Supplementary Material

A. Implementation Details

A.1. Deformation Module

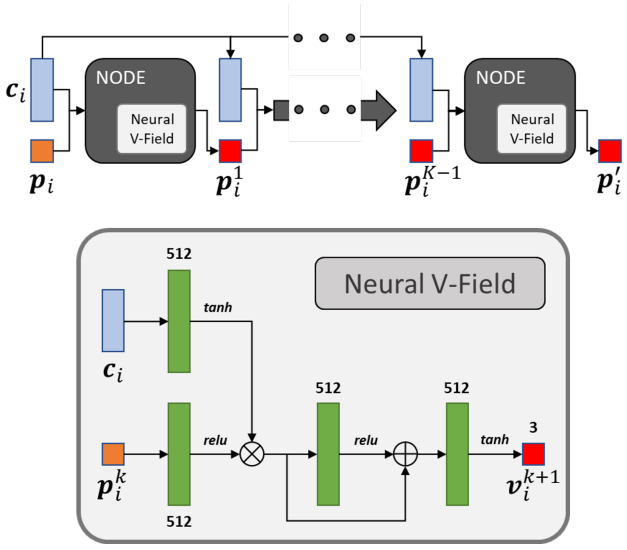


Figure 1. Deformation Module

In implementation, we realize the deformation module with several concatenated NODE blocks with neural velocity field as the dynamics function. As shown in Fig. 1, the integral of the last NODE block, which is the intermediate deformed position, will be the input of the next NODE block, together with the unchanged latent code of the shape. $\boldsymbol{p}_i$ are the original points in the SDF of shape X_i, $\boldsymbol{p}_i^k$ is the deformed positions of $\boldsymbol{p}_i$ after the k-th NODE block and $\boldsymbol{p}_i^{'}$ is the final deformed position of $\boldsymbol{p}_i$ in the template space.

The neural velocity field is simply residual blocks based on fully connected layers, which are represented as green blocks in Fig. 1. The hidden feature size are set to be 512 and the final activate function is set to be *tanh* because we want the deformed positions are within the normalized range [-1, 1]. In Fig. 1, we multiply the point features and shape features, but it is not necessary. There should be no significant difference if the point features and shape features are concatenated instead.

A.2. Training and Inference Setting

During training and inference, the tolerance of NODE solver is set as $1e^{-5}$. We set the length of deform code as 256 and the points sampled from each shape objects are 8000, of which 4000 are inside points and the others are outside points. In inference, the points sampled for optimizing deform code could be partial or complete. For each organ class, We train a NDF for 2000 epochs using Adam [4] with a learning rate $5e-4$ and batch size of 8. Also, we jointly optimize the deform codes along with NDF training using Adam with a learning rate $1e-3$. The hyper-parameters ϵ and λ follows the setting of [7] and regularization loss weight λ_{pw} is set to be 1. During inference, the deform codes is optimized for 2400 iterations with a learning rate $5e-2$.

DIT uses the exactly same training and inference setting as NDF. We train DeepSDF, DIF and AtlasNet for all organs with the default settings for airline shapes provided by their official implementations.

B. Data Source and Data Preparation

Pancreas CT dataset contains 82 abdominal contrast enhanced 3D CT scans with pancreas labelled. We split 61 samples for training and 21 for testing.

Multi-Modality Whole Heart Segmentation (MMWHS) challenge [8] provided 60 labelled multi-modality medical images with 7 whole heart substructure annotated, including left and right ventricle blood cavity, left and right atrium blood cavity, myocardium of left ventricle, ascending aorta as well pulmonary artery. In our experiments, we construct the whole heart label with the union of the annotations of blood cavity and myocardium of left ventricle. The ratio between training set and testing set is $3:1$.

Lung dataset combines the 85 chest data from [2] and 51 lobe segmentation data from [6]. The six sub-structures labelled by lobe segmentation data are upper left lung, middle left lung, lower middle lung, upper right lung, lower right lung and trachea. The organ shape we learn to reconstruct is the union of lung and trachea for both chest data ans lobbe segmentation data. We put 85 chest data and 17 lobe segmentation data into training set and 34 lobe segmentation data in testing set.

Liver dataset collects 190 samples coming from [2] with liver annotations. We split them into 145 and 45 shape instances into training and testing set respectively.

Since all these organs are labelled as 3D volumes, we first extract the organ surface using Marching Cube, then follow [5] to sample normalized SDF points near the mesh surface.

C. Experiments

C.1. Seen Shape Representation

From Tab. 1, we can see DIF-Net [3] performs best in terms of both CD and NC. As we have discussed in the main paper, we believe it is because of the extra points sampled on the shape surface, which could provide you more accurate representation about the zero level set of SDF. But

Model / Organ	CD Mean (↓)				CD Median (↓)				NC Mean (↑)				NC Median (↑)			
	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart
AtlasNet_Sph	4.5	1.76	3.64	5.03	4.08	1.39	3.21	4.64	0.733	0.836	0.82	0.817	0.736	0.841	0.824	0.822
AltlasNet_25	5.48	1.9	8.97	3.08	3.06	0.985	1.86	2.32	0.674	0.833	0.828	0.827	0.684	0.835	0.837	0.83
DeepSDF	0.34	0.232	0.247	0.375	0.335	0.226	0.244	0.359	0.927	0.876	0.933	0.936	0.931	0.877	0.933	0.94
DIF-Net	0.568	**0.122**	**0.122**	**0.243**	**0.205**	**0.102**	**0.118**	**0.245**	**0.979**	**0.894**	0.856	**0.961**	**0.981**	**0.895**	0.856	**0.965**
DIT	0.349	0.303	0.682	0.632	0.343	0.287	0.376	0.583	0.929	0.878	0.931	0.934	0.935	0.878	0.934	0.94
Ours	**0.315**	0.291	0.351	0.479	0.309	0.281	0.343	0.45	0.933	0.883	**0.939**	0.944	0.937	0.884	**0.94**	0.949

