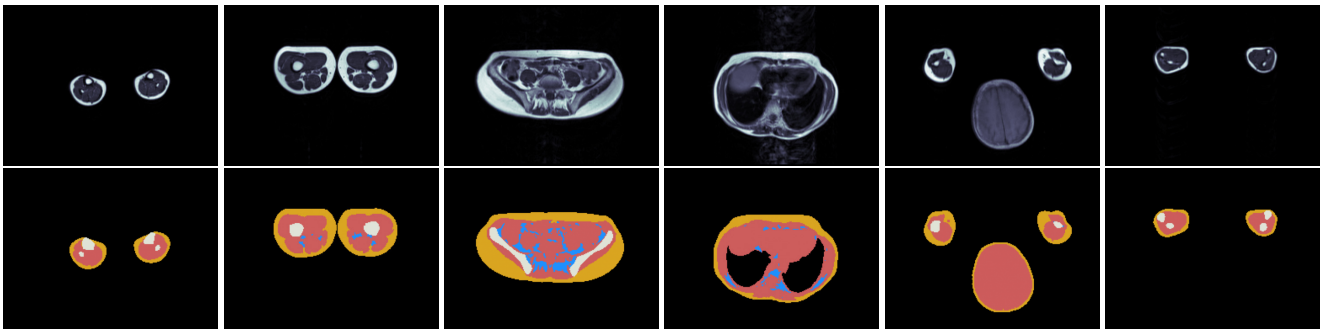

Figure 2. First row: input MRI images. Second row: segmentation results from the nnU-Net. Tissues color-code: bone (white), lean (red), subcutaneous adipose (yellow), intra-muscular and visceral adipose (blue), empty (black). From left to right: calf, thighs, hips, chest, head and arms, forearms.

Subcutaneous Adipose Tissue (SAT), or Empty (E) (see Fig. 2 with segmentation examples). From the segmented volume, we extract the subject's body surface as a point cloud and fit the SMPL [41] body model to it. We also compute tight fits of the SMPL body mesh that capture the flattened body shape extracted from the MRI (see Fig. 3). In this way, we create a dataset of paired observations of the inner body tissues together with the human body surface.

3.1. MRI segmentation

MRI scans dataset. We work with 442 scans (260 female, 182 male) acquired with a 1.5 T scanner (Magnetom Sonata, Siemens Healthcare) following a standardized protocol for whole body adipose tissue topography mapping [42]. All subjects gave prior informed written consent and the study was approved by the local ethics board. Each scan has around 110 slices, slightly varying depending on the height of the subject. The slice resolution is 256×192, with an approximate voxel size of $2 \times 2 \times 10$ mm.

Tissue definitions. Given an input MRI image (slice), our goal is to classify the tissue type of each pixel. For the Bone Tissue (BT) we focus on the long bones: femur, tibia, fibula, humerus, ulna, radius, and hips. We do not segment smaller bones, such as vertebrae, ribs, or phalanges, as, with the limited resolution of the MRI images, it is difficult to consistently identify them in the images. The muscles and organs are segmented as Lean Tissue (LT). The Subcutaneous Adipose Tissue (SAT) and the Intra-Muscular and Visceral Adipose Tissue (IMVAT) denote the human fat; SAT is located directly under the skin, whereas IMVAT is located inside the muscles and around the organs. MRI pixels where no tissue is detected are classified as Empty (E). Empty areas include the background outside the body, the lungs, skull cortical bone, and other cavities inside the body.

Human tissue segmentation. To segment the whole MRI dataset into tissues, we use a *human-in-the-loop* approach similar to SAM [35]. We leverage initial automatic segmentations [73] and manual annotations to train and refine a nnU-Net [27] with the help of human supervision. The full description of the segmentation process is provided in Sup. Mat. 1. In the remainder of the paper, these segmentations are treated as the ground truth internal tissues. To obtain the external shape, we identify the body contour in the segmented images and by using the MRI 3D metadata, we extract a *skin* point cloud $\mathcal{S}$.

3.2. SMPL fits to the skin point cloud

For each subject i, we fit the SMPL body mesh [41] to the *skin* point cloud $\mathcal{S}_i$. The benefits of doing so are two-fold. First, we obtain the per-subject SMPL mesh $\mathbf{S}_i$ and shape and pose parameters ($\boldsymbol{\beta}_i \in \mathbb{R}^{10}$,$\boldsymbol{\theta}_i \in \mathbb{R}^{69}$). The pose $\boldsymbol{\theta}_i$ will be used to unpose the MRI into a canonical space, and the shape $\boldsymbol{\beta}_i$ to model the subject's shape. Second, by allowing the SMPL body mesh to deform and capture the flattened subject's shape in the MRI, we can quantify the displacement field $\mathbf{d}_i$ that the body undergoes when lying in the MRI. This 3D field allows us to effectively model compression.

While SMPL is commonly fitted to a wide variety of data, our situation is unique. SMPL is learned using body scans of people in upright standing poses, thus the learned shape and pose space can not capture the soft-tissue deformation that the body undergoes in the prone position created by lying on the MRI table. As a consequence, directly fitting the SMPL body model to the MRI point cloud leads to an incorrect subject shape. To solve this problem we propose a two-step fitting approach. First, we compute an initial SMPL tight fit that allows us to compute the volume of the subject. Then, with a volume-preserving constraint, we compute the best SMPL model pose and shape parameters ($\boldsymbol{\beta}_i$,$\boldsymbol{\theta}_i$) that match the MRI point cloud, obtaining $\mathbf{S}_i$. We then allow the vertices to deform, resulting in a deformed mesh $\mathbf{F}_i$ that tightly fits the MRI surface. The two-step optimization details are provided in Sup. Mat. 2 and Fig. 3 illustrates the obtained results.

