

# Topology-Preserving Shape Reconstruction and Registration via Neural Diffeomorphic Flow

Shanlin Sun, Kun Han, Deying Kong, Hao Tang, Xiangyi Yan, Xiaohui Xie  
 University of California, Irvine

{shanlins, khan7, deyingk, htang6, xiangyy4, xhx}@uci.edu

## Abstract

*Deep Implicit Functions (DIFs) represent 3D geometry with continuous signed distance functions learned through deep neural nets. Recently DIFs-based methods have been proposed to handle shape reconstruction and dense point correspondences simultaneously, capturing semantic relationships across shapes of the same class by learning a DIFs-modeled shape template. These methods provide great flexibility and accuracy in reconstructing 3D shapes and inferring correspondences. However, the point correspondences built from these methods do not intrinsically preserve the topology of the shapes, unlike mesh-based template matching methods. This limits their applications on 3D geometries where underlying topological structures exist and matter, such as anatomical structures in medical images. In this paper, we propose a new model called Neural Diffeomorphic Flow (NDF) to learn deep implicit shape templates, representing shapes as conditional diffeomorphic deformations of templates, intrinsically preserving shape topologies. The diffeomorphic deformation is realized by an auto-decoder consisting of Neural Ordinary Differential Equation (NODE) blocks that progressively map shapes to implicit templates. We conduct extensive experiments on several medical image organ segmentation datasets to evaluate the effectiveness of NDF on reconstructing and aligning shapes. NDF achieves consistently state-of-the-art organ shape reconstruction and registration results in both accuracy and quality. The source code is publicly available at [https://github.com/Siwensun/Neural\\_Diffeomorphic\\_Flow--NDF](https://github.com/Siwensun/Neural_Diffeomorphic_Flow--NDF).*

## 1. Introduction

3D geometry representation is fundamental to many downstream tasks in computer vision such as 3D model understanding, reconstruction, and matching. In particular, shape representation is of vital importance to many

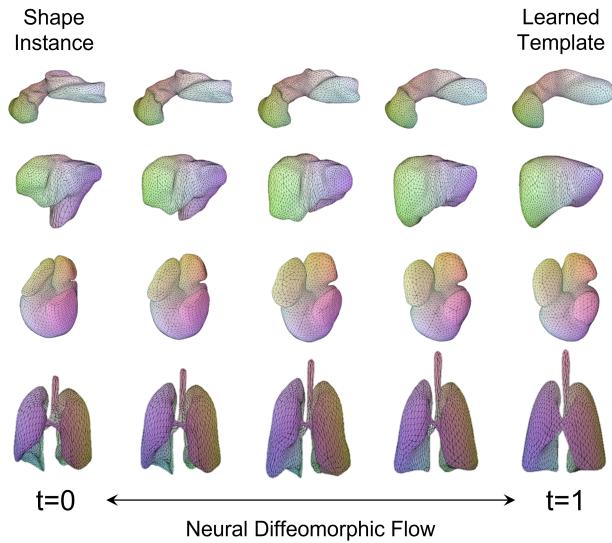


Figure 1. Our model deforms back and forth between shape instances and learned templates through neural diffeomorphic flow, in a invertible and progressive manner. Colors on the shape surface show the correspondences and edges show the topologies.

medical image applications such as organ segmentation [10, 44, 45, 51, 53, 56, 57], medical image reconstruction [52, 54, 55], shape abnormality detection and surgical navigation [31, 43].

Recently deep implicit functions (DIFs) have emerged as an effective and efficient tool for modeling 3D objects [12, 36, 40, 42, 50]. Compared to traditional representations such as voxel grids, point clouds, and polygon meshes, DIF-based 3D representations have the advantages of being compact while at the same time enjoying strong representation power, making it more suitable for modeling complex shapes with fine geometric details. However, DIFs do come with a strong drawback - it is difficult to establish correspondences between two shapes, unlike the traditional method such as meshes. This drawback limits the application of DIFs for shape analysis, especially in many medical image applications, where being able to map and compare

shapes is often a necessity.

A number of methods have been proposed to address the limitation of DIFs. The DIT (Deep Implicit Templates) [59] and DIF-Net (Deformed Implicit Field), build upon DeepSDF [40], formulates DIFs as conditional deformations of a template deep implicit function, and uses a spatial warping module to explicitly model the conditional deformations and infer point-wise transformations. [32] learns dense 3D shape correspondence from semantic part embedding by introducing an inverse implicit function to BAE-Net [11].

However, a common drawback of the above methods is that the conditional deformation modeled by these methods (e.g., LSTM in DIT) is agnostic of the topology of the shapes. This will be problematic in situations where two shapes share the same topology and we want the topology to be preserved after deformation. Applications falling into this category include modelling 3D shapes of human body, anatomical structures in medical images, and other objects with fixed topologies. What's more, considering the small number of anatomical shapes available for training, it is challenging to generalize the learned deformations to unseen data if no shape prior is utilized.

In this work, we propose a new formulation of DIFs called Neural Diffeomorphic Flow (NDF) for representing 3D shapes. Similar to DIT and DIF-Net, NDF models shapes as conditional deformations of a template DIF. But different from DIT, the conditional deformation is intrinsically diffeomorphic, thereby ensuring that the resulting deformation is topology preserving. NDF is also different from AtlasNet [23], which can preserve topology but requires predefined fixed topology as its shapes are modeled by meshes.

Our main contributions are summarized as follows:

- We introduce invertible NDF to match a shape to its implicit template. It can align point clouds or meshes without sacrificing accuracy while guaranteeing topology preservation.
- We design a quasi time-varying velocity field to learn diffeomorphic flows based on neural ODEs, allowing us to model shape deformation in a progressive and time-invertible manner.
- We tested NDF on multiple organ datasets and demonstrated that it leads to state-of-the-art shape reconstruction and registration results on both existing and new shapes. On shape registration, NDF generates one or several order of magnitude fewer unpleasant faces .