Table 1. **Shape Representation** – We demonstrate the reconstruction results of different representation methods on four organ categories. AtlasNet_Sph and AtlasNet_25 are AtlasNet using 3D sphere mesh and 25 square patches as the template shape respectively. Lower is better for chamfer distance ($\times 10^3$) and worse for normal consistency. Bold numbers are the best and the underlined are the second best.

as shown in main paper, the reconstruction performance of DIF-Net is not competitive compared to other deep implicit functions like DeepSDF [5] and DIT [7]. Involving many surface points in training will make deep implicit functions too "explicit" to handle the noise in the raw data. In general, DeepSDF performs better than our work and DIT on seen shape representation in terms of CD. It is because, DeepSDF has latent code with more freedom that it is said to represent one shape. But in DIT and NDF, the latent code controls the deformations of one shape. Thus, DeepSDF is more likely to train and overfit but relies less on the understanding of shape priors. Therefore, under a small number of training samples, we can witness DeepSDF is inferior to DIT and NDF on unseen shape reconstruction. What's more, compared to DIT, our model is better in all organs and both evaluation metrics.

C.2. Seen Shape Registration

As the unseen shape registration results, our model can achieve the best performance with great advantages over the other methods. AtlasNet_Sph performs good at the number of unpleasant faces and even better than NDF on Lung and heart in terms of E-NMF and SI ratio. But they are very poor in shape registration accuracy, which reveals that their shape reconstruction results are over-smoothened. Compared to DIT and DIF-Net, our results are mostly better, especially in terms of the unpleasant faces numbers.

C.3. Label Transfer

Here, we will provided quantitative results of label transfer realized by DIT and NDF. We investigate the label transfer IOU performance on the sub-structures labelled in MMWHS dataset and lobe segmentation dataset. For this task, we first choose 5 source meshes from labelled training samples and apply point correspondence towards all test samples based on these 5 source meshes. Now we have the align points of each test shapes with labels. Lastly, we label each vertex of test mesh with the nearest labeled aligned points. We didn't compare the results of DIF and AtlasNet because they are not even close to the registration performance of our method.

Tab. 3 reflects that our method can out-perform DIT on both MMWHS and lobe segmentation datasets in the label transfer task. The high IOU score reveals that our work can distinguish the semantic sub-regions regardless of the shape variances and has the potential to be applied in few-shot 3D shape segmentation learning.

C.4. Point-to-point Error

We conducted an experiment on a motion sequence (chicken wings dance) from the D-FAUST dataset [1], containing dense point correspondence. We used the point-to-point euclidean distance to evaluate the registration accuracy and compared our method to DiT (the previous state-of-the-art). Our method yields significantly better results in both accuracy (L2 distance) and quality (SI, self-intersection ratio) (Tab. 4).

C.5. Qualitative Results

We encourage readers to zoom in the following illustrations for more detailed comparisons.

Neural Diffeomorphic Flow. Fig. 2 demonstrate how two shapes find their correspondence via the learned template and show the intermediate results.

Template Shapes. Fig. 3 presents the template shapes learned by DIF, DIT and NDF.

Seen/Unseen Shape Representation/Reconstruction. In Fig. 4, we select five cases for each class to demonstrate the seen shape representation results of six methods we compared in the main paper. In Fig. 5, we select five cases for each class to demonstrate the unseen shape reconstruction results of six methods we compared in the main paper.

Seen and Unseen Shape Registration.

We sample 5 training cases and 5 test cases for the supplementary visual results. From Fig. 6-9, we can see our NDF is consistently better than the other two methods regarding both accuracy and fidelity when doing shape registration with verying-topology template shapes.

# of Vertices	Model / Organ	CD Mean				NC Mean				E-NMF Ratio Mean				SI Ratio Mean			
		Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart
2500	AtlasNet_Sph	4.5	1.76	3.64	5.03	0.733	0.836	0.82	0.817	26.7	1.13	**1.36**	**0**	5820	289	19.8	**0**
	AtlasNet_25	5.48	1.9	8.97	3.08	0.674	0.833	0.828	0.827	55.3	36.9	68.2	71.2	24700	25600	24800	26300
	DIF-Net	3.69	0.584	**0.372**	1.08	0.847	0.868	0.926	0.893	407	77.6	77.1	107	6870	32.6	933	1450
	DIT	0.377	0.312	0.848	0.678	0.92	0.875	0.922	0.931	11.4	1.41	20.6	25.7	5.91	9.45	410	**0**
	Ours	**0.326**	**0.309**	0.406	**0.528**	**0.93**	**0.881**	**0.934**	**0.941**	**0.492**	**0.221**	19.8	30.4	**0.656**	**0.276**	**15.9**	**0**
5000	DIF-Net	3.65	0.561	**0.345**	1.02	0.849	0.871	0.933	0.894	435	16.2	54.9	121	6000	79.2	1380	1890
	DIT	0.367	0.312	0.882	0.678	0.922	0.875	0.924	0.931	28.2	1.41	17.9	25.7	33.1	9.45	406	**0**
	Ours	**0.32**	**0.299**	0.387	**0.502**	**0.932**	**0.882**	**0.936**	**0.943**	**0.246**	**0.145**	**8**	**15.6**	**0**	**0**	**9.21**	**0**

