SMPL-A: Modeling Person-Specific Deformable Anatomy

Hengtao Guo[1,2], Benjamin Planche[1], Meng Zheng[1], Srikrishna Karanam[1], Terrence Chen[1], Ziyan Wu[1]

[1]United Imaging Intelligence, Cambridge MA, USA

[2]Rensselaer Polytechnic Institute, Troy NY, USA

{first.last}@uii-ai.com, guoh9@rpi.edu

Abstract

A variety of diagnostic and therapeutic protocols rely on locating in vivo target anatomical structures, which can be obtained from medical scans. However, organs move and deform as the patient changes his/her pose. In order to obtain accurate target location information, clinicians have to either conduct frequent intraoperative scans, resulting in higher exposition of patients to radiations, or adopt proxy procedures (e.g., creating and using custom molds to keep patients in the exact same pose during both preoperative organ scanning and subsequent treatment. Such custom proxy methods are typically sub-optimal, constraining the clinicians and costing precious time and money to the patients. To the best of our knowledge, this work is the first to present a learning-based approach to estimate the patient's internal organ deformation for arbitrary human poses in order to assist with radiotherapy and similar medical protocols. The underlying method first leverages medical scans to learn a patient-specific representation that potentially encodes the organ's shape and elastic properties. During inference, given the patient's current body pose information and the organ's representation extracted from previous medical scans, our method can estimate their current organ deformation to offer guidance to clinicians. We conduct experiments on a well-sized dataset which is augmented through real clinical data using finite element modeling. Our results suggest that pose-dependent organ deformation can be learned through a point cloud autoencoder conditioned on the parametric pose input. We hope that this work can be a starting point for future research towards closing the loop between human mesh recovery and anatomical reconstruction, with applications beyond the medical domain.

1. Introduction

Many medical procedures rely on precise location and shape information of target tissues. Since radioactive scans are invasive, there is a maximum number that a patient can safely undergo per specific time period [10, 12]. Due to this
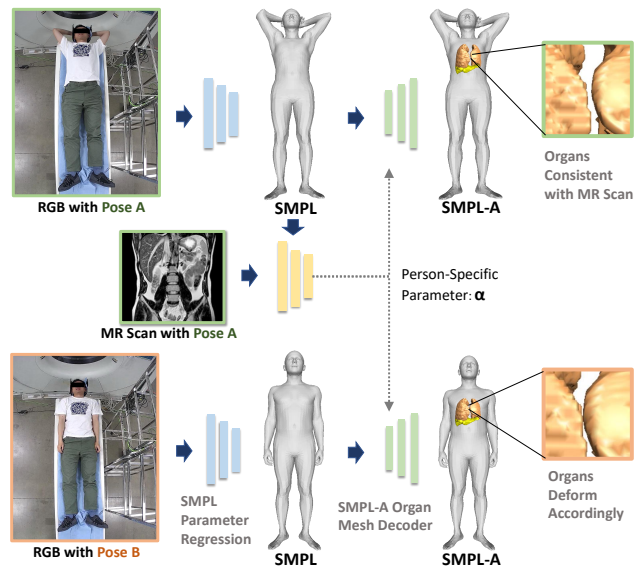

Figure 1. Application overview. The information from previous medical scans can facilitate estimating current organ's shape and deformation.

constraint, doctors have to carefully plan imaging protocols, weighing risks for the patient versus spatial and temporal inconsistency that may arise from too coarse scan-based investigation. Any solution to help determine the evolution of an illness or the impact of a treatment without excessive imaging would greatly help clinicians and reduce patients' risks. Currently, complex protocols are used to precisely target organ regions while minimizing potential radioactive damage. For example, in radiotherapy (a cancer treatment that uses high doses of radiation to kill cancerous cells and shrink tumors [8]), it is common practice to create and use patient-specific molds (usually made of plastic or plaster) to keep patients in the exact same pose during both preoperative organ scanning and during subsequent treatment sessions. Such customized molds help ensure the desired organ region is targeted accurately without having to re-scan the patient and re-localize the tumors before each session. This custom proxy method—like other

methods specific to different scan-reliant treatments—are typically sub-optimal, constraining the clinicians and costing precious time and money to the patients. These protocols could however be relaxed with appropriate organ pose estimation, thereby removing the need for patient-specific molds and decreasing cost and delays of such vital procedures (*i.e.*, more generic and thus less costly molds could be adopted with the sole purpose of keeping patients still during treatment). Other less sensitive domains could also benefit from cheaper and computationally-lighter organ pose estimation. Existing simulation tools mostly rely on finite-element modeling (FEM) [11, 16, 23] to predict the deformation of soft tissues—such as organs—based on input load forces. Their inherent computational footprint and abstraction makes them unpractical for real-time and artistic applications (*e.g.*, movie or video-game industries). In this paper, we not only propose a modeling function that relies on light-weight operations during inference, but we also condition it on more concrete parameters, *i.e.*, the pose of the target person; therefore making this kind of simulation accessible to a larger audience.

In order to learn the latent correspondences between human contours and in vivo organ deformations, we need a representation of the human body model. While the popular skinned multi-person linear (SMPL) model [22] and its variants [27, 33] provide a flexible statistical representation to capture human pose and shape deformations, they do not have the necessary information to model in vivo organ deformations. In our work, we propose a novel approach to address this issue. While retaining the computational flexibility offered by the family of SMPL models, we take one new step: modeling how the pose parameters affect internal organ deformations. Our intuition is that when a patient changes pose (*e.g.*, arms stretched over head as in Figure 2), internal organs deform accordingly, and we seek to capture all these changes (pose and internal deformations) jointly. Concretely, while the SMPL family defines two parameters, θ for pose and β for shape, our method adds one additional parameter for organ deformation: α. Consequently, whereas the SMPL model learns the function mapping (β, θ) to the vertices of the human mesh, our method seeks to additionally capture organ deformations from (β, θ, α). In particular, we propose a framework that can (1) extract patient-specific organ representation from previous medical scans and (2) estimate the organ deformation conditioned on arbitrary human poses for any scans in the future. To the best of our knowledge, there has been little previous work studying similar topics, despite the need transpiring from the medical community.