## 2. Related works

**Deep Implicit Functions.** Traditional implicit function is defined in the grid space and extracts the explicit shape sur-

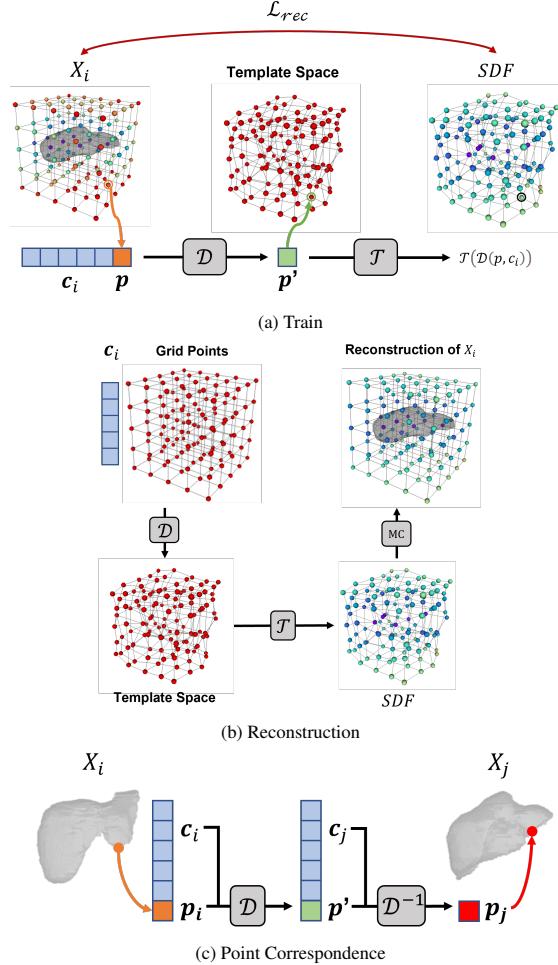


Figure 2. **Overview of NDF** - (a): We train deform code  $c_i$ , deformation module  $D$  and deep sdf representations  $T$  jointly.  $p$  is a sampled 3D position and  $p'$  is the deformation position of  $p$  in the template space. We sample points from the continuous 3D space, but we draw grids here only for illustration; (b): Suppose we have already optimized  $c_i$ , to reconstruct an unseen shape, we input grid points into our model and the reconstructed mesh is obtained via marching cube (MC) model output; (c): illustration of point correspondences as in Eq.13.

face from its zero-level set. Deep implicit function is the extension of traditional implicit function to represent shapes in continuous 3D space and have shown great representation capacity. DeepSDF [40] is an example of auto-decoder models representing continuous SDF. Many works are developed based on it, among which [6, 28, 46] try to depict finer structures by modelling SDF in the unit of local regions. Additionally, DualSDF [25] designs dual pathways (primitive and accurate) to represent SDF with VAD framework and C-DeepSDF [18] intends to improve the training strategy via curriculum learning. Occupancy Network [36] represents another branch of deep implicit function that constructs the solid mesh via classifying 3D points whether

are included in mesh or not. Occupancy Flow [39] follows similar ideas of Occupancy Network in shape representation but extend it to 4D with a continuous vector field in time and space.

**Point Correspondence and Shape Registration.** There are several ways to achieve point correspondence such as template learning [23, 27, 29, 33, 49], elementary representation [16, 21, 22], deformation field-based methods [35, 38] and so on. Mesh-based templates are very popular in representing similar shapes such human body, face and hand, where the templates topology are fixed, so they cannot deal with topological changes. Element-based methods can only capture structure-level features because they aim to describe complex shapes with simple elements. DIF-Net [15] and DIT [59] are typical deformation field-based methods and DIT can generate smoother deformation because it applies LSTM to do deformations. Our work can be seen as deformation field-based methods but our deformation field is topology-preserving and invertible.

**Diffeomorphic Transformation.** A diffeomorphism is an invertible mapping where the forward and backward transformations are smooth. It is widely used in nonrigid registration and shape analysis. Model complexity and computation makes it hard to be incorporated with deep learning solution. [17] is the first paper that built diffeomorphic image transformations into a deep classification model via CPAB transformations [20]. In terms of medical registration problem, people usually assume the velocity field is stationary and defined in the grid space [2]. Many works [3, 4, 13, 14, 30] used scaling and squaring method [1] to do fast integration of stationary velocity field. Recently, with the power of neural ordinary differential equation (NODE) solver [8, 9], optimizing a neural diffeomorphic flow efficiently became possible. Occupancy Flow [39], based on Occupancy Network, learns 4D reconstruction with implicit correspondences by modelling a temporally and spatially continuous vector field. Neural Mesh Flow [24] focuses on generating manifold mesh from images or point clouds via conditional continuous diffeomorphic flow.

### 3. Method

Our Deep Implicit Function representation via NDF follows the formulation of Deep Implicit Templates (DIT) [59], which decompose a Coded Shape DeepSDF [40] into a Single Shape DeepSDF and a conditional spatial warping function. Our work depicts this spatial warping function as a conditional diffeomorphic flow under which topology is preserved. In this section, we first review Deep Implicit Templates and then introduce our proposed conditional spatial deformation.

### 3.1. Review of Deep Implicit Templates

Deep Implicit Templates, same as DeepSDF, represent a 3D shape  $X_i$  with a continuous signed distance field (SDF)  $\mathcal{F}$ . Given a random 3D point  $p$  and deform code  $c_i$  of length  $k$ ,  $\mathcal{F}$  outputs the point's distance to the closest surfaces, whose sign indicates whether the point lies inside or outside the underlying shape surface:

$$\mathcal{F}(p, c_i) = s, \quad \text{where } p \in \mathbb{R}^3, c_i \in \mathbb{R}^k, s \in \mathbb{R} \quad (1)$$

During training, each shape code is paired with one training shape  $X_i$ . During inference, the deform code corresponding to a new shape is obtained via optimization. The underlying shape surface is implicitly expressed as the zero-level set surface of  $\mathcal{F}$ , obtained with, for example, Marching Cubes [34].

Different from DeepSDF where the physical meaning of latent code  $c$  is ambiguous, DIT treats  $c$  as a variable controlling how each shape deforms to a template shape, so that the conditional continuous SDF  $\mathcal{F}$  can be decompose into  $\mathcal{T} \circ \mathcal{D}$

$$\mathcal{F}(p, c_i) = \mathcal{T}(\mathcal{D}(p, c_i)) \quad (2)$$

where  $\mathcal{D} : \mathbb{R}^3 \times \mathbb{R}^k \mapsto \mathbb{R}^3$  is the conditional spatial deformation module that maps the coordinate of  $p$  of shape  $X_i$  to a canonical position  $p'$  given deform code  $c_i$  and  $\mathcal{T}$  is essentially a single shape DeepSDF modeling the implicit template. By this design, it builds up point correspondences between the learned template and each shape instance, on top of which correspondences across all shapes within one category are achieved.