3.3. Human Implicit Tissues (HIT) Dataset

The new HIT dataset contains, for each subject i: a) the volumetric image segmented into BT, LT, SAT, IMVAT, and E, b) the image MRI center and per-pixel spacing, to transform indices from the volumetric image into 3D metric locations, and c) the

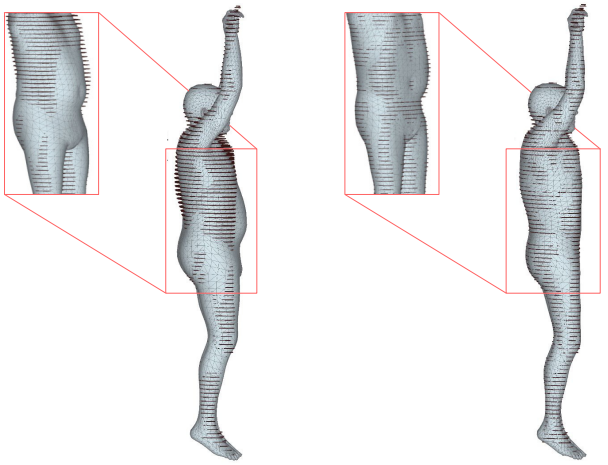

Figure 3. SMPL fits to MRI point clouds. The SMPL fit $\mathbf{S}_i$ (left) does not capture the flattened shape, the tight fit $\mathbf{F}_i$ (right) does.

skin point cloud $\mathcal{S}_i$, the fitted SMPL model mesh $\mathbf{S}_i$ represented by its parameters $(\boldsymbol{\theta}_i, \boldsymbol{\beta}_i)$, as well as the SMPL tight fit $\mathbf{F}_i$. From b) we compute the compression displacements $\mathbf{d}_{\text{comp}} \in \mathbb{R}^{N_v}$, between the $\mathbf{S}_i$ and $\mathbf{F}_i$ vertices, where $N_v = 6890$ are the number of SMPL vertices. Note that the original MRI images are not included. This dataset is made available for academic research.

Final three-layer representation. The Intra-Muscular and Visceral Adipose Tissue (IMVAT) segmented in the MRI images is sparsely located around the muscles and abdominal organs (blue in Figs. 1 and 2). As its precise 3D location highly varies among people, we leave the precise localization of IMVAT for future work and infer 3 layers of tissue: Bone Tissue (BT), Lean Tissue and Intra-Muscular and Visceral Adipose Tissue (LT + IMVAT) and Subcutaneous Adipose Tissue (SAT), which are used for anatomic digital twins [34, 54, 56, 65]. That is, in the remainder of the paper, we merge IMVAT with the surrounding LT structures and refer to them together as LT.

4. HIT method

Problem statement. We formalize the inference of the tissues inside the body as a 4-tissue classification problem (BT, LT (+ IMVAT), SAT, E). HIT learns an implicit function that takes as input SMPL shape and pose parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ and a 3D point $\mathbf{x}$, and outputs the tissue class at that point.

4.1. HIT spaces

To learn the tissue occupancy, HIT warps the data from the input MRI space into a canonical space. These warps are defined between four spaces, illustrated in Fig. 4. The *canonical space* is where the SMPL template mesh, $\mathbf{T}$, in a "star" pose, is defined. Points in the canonical space are indexed by $\mathbf{x}^c$. The *shaped space* is where additive offsets, $\mathbf{d}_\beta \in \mathbb{R}^{N_v \times 3}$, controlled by the shape $\boldsymbol{\beta}$ of the subject, are applied to the template. We

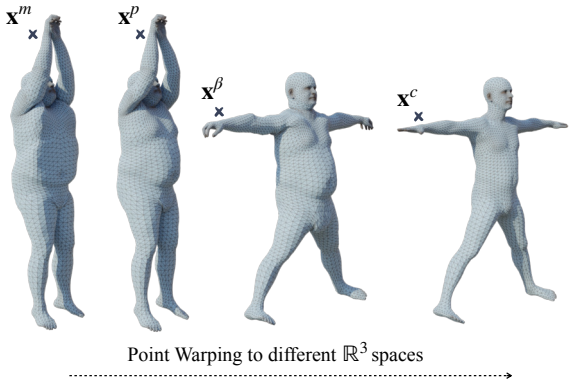

Figure 4. HIT defines four $\mathbb{R}^3$ spaces. A point $\mathbf{x}^m$ in the original MRI space corresponds to $\mathbf{x}^p$ in the posed space, $\mathbf{x}^\beta$ in the shaped space, and $\mathbf{x}^c$ in the canonical space.

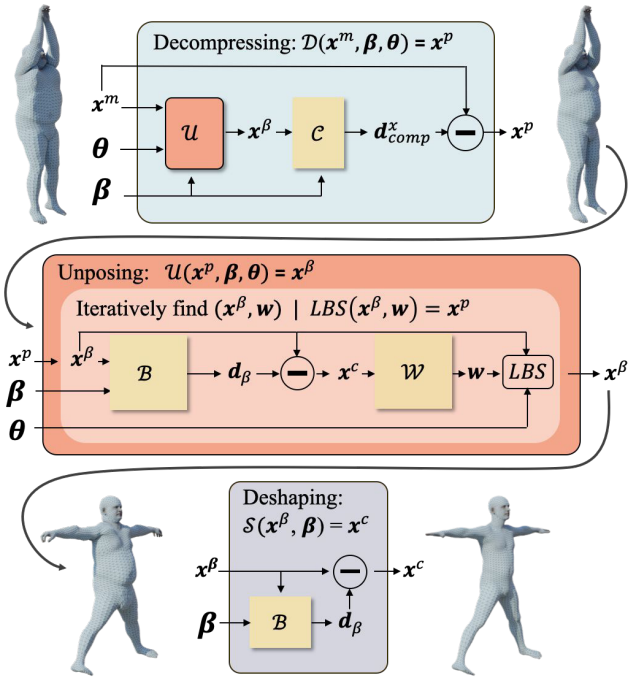

Figure 5. HIT modules $(\mathcal{D}, \mathcal{U}, \mathcal{S})$ and networks $(\mathcal{C}, \mathcal{B}, \mathcal{W})$ to warp points between spaces.