Table 2. **Shape Registration on seen Shape Instances**

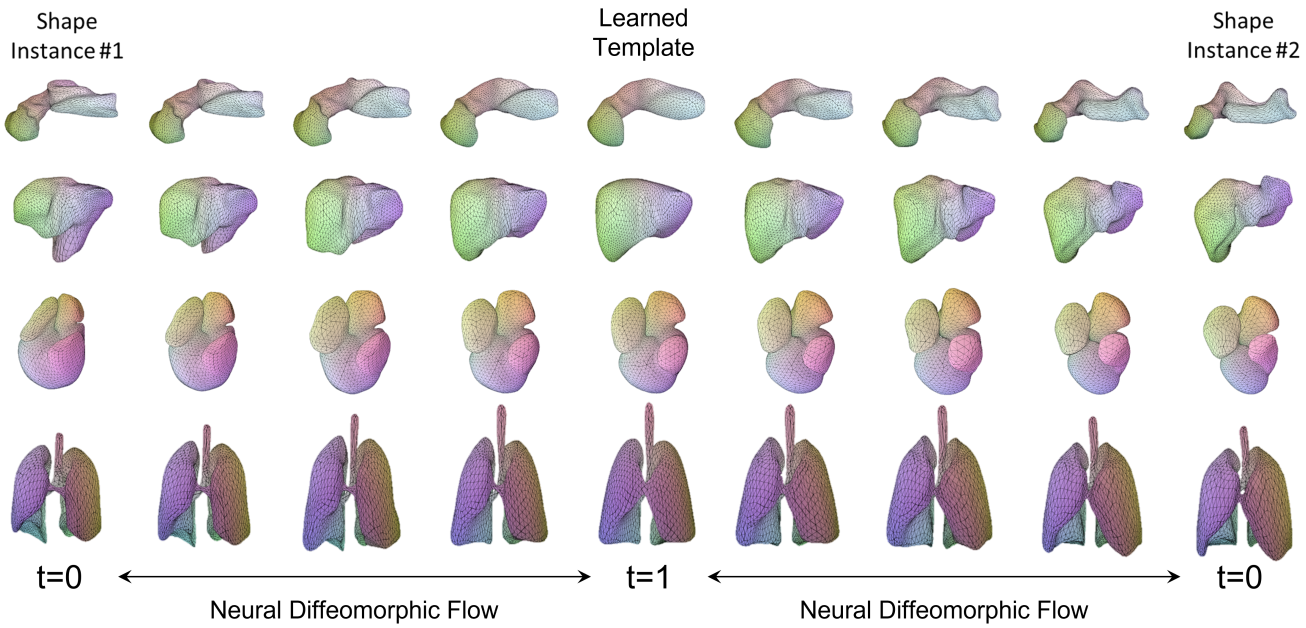


Figure 2. Demonstration of our representation

Model	MMWHS	LOBE	Mean
DIT	89.94	76.63	83.12
Ours	**92.38**	**76.66**	**84.52**

Table 3. Label IOU on the label transfer task

	train		test	
Model	L2 (↓)	SI (↓)	L2 (↓)	SI (↓)
DiT	0.0124	0.0646	0.0121	0.0990
ours	**0.0028**	**0.0215**	**0.0032**	**0.0222**

Table 4. Registrations results on D-FASUT sequence

References

[1] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J Black. Dynamic faust: Registering human bodies in motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6233–6242, 2017. 2

[2] Xuming Chen, Shanlin Sun, Narisu Bai, Kun Han, Qianqian Liu, Shengyu Yao, Hao Tang, Chupeng Zhang, Zhipeng Lu, Qian Huang, et al. A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy. *Radiotherapy and Oncology*, 160:175–184, 2021. 1

[3] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10286–10296, 2021. 1

[4] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[5] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2019. 1, 2