To summarize, our contributions are three-fold:

1. We unify the shape representation for one specific organ across different patients by automatically setting up anatomical correspondences. Specifically, we ap-
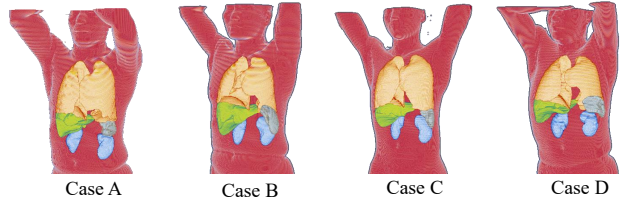

Figure 2. Four data samples in our dataset. For each subject, the segmentation of lung (yellow), livers (green), spleen (grey) and kidneys (blue) are available, alongside with full-body skin segmentation (red).

ply a 3D deformable registration between a mean organ binary segmentation and any given segmentation in the dataset. The generated deformation field can warp the mean organ's 3D mesh into any shape in the dataset such that every organ's representation can be unified into a fixed-length point cloud representation with anatomical correspondences, as in Figure 3.

2. We propose SMPL-A ("A" for anatomy), an extenstion to the SMPL model for pose/shape-dependent organ deformation estimation. SMPL-A encodes the organ's shape into a low-dimensional representation α and can reconstruct a deformed organ shape conditioned on different poses during inference. In a novel fashion, the proposed framework correlates a patient's external pose change to internal organ deformation, which may significantly contribute to surgical guidance;

3. Our work suggests that the organ's shape and elastic properties can indeed be encoded into lower-dimensional representations α. The experimental results also demonstrate that conditioned on the patient's exterior body movement, a deformed organ mesh can be reconstructed with the patient-specific organ representation that was extracted from previous scans.

2. Related Work

Modeling Human Body. Person-specific 3D models can be obtained through a variety of means. Most state-of-the-art solutions start from a mean/template 3D mesh (set of vertices encoding the body surface information) and apply adequate deformations so to fit the person's appearance [24]. Such paradigm resulted in the creation of various statistical models of the 3D human body [1, 14, 22]. The most predominant in the literature, the skinned multi-person linear (SMPL) model [22] is based on skinning and blend shapes, and is learned from thousands of 3D body scans. SMPL originally uses rotation vectors for 24 joints—encoded as a vector $\theta \in \mathbb{R}^{3 \times 24}$—and a shape parameter $\beta \in \mathbb{R}^{10}$ to represent a human mesh at a specific pose. As β domain is defined via dimensional reduction over the

dataset of body scans (principal component analysis), each of its dimension maps to specific morphological changes, *e.g.*, the first dimension controls the body height and thickness, the second dimension mostly impacts waist size, *etc*. Simple yet highly expressive, this model has been widely adopted by the community [5, 17, 20, 28, 36] and many extensions have been proposed since, *e.g.*, adding different parameters to articulates the hands [33], the head (eyeball, jaw, and neck) [27], *etc*. Based on all these parameters, SMPL can model realistic surface deformations linked to body shape and pose (*e.g.*, visible contraction of muscles and deformation of fat pad w.r.t. some poses). However, it has no notion of the underlying anatomy, *i.e.*, volumetric content under the skin surface. In this work, we propose to adopt SMPL's methodology and extend its model with volumetric information on patient-specific anatomy.

Learning Tissue Deformation. At the crossroad between computer science and medicine, the modeling of anatomical elements and their behavior under various conditions has gained a lot of interest in recent years [31], driven by extensive developments in machine learning. Data-driven models based on artificial neural networks, applied to the simulation of deformable organs, have achieved commendable accuracy [25, 29, 30, 34]. For example, Pfeiffer *et al*. [30] have demonstrated that a convolutional neural network (CNN), trained on synthetic data, is able to extrapolate the displacement field of all points inside an organ based on a partial input surface force field. Similarly, Salehi *et al*. [34] leverages a graph convolutional network (GCN) to better induce the deformation of organs' 3D structures caused by force loads applied to surface vertices. Most of these learning-based methods are proposed as alternative to the more traditional but much heavier FEM [13]. Yeo *et al*. [38] proposed an interesting approach which uses learned pose/shape factors to facilitate segmenting adipose tissue. Unlike these previous models that express the shape of organs as functions of local, abstract load forces, our goal is to estimate the deformation of internal organs based on the pose and shape of the patients themselves, as well as past observations (scans) providing patient-specific organ information.

Point Cloud Registration. In order to morph a template 3D shape into custom ones (*e.g.*, person-specific body mesh), a unified representation needs to be applied to the considered samples. For example, SMPL defines the human body as a mesh composed of 13,776 triangular surfaces supported by 6,890 vertices, whose displacements from one individual to the other can be controlled by a direct mapping [22]. We adopt a similar, common shape representation for the organs across all patients, homogenizing the number and organization of their surface vertices, as highlighted in Figure 3. Opting for a vertex-based representation allows us to leverage the extensive efforts in recent years
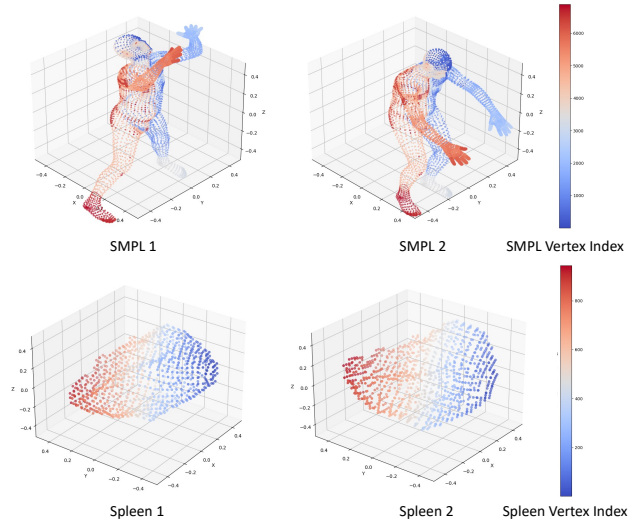