### 3.2. Neural Diffeomorphic Flow

**Diffeomorphic Flow.** We intend to establish dense point correspondences between each shape object and the template shape, and keep the desired geometric topology using diffeomorphic flow. Let  $\Phi_i(p, t) : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$  describe the continuous trajectory of a 3D point ( $p$ ) during the time interval  $[0, 1]$  where the starting points and destination points respectively located in the SDF of shape  $X_i$  and the template shape. And let  $\mathbf{v}_i(p, t) : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$  define the velocity field of 3D points with respect to shape  $X_i$  in time interval  $[0, 1]$ . The diffeomorphic flow  $\Phi_i$  of shape  $X_i$  is the solution of the initial value problem (IVP) of an ordinary differential equation (ODE) as below:

$$\frac{\partial \Phi_i}{\partial t}(p, t) = \mathbf{v}_i(\Phi_i(p, t), t) \quad \text{s.t.} \quad \Phi_i(p, 0) = p \quad (3)$$

where  $p \in \mathbb{R}^3$  is a 3D position on the SDF of shape  $X_i$ . Thus, the diffeomorphic deformations module  $\mathcal{D}$  conditioning on the deform code of shape  $X_i$  is given by:

$$\mathcal{D}(p, c_i) = \Phi_i(p, 1) \quad (4)$$

If velocity field  $\mathbf{v}_i(\cdot, \cdot)$  is globally Lipschitz continuous, the solution for the IVP exists and is unique in the interval  $[0, 1]$ , which means any two ODE trajectories do not cross each other [19]. This can provide the diffeomorphic flow with the property of topology preservation to maintain structure consistency.

Diffeomorphic flow is invertible, thus the inverse flow  $\Psi_i : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$  from template shape to shape  $X_i$  can be obtained by solving the following ODE:

$$\frac{\partial \Psi_i}{\partial t}(\mathbf{p}, t) = -\mathbf{v}_i(\Psi_i(\mathbf{p}, t), t) \quad \text{s.t.} \quad \Psi_i(\mathbf{p}, 0) = \mathbf{p} \quad (5)$$

$$\mathcal{D}^{-1}(\mathbf{p}, \mathbf{c}_i) = \Psi_i(\mathbf{p}, 1) \quad (6)$$

where  $\mathbf{p}$  is a 3D point on the SDF of template shape and  $\mathcal{D}^{-1}$  denotes the inverse diffeomorphic field. So far, the invertible deformation between any shape instance and the templates of its category can be described as the integral of the velocity field.

**Conditional Quasi Time-varying Velocity Field.** Our goal is to learn a neural network that parameterizes the velocity field to capture the dense topology-preserved point correspondence across shape objects. In this section, we will describe how we design it. We denote  $\mathbf{v}$  as the neural network representing the velocity field in time and space, and  $\mathbf{v}_i$  describes the neural velocity field with respect to the shape  $X_i$ , i.e.,  $\mathbf{v}_i(\mathbf{p}, t) = \mathbf{v}(\mathbf{p}, \mathbf{c}_i, t)$ , where  $\mathbf{c}_i$  a vector controls how neural velocity field deforms points in the SDF of  $X_i$ .

$\mathbf{v}_i(\mathbf{p}, t)$  is a general expression of velocity field depending on both time and position, here we call it time-varying velocity field to distinguish with stationary velocity field  $\mathbf{v}_i(\mathbf{p})$ , where the velocity of a point in the field only decided by its position. In our case, training time-varying velocity field might be difficult because unlike 4D reconstruction [39] having adequate training samples of multiple frames, our model can only be supervised on  $t = 0$  and  $t = 1$ . In dealing with the medical registration problem defined in regular grid space, people often assume the deformable transformation is based on a stationary velocity field because it can be efficiently integrated through scaling and squaring technique [26, 37].

Our insight is to design a quasi time-varying velocity field composed of several subsequent stationary velocity fields to realize progressive deformations. Concretely, we hope the first several diffeomorphic flows are able to roughly align the shape instance to its template shape while the last several diffeomorphic flows take the minor adjustment in geometric details. Suppose this quasi time-varying velocity field is made up with  $K$  stationary velocity fields. Let  $\mathbf{v}_i^k(\cdot) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  describe the  $k$ -th stationary velocity field of shape  $X_i$  and  $\chi_A(\cdot)$  is an indicator function of  $A$ .

The velocity field is governed by the following formula

$$\mathbf{v}_i(\Phi_i(\mathbf{p}, t), t) = \sum_{k=0}^K \chi_{[\frac{k}{K}, \frac{k+1}{K}]}(t) \cdot \mathbf{v}_i^k(\Phi_i(\mathbf{p}, t)) \quad (7)$$

As shown in Eq.7, the quasi time-varying velocity field is basically a step function regarding time  $t$ . We can further derive the diffeomorphic flow by integrating  $\mathbf{v}_i(\cdot, \cdot)$

$$\Phi_i(\mathbf{p}, t) = \Phi_i(\mathbf{p}, \frac{k}{K}) + \int_{\frac{k}{K}}^{t-\frac{k}{K}} \mathbf{v}_i^k(\Phi_i(\mathbf{p}, t)) dt \quad (8)$$

where  $t \in [\frac{k}{K}, \frac{k+1}{K}]$  and  $k \in \{0, 1, \dots, K-1\}$ . This equation can be solved with a neural ordinary differential equation (NODE) solver [9]. In other words,  $\Phi_i(\mathbf{p}, t)$  is the output of a NODE block receiving  $\mathbf{v}(\mathbf{p}, \mathbf{c}_i, t)$  as dynamic function. To make this conditional velocity field neural network compatible with NODE, the conditional parameters (deform codes  $c$ ) stay unchanged when solving the integral.

We use residual MLP architecture similar to [24, 39] to represent the velocity field. For each shape  $X_i$ , we initialize the deform code  $\mathbf{c}_i$  as [40] suggests, which can be either concatenated or multiplied to the point features.