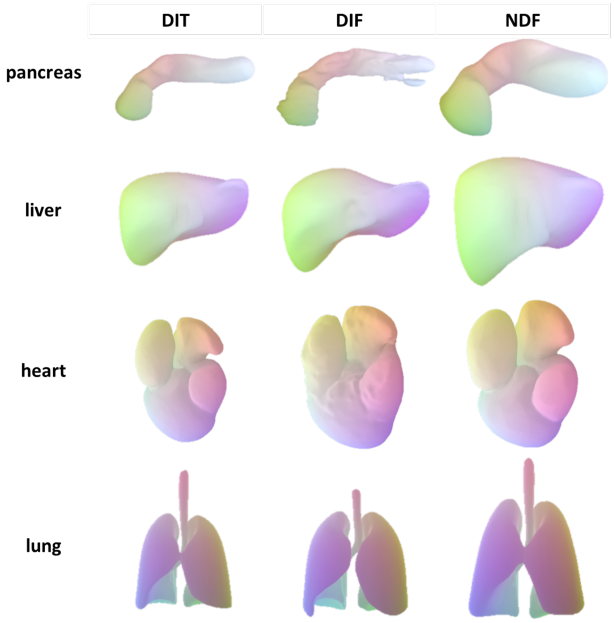
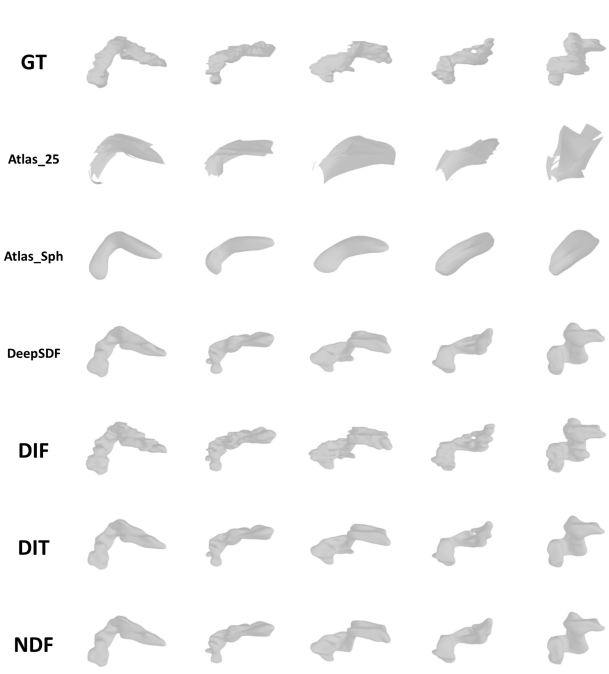


Figure 3. Template shapes learned by DIF, DIT and NDF

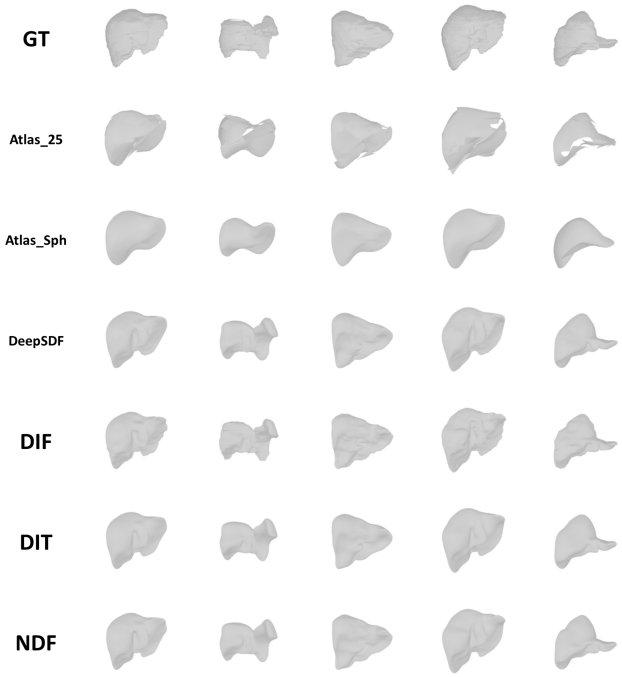
[6] Hao Tang, Chupeng Zhang, and Xiaohui Xie. Automatic pulmonary lobe segmentation using deep learning. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1225–1228. IEEE, 2019. 1

[7] Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1429–1439, 2021. 1, 2

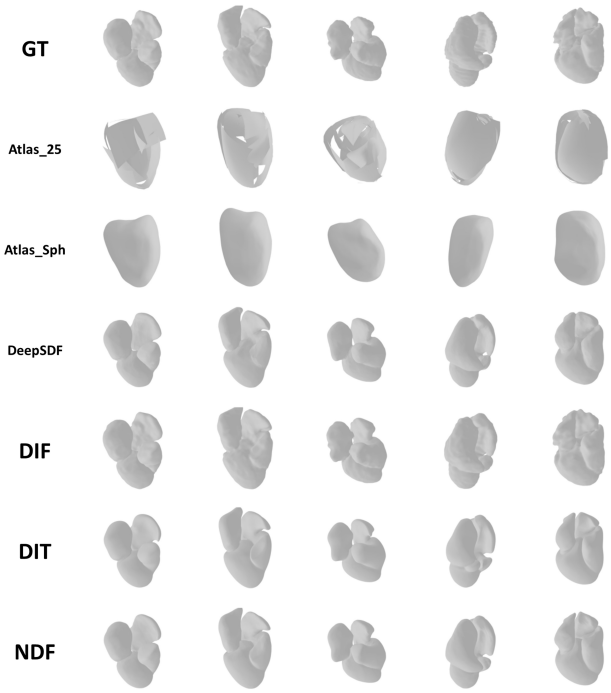
[8] Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016. 1