Figure 3. Point correspondences across samples. The idea of unifying the shape representation of organs originates from the SMPL model. However, instead of creating an artistic template model as in SMPL, we use Elastix registration to set up point-to-point correspondence for organs' surface [19], *e.g.*, using 944 points to represent the surface of the spleen. The color coding indicates the vertex's index in the points cloud: two points with the same color are locating on the corresponding anatomical structures.

towards the efficient processing of 3D point-cloud structures. Among other operations, the registration of point clouds is the basis for 3D reconstruction and scenes perception [6, 15]. It helps set up correspondence between two sets of 3D points. The same-source points cloud registration can be further divided into optimization-based registration methods [21, 37], feature-learning methods [9], end-to-end learning registration [35]. The vast majority of works on points cloud registration focus on laser-scanning captured data and aim for natural scenery understanding. One recent work [3] proposed a global feature extraction module with PointNet [32] backbone to fuse ultrasound and magnetic resonance images for surgery guidance, by generating 3D displacement vectors for each input point. The aforementioned methods can all serve as candidates to set up point-to-point correspondence.

3. Methodology

In this section, we detail our methodology for pose-dependent organ deformation estimation. We start by formalizing the problem, and then justify the design of each component of the proposed , showing how to condition the deformed organ mesh reconstruction on the patient's external body shape and pose.
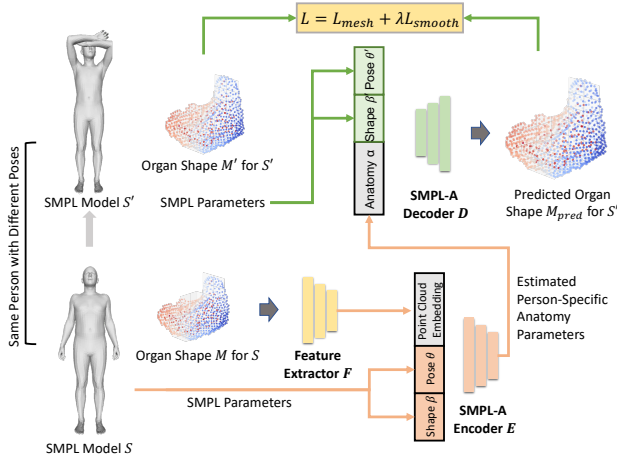
Figure 4. Overview of SMPL-A Let M be the 3D shape of a target organ scanned from a patient with pose θ and body shape β. We jointly learn a shape-encoding function E to extract patient-specific organ properties α, as well as a pose-dependent decoding function D that predicts proper organ deformation from new combinations of the three parameter sets (θ, β, and α).

3.1. Formalization

For a given patient, we assume that a custom SMPL model is provided (*i.e.*, optimizing β according to pictures of the patient), as well as a set of annotated computed tomography (CT) scans I captured at known patient poses θ_{CT} (we assume that 3D model and scans are captured not too long apart so that the patient's shape β can be assumed constant between the two modalities). Originally inspired by the human SMPL model, we represent organs as 3D point clouds $M \in \mathbb{R}^{3 \times m}$ (m number of points) in order to capture the organs' complex topology (see Figure 3 and details in Subsection 4.2). After converting the 3D CT scans I into point clouds, our goal is thus to define a parametric set of functions to model their patient-specific properties.

From correspondences between these modalities (SMPL models and scanned point clouds), we propose to learn a low-dimensional vector $\alpha \in \mathbb{R}^k$ (*i.e.*, composed of k_α real-valued parameters) that encodes patient-specific properties of the target organ(s), partially disentangled from the patient's shape and pose. This feature vector can later guide the inference of the 3D shape of target organ(s) under different body poses θ', *e.g.*, if during treatment the patient should assume a pose $\theta' \neq \theta_{\text{CT}}$ (we later discuss how our scheme can also be extended to predicting organ deformation for different body shape β', *e.g.*, if the patient lost or gained weight since the last scans).

To that end, as shown in Figure 4, the proposed solution consists of a learning-based function $E(\theta, \beta, M) = \alpha$ that encodes the patient-specific organ characteristics, and a decoding function $D(\theta', \beta', \alpha) = M_{pred}$, *i.e.*, predicting the deformed organ shape when the patient takes a pose to θ'.

The design of these two models E and D, as well as the

methodology w.r.t. to their training and data preprocessing, are subsequently detailed.

3.2. Model Definition

The aforementioned patient-specific organ shape encoder E and pose-conditioned organ shape decoder D are defined as artificial neural networks, trained jointly in an end-to-end way.

Patient-Specific Organ Shape Encoder E aims to map the organ's 3D mesh (represented as points cloud) to a low-dimensional feature representation α. Similar to how the shape parameter β in SMPL encodes complex person-specific morphological information, α represents here the complex shape information of target organ, encoded into a lower-dimension space.