### 3.3. Training

We employ two modules to represent continuous SDF: a conditional deformation module  $\mathcal{D}$  and a single shape DeepSDF  $\mathcal{T}$ . Like the other auto-decoder models, these two modules and deform code  $c$  are trained jointly (as illustrated in Fig. 2a) with a reconstruction loss and a regularization loss:

$$\mathcal{L} = \mathcal{L}_{rec} + \lambda_{reg} \mathcal{L}_{reg} \quad (9)$$

Since the deformation between a shape and the template shape performs in a progressive manner, we choose to use the curriculum learning strategy same as [18, 59]. [18] set different curriculum learning hyper-parameters at different training stages and [59] set different hyper-parameters for different warping stages. In our work, we will count the deformations of different timestamps into curriculum learning. To this end, the reconstruction loss can be written as:

$$\mathcal{L}_{rec} = \sum_{t \in T} \sum_{i=1}^N \sum_{j=1}^S L_{\epsilon_t, \lambda_t} (\mathcal{T}(\Phi_i(\mathbf{p}_j, t)), s_{i,j}) \quad (10)$$

where  $T$  is the set of evaluating timestamps,  $N$  is the number of training shapes,  $S$  is the number of SDF samples for one shape,  $s_{i,j}$  is the ground truth SDF of the  $j$ -th samples point  $\mathbf{p}_j$  from the  $i$ -th shape and  $\Phi_i(\mathbf{p}, t)$  is defined as in Eq.8.  $L_{\epsilon_t, \lambda_t}$  is the curriculum training loss where  $\epsilon$  controls the width of the tolerance zone and  $\lambda$  controls the importance of the hard and semi-hard examples. The evaluation timestamps  $T$  is set to be  $\{0.25, 0.5, 0.75, 1.0\}$  in

practice. For more details about curriculum learning, please refer to [18].

Our regularization loss is very concise because NODE solver can largely prevent self-intersection [19, 58] with no explicit regularization. In other words, we don't need to design the point pair regularization [59] or deformation smoothness prior [15] to avoid the local distortion. In [24], the authors showed a toy example to demonstrate the "regularfizar's dilemma" that introducing strong regularization in training might lead to an unpleasant mesh reconstruction results. In our setting, we only need to constrain the magnitude of deformation field as well as learned deform codes.

$$\mathcal{L}_{reg} = \sum_{t \in T} \sum_{i=1}^N \sum_{j=1}^S \mathcal{L}_{0.25} \left( \|\Phi_i(\mathbf{p}_j, t) - \mathbf{p}_j\|_2 \right) + \sum_{k=1}^K \|\mathbf{c}_k\|_2^2 \quad (11)$$

where  $\mathcal{L}_{0.25}$  is the Huber loss with  $\delta = 0.25$ . The goal of this point-wise regularization loss is to prevent the model from learning an over-simplified template but to look for a template shape which owns the most common structures that all shape instances within one category share.

### 3.4. Inference

At inference, after fixing the trained parameters of NDF, a deform code  $\mathbf{c}_i$  for a new shape  $X_i$  should be obtained via optimization in the first place. In our work, shape reconstruction is to extract the zero-level set surface from the SDF of an unseen shape object, which is predicted by our model given the optimized deform code. The correspondence between two shape objects  $X_i$  and  $X_j$  can be found via forward diffeomorphic deformations from  $X_i$  to template space and then backward diffeomorphic deformations from template space to  $X_j$ , given  $\mathbf{c}_i$  and  $\mathbf{c}_j$  respectively.

**Learn deform code.** Same as DeepSDF [40], a deform code  $\mathbf{c}_i$  of shape  $X_i$  is the Maximum-a-Posterior estimation as:

$$\hat{\mathbf{c}}_i = \arg \min_{\mathbf{c}_i} \sum_{(\mathbf{p}, s) \in X_i} \ell(\mathcal{F}(\mathbf{p}, \mathbf{c}_i), s) + \frac{1}{\sigma^2} \|\mathbf{c}_i\|_2^2 \quad (12)$$

Different from the progressive reconstruction loss we designed for training,  $\ell(\cdot, \cdot)$  here is the absolute error between model outputs and ground truth.

**Reconstruction.** Having learned the deform code  $\mathbf{c}_i$ , shape  $X_i$  comes from the zero-level set surface of  $\mathcal{F}(\mathbf{p}_{grid}, \mathbf{c}_i)$  (shown in Fig. 2b), where  $\mathbf{p}_{grid} \in \mathbb{Z}_+^3$ . The resolution of reconstructed mesh can be manipulated by the number of grid points. In practice, we sample  $256^3$  grid points for all deep implicit functions for comparison.

**Point correspondence and Shape Registration.** In AtlasNet [23], the points that could find correspondence are

only template mesh vertices. As for implicit template-based methods such as [15, 59], they learn dense point correspondence, which is approximated since it is built by nearest neighbour search in the template space.

In our work, suppose we have a 3D point  $\mathbf{p}_i$  in shape  $X_i$ , its correspondence point  $\mathbf{p}_j$  in shape  $X_j$  can be found by (also shown in Fig. 2c):

$$\mathbf{p}_j = \mathcal{D}^{-1}(\mathcal{D}(\mathbf{p}_i, \mathbf{c}_i), \mathbf{c}_j) \quad (13)$$

where  $\mathbf{c}_i$  and  $\mathbf{c}_j$  are the optimized deform codes of shape  $X_i$  and  $X_j$  respectively.

Compared to point correspondence, shape registration not only seeks the aligned point set but also the aligned mesh. That means, given the source mesh  $\mathcal{M}_s$  with vertices  $\mathcal{V}_s$  and edges  $\mathcal{E}_s$ , the target aligned mesh  $\mathcal{M}_t$  is  $(\mathcal{V}_t, \mathcal{E}_s)$ , where  $\mathcal{V}_t$  is the correspondence point set of  $\mathcal{V}_s$ .

## 4. Experiments

**Datasets.** NDF focuses on reconstructing shapes with common intrinsic topology, we choose to demonstrate our results on four medical datasets: Pancreas CT [41], Multi-Modality Whole Heart Segmentation [60], Lung and Liver, since these four types of organ have clear common topology but are of fair shape variation as can be learned from the difference between learned templates and shape instances. For more details about data source and preparation, please refer to supplementary material.

**Experimental Setup.** We conduct two types of experiments to support the effectiveness of NDF. First, we investigate the representation power of our diffeomorphic flow-based methods on training samples and the reconstruction power on unseen shapes. We then evaluate the quality of the learned correspondence between two meshes (shape registration).

