Hybrid Neural Diffeomorphic Flow for Shape Representation and Generation via Triplane

A. Data Source and Data Preparation

For the Pancreas-CT and Inhouse Liver datasets, we follow the same data splitting strategy as described in NDF. For the Pancreas-CT dataset, 61 out of 81 cases are used for training, while the remaining 21 cases are used for testing the generalization capability in the reconstruction tasks. As for the Inhouse Liver dataset, we use 145 instances for training and 45 instances for testing.

For the generation tasks, we utilize the Abdomen1k dataset, which contains 573 valid liver instances and 693 valid pancreas instances after filtering out incomplete shapes or shapes with large spacings. For our methods, DeepSDF, NDF and NFD, we use 300 cases for each organ to train the deformation module and template module. After training, we perform reconstruction for all shapes to obtain the training data for the diffusion models. For other generative models, we use all available data to train the models. For the heart and lung data, we follow the similar processor for data processing and training.

The initial data format for each instance is a mask. We start by extracting the mesh from the ground truth mask using the marching cubes algorithm, following the approach in NDF. To improve the mesh quality, we apply laplacian smoothing to remove artifacts. Once the mesh is obtained, we sample the signed distance values. Specifically, we uniformly sample 20 percent of the values throughout the entire space, while the remaining values are sampled near the surface of the mesh.

B. Training and Inference Settings

Each shape in our approach is represented by triplane features of size $L \times L \times 3C$, where L represents the dimensions (in our case, $L = 96$) and C represents the number of channels (in our case, $C = 4$). These triplane features are optimized during both the training and inference stages of our model.

During training, we use an initial learning rate of 0.005 for the deformation module $\mathcal{D}$ and the template module $\mathcal{T}$, while the initial learning rate for the triplane feature is set to 0.001. After every 500 epochs, all learning rates are multiplied by a factor of 0.5. We optimize the parameters using the Adam optimizer. For reconstruction of unseen shapes, we fix the deformation module $\mathcal{D}$ and the template module $\mathcal{T}$, and only optimize the per-object triplane features with 1600 iterations. We use an initial learning rate of 0.0005 and then halve the learning rate every 800 iterations.

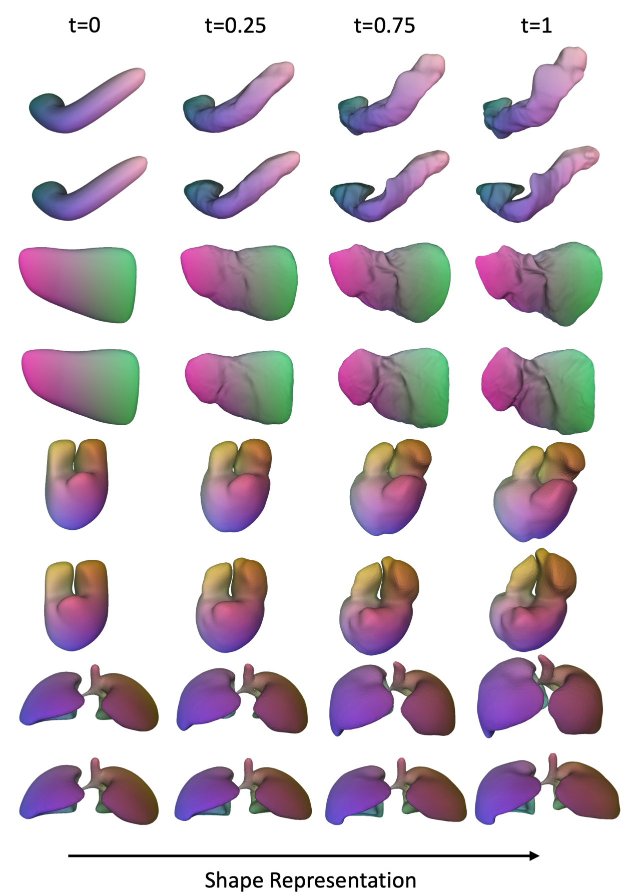

Figure 1. Additional **Shape Reconstruction** result on unseend data.