Since the shape of anatomic structures is intrinsically dependent on the pose and shape parameters of the subject him/herself, we also provide E with the corresponding θ and β parameters as input, so that the network can learn to decouple the organ shape information from them. To guide the shape encoding, a point-cloud feature extractor F first takes the processed organ point cloud (*cf*. Section 4.2) as input and produce a latent feature vector $z_M \in \mathbb{R}^{k_z}$. Fixing a reasonable value for k_z (set to 128 in our experiments) allows to balance the contributions of M compared to the contributions of low-dimensional SMPL parameters, during organ shape encoding. Afterwards, a multi-layer perceptron (MLP) takes the concatenated organ feature vector z_M and the SMPL parameters θ and β as input to encode a patient-specific, pose-agnostic organ representation α. Overall, the patient-specific organ shape encoder can be denoted as:

$$E(\theta, \beta, M) = \text{MLP}\left(\theta, \beta, F(M)\right) = \alpha. \qquad (1)$$

Pose-Conditioned Organ Shape Decoder D aims to reconstruct the deformed organ mesh conditioned on different poses for the same patient. It takes the patient's organ shape prior α, SMPL shape parameter β and a different pose parameters θ' as input to infer the deformation of the organ under the new body pose θ', *i.e.*:

$$D(\theta', \beta, \alpha) = M_{pred} \qquad (2)$$

where $M_{pred} \in \mathbb{R}^{3 \times m}$ is deformed organ mesh prediction to obtain M', organ shape of the patient under body pose θ'.
Training: We propose to train the aforementioned system in an end-to-end manner *cf*. literature on autoencoders. We compute the reconstruction loss, named L_{mesh}, directly on the set of vertices, as follows:

$$L_{mesh} = \|M' - M_{pred}\|_2. \qquad (3)$$

Optimizing the parameters of E and D based on this criterion will encourage E to extract patient-specific organ prior α disentangled from the reparameterized variable θ/θ', and
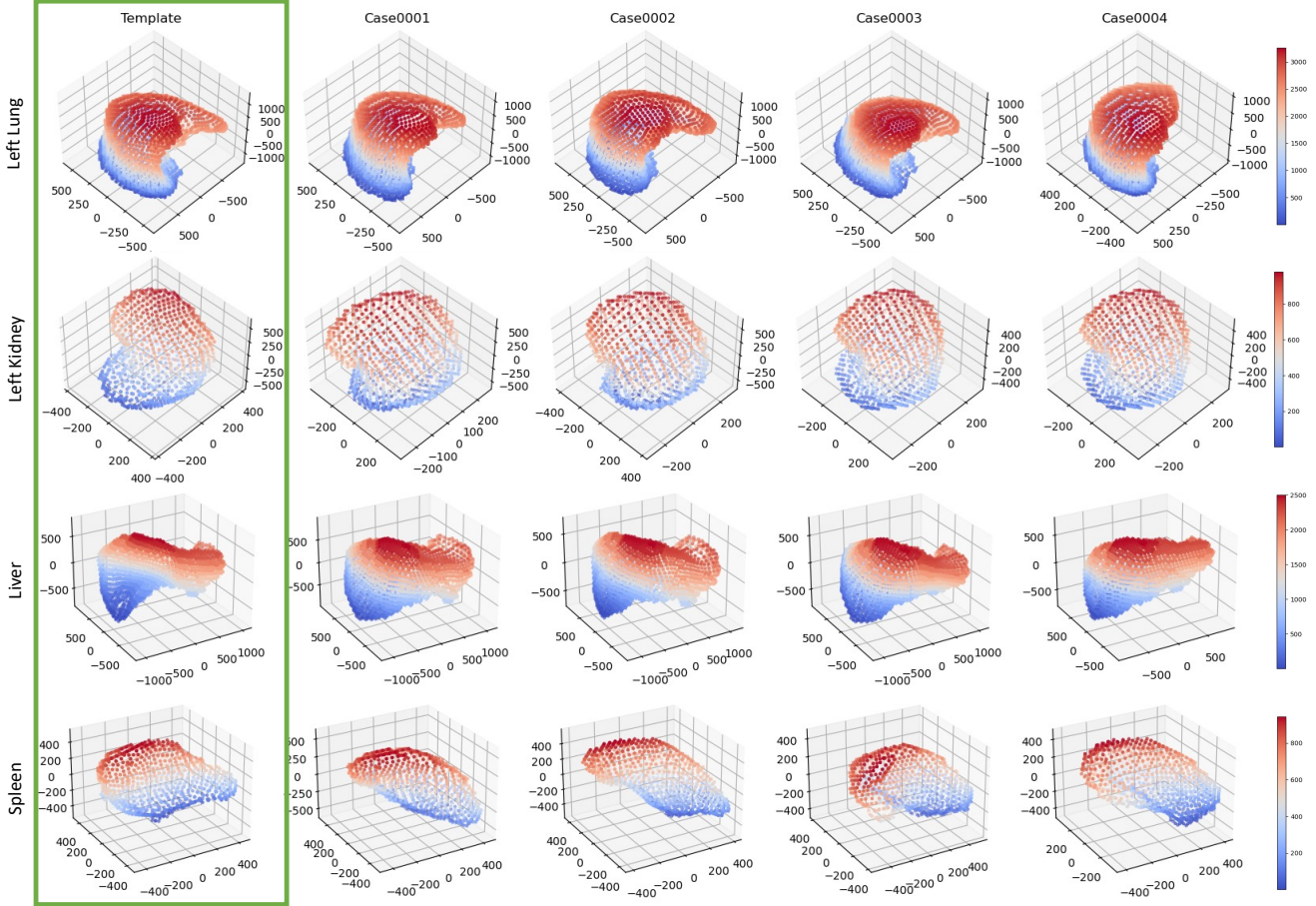
Figure 5. SMPL-A dataset overview. The color bar indicates the index of the vertex in the points cloud, which shows the point-to-point correspondence across different subjects.

will encourage D to leverage all available information to predict proper organ displacement. Note, however that while this objective encourages M_{pred} to approximate M', D may generate non-smooth displacement fields $\delta M = M_{pred} - M$ that are physically unrealistic. To further impose a constraint on the smoothness of the deformed organ surface, a regularization term L_{smooth}, weighted by a factor λ, is considered. Borrowed from [2], this term approximates the spatial gradients of δM using differences between neighboring points (see paper for further details). The full loss function for training is, therefore, denoted as:

$$L = L_{mesh} + \lambda L_{smooth}. \qquad (4)$$

For training, our framework thus needs pairs of segmented organ shapes M and M', along with their corresponding patient's pose parameters θ and θ', and supposedly constant patient's body shape β. To relax these data requirements, we present in the next section how to use non-rigid registration to unify the organs' representations and how to apply FEM to effectively augment the data with synthetic deformation pairs (θ, M) and (θ', M').