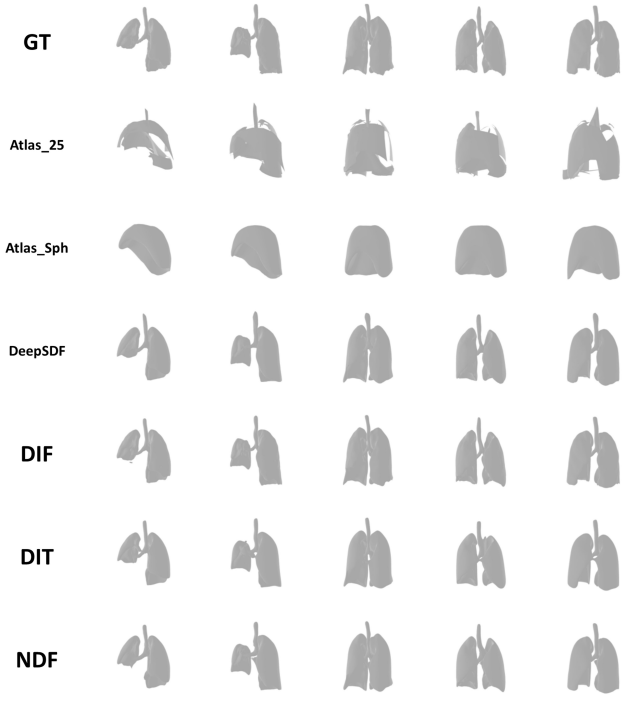
(a) Pancreas



(b) Liver

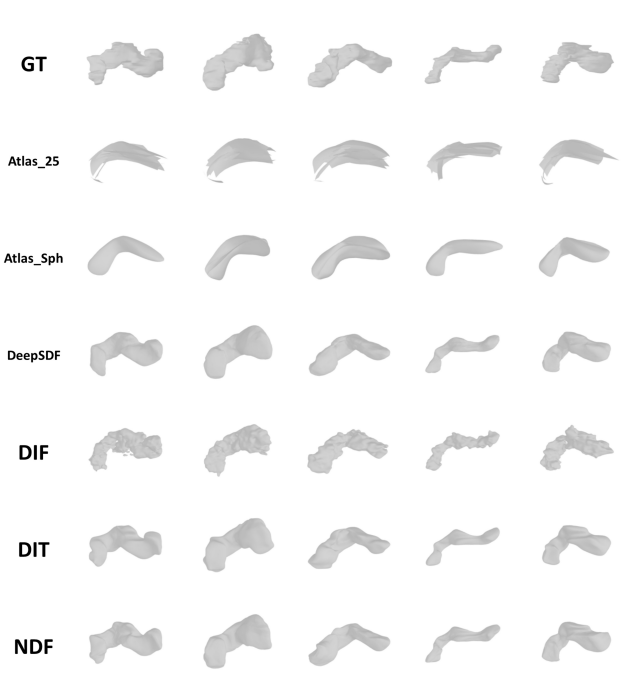


(c) Heart

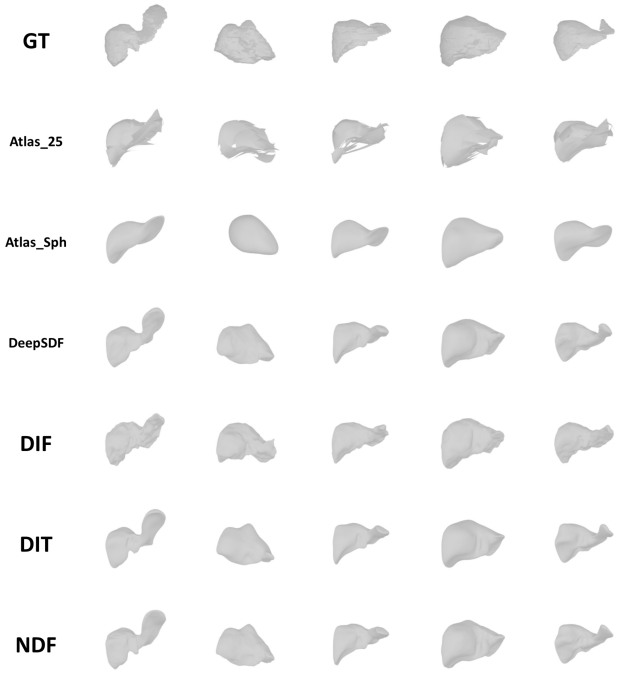


(d) Lung

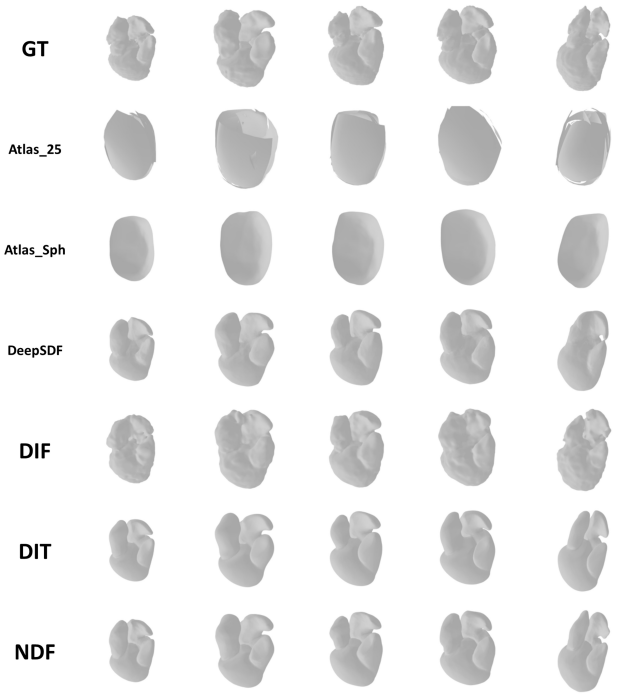
Figure 4. Seen Shape Representation Examples