	CD Mean(↓)				NC Mean(↑)			
Model/Data	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung
DeepSDF	0.34	0.232	0.375	0.247	0.927	0.876	0.936	0.933
NFD	0.20	0.168	0.254	0.252	0.969	0.884	0.946	**0.944**
AtlasNet	4.5	1.76	5.03	3.64	0.733	0.836	0.817	0.82
DIT	0.349	0.303	0.632	0.682	0.929	0.878	0.934	0.931
DIF-Net	0.568	**0.122**	**0.243**	**0.122**	**0.979**	**0.894**	**0.961**	0.856
NDF	0.315	0.291	0.315	0.351	0.933	0.883	0.944	0.939
Ours	**0.133**	0.266	0.319	0.346	0.965	0.889	0.939	0.937

Table 1. **Shape Representation** results on training shapes. DIF-Net achieves the best result on the shape representation task. However, for the following shape reconstruction and registration on the unseen data, DIF works poorly compared to other methods, which demonstrates the overfitting of the training data for DIF-Net.

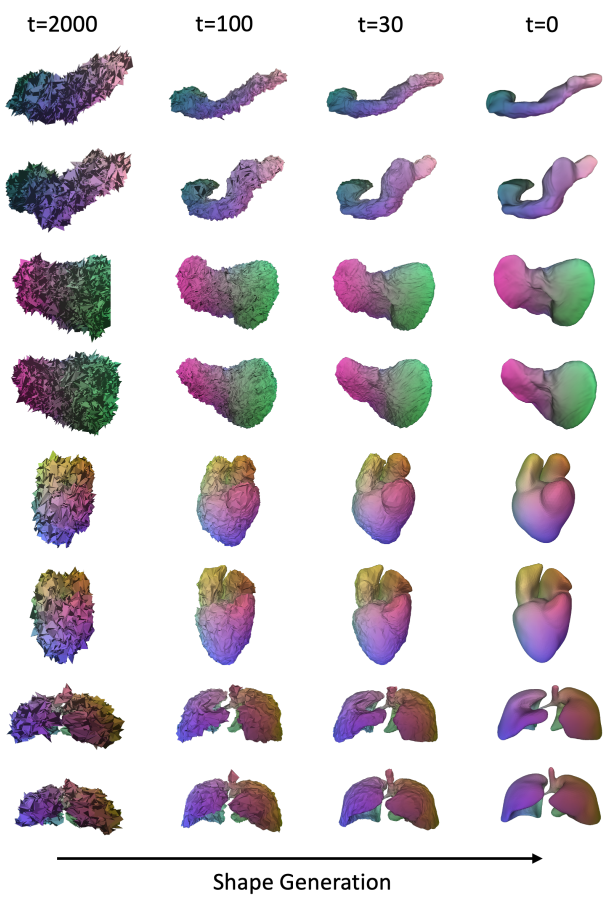

Figure 2. Additional **Shape Generation** results from generated triplane features.

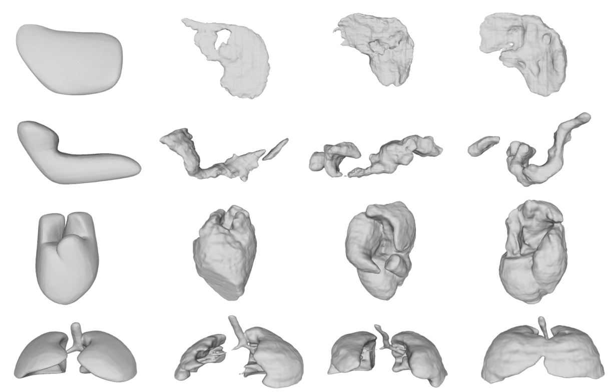

Figure 3. **NFD** ignores the underlying topology and point correspondence, therefore, the generated shape may be degraded and unrealistic.