4. Data Preparation

Due to the complexity and heterogeneity of the data modalities considered in this paper, and for reproducibility purposes, we share details on data pre-processing in the following section.

4.1. SMPL Pose/Shape Annotation

Every patient in our dataset has their CT scan processed into full-body skin segmentation that highlights the contour of the patient, as in Figure 2. To find the corresponding pose parameter θ and shape parameter β, we manually generate the SMPL model that matches the skin best. We use the SMPL-X [27] add-on in Blender [7] to adjust the pose joints by joints and alter the shape parameters to match the body contour. Manually generating the matching SMPL model for the skin segmentation is considered more accurate for downstream analysis, though this step can also benefit from automatic human modeling process, $e.g.$, applying the Implicit Part Network (IPNet) [4] or similar solutions.

4.2. Organ Correspondence

The classical SMPL model, sculpted by an artist, has a fixed-length design which unifies the representation of different human subjects, even given different body shapes and poses. Another benefit is that, for different human subjects, the i-th vertex denotes the same anatomical structure. In our task, we also need to parameterize the shape of organs across different patients, by homogenizing the size of their representation as in Figure 3.

Specifically, we create one mean mesh model for each type of organ and then use it as a template to warp into other shapes for a uniform representation (Figure 5). Taking the spleen as an example: we first gather a dataset that contains 3D spleen segmentation $\{I_n | n = 1, ..., N\}$, where I_n denotes the n-th patient's binary spleen segmentation image. We first compute a mean segmented 3D scan $\hat{I}$ from all the cases in the dataset through simple averaging, and then evenly sample m vertices on its surface to obtain a mean organ mesh $\hat{M}$. The organ mean shape $\hat{M}$ serves as template to unify the representation of the organ from different subjects, with the j-th vertex assigned to the same anatomical location across the patients.

To obtain such unified representation, we apply Elastix [19]—an intensity-based medical image registration method—to set up the correspondence between organs from different subjects. The optimization function is denoted as

$$\widehat{\mu}_k = \arg\min_{\mu} C(I_k, T(\hat{I}; \mu)), \qquad (5)$$

where we aim to find a transformation T_μ to non-rigidly align the 3D template image $\hat{I}$ to any give target fixed image I_n, in order to minimize the selected cost function C. The optimization target μ contains the transformation parameters of a linear transformation and a B-spline deformation field. We use here the mean square difference (MSD) for C. The optimized transformation $T_{\hat{\mu}}$ can then be applied to the 3D template mesh $\hat{M}$, such that we could use the transformed template mesh $T(\hat{M}; \hat{\mu})$ to represent any other cases in the dataset. Considering the template mesh $\hat{M}$ has m vertices, every other case in the dataset can all be represented by m vertices correspondingly.

4.3. Finite Element Modeling for Deformation

Our work aims to model the organ's deformation under different poses. To learn such pattern, multiple scans are needed for each patient under various pose conditions. As a matter of fact, such data is extremely difficult to require as it may introduce additional radiation to the subjects during acquisition. To make up for the lack of data, we run FEM simulations [11,16,23] to create organ deformations, by applying different amounts of load to the organ mesh model.

We first study the statistics of the patients' pose information collected from the annotated SMPL dataset. The ma-
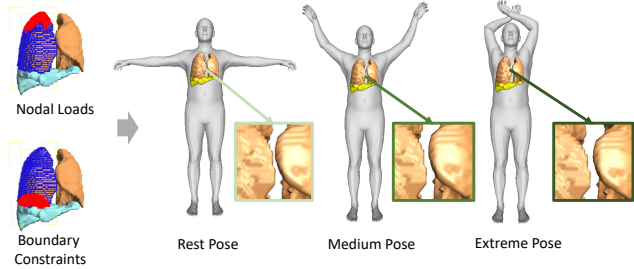

Figure 6. Synthetic data augmentation. We correlate some body movements (*e.g.*, arms lifting) to the amount of internal organs stretching for synthetic data generation, using FEM.

jority of subjects are lifting their arms to guarantee a better imaging purpose, as in Figure 2. To match such scenario, we simulate the relationship between pose and organ deformation with a focus on the rotation angles of collar and shoulder joints, *i.e.*, the more lifted the arms are, the larger the stretching forces—and therefore deformation—applied to the organs are, as shown in Figure 6. We thus correlate the collars/shoulders' joints rotations parameters to the loads applied during finite element modeling, to produce a reasonable amount of deformation to the organ. Specifically, we assign nodal loads to the upper surface of the organs and place boundary constraints to stabilize the lower surface. The applied load will slightly stretch the organ in the positive Z-direction, so as to simulate how the organs deform while raising arms. For each case, we run FEM to generate a deformation sequence length of 10, from a resting pose to an extreme pose with corresponding amount of stretch in the organs. During the downstream analysis, we expect the trained network to recover such simulated relationship between the pose parameters and the organ deformation for a prove of theory. Used only during training (when computational constraints can be relaxed), the FEM method is a compensation for the lack of data samples. Even once multiple scans from each patient end up being collected, this simulation could be preserved to further augment the dataset.