denote points there as $\mathbf{x}^\beta = \mathbf{x}^c + \mathbf{d}_\beta$. The shaped points can then be posed through linear blend skinning into the *posed space* $\mathbf{x}^p = \text{LBS}(\mathbf{x}^\beta, \mathbf{w}, \theta)$, where $\mathbf{w} \in \mathbb{R}^{N_p}$ is a vector of blend-weights, $N_p = 24$ is the number of SMPL body parts, and θ represents the pose parameters. Finally, to model the MRI table compression on the body, we define volumetric offsets $\mathbf{d}^x_{\text{comp}} \in \mathbb{R}^3$ and denote points in the *original MRI space* $\mathbf{x}^m = \mathbf{x}^p + \mathbf{d}^x_{\text{comp}}$. Note that points in all spaces live in $\mathbb{R}^3$ and can be inside, outside, or on the SMPL surface.

4.2. HIT architecture

Our architecture is composed of 4 building blocks. Three modules enable warping an MRI point into the canonical space $\mathbf{x}^c = (\mathcal{S} \circ \mathcal{U} \circ \mathcal{D})(\mathbf{x}^m, \boldsymbol{\beta}, \boldsymbol{\theta})$ by Decompressing ($\mathcal{D}$), Unposing ($\mathcal{U}$) and Deshaping ($\mathcal{S}$). The warping architectures are illustrated in Fig. 5. Once the warped point is in the canonical space, the network $\mathcal{T}(\mathbf{x}^c)$ predicts its tissue class.

Deshaping module. Given a shape $\boldsymbol{\beta}$, the Deshaping module $\mathcal{S}$ transforms shaped points into canonical points, i.e. $\mathcal{S}(\mathbf{x}^\beta, \boldsymbol{\beta}) = \mathbf{x}^c$. In the module, the function $\mathcal{B}$ predicts the offsets $\mathcal{B}(\mathbf{x}^\beta, \boldsymbol{\beta}) = \mathbf{d}_\beta$, which are subtracted from $\mathbf{x}^\beta$ to obtain $\mathbf{x}^c$ (see Fig. 5 bottom diagram).

Unposing module. Given shape and pose parameters, the Unposing module $\mathcal{U}$ warps points from the *posed space* into the *shaped space*: $\mathcal{U}(\mathbf{x}^p, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{x}^\beta$. Similar to Chen et al. [12], this module uses two MLPs: $\mathcal{B}(\mathbf{x}^\beta, \boldsymbol{\beta}) = \mathbf{d}_\beta$ defined in the previous paragraph and the function $\mathcal{W}(\mathbf{x}^c) = \mathbf{w}$ which predicts the skinning weights $\mathbf{w}$ of a point in the canonical space (see Fig. 5 middle diagram). Unposing points inside or outside a posed SMPL mesh is challenging because the skinning weights are only defined on the SMPL surface. Chen et al. use a root-finding algorithm [9] that finds candidate points $\{\mathbf{x}_i^\beta\}$ for a given posed point $\mathbf{x}^p$. Then their SMPL occupancy prediction is used to decide on the best candidate. However, this is not applicable in our multi-tissue case; i.e. if a point has two roots, one in LT and one in BT, there is no way to know which one is correct. To overcome this limitation, we initialize the root finding with skinning weights $\mathbf{w}$, fetched from the closest SMPL vertex. This allows the iterative algorithm to converge to the skinning weights that properly unpose the point.

Decompression module. To model the body deformation displacement induced by the MRI table, we learn a Decompression module $\mathcal{D}$ that maps points in the MRI space to the posed space: $\mathcal{D}(\mathbf{x}^m, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{x}^p$. We do so by learning to predict the volumetric compression displacements $\mathbf{d}^x_{\text{comp}}$, which generalize the computed $\mathbf{d}_{\text{comp}}$ on the SMPL surface. However, learning the volumetric body decompression is challenging: a 3D point $\mathbf{x}^m$ can represent a different anatomic region for two different subjects. Thus, instead of learning to predict displacements in the MRI space, we first unpose $\mathbf{x}^m$ into the *shaped space* $\mathbf{x}^\beta$ and predict $\mathcal{C}(\mathbf{x}^\beta, \boldsymbol{\beta}) = \mathbf{d}^x_{\text{comp}}$ so that $\mathbf{x}^m + \mathbf{d}^x_{\text{comp}} = \mathbf{x}^p$ (see Fig. 5 top diagram). The *shaped space* has a natural shape consistency which helps predict the compression.

Multi-Tissue network. Once points are in the *canonical space*, HIT uses an MLP to predict the point tissue class $\mathcal{T}(\mathbf{x}^c) = \{\text{E,LT,SAT,BT}\}$.

Sup. Mat. 3 provides the network implementation details.

4.3. Training, losses and sampling

Training. To train HIT we proceed in 3 steps. First $\mathcal{B}$ and $\mathcal{W}$ are pre-trained by randomly sampling shaped and posed SMPL bodies. In parallel, $\mathcal{C}$ is pre-trained using the computed $\mathbf{d}_{\text{comp}}$ (see Sec. 3). Then, the weights of $\mathcal{C}$ are frozen, and $\mathcal{B}$, $\mathcal{W}$, $\mathcal{T}$ and $\mathcal{M}$ are jointly trained on the HIT dataset. We note the network's trainable weights as ψ_* and use the subscript $*$ to refer to the network name, i.e. $\psi_\mathcal{B}$ are the weights of the network $\mathcal{B}$. To train our architecture we minimize the following losses.

Deshaping loss. Let $\mathbf{x}_v^\beta$ be a vertex of the shaped mesh SMPL($\boldsymbol{\beta}$), and $\mathbf{x}_v^c$ be the corresponding vertex on the SMPL template mesh. To train the weights $\psi_\mathcal{B}$, we enforce the predicted displacement to match the reversed SMPL's $\boldsymbol{\beta}$ offset at the body surface level by minimizing

$$l_s(\psi_\mathcal{B}) = \text{MSE}_v(\mathcal{B}(\mathbf{x}_v^\beta, \boldsymbol{\beta}) - (\mathbf{x}_v^c - \mathbf{x}_v^\beta)), \tag{1}$$

where MSE_v is the mean square error over sampled points.