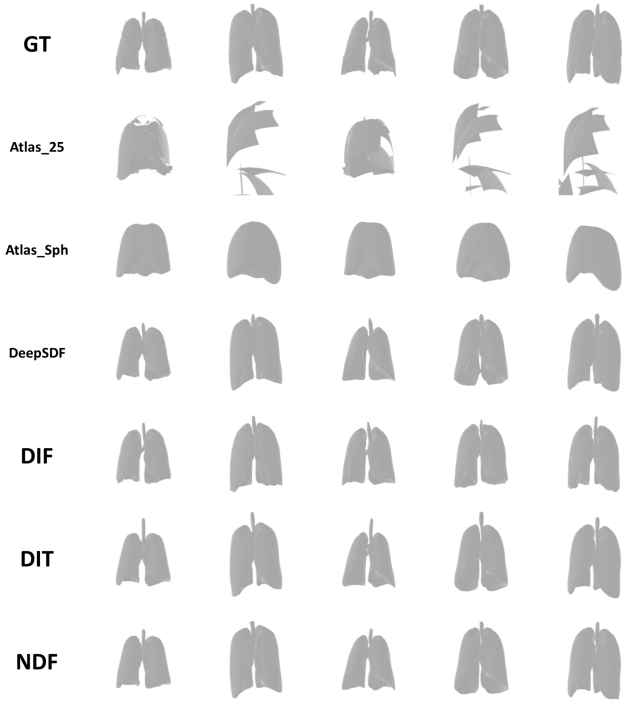
(a) Pancreas



(b) Liver

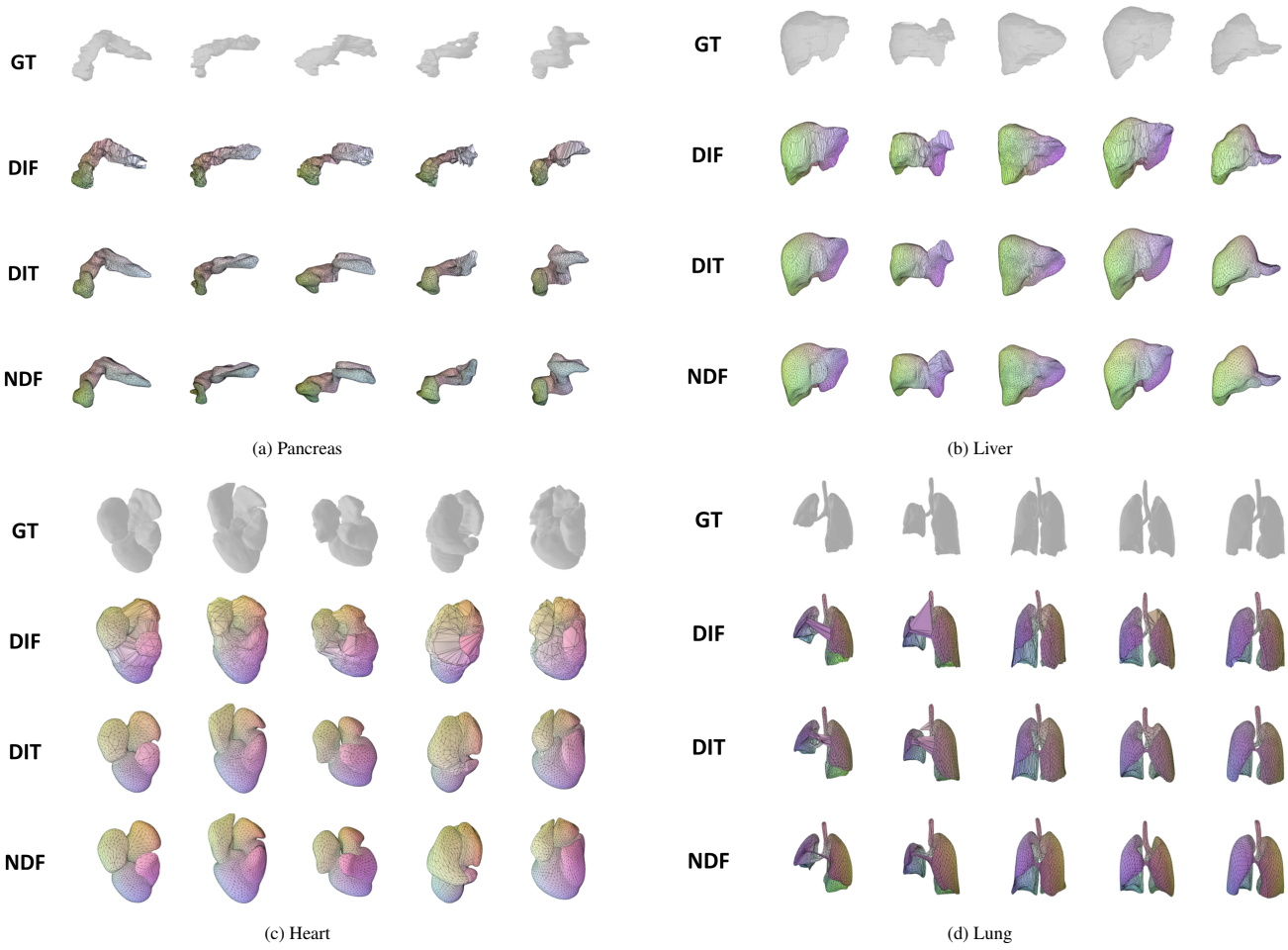


(c) Heart



(d) Lung

Figure 5. Unseen Shape Reconstruction Examples



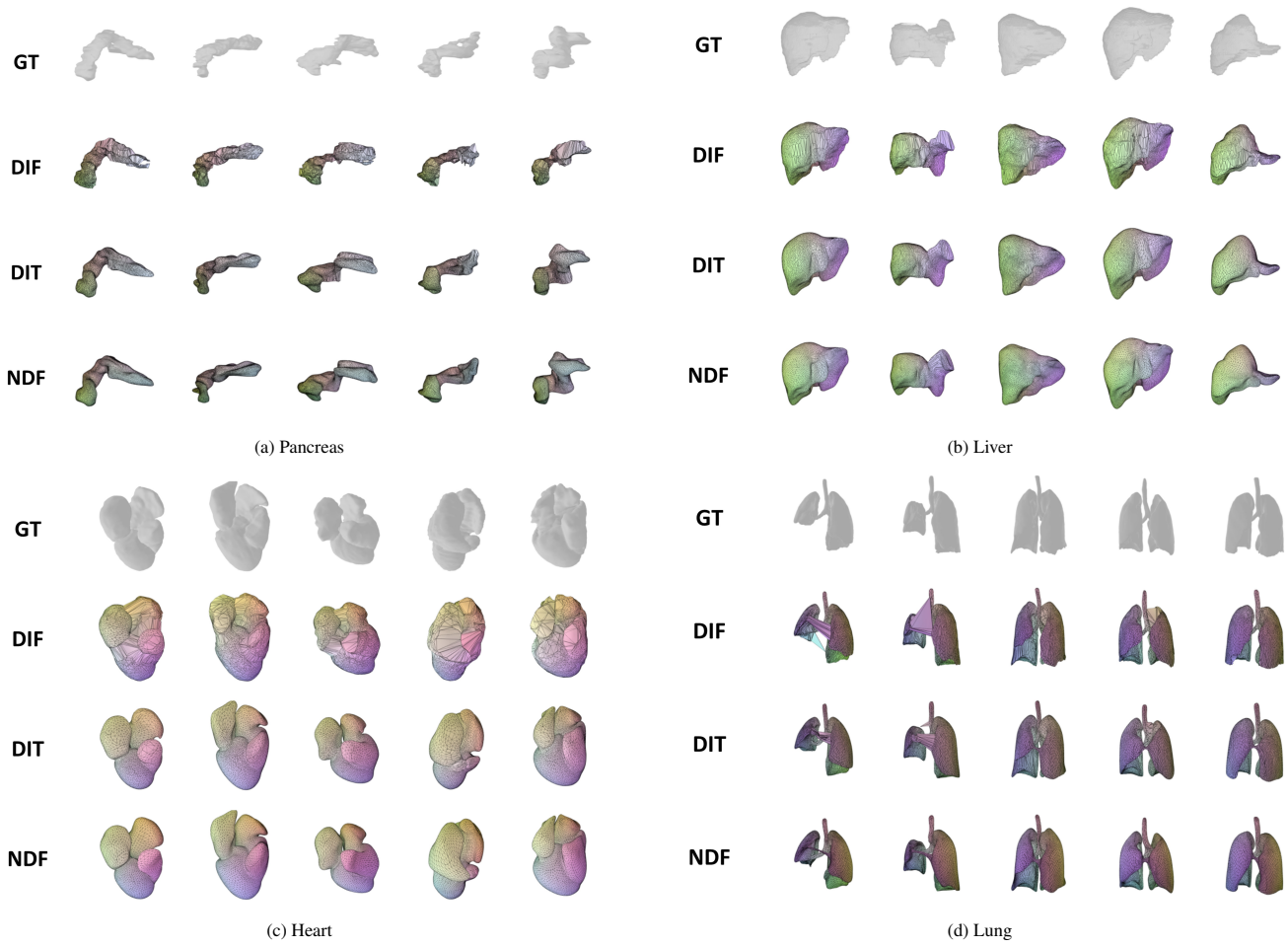
(a) Pancreas

(b) Liver

(c) Heart

(d) Lung

Figure 6. Seen Shape Registration Examples with 2500-vertex Template Shapes