5. Experiments

5.1. Implementation Details

We collected 30 cases from IRB-approved clinical trial, each with the skin, lungs, kidneys, liver, and spleen segmentation. We follow the aforementioned data preparation (*cf*. Section 4.3) to generate the SMPL models for each patient by aligning the template model to the skin segmentation, and to further augment the dataset using FEM to create a sequence of deformed organs. We opt for different mesh size m for the template of each organ due to the varying volume size of each organ (*e.g.*, $m = 944$ vertices for the spleen, 3,184 vertices for the left lung, *etc*., as illustrated in Fig-
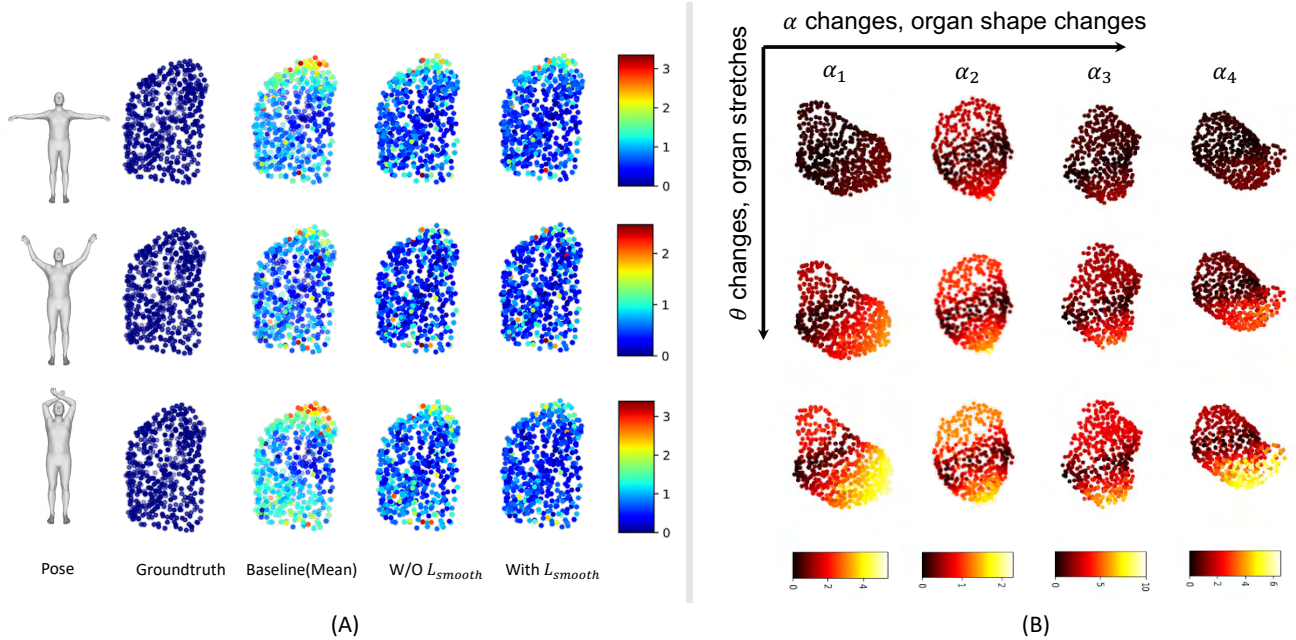
Figure 7. (A) Predicted deformation error (color-coded, in mm) for the left lung; (B) Impact of patients' pose parameter θ and organ parameter α on deformed spleen shape reconstruction (color-coded by the stretch amount w.r.t. the first frame in each sequence, in mm).

ure 5). During training, we subsample 512 equally spaced points to guarantee computational efficiency and maintain the point-to-point correspondence. We use 24/3/3 sequence splits respectively for training, validation, and testing.

The network structure of feature extractor F and organ shape decoder D are both multi-layer perceptrons, implemented using PyTorch [26]. Note that our solution is meant to be network-agnostic, *e.g.*, the backbone for F could be switched to any other network structures such as Point-Net [32]. For each organ, we train the network end-to-end for 300 epochs with batch size $K = 8$ using Adam optimizer [18] with an initial learning rate of $1{\times}10^{-5}$, which decays by 0.9 after every five epochs.

5.2. Results and Discussions

For each testing subject with FEM simulated deformation sequence $\{(\theta_i, \beta, \alpha_i, M_i)|i = 0, ..., 10\}$, we evaluate the proposed SMPL-A by reconstructing the 3D deformed meshes $\{M_{pred,i}|i = 1, ..., 10\}$ from the first frame's parameters α_0 and θ_i, β_i. Since there is little previous references for comparison, we compute a mean mesh for each sequence that serves as baseline method. Since the surface of the organ have point-to-point correspondence, we compute the surface distance error in millimeters as metric.

The results for multi-organ reconstruction are shown in Table 1. We run a paired t-test for statistical analysis with a confidence interval of 0.05. The results suggest that the reconstruction error of the proposed SMPL-A is significantly lower than the baseline error. This indicates that the proposed SMPL-A is able to reconstruct the same person's 3D

organ shape, stretched by a different pose.

Figure 7-A shows the results for one left lung deformation sequence (refer to supplemental materials for more). Lifting of the arms is gradually increased from first to last row, leading to the stretching of the internal organs. The ground-truth column shows the organ deformed by the FEM. The color bar encodes the point-wise surface error (in mm), compared to the ground-truth. Comparing the two rightmost columns, we can observe how the smoothness constraint L_{smooth} prevents the organ surface to become uneven (with points deviating from the expected surface planes) by enforcing smooth local deformations.

Figure 7-B shows how different θ and α impacts our SMPL-A model. We reconstruct the spleen shape for different patients (*i.e.*, varying α across columns) and different poses (*i.e.*, varying θ across rows). The morphological differences indicate that SMPL-A can encode patient-specific organ information into α and recover the 3D point cloud from it, while also learning a correspondence between the patient pose information θ and predicted deformation. Additionally, Figure 8 shows the left lung mesh reconstruction for a different patient. We plot the errors onto the mesh surface for visual purposes.