Skinning weight loss. To train the $\psi_\mathcal{W}$ weights we enforce the predicted skinning weights to be consistent with the SMPL ones by minimizing

$$l_w(\psi_\mathcal{W}) = \text{MSE}_v(\mathcal{W}(\mathbf{x}_v^c) - \mathbf{w}_v), \tag{2}$$

where $\mathbf{w}_v \in \mathbb{R}^{N_p}$ denotes the SMPL's skinning weights of the SMPL template vertex $\mathbf{x}_v^c$.

Decompression loss. To train $\mathcal{C}$ to predict a displacement that is similar to the one between $\mathbf{x}_v^m$ and $\mathbf{x}_v^p$, we minimize

$$l_c(\psi_\mathcal{C}) = \text{MSE}_v(\mathcal{C}(\mathcal{U}(\mathbf{x}_v^m, \boldsymbol{\beta})), \mathbf{x}_v^p - \mathbf{x}_v^m). \tag{3}$$

Multi-tissue loss. Given a point sampled inside the compressed body $\mathbf{x}_k^m$ and canonicalized to $\mathbf{x}_k^c$, we train $\mathcal{T}$ to predict the correct tissue label. This is done by optimizing the weighted cross-entropy loss between the tissue predictions and the training data, where weights are inversely proportional to the tissue sample size.

Sampling strategy. The MRI scans have a discrete volumetric representation. To avoid aliasing artifacts, we train HIT with points $\mathbf{x}_k^m$ sampled at the center of voxels. Note that this only holds for querying the GT volume; once the implicit function is learned, one can sample arbitrary locations inside the body and extract smooth tissue volumes.

Uniformly sampling the MRI voxel centers means that the canonical space is not uniformly sampled, e.g. the space between the legs is wider in the canonical space. Thus, we also uniformly sample points outside the SMPL template mesh in the canonical space and classify them as E .

As the MRI resolution is low for hands, we force these parts to be always predicted as LT by uniformly sampling points in the canonical space inside the hands' bounding boxes.

5. Experiments

To evaluate HIT we split the data into 80% train, 10% validation and 10% test sets (females 201/25/26, males 136/17/16), and we train separate models for males and females as the literature reports significant differences in body composition [8].

	Female						Male					
	LT		SAT		BT		LT		SAT		BT	
	D.S. ↑	Δ% ↓	D.S. ↑	Δ% ↓	D.S. ↑	Δ% ↓	D.S. ↑	Δ% ↓	D.S. ↑	Δ% ↓	D.S. ↑	Δ% ↓
Chance	51.4	4.9	40.2	6.7	3.9	0.7	60.3	5.1	31.8	6.5	4.1	0.8
HIT	77.8	4.0	57.7	9.4	45.5	0.6	81.0	5.1	54.7	6.5	52.2	0.7

Table 1. Quantitative evaluation for females and males on the three tissues (LT, SAT, BT). DICE Score (D.S.) - higher is better and Δ % is the relative difference in the quantity prediction in percent - lower is better.

5.1. Internal tissues evaluation

Since we address a novel problem, to the best of our knowledge, no prior work can be directly used for comparison: e.g. OSSO [32] solely predicts the bone structures. To have a numeric calibration for the multi-tissue problem, we propose a *Chance* baseline, which for each queried point predicts the tissues [E, LT, SAT, BT] with probabilities 0.03, 0.52, 0.41, 0.04 and 0.04, 0.60, 0.32, 0.04 for females and males respectively. These values follow the average percentage of each tissue in the training set.

To quantitatively evaluate the HIT architecture, we report mean Dice scores [18] for each predicted tissue on the test set. Additionally, we compute the relative error of the predicted tissue quantity, by computing $\Delta = |V_{pred} - V_{GT}|/V_B$, where $V_{pred}, V_{GT}, V_B \in \mathbb{R}$ are the volumes of the predicted tissues, ground-truth tissues, and full body respectively.

Table 1 shows that the HIT Dice scores are significantly better than the Chance baseline. Additionally, Fig. 6, Fig. 7, and Sup. Mat. 5.2 present qualitative results of the tissue predictions on transverse planes and in 3D. As visible in Fig. 6 the predicted inner tissues are consistent across the body and exhibit plausible compression. In addition, most errors arise at the tissue interfaces. The 3D visualization in Fig. 7 and Sup. Mat. 5.3 also shows that the obtained 3D meshes are visually consistent with the GT ones. Let us note that extracting per-tissue 3D meshes from a multi-label function is not trivial. In Sup. Mat. 4 we detail how we do this.

To further put HIT bone predictions in context, we numerically compare to OSSO [32] by measuring the distance between the segmented bones and the predicted ones. We show in Sup. Mat. 5.4 that HIT outperforms OSSO in terms of mean absolute error.

Regarding the metric Δ in Tab. 1, measuring the predicted tissue volume, it is interesting to note that HIT is on pair with the Chance baseline (or even under-performs it for female SAT). The Chance Δ metric is quantifying the error of predicting the mean volume, i.e. the variability in the dataset volumes. The similar HIT Δ metric points out that, in fact HIT is predicting an average tissue quantity. This is not surprising, as HIT predictions are conditioned on 10 SMPL shape parameters, which cannot capture all individual shape details. In Sec. 6 we discuss how future work could improve the current predictions.