The natural baselines for shape registration are DIT [59] and DIF-Net [15] because we share the similar shape representation formula based on deep implicit function. We also compare our model with AtlasNet [23] which reconstructs shape using explicit mesh parameterization. To make our comparisons fair, we build our model based on the implementation of DeepSDF and choose it as the baseline for 3D shape representation besides DIT, DIF-Net and AtlasNet.

### 4.1. Shape Representation and Reconstruction

We use chamfer distance (CD) and normal consistency (NC) as the metrics to evaluate the quality of shapes representations and reconstructions by all methods. In the scope of this work, shape representation is to represent seen shape objects given the trained deform code and shape reconstruction is to represent unseen shape instances after optimizing

Model / Organ	CD Mean ( $\downarrow$ )				CD Median ( $\downarrow$ )				NC Mean ( $\uparrow$ )				NC Median ( $\uparrow$ )			
	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart	Pancreas	Liver	Lung	Heart
AtlasNet.Sph [23]	8.08	3.46	5.01	7.55	7.44	2.46	3.76	7.38	0.703	0.823	0.824	0.808	0.7	0.829	0.826	0.814
AtlasNet.25 [23]	6.05	2.48	207	4.86	5.64	1.72	4.54	3.86	0.65	0.818	0.772	0.824	0.643	0.823	0.791	0.823
DeepSDF [40]	0.711	0.539	0.669	<b>0.951</b>	0.675	0.536	0.661	<b>0.898</b>	0.898	0.866	0.928	0.913	0.903	0.868	0.929	0.92
DIF-Net [15]	4.18	1.58	1.86	2.23	3.97	1.25	1.67	1.83	0.756	0.832	0.882	0.838	0.768	0.837	0.885	0.838
DIT [59]	0.63	0.509	0.712	1.05	0.658	0.505	0.693	0.976	0.903	0.87	0.934	0.919	0.904	0.873	0.934	0.93
<b>Ours</b>	<b>0.512</b>	<b>0.476</b>	<b>0.643</b>	0.993	<b>0.515</b>	<b>0.479</b>	<b>0.631</b>	0.925	<b>0.917</b>	<b>0.873</b>	<b>0.937</b>	<b>0.923</b>	<b>0.918</b>	<b>0.875</b>	<b>0.937</b>	<b>0.932</b>

Table 1. **Shape Reconstruction** – 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. The chamfer distance results shown above are multiplied by  $10^3$ .  $\uparrow$  means higher is better and  $\downarrow$  means lower is better. Here we use "Lung" to denote the union shape of lung and trachea organ and use "Heart" to denote the union structures of blood cavities.

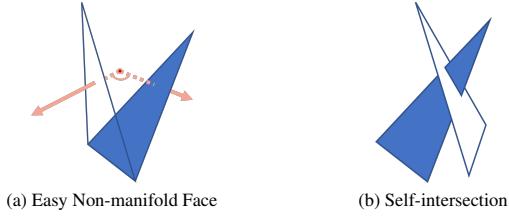


Figure 3. **Unpleasant Faces** - (a): one face is a E-NMF if its normal direction is significant different from that of its adjacent faces; (b): SI faces cross other faces in the same mesh.

the deform code. So, shape representation tells the effectiveness of representation methods while shape reconstruction reflects the generability. Due to space limitation, we only report the complete shape representation results for all datasets in supplementary material.

Generally speaking, DIF-Net shows the strongest shape representation ability but very poor shape reconstruction performance. We believe the overfitting is a result of its point sampling strategy that many surface points are involved in training and inference. In Tab. 1, we can observe NDF achieves the best reconstruction results in terms of both chamfer distance and normal consistency for almost all datasets, compared to the state-of-the-art methods.

## 4.2. Shape Registration

We have described how we realize shape registration on top of dense point correspondence in Sec. 3.4. In our experiments of shape registration, the source meshes  $\mathcal{M}_s$  are basically the template meshes and the target meshes  $\mathcal{M}_t$  are all shape instances. To make different methods comparable, we apply Approximated Centroidal Voronoi Diagrams (ACVD) [47, 48] to the template meshes of DIT, DIF-Net and NDF to make their template meshes meet the same resolution and similar topology. Specifically, we re-mesh these template meshes into  $\mathcal{M}_s$  with 2500 vertices (clusters) or 5000 vertices.

Apart from CD and NC, we design two more metrics to evaluate the geometrical fidelity of registration results: easy non-manifold face (E-NMF) ratio and self-intersection (SI)

ratio, as shown in Fig. 3, NMFs are such faces that have opposite normal direction to their adjacent faces. But in our scenario, this definition is too harsh because for organs such as lung and liver, the sudden change of face normal directions might take place in some local regions. As a result, one face will be defined as an E-NMF if the cosine similarity between normal directions of any of its adjacent faces and itself is less than  $\delta$ . It is set to be 0 when we evaluate heart and pancreas organs, and set to be -0.5 and -0.8 for lung and liver respectively.

Tab. 2 strongly supports that our NDF can densely align points across shapes while maintaining the topology. NDF achieves the best results in accuracy whatever experiment settings and organ classes are. Also, as for E-NMF ratio and SI ratio, our model can also outperform the other methods in most cases. AtlasNet.Sph beats us on liver, lung and heart regarding E-NMF ratio in a price of the over-smoothed reconstruction results. Notably, in comparison with DIF-Net and DIT that are very competitive in shape representation and reconstruction, our method obviously outperforms them in all metrics due to the properties of deep diffeomorphic flow (Sec. 3.2). In summary, NDF is superior to the other state-of-the-art methods with respect to shape registration accuracy as well as fidelity by a great margin. We also report shape registration results on seen shape objects in supplementary material.