5.3. Ablation Studies

To demonstrate that the low-dimensional representation α can encode the organ's shape information (similar to the 10-D shape parameter β for the SMPL model), we train a 3D points cloud autoencoder with various sizes of latent space, noted as k_α. This experiment does not involve or-
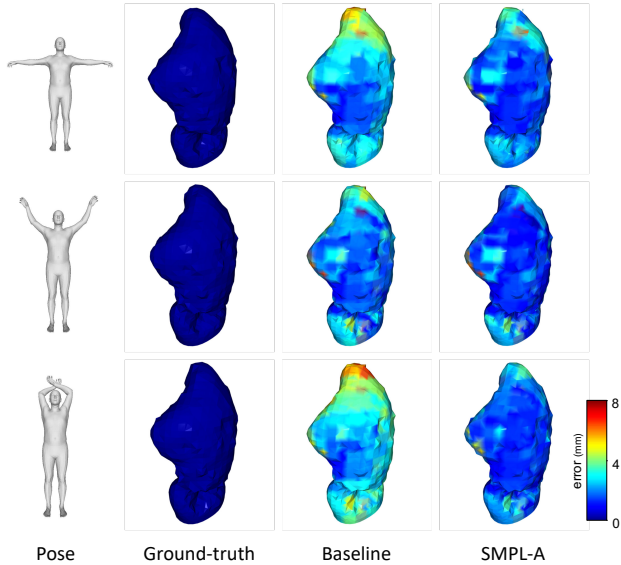
Figure 8. Predicted deformation error (color-coded, in mm) for a left lung, using 3D mesh as target modality.

Table 1. Reconstruction error (mm) for the deformation of six organ parts. The baseline method computes the error between mean and ground-truth shape.

	Baseline Error	SMPL-A Error
Spleen	2.5353 ± 0.1088	1.4496 ± 0.0949
Liver	2.9810 ± 0.1395	1.7832 ± 0.1012
Left Kidney	2.2364 ± 0.1174	1.2694 ± 0.0974
Right Kidney	2.4608 ± 0.1207	1.2504 ± 0.1088
Left Lung	4.4784 ± 0.2586	2.2475 ± 0.1688
Right Lung	4.9846 ± 0.2407	2.3716 ± 0.1942

Table 2. We train a point-cloud autoencoder to reconstruct the organ's shape for α's ablation study, without pose-conditioned deformation.

k_α	Reconstruction Error (mm)			
	Spleen	Liver	Lung_left	Kidney_left
5	1.45 ± 0.23	2.45 ± 0.23	2.65 ± 0.21	1.65 ± 0.24
10	1.09 ± 0.17	1.76 ± 0.20	1.75 ± 0.23	1.33 ± 0.14
20	0.78 ± 0.13	1.21 ± 0.15	1.22 ± 0.18	0.84 ± 0.12
40	$\mathbf{0.74 \pm 0.16}$	$\mathbf{1.05 \pm 0.19}$	$\mathbf{1.15 \pm 0.16}$	$\mathbf{0.82 \pm 0.14}$

gan deformation and only concentrates on the shape encoding. From Table 2 we can see that, for any organ mesh $M \in \mathbb{R}^{3 \times 512}$, we can still reconstruct its shape from its latent representation α with relatively low reconstruction error, even when k_α is as low as 5. The larger the size of α, the more expressive it can be, which matches the intuition.

We also try to understand the potential meaning behind each dimension in the latent representation α. In Figure 9, we gradually increase the first three values within the α vector to see their respective impacts on the organ shape reconstruction. The first value $\alpha[1]$ seems to correlate with
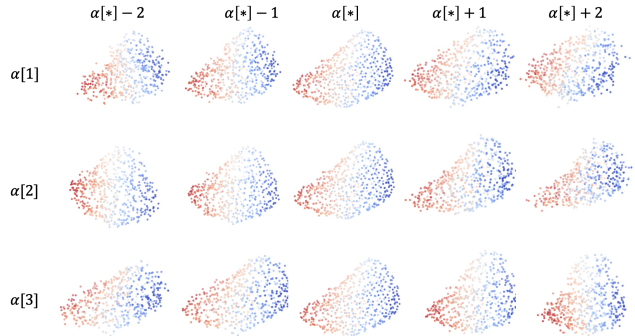

Figure 9. Impact of the first three values of α in reconstruction.

the orientation of the organ; the second value $\alpha[2]$ correlates to the thickness in the vertical direction; the third value $\alpha[3]$ correlates more to the horizontal thickness. Similar to how changing the β values results in SMPL models [22] that have different height, fatness and proportions; changing the α values results in our SMPL-A model recovering organs showing a variety of shapes.

6. Conclusion

In this work, we propose SMPL-A, a model for in vivo organ deformation based on the subject's different poses, as an extension to the SMPL family. We first set up the organ mesh correspondences through deformable image registration, followed by FEM to simulate organ deformation caused by lifting the arms. The generated organ mesh representations and corresponding pose/shape parameters are used for training the SMPL-A network. Upon convergence, the SMPL-A network can extract the organ's shape representation α specific to each patient, and predict its deformed shape, conditioned on different pose parameters. During inference, given the organ shape representation α of a person in record, and his/her current pose/shape parameters, SMPL-A can predict the current organ shape for various applications such as surgery guidance and radiotherapy.

To reflect on our method, we use FEM to simulate pose-dependent organ deformation as a compensation to data scarcity. The simulation may not perfectly match real-life data, yet experimental results demonstrate that proposed SMPL-A can recover such complex, simulated relationship. Given sufficient data, we could further disentangle θ (e.g., for body articulations not covered yet by SMPL) and also β (requiring, e.g., time-series morphological information) from the organ's information, and reveal the intrinsic relationship between body pose/shape and in vivo organ's deformation.

As the first study of this kind (leveraging cross-modal correspondences to tie human pose and organ deformation), our work aims to demonstrate the soundness and feasibility of such an approach, and to inspire further studies, as well as further data collection effort to enable them.