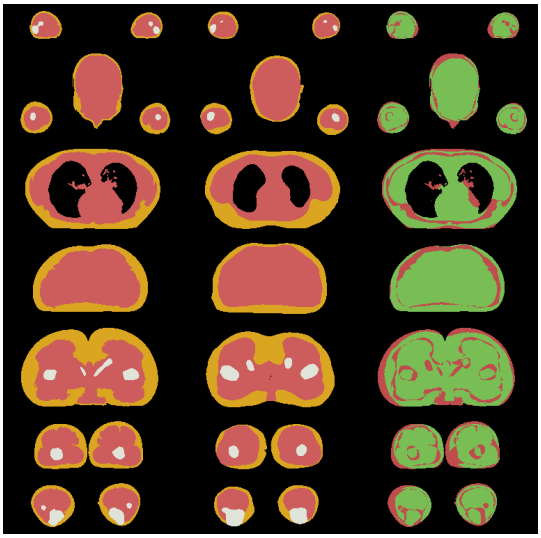

Figure 6. Transverse slices (female): (left) GT tissues, (middle) HIT predictions, (right) accuracy (green correct, red otherwise).

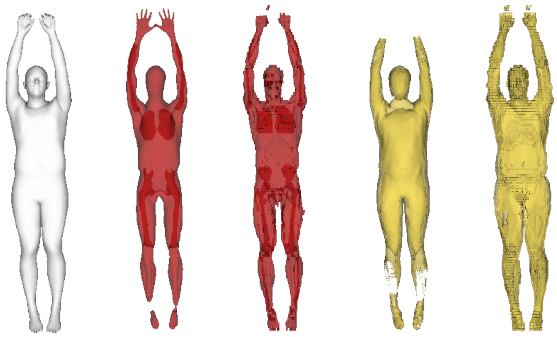

Figure 7. From left to right: SMPL fit $\mathbf{S}$ (gray), HIT LT prediction, GT LT, HIT SAT prediction, GT SAT.

5.2. Generalisation to new body shapes and poses

To generalize to new body shapes and poses, correctly modeling the compression is key. We thus ablate the HIT compression module by learning HIT_{ncmp}, a HIT variant without the compression module. As visible in Fig. 8, HIT_{ncmp} can not generalize and generates compressed results for standing bodies. See Sup. Mat. 5.1 for a visualization of the HIT learned

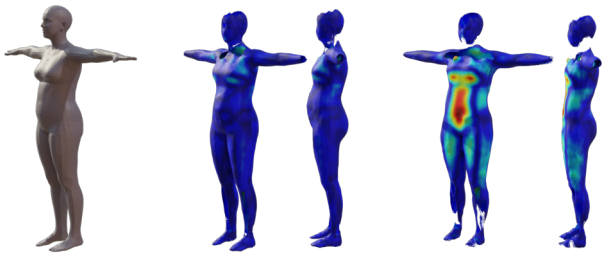

Figure 8. Prediction of the SAT occupancy for the mean SMPL body in T-pose. Left: SMPL mesh, middle HIT, right HIT_{ncmp}. Note how the compression remains in the inference for HIT_{ncmp}. Color code: distance to the SMPL mesh (blue=0 cm, red=5cm).

Figure 9. Prediction of the lean tissues for different body shapes. Varying (left) the first component of SMPL - related to size on females - and (right) the second one - related to weight - for males in the range {+2, -2}. The predicted tissues consistently adapt to the new shapes, leading to visually plausible predictions.

displacement and compression fields.

New shapes. To explore how the HIT tissue predictions generalize to new body shapes we vary the input SMPL shape components related to height and weight [41], in the range of $[-2, +2]$. Figure 9 shows that HIT predicts plausible tissues that vary in accordance with the person's shape.

New poses. Learning in the SMPL canonical space enables querying skinning weights for each point inside the body, and thus the inferred tissue volumes can be easily reposed. Figure 1 shows the reposed tissue volumes for two different poses. Note that at inference time, the compression network is bypassed to yield non-compressed body shapes.

6. Discussion and Conclusion

HIT introduces the new problem of inferring the human tissues inside of a body from a surface observation only. This work is relevant for medicine, sports science, biomechanics, and computer graphics as it can ease the creation of personalized anatomic digital twins. We formulate the problem as a multi-tissue classification task and learn an implicit function that takes, as input, a query point, and SMPL pose and shape parameters and predicts its tissue class. To learn HIT, we create a dataset of paired full-body volumetric segmented MRI scans and SMPL meshes capturing the body surface shape. We evaluate and ablate the proposed model on the created dataset, showing the quality of the HIT reconstructed tissues.

To the best of our knowledge, this is the first work to predict the volumetric composition of the tissues inside the body from an outer surface observation. We show that it is possible to predict health-relevant tissues inside the body, and most importantly, we give a first quantification of the accuracy of these predictions against medical data. To foster future research on this topic, the dataset and HIT model are made available for academic purposes at https://hit.is.tue.mpg.de.

Limitations. While we learn to uncompress the soft tissue from the body surface, we can not validate the behavior of this uncompression inside the body for now. This would require specific data or simulation. Also, the current use of HIT for a precise prediction of the tissue percentage remains limited. Our work does not explore the precise location of Intra-Muscular and Visceral Adipose Tissue (IMVAT). Its structure is very sparse and, while its volume quantification is relevant in medicine, it is not clear whether its exact pixel-wise location is. Still, the HIT dataset provides the IMVAT segmentation masks, which will enable future exploration.

Future Work. The distribution of the lean and subcutaneous adipose tissues inside the body is relatively well structured; i.e. neighbouring points inside the body will, most of the time, be of the same tissue. While these structures naturally emerge in our results, we did not use any explicit loss to enforce the tissue's spatial consistency. One direction to improve the accuracy of the predictions could be to study structural losses, for instance, adversarial networks on slices of the predicted volume, enforcing the structure of multiple points at once. Moreover, our approach uses a SMPL body mesh for inference and does not include the individual features present in the point cloud of the scan. Future methods could explore how to integrate this information to improve the predictions.

Risks. Our work has associated privacy risks. Today, many methods can estimate accurate SMPL bodies from images [15, 61], and HIT can infer their internal tissues. As a good estimate of the body composition relates to health risks [22, 36, 62], HIT could allow the estimation of health risks from a single image of a person. This is valuable as an early diagnostic tool when used with the person's knowledge but could turn into a risk if it is used without consent.