## 4.3. Qualitative Results

Fig. 1 demonstrates how NDF deforms shape instances to the learned templates in a coarse to fine manner while preserving the topology. Fig. 4 shows an example of pancreas shape representation and registration that can suggest why NDF stands out in Tab. 2.. From Fig. 4a, we can see the pancreas template learned by DIF-Net is problematic that the structure marked by a red circle has no anatomical meaning. Different from DIF-Net and DIT which use nearest neighbours searching to match points, NDF is invertible and topology-preserved. Therefore, our shape registration results will have the comparable quality of shape reconstruction results. In Fig. 4a and Fig. 4b, there are many

# 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	8.08	3.46	5.01	7.55	0.703	0.823	0.824	0.808	31	<b>0.391</b>	<b>1.65</b>	<b>0</b>	5860	29.5	13.8	<b>0</b>
	AtlasNet_25	6.05	2.48	207	4.86	0.65	0.818	0.772	0.824	48.4	42.4	43.3	73.4	24500	25100	18700	24800
	DIF-Net	7.44	1.74	1.98	2.44	0.736	0.834	0.879	0.842	540	16.8	63.9	76.1	4990	595	553	642
	DIT	0.682	0.543	0.758	1.09	0.893	0.867	0.928	0.917	24.4	2	13.6	27.6	149	6.23	1.76	<b>0</b>
	Ours	<b>0.53</b>	<b>0.507</b>	<b>0.704</b>	<b>1.06</b>	<b>0.915</b>	<b>0.872</b>	<b>0.935</b>	<b>0.922</b>	<b>0.191</b>	1.02	8.58	25.4	<b>0</b>	<b>0.89</b>	<b>0</b>	<b>0</b>
5000	DIF-Net	10.5	2.06	1.94	2.42	0.694	0.832	0.881	0.838	276	8.27	45.2	98.1	2560	4.61	786	1090
	DIT	0.677	0.528	0.736	1.07	0.893	0.868	0.931	0.918	66.6	2.18	7.06	14.8	346	11.8	2.06	<b>0</b>
	Ours	<b>0.518</b>	<b>0.49</b>	<b>0.67</b>	<b>1.02</b>	<b>0.916</b>	<b>0.873</b>	<b>0.936</b>	<b>0.923</b>	<b>0.191</b>	<b>0.378</b>	<b>2.68</b>	<b>14.1</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>

Table 2. **Shape Registration on Unseen Shape Instances** – We align all unseen shape instance to the source mesh  $\mathcal{M}_s$  for four organ categories. For AtlasNet,  $\mathcal{M}_s$  is defined as their explicit template mesh. For DIF-Net, DIT and our NDF,  $\mathcal{M}_s$  is defined as the re-meshed learned template mesh. We compare AtlasNet\_Sph and AtlasNet\_25 to the other methods having 2500 vertices in  $\mathcal{M}_s$  after re-mesh. We also compare DIF-Net, DIT and NDF provided  $\mathcal{M}_s$  is of 5000 vertices after remesh. Even though the number of vertices in  $\mathcal{M}_s$  of AtlasNet\_Sph is not 2500 (very close to), we still put them in comparison with methods having 2500-vertex  $\mathcal{M}_s$ . The E-NMF ratio results shown above are multiplied by  $10^5$

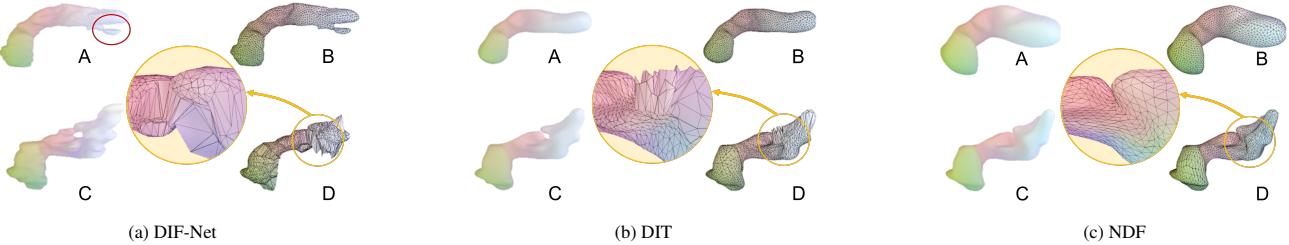


Figure 4. **Qualitative Results Comparison** - For each method, they have four type of mesh shown above. Mesh A is the learned template mesh, Mesh B is the re-meshed template mesh with 2500 vertices, Mesh C is a reconstructed shape and Mesh D is the shape registration result. We also zoom in Mesh D for a better qualitative comparison.

Exp.	CD Mean	NC Mean
1 SV	1.42	0.907
2 QTV <sub>4</sub>	<b>0.516</b>	0.914
3 QTV <sub>4</sub> + CL	0.51	0.916
4 QTV <sub>8</sub> + CL	0.51	0.916
5 QTV <sub>4</sub> + CL + w/o pp	0.512	0.917
6 TV	1.65	0.902
7 TV + CL	125	0.885

Table 3. **Ablation Study in Shape Reconstruction** – The notations of experiments are described in Sec. 4.4. The subscript of QTV is the number of progressive reconstruction steps. These experiments are conducted on the unseen pancreas shapes.

unpleasant triangles Fig. (3) in the local regions where the shape distortions between the shape instance and template are large. On the contrary, the registration result generated by NDF is clean, smooth and accurate. More qualitative results can be found in supplementary material.

#### 4.4. Ablation Study

Our ablation study is developed on pancreas shapes to investigate the effects of three designs in our approach. In this section, "SV", "QTV", "TV" sequentially stands for stationary, our quasi time-varying and time-varying velocity field, "CL" is the short term for curriculum learning and "pp" denotes the point pair loss, which acts as a smooth

Exp.	CD Mean	NC Mean	E-NMF Mean	SI Mean
1 SV	1.46	0.906	1.43	0
2 QTV <sub>4</sub>	0.521	0.914	0	0
3 QTV <sub>4</sub> + CL	0.52	0.915	0	0
4 QTV <sub>8</sub> + CL	0.521	0.915	2.38	0
5 QTV <sub>4</sub> + CL + w/o pp	<b>0.518</b>	0.916	1.91	0
6 TV	1.65	0.902	0.953	0
7 TV + CL	162	0.884	12.9	0.103

Table 4. **Ablation Study in Shape Registration** – All these experiments are conducted on unseen pancreas shapes given the  $\mathcal{M}_s$  with 5000 vertices.

regularization. Our final approach is labeled as Exp.5 in Tab. 3 and Tab. 4.