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005. 2

[2] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019. 5

[3] Z. M. Baum, Y. Hu, and D. C. Barratt. Real-time multimodal image registration with partial intraoperative point-set data. *Medical image analysis*, 74:102231, 2021. 3

[4] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. Combining implicit function learning and parametric models for 3d human reconstruction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 311–329. Springer, 2020. 5

[5] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In *European conference on computer vision*, pages 561–578. Springer, 2016. 3

[6] L. Cheng, S. Chen, X. Liu, H. Xu, Y. Wu, M. Li, and Y. Chen. Registration of laser scanning point clouds: A review. *Sensors*, 18(5):1641, 2018. 3

[7] B. O. Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 5

[8] G. Delaney, S. Jacob, C. Featherstone, and M. Barton. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(6):1129–1137, 2005. 1

[9] H. Deng, T. Birdal, and S. Ilic. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 195–205, 2018. 3

[10] A. J. Einstein. Effects of radiation exposure from cardiac imaging: how good are the data? *Journal of the American College of Cardiology*, 59(6):553–565, 2012. 1

[11] S. M. Finley, D. S. Brodke, N. T. Spina, C. A. DeDen, and B. J. Ellis. Febio finite element models of the human lumbar spine. *Computer methods in biomechanics and biomedical engineering*, 21(6):444–452, 2018. 2, 6

[12] J. D. Hall, M. Godwin, and T. Clarke. Lifetime exposure to radiation from imaging investigations. *Canadian Family Physician*, 52(8):976–977, 2006. 1

[13] K. Halonen, C. M. Dzialo, M. Mannisi, M. Venäläinen, M. de Zee, and M. S. Andersen. Workflow assessing the effect of gait alterations on stresses in the medial tibial cartilage-combined musculoskeletal modelling and finite element analysis. *Scientific reports*, 7(1):1–14, 2017. 3

[14] D. A. Hirshberg, M. Loper, E. Rachlin, and M. J. Black. Coregistration: Simultaneous alignment and modeling of articulated 3d shape. In *European conference on computer vision*, pages 242–255. Springer, 2012. 2

[15] X. Huang, G. Mei, J. Zhang, and R. Abbas. A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*, 2021. 3

[16] S. F. Johnsen, Z. A. Taylor, M. J. Clarkson, J. Hipwell, M. Modat, B. Eiben, L. Han, Y. Hu, T. Mertzanidou, D. J. Hawkes, et al. Niftysim: A gpu-based nonlinear finite element package for simulation of soft tissue biomechanics. *International journal of computer assisted radiology and surgery*, 10(7):1077–1095, 2015. 2, 6

[17] S. Karanam, R. Li, F. Yang, W. Hu, T. Chen, and Z. Wu. Towards contactless patient positioning. *IEEE transactions on medical imaging*, 39(8):2701–2710, 2020. 3

[18] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[19] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009. 3, 6

[20] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M. J. Black, and P. V. Gehler. Unite the people: Closing the loop between 3d and 2d human representations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6050–6059, 2017. 3

[21] H. M. Le, T.-T. Do, T. Hoang, and N.-M. Cheung. Sdrsac: Semidefinite-based randomized approach for robust point cloud registration without correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 124–133, 2019. 3

[22] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015. 2, 3, 8

[23] S. A. Maas, B. J. Ellis, G. A. Ateshian, and J. A. Weiss. Febio: finite elements for biomechanics. *Journal of biomechanical engineering*, 134(1), 2012. 2, 6

[24] A. Mazier, S. Ribes, R. Testylier, F. Van Meer, B. Gilles, F. Faure, and S. Bordas. A rigged model of the breast for preoperative surgical planning. *arXiv preprint arXiv:2105.00763*, 2021. 2

[25] A. Mendizabal, P. Márquez-Neila, and S. Cotin. Simulation of hyperelastic materials in real-time using deep learning. *Medical image analysis*, 59:101569, 2020. 3

[26] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019. 7

[27] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. Osman, D. Tzionas, and M. J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10975–10985, 2019. 2, 3, 5

[28] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis. Learning to estimate 3d human pose and shape from a single color image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 459–468, 2018. 3

[29] M. Pfeiffer, C. Riediger, S. Leger, J.-P. Kühn, D. Seppelt, R.-T. Hoffmann, J. Weitz, and S. Speidel. Non-rigid vol-

ume to surface registration using a data-driven biomechanical model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 724–734. Springer, 2020. 3

[30] M. Pfeiffer, C. Riediger, J. Weitz, and S. Speidel. Learning soft tissue behavior of organs for surgical navigation with convolutional neural networks. *International journal of computer assisted radiology and surgery*, 14(7):1147–1155, 2019. 3

[31] R. Phellan, B. Hachem, J. Clin, J.-M. Mac-Thiong, and L. Duong. Real-time biomechanics using the finite element method and machine learning: Review and perspective. *Medical Physics*, 48(1):7–18, 2021. 3

[32] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 3, 7

[33] J. Romero, D. Tzionas, and M. J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics (ToG)*, 36(6):1–17, 2017. 2, 3

[34] Y. Salehi and D. Giannacopoulos. Physgnn: A physics-driven graph neural network based model for predicting soft tissue deformation in image-guided neurosurgery. *arXiv preprint arXiv:2109.04352*, 2021. 3

[35] L. Wang, J. Chen, X. Li, and Y. Fang. Non-rigid point set registration networks. *arXiv preprint arXiv:1904.01428*, 2019. 3

[36] Y. Xu, W. Wang, T. Liu, X. Liu, J. Xie, and S.-C. Zhu. Monocular 3d pose estimation via pose grammar and data augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 3

[37] H. Yang and L. Carlone. A polynomial-time solution for robust registration with extreme outlier rates. *arXiv preprint arXiv:1903.08588*, 2019. 3

[38] S. Yeo, J. Romero, M. Loper, J. Machann, and M. Black. Shape estimation of subcutaneous adipose tissue using an articulated statistical shape model. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(1):51–58, 2018. 3