Acknowledgments. Marilyn Keller was supported by the International Max Planck Research School for Intelligent Systems. Shivam Chandhok and Sergi Pujades' work was funded by the ANR JCJC SEMBA project. We thank Kate Duquesne and Emmanuel Audenaert for the initial interactive bone segmentations. MJB CoI Discolosure: https://files.is.tue.mpg.de/black/CoI_CVPR_2024.txt

References

[1] Michael J. Ackerman. The visible human project. *Proceedings of the IEEE*, 86(3):504–511, 1998. 3

[2] Irene Albrecht, Jörg Haber, and Hans-Peter Seidel. Construction and animation of anatomically based human hand models. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 98–109, 2003. 3

[3] Dicko Ali-Hamadi, Tiantian Liu, Benjamin Gilles, Ladislav Kavan, François Faure, Olivier Palombi, and Marie-Paule Cani. Anatomy transfer. *ACM TOG*, 32(6):1–8, 2013. 3

[4] Thiemo Alldieck, Hongyi Xu, and Cristian Sminchisescu. imGHUM: Implicit generative models of 3D human shape and articulated pose. In *CVPR*, pages 5461–5470, 2021. 3

[5] Armelle Bauer. *Modélisation anatomique utilisateur-spécifique et animation temps-réel. Application à l'apprentissage de l'anatomie.* Theses, Université Grenoble Alpes, 2016. 3

[6] Bharat Lal Bhatnagar, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. LoopReg: Self-supervised learning of implicit surface correspondences, pose and shape for 3D human mesh registration. *Advances in Neural Information Processing Systems*, 33:12909–12922, 2020. 2, 3

[7] Oliver Boyne, James Charles, and Roberto Cipolla. FIND: An unsupervised implicit 3D model of articulated human feet. In *BMVC*, 2022. 3

[8] Miriam A. Bredella. Sex differences in body composition. *Sex and gender factors affecting metabolic homeostasis, diabetes and obesity*, pages 9–27, 2017. 6

[9] Charles G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19: 577–593, 1965. 6

[10] Jessa M. Buchman-Pearle and Stacey M. Acker. Estimating soft tissue artifact of the thigh in high knee flexion tasks using optical motion capture: Implications for marker cluster placement. *Journal of Biomechanics*, 127:110659, 2021. 2

[11] Xu Chen, Yufeng Zheng, Michael J. Black, Otmar Hilliges, and Andreas Geiger. SNARF: Differentiable forward skinning for animating non-rigid neural implicit shapes. In *ICCV*, 2021. 3

[12] Xu Chen, Tianjian Jiang, Jie Song, Jinlong Yang, Michael J. Black, Andreas Geiger, and Otmar Hilliges. gDNA: Towards generative detailed neural avatars. In *CVPR*, 2022. 2, 3, 6

[13] Xu Chen, Tianjian Jiang, Jie Song, Max Rietmann, Andreas Geiger, Michael J. Black, and Otmar Hilliges. Fast-SNARF: A fast deformer for articulated neural fields. *IEEE TPAMI*, 45(10): 11796–11809, 2023. 3

[14] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, pages 5939–5948, 2019. 2

[15] Vasileios Choutas, Lea Müller, Chun-Hao P. Huang, Siyu Tang, Dimitrios Tzionas, and Michael J. Black. Accurate 3D body shape regression using metric and semantic attributes. In *CVPR*, pages 2718–2728, 2022. 8

[16] Enric Corona, Albert Pumarola, Guillem Alenya, Gerard Pons-Moll, and Francesc Moreno-Noguer. Smplicit: Topology-aware generative model for clothed people. In *CVPR*, pages 11875–11885, 2021. 3

[17] Boyang Deng, John P. Lewis, Timothy Jeruzalski, Gerard Pons-Moll, Geoffrey Hinton, Mohammad Norouzi, and Andrea Tagliasacchi. NASA: Neural articulated shape approximation. In *ECCV*. Springer, 2020. 3

[18] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. 7

[19] Yao Fen, Jinlong Yang, Marc Pollefeys, Michael J. Black, and Timo Bolkart. Capturing and animation of body and clothing from monocular video. In *ACM TOG*, 2022. 3

[20] Nikhil Gagvani and Deborah Silver. Animating volumetric models. *Graphical models*, 63(6):443–458, 2001. 3

[21] Benjamin Gilles, Lionel Reveret, and Dinesh K. Pai. Creating and animating subject-specific anatomical models. *Computer Graphics Forum*, 29(8):2340–2351, 2010. 2

[22] Scott M. Grundy. Obesity, metabolic syndrome, and cardiovascular disease. *The Journal of Clinical Endocrinology & Metabolism*, 89(6):2595–2600, 2004. 2, 8

[23] Hengtao Guo, Benjamin Planche, Meng Zheng, Srikrishna Karanam, Terrence Chen, and Ziyan Wu. SMPL-A: Modeling person-specific deformable anatomy. In *CVPR*, pages 20814–20823, 2022. 2, 3

[24] Edgar Heather J. H., Shamsi Daneshvari Berry, Emily Moes, Natalie L. Adolphi, Patrick G. Bridges, and Kurt B. Nolte. New mexico decedent image database. Office of the Medical Investigator, University of New Mexico, 2020. 3

[25] Zeng Huang, Yuanlu Xu, Christoph Lassner, Hao Li, and Tony Tung. Arch: Animatable reconstruction of clothed humans. In *CVPR*, pages 3093–3102, 2020. 3

[26] Alexandru-Eugen Ichim, Petr Kadleček, Ladislav Kavan, and Mark Pauly. Phace: Physics-based face modeling and animation. *ACM TOG*, 36(4):1–14, 2017. 3

[27] 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 segmentation. *Nature Methods*, 18:203 – 211, 2020. 2, 4

[28] James Jacobs, Jernej Barbic, Essex Edwards, Crawford Doran, and Andy van Straten. How to build a human: Practical physics-based character animation. In *Proceedings of the 2016 Symposium on Digital Production*, pages 7–9, 2016. 2