**Quasi Time-varying Velocity Field.** From the comparisons among Exp.1, 2 and 6 in Tab. 3, our quasi time-varying velocity field wins in all aspects. Time-varying velocity field should be a natural choice but without enough temporal information, the generability will be questionable. Stationary velocity field is widely applied in the medical image/surface registration problem, but it turns out to be hard to get the optimal if assume the continuous velocity field is independent of time. We also explore the effect of progressive representation steps by comparing Exp.4 and 8, we have not observed some extra improvements resulting

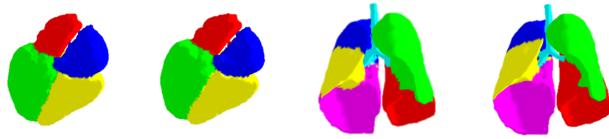


Figure 5. **Label Transfer** - *LeftMost*: Sub-heart ground truth; *MiddleLeft*: Label Transfer; *MiddleRight*: Lobe ground truth; *RightMost*: Label Transfer

from more representation steps. Fig. 1 also indicates most deformations have been done in the starting phrases.

**Curriculum learning.** The improvements earned from curriculum learning are significant in other work [18, 59]. While in our work, it is beneficial but not essential, as can be seen in Tab. 3 and Tab. 4. Furthermore, it is even harmful when used together with a time-varying velocity field.

**Smooth Regularization.** Our opinion that the smooth regularization term is not needed when training deep diffeomorphic flow is supported. In Tab. 3 and 4, we can see Exp.5 and Exp.6 get the very close reconstruction accuracy and both of them generate very few unpleasant faces in shape registration. In summary, even with the most basic training loss and training strategy, our design of model can get a very competitive performance in shape reconstruction and registration.

## 5. Applications and Limitations

### 5.1. Applications

NDF keeps most of the benefits of DeepSDF such as shape completion and shape interpolation. As for medical meaning, NDF can help post-process the segmentation results and do plausible data augmentation.

Our model can also help transfer labels from seen shape to unseen shape. We choose 5 samples from the training set and transfer their labels to the target shape separately. The final label is the majority voting results. Fig. 5 shows two examples of labels transferred by NDF.

The ambition of our work is to boost the shape analysis in medical imaging by helping establish organ shape dataset having dense topology-preserving point correspondences. Specifically, as long as our model is trained on one class of organ shapes and implicit template mesh  $\mathcal{M}_T = (\mathcal{V}_T, \mathcal{E}_T)$  is labelled, the organ mesh of the same class could be aligned as Sec. 3.4 explains. Given such dataset, we can learn a model to parameterize shapes as SMPL [33].

### 5.2. Limitations

Our model concentrates on reconstructing and matching a group of shapes sharing common structures, so we haven't applied it to the popular 3D shape datasets like

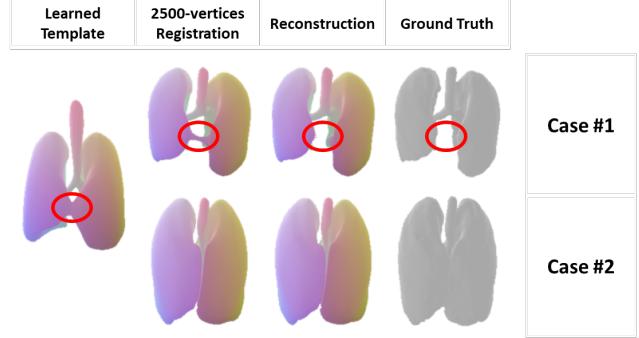


Figure 6. **Limitations** - Case #1 has different local structure with learned template. Case #2 is a successful example.

ShapeNet [7]. As can be seen in Fig. 6, our learned template has structures (marked by red circle) that don't exist in case #1, then our shape reconstruction and registration results of case #1 shape is negatively affected by the extra structures. This issue can be partially addressed by introducing the correction module [15] or considering shapes as groups of sub-structures [32] that are individually topology-preserving or non-existent.

To our best knowledge, there is no medical dataset having structures as well as dense point correspondences annotated. Thus, we cannot evaluate shape registration results in terms of point-to-point error. In the future, we will explore the potentials of our model on some synthetic data like D-FAUST [5], with which we can do point-to-point analysis. We will present The inference runtime of our method is indeed longer than competing methods such as DiT. The main bottleneck of our method is the neural ODE (NODE) module, which requires repeated functional evaluation to solve ODEs within a given error tolerance.

## 6. Conclusions

In this paper, we propose a novel deep implicit function based on neural diffeomorphic flow (NDF) for topology-preserving shape representation. Our experimental results demonstrate that explicitly considering topology preservation leads to significant improvements on shape representation and registration, as illustrated on medical images, where topology preservation is often a necessary requirement. We also propose a conditional quasi time-varying approach to model NDF through an auto-decoder model consisting of multiple neural ODE blocks, allowing us to model the shape deformation in a progressive manner.

## References

- [1] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer, 2006. 3
- [2] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. 3
- [3] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008. 3
- [4] Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800, 2019. 3
- [5] 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. 8
- [6] Rohan Chabra, Jan E Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *European Conference on Computer Vision*, pages 608–625. Springer, 2020. 2
- [7] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenett: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 8
- [8] Ricky T. Q. Chen, Brandon Amos, and Maximilian Nickel. Learning neural event functions for ordinary differential equations. *International Conference on Learning Representations*, 2021. 3
- [9] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 2018. 3, 4
- [10] 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
- [11] Zhiqin Chen, Kangxue Yin, Matthew Fisher, Siddhartha Chaudhuri, and Hao Zhang. Bae-net: branched autoencoder for shape co-segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8490–8499, 2019. 2
- [12] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. 1
- [13] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer, 2018. 3
- [14] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis*, 57:226–236, 2019. 3
- [15] 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. 3, 5, 6, 8
- [16] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. *arXiv preprint arXiv:1908.04725*, 2019. 3
- [17] Nicki Skafte Detlefsen, Oren Freifeld, and Søren Hauberg. Deep diffeomorphic transformer networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4403–4412, 2018. 3
- [18] Yueqi Duan, Haidong Zhu, He Wang, Li Yi, Ram Nevatia, and Leonidas J Guibas. Curriculum deepsdf. In *European Conference on Computer Vision*, pages 51–67. Springer, 2020. 2, 4, 5, 8
- [19] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. *arXiv preprint arXiv:1904.01681*, 2019. 4, 5
- [20] Oren Freifeld, Søren Hauberg, Kayhan Batmanghelich, and Jonn W Fisher. Transformations based on continuous piecewise-affine velocity fields. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2496–2509, 2017. 3
- [21] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas A Funkhouser. Deep structured implicit functions. 2019. 3
- [22] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7154–7164, 2019. 3
- [23] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 216–224, 2018. 2, 3, 5, 6
- [24] Kunal Gupta. *Neural Mesh Flow: 3D Manifold Mesh Generation via Diffeomorphic Flows*. University of California, San Diego, 2020. 3, 4, 5
- [25] Zekun Hao, Hadar Averbuch-Elor, Noah Snavely, and Serge Belongie. Dualsdf: Semantic shape manipulation using a two-level representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7631–7641, 2020. 2
- [26] Nicholas J Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications*, 26(4):1179–1193, 2005. 4