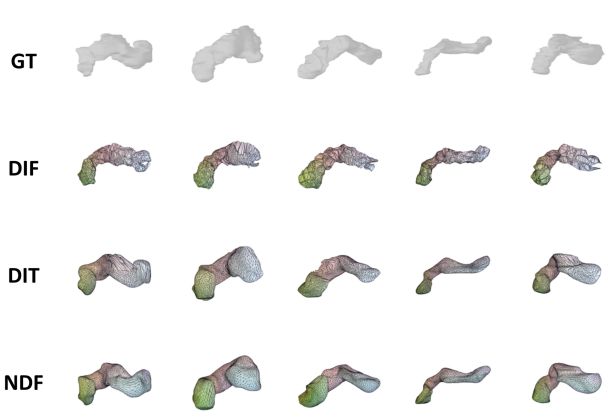
(a) Pancreas

(b) Liver

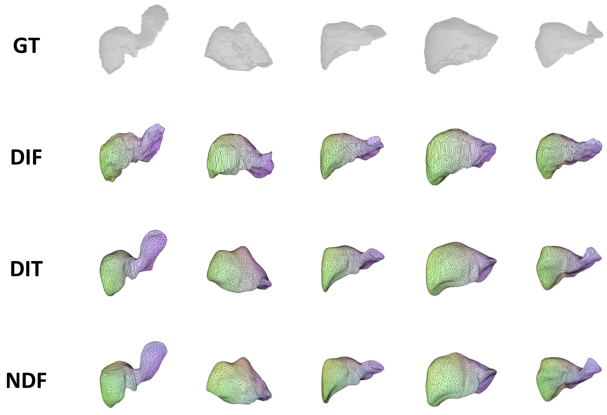
(c) Heart

(d) Lung

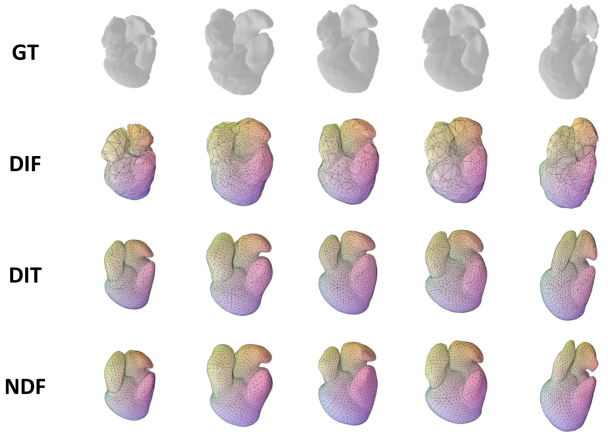
Figure 7. Seen Shape Registration Examples with 5000-vertex Template Shapes



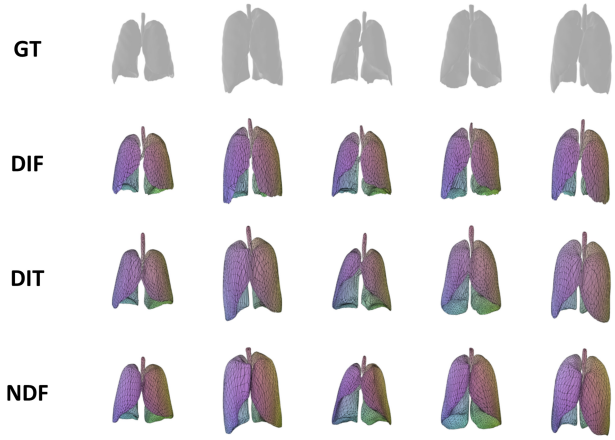
(a) Pancreas



(b) Liver