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

[30] Korrawe Karunratanakul, Jinlong Yang, Yan Zhang, Michael J. Black, Krikamol Muandet, and Siyu Tang. Grasping field: Learning implicit representations for human grasps. *Int. Conf. on 3D Vision (3DV)*, pages 333–344, 2020. 3

[31] Korrawe Karunratanakul, Adrian Spurr, Zicong Fan, Otmar Hilliges, and Siyu Tang. A skeleton-driven neural occupancy representation for articulated hands. In *Int. Conf. on 3D Vision (3DV)*, pages 11–21. IEEE, 2021. 3

[32] Marilyn Keller, Silvia Zuffi, Michael J. Black, and Sergi Pujades. OSSO: Obtaining skeletal shape from outside. In *CVPR*, pages 20492–20501, 2022. 1, 2, 3, 7

[33] 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. In *ACM TOG*, 2023. 2, 3

[34] Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Seungbae Bang, Jinwook Kim, Michael J. Black, and Sung-Hee Lee. Data-driven physics for human soft tissue animation. *ACM TOG*, 36(4):54:1–54:12, 2017. 2, 3, 5

[35] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *ICCV*, 2023. 4

[36] Marcus D. R. Klarqvist, Saaket Agrawal, Nathaniel Diamant, Patrick T. Ellinor, Anthony Philippakis, Kenney Ng, Puneet Batra, and Amit V. Khera. Silhouette images enable estimation of body fat distribution and associated cardiometabolic risk. *NPJ Digital Medicine*, 5(1):1–9, 2022. 2, 3, 8

[37] Paul G. Kry and Dinesh K. Pai. Interaction capture and synthesis. *ACM TOG*, 25(3):872–880, 2006. 3

[38] Jianning Li, Antonio Pepe, Gijs Luijten, Christina Schwarz-Gsaxner, Jens Kleesiek, and Jan Egger. Anatomy completor: A multi-class completion framework for 3D anatomy reconstruction. In *International Workshop on Shape in Medical Imaging*, pages 1–14. Springer, 2023. 3

[39] Zhe Li, Tao Yu, Chuanyu Pan, Zerong Zheng, and Yebin Liu. Robust 3D self-portraits in seconds. In *CVPR*, pages 1344–1353, 2020. 3

[40] Matthew Loper, Naureen Mahmood, and Michael J. Black. MoSh: Motion and shape capture from sparse markers. *ACM TOG*, 33(6):1–13, 2014. 2

[41] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM TOG*, 34(6):1–16, 2015. 2, 4, 8

[42] Jürgen Machann, Claus Thamer, Birgit Schnoedt, Michael Haap, Hans-Ulrich Haring, Claus D Claussen, Michael Stumvoll, Andreas Fritsche, and Fritz Schick. Standardized assessment of whole body adipose tissue topography by MRI. *Journal of Magnetic Resonance Imaging*, 21(4):455–462, 2005. 3, 4

[43] Maulik D. Majmudar, Siddhartha Chandra, Kiran Yakkala, Samantha Kennedy, Amit Agrawal, Mark Sippel, Prakash Ramu, Apoorv Chaudhri, Brooke Smith, Antonio Criminisi, Steven B. Heymsfield, and Fatima Cody Stanford. Smartphone camera based assessment of adiposity: A validation study. *NPJ Digital Medicine*, 5(79), 2022. 2, 3

[44] Marko Mihajlovic, Yan Zhang, Michael J. Black, and Siyu Tang. LEAP: Learning articulated occupancy of people. In *CVPR*, pages 10461–1047, 2021. 3

[45] Marko Mihajlovic, Aayush Bansal, Michael Zollhoefer, Siyu Tang, and Shunsuke Saito. KeypointNerf: Generalizing image-based volumetric avatars using relative spatial encoding of keypoints. In *ECCV*, pages 179–197. Springer, 2022. 3

[46] Marko Mihajlovic, Shunsuke Saito, Aayush Bansal, Michael Zollhoefer, and Siyu Tang. Coap: Compositional articulated occupancy of people. In *CVPR*, pages 13201–13210, 2022. 3

[47] Manfred J. Müller, Merit Lagerpusch, Janna Enderle, Britta Schautz, Martin Heller, and Anja Bosy-Westphal. Beyond the body mass index: tracking body composition in the pathogenesis of obesity and the metabolic syndrome. *Obesity Reviews*, 13: 6–13, 2012. 2

[48] Bennett K. Ng, Benjamin J. Hinton, Bo Fan, Alka M. Kanaya, and John A. Shepherd. Clinical anthropometrics and body composition from 3D whole-body surface scans. *European journal of clinical nutrition*, 70(11):1265–1270, 2016. 2, 3

[49] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4D reconstruction by learning particle dynamics. In *CVPR*, pages 5379–5389, 2019. 3

[50] Ahmed A. A. Osman, Timo Bolkart, Dimitrios Tzionas, and Michael J. Black. SUPR: A sparse unified part-based human representation. In *ECCV*. Springer International Publishing, 2022. 3

[51] Dinesh K. Pai, Austin Rothwell, Pearson Wyder-Hodge, Alistair Wick, Ye Fan, Egor Larionov, Darcy Harrison, Debanga Raj Neog, and Cole Shing. The human touch: Measuring contact with real human soft tissues. *ACM TOG*, 37(4), 2018. 3

[52] Pablo Palafox, Aljaž Božič, Justus Thies, Matthias Nießner, and Angela Dai. NPMs: Neural parametric models for 3D deformable shapes. In *CVPR*, pages 12695–12705, 2021. 3

[53] Renzo Phellan, Bahe Hachem, Julien Clin, Jean-Marc Mac-Thiong, and Luc Duong. Real-time biomechanics using the finite element method and machine learning: Review and perspective. *Medical Physics*, 48(1):7–18, 2021. 3