- [27] Haibin Huang, Evangelos Kalogerakis, and Benjamin Martin. Analysis and synthesis of 3d shape families via deep-learned generative models of surfaces. In *Computer Graphics Forum*, volume 34, pages 25–38. Wiley Online Library, 2015. 3
- [28] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6001–6010, 2020. 2
- [29] Vladimir G Kim, Wilmot Li, Niloy J Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. Learning part-based templates from large collections of 3d shapes. *ACM Transactions on Graphics (TOG)*, 32(4):1–12, 2013. 3
- [30] Julian Krebs, Hervé Delingette, Boris Mailhé, Nicholas Ayache, and Tommaso Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 38(9):2165–2176, 2019. 3
- [31] Helko Lehmann, Reinhard Kneser, Mirja Neizel, Jochen Peters, Olivier Ecabert, Harald Kühl, Malte Kelm, and Jürgen Weese. Integrating viability information into a cardiac model for interventional guidance. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 312–320. Springer, 2009. 1
- [32] Feng Liu and Xiaoming Liu. Learning implicit functions for topology-varying dense 3d shape correspondence. *arXiv preprint arXiv:2010.12320*, 2020. 2, 8
- [33] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015. 3, 8
- [34] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987. 3
- [35] Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter. Gaussian process morphable models. *IEEE transactions on pattern analysis and machine intelligence*, 40(8):1860–1873, 2017. 3
- [36] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4460–4470, 2019. 1, 2
- [37] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003. 4
- [38] Andriy Myronenko and Xubo Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010. 3
- [39] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5379–5389, 2019. 3, 4
- [40] 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, 3, 4, 5, 6
- [41] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrin B Turkbey, and Ronald M Summers. Data from pancresa-ct. <https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU>, 2016. 5
- [42] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33, 2020. 1
- [43] Avan Suinesiaputra, Pierre Ablin, Xenia Alba, Martino Alessandrini, Jack Allen, Wenjia Bai, Serkan Cimen, Peter Claes, Brett R Cowan, Jan D’hooge, et al. Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. *IEEE journal of biomedical and health informatics*, 22(2):503–515, 2017. 1
- [44] Hao Tang, Xingwei Liu, Kun Han, Xiaohui Xie, Xuming Chen, Huang Qian, Yong Liu, Shanlin Sun, and Narisu Bai. Spatial context-aware self-attention model for multi-organ segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 939–949, 2021. 1
- [45] Hao Tang, Xingwei Liu, Shanlin Sun, Xiangyi Yan, and Xiaohui Xie. Recurrent mask refinement for few-shot medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3918–3928, 2021. 1
- [46] Edgar Treitschke, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Carsten Stoll, and Christian Theobalt. Patchnets: Patch-based generalizable deep implicit 3d shape representations. In *European Conference on Computer Vision*, pages 293–309. Springer, 2020. 2
- [47] Sébastien Valette and Jean-Marc Chassery. Approximated centroidal voronoi diagrams for uniform polygonal mesh coarsening. In *Computer Graphics Forum*, volume 23, pages 381–389. Wiley Online Library, 2004. 6
- [48] Sébastien Valette, Jean Marc Chassery, and Rémy Prost. Generic remeshing of 3d triangular meshes with metric-dependent discrete voronoi diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 14(2):369–381, 2008. 6
- [49] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 52–67, 2018. 3
- [50] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. *arXi @inproceedingschen2019bae, title=BAE-NET: branched autoencoder for shape co-segmentation, author=Chen, Zhiqin and Yin, Kangxue and Fisher, Matthew and Chaudhuri, Siddhartha and Zhang, Hao, booktitle=Proceedings of the IEEE/CVF International Conference on Computer Vision, pages=8490–8499, year=2019 v preprint arXiv:1905.10711*, 2019. 1

- [51] Xiangyi Yan, Hao Tang, Shanlin Sun, Haoyu Ma, Deying Kong, and Xiaohui Xie. After-unet: Axial fusion transformer unet for medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3971–3981, 2022. 1
- [52] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Transactions on Medical Imaging*, 39(1):188–203, 2019. 1
- [53] Chenyu You, Junlin Yang, Julius Chapiro, and James S. Duncan. Unsupervised wasserstein distance guided domain adaptation for 3d multi-domain liver segmentation. In *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, pages 155–163. Springer International Publishing, 2020. 1
- [54] Chenyu You, Linfeng Yang, Yi Zhang, and Ge Wang. Low-Dose CT via Deep CNN with Skip Connection and Network in Network. In *Developments in X-Ray Tomography XII*, volume 11113, page 111131W. International Society for Optics and Photonics, 2019. 1
- [55] Chenyu You, Qingsong Yang, Lars Gjesteby, Guang Li, Shenghong Ju, Zhuiyang Zhang, Zhen Zhao, Yi Zhang, Wenxiang Cong, Ge Wang, et al. Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. *IEEE Access*, 6:41839–41855, 2018. 1
- [56] Chenyu You, Ruihan Zhao, Lawrence Staib, and James S Duncan. Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation. *arXiv preprint arXiv:2105.07059*, 2021. 1
- [57] Chenyu You, Yuan Zhou, Ruihan Zhao, Lawrence Staib, and James S. Duncan. Simcvd: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. *arXiv preprint arXiv:2108.06227*, 2021. 1
- [58] Han Zhang, Xi Gao, Jacob Unterman, and Tom Arodz. Approximation capabilities of neural odes and invertible residual networks. In *International Conference on Machine Learning*, pages 11086–11095. PMLR, 2020. 5
- [59] 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. 2, 3, 4, 5, 6, 8
- [60] Xiahai Zhuang and Juan Shen. Multi-scale patch and multimodality atlases for whole heart segmentation of mri. *Medical image analysis*, 31:77–87, 2016. 5