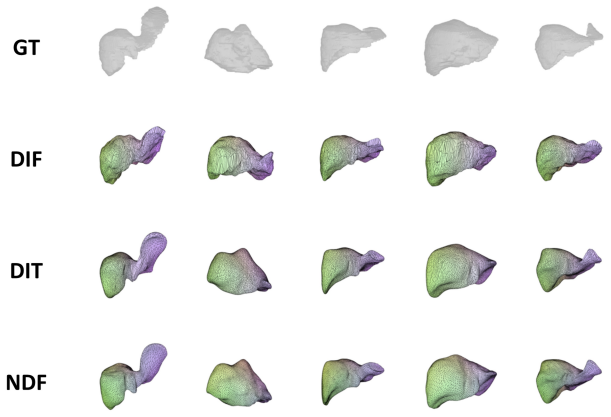
(c) Heart



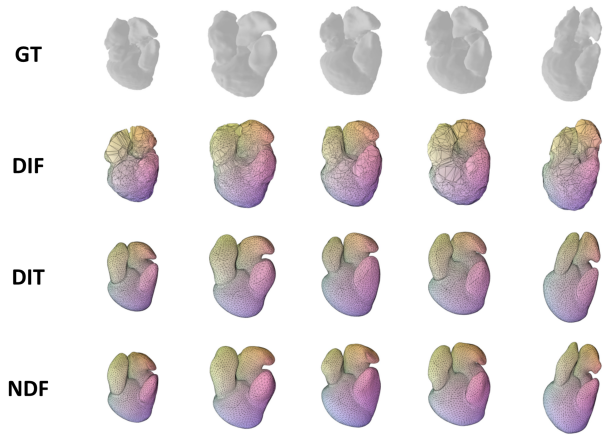
(d) Lung

Figure 8. Unseen Shape Registration Examples with 2500-vertex Template Shapes

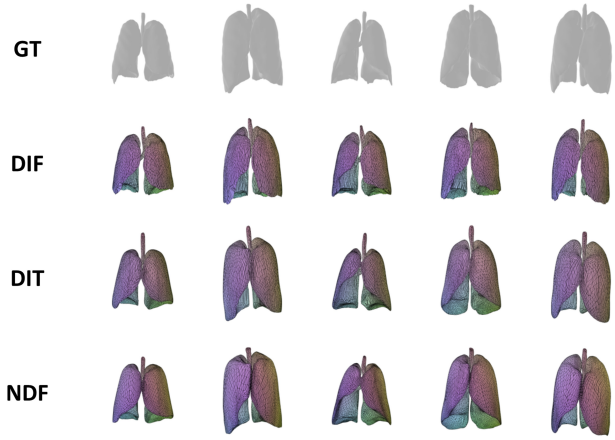
(a) Pancreas



(b) Liver



(c) Heart



(d) Lung

Figure 9. Unseen Shape Registration Examples with 5000-vertex Template Shapes