[54] Pablo Ramón, Cristian Romero, Javier Tapia, and Miguel A Otaduy. Flsh–friendly library for the simulation of humans. In *ACM TOG*, 2023. 2, 3, 5

[55] Taehyun Rhee, J.P. Lewis, Ulrich Neumann, and Krishna Nayak. Scan-based volume animation driven by locally adaptive articulated registrations. *IEEE TVCG*, 17(3):368–379, 2011. 3

[56] Cristian Romero, Miguel A. Otaduy, Dan Casas, and Jesus Perez. Modeling and estimation of nonlinear skin mechanics for animated avatars. *Computer Graphics Forum*, 39(2):77–88, 2020. 3, 5

[57] Shunsuke Saito, Zi-Ye Zhou, and Ladislav Kavan. Computational bodybuilding: Anatomically-based modeling of human bodies. *ACM TOG*, 34(4):1–12, 2015. 2, 3

[58] Shunsuke Saito, Jinlong Yang, Qianli Ma, and Michael J. Black. SCANimate: Weakly supervised learning of skinned clothed avatar networks. In *CVPR*, pages 2886–2897, 2021. 3

[59] Igor Santesteban, Miguel A. Otaduy, Nils Thuerey, and Dan Casas. ULNeF: Untangled layered neural fields for mix-and-match virtual try-on. In *Advances in Neural Information Processing Systems, (NeurIPS)*, 2022. 3

[60] Robert Schleicher, Marlies Nitschke, Jana Martschinke, Marc Stamminger, Björn Eskofier, Jochen Klucken, and Anne Koelewijn. BASH: Biomechanical Animated Skinned Human for Visualization of Kinematics and Muscle Activity. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, pages 25–36, 2021. 3

[61] Akash Sengupta, Ignas Budvytis, and Roberto Cipolla. Probabilistic 3D human shape and pose estimation from multiple unconstrained images in the wild. In *CVPR*, pages 16094–16104, 2021. 8

[62] John A. Shepherd, Bennett K. Ng, Bo Fan, Ann V. Schwartz, Peggy Cawthon, Steven R. Cummings, Stephen Kritchevsky, Michael Nevitt, Adam Santanasto, and Timothy F. Cootes. Modeling the shape and composition of the human body using dual energy X-ray absorptiometry images. *PLOS One*, 12(4): e0175857, 2017. 8

[63] 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(C), 2023. 3

[64] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3):e1001779, 2015. 3

[65] Javier Tapia, Cristian Romero, Jesús Pérez, and Miguel A. Otaduy. Parametric skeletons with reduced soft-tissue deformations. *Computer Graphics Forum*, 40(6):34–46, 2021. 2, 3, 5

[66] Joseph Teran, Eftychios Sifakis, Silvia S. Blemker, Victor Ng-Thow-Hing, Cynthia Lau, and Ronald Fedkiw. Creating and simulating skeletal muscle from the visible human data set. *IEEE TVCG*, 11(3):317–328, 2005. 3

[67] Garvita Tiwari, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. Neural-GIF: Neural generalized implicit functions for animating people in clothing. In *CVPR*, pages 11708–11718, 2021. 3

[68] Bohan Wang, George Matcuk, and Jernej Barbič. Hand modeling and simulation using stabilized magnetic resonance imaging. *ACM TOG*, 38(4):1–14, 2019. 3

[69] Shaofei Wang, Marko Mihajlovic, Qianli Ma, Andreas Geiger, and Siyu Tang. MetaAvatar: Learning animatable clothed human models from few depth images. In *Advances in Neural Information Processing Systems*, 2021. 3

[70] Shaofei Wang, Katja Schwarz, Andreas Geiger, and Siyu Tang. ARAH: Animatable volume rendering of articulated human SDFs. In *ECCV*, pages 1–19. Springer, 2022. 3

[71] Michael C. Wong, Bennett K. Ng, Isaac Tian, Sima Sobhiyeh, Ian Pagano, Marcelline Dechenaud, Samantha F. Kennedy, Yong E. Liu, Nisa N. Kelly, Dominic Chow, Andrea K. Garber, Gertraud Maskarinec, Sergi Pujades, Michael J. Black, Brian Curless, Steven B. Heymsfield, and John A. Shepherd. A pose-independent method for accurate and precise body composition from 3D optical scans. *Obesity*, 29(11):1835–1847, 2021. 3

[72] Michael C. Wong, Jonathan P. Bennett, Lambert T. Leong, Isaac Y. Tian, Yong E. Liu, Nisa N. Kelly, Cassidy McCarthy, Julia M.W. Wong, Cara B. Ebbeling, David S. Ludwig, Brian A. Irving, Matthew C. Scott, James Stampley, Brett Davis, Neil Johannsen, Rachel Matthews, Cullen Vincellette, Andrea K. Garber, Gertraud Maskarinec, Ethan Weiss, Jennifer Rood, Alyssa N. Varanoske, Stefan M. Pasiakos, Steven B. Heymsfield, and John A. Shepherd. Monitoring body composition change for intervention studies with advancing 3D optical imaging technology in comparison to dual-energy x-ray absorptiometry. *The American Journal of Clinical Nutrition*, 2023. 2, 3

[73] 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. 2, 3, 4

[74] Yuliang Xiu, Jinlong Yang, Dimitrios Tzionas, and Michael J. Black. ICON: Implicit Clothed humans Obtained from Normals. In *CVPR*, 2022. 3

[75] Yufei Ye, Abhinav Gupta, and Shubham Tulsiani. What's in your hands? 3D reconstruction of generic objects in hands. In *CVPR*, pages 3895–3905, 2022. 3

[76] Lifeng Zhu, Xiaoyan Hu, and Ladislav Kavan. Adaptable anatomical models for realistic bone motion reconstruction. *Computer Graphics Forum*, 34(2):459–471, 2015. 3

[77] Gaspard Zoss, Thabo Beeler, Markus Gross, and Derek Bradley. Accurate markerless jaw tracking for facial performance capture. *ACM TOG*, 38(4):1–8, 2019. 3