Hybrid Neural Diffeomorphic Flow for Shape Representation and Generation via Triplane

Kun Han[1] Shanlin Sun[1] Thanh-Tung Le[1] Xiangyi Yan[1] Haoyu Ma[1] Chenyu You[2] Xiaohui Xie[1]
[1]University of California, Irvine, USA [2] Yale University, USA
{khan7,shanlins,thanhtul,xiangyy4,haoyum3,xhx}@uci.edu
{chenyu.you}@yale.edu

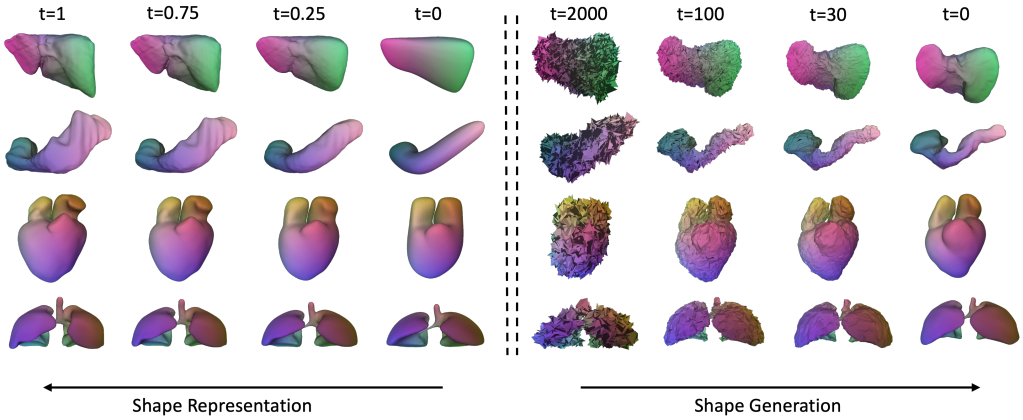

Figure 1. **Shape Representation** and **Shape Generation**. **Left** half presents the diffeomorphic deformation from the learned **template** (t=0) to instance shapes (t=1), with color highlighting the dense correspondence captured by triplane features. **Right** half presents the denoising process for shape generation. The shapes are generated as deformed templates and the 3D deformation is controlled by the generated triplane features from diffusion.

Abstract

Deep Implicit Functions (DIFs) have gained popularity in 3D computer vision due to their compactness and continuous representation capabilities. However, addressing dense correspondences and semantic relationships across DIF-encoded shapes remains a critical challenge, limiting their applications in texture transfer and shape analysis. Moreover, recent endeavors in 3D shape generation using DIFs often neglect correspondence and topology preservation. This paper presents HNDF (Hybrid Neural Diffeomorphic Flow), a method that implicitly learns the underlying representation and decomposes intricate dense correspondences into explicitly axis-aligned triplane features. To avoid suboptimal representations trapped in local minima, we propose hybrid supervision that captures both local and global correspondences. Unlike conventional approaches that directly generate new 3D shapes, we further explore the idea of shape generation with deformed template shape via diffeomorphic flows, where the deformation is encoded by the generated triplane features. Leveraging a pre-existing 2D diffusion model, we produce high-quality and diverse 3D diffeomorphic flows through generated triplanes features, ensuring topological consistency with the template shape. Extensive experiments on medical image organ segmentation datasets evaluate the effectiveness of HNDF in 3D shape representation and generation.

1. Introduction

3D geometry representation is critical for numerous computer vision tasks, including 3D model reconstruction, matching and manipulation. Deep implicit functions (DIFs) have emerged as promising alternatives to traditional representation methods such as voxel grids, point clouds and polygon meshes. DIFs offer several advantages such as compactness, continuity, and the ability to capture fine geometric details. They enable efficient computation while leveraging deep neural networks for end-to-end training, enhancing shape representation and understanding.

However, despite the promising results in direct object modeling using DIFs, it is important to consider the common shape features and semantic correspondences shared

among objects. Conventional DIFs face challenges in establishing correspondences between different shapes, limiting their applicability in domains like medical image segmentation [12, 18, 25, 59] and texture transfer [7, 30]. Previous methods [5, 44, 61] have proposed shape modeling as conditional deformations of a template DIF to address this limitation. However, these methods still have limitations, such as being topology-agnostic or lacking the capability to capture correspondences for local details.

Recent researches have also explored the integration of DIFs for the 3D shape generation [33, 36, 42, 60]. Compared to point clouds and polygon meshes, DIF-based generation offers continuous representations with high quality and resolution. However, existing approaches primarily focus on direct shape generation without considering underlying point correspondence and topology preservation.

To overcome these challenges, we introduce Hybrid Neural Diffeomorphic Flow (HNDF) for shape representation and generation. HNDF models shapes as conditional deformations of a template DIF, similar to previous work [5, 44, 48, 61]. However, HNDF encodes diffeomorphic deformations into axis-aligned triplane features to enhance representation capability. Local deformations are controlled through interpolation of triplane features with a shared feature decoder. Nevertheless, the direct application of triplanes may lead to local optimization issues and defective deformations, resulting in inaccurate representations. To address this, we propose a hybrid supervision approach that considers both local and global correspondences, along with additional modifications and regularization to preserve the diffeomorphism property of the represented deformations. This combination of triplane feature exploration and supervision enables high representation capabilities and accurate dense correspondences.

Unlike conventional 3D shape generation works which primarily focus on direct shape generation, we explore the idea of deformation-based shape generation, where the template shape is deformed based on newly generated diffeomorphic deformations. This approach ensures that the newly generated shapes maintain the same topology as the template shape, preserving topological consistency while offering a wide range of diverse shapes. To achieve this, we represent deformations using optimized per-object triplane features, which encode diffeomorphic deformations as three axis-aligned 2D feature planes. We concatenate the triplane features as multi-channel images and leverage the existing 2D diffusion models to generate new triplane features. By applying the new diffeomorphic deformations encoded in the triplane features, we deform the template shape to generate novel 3D shapes while preserving their topological characteristics.

The contributions of this paper are as follows:

1. We propose HNDF, which leverages axis-aligned triplane features to provide high representation capability and capture dense correspondences accurately.

2. We demonstrate that hybrid supervision and regularization are essential for ensuring correct deformation representation and preventing the representation from local optima.

3. Rather than directly generating 3D shapes, we explore the concept of shape generation through diffeomorphic deformations and provide a baseline method utilizing 2D diffusion model. The topology and correspondences are preserved in newly generated 3D shapes.

2. Related Works

Deep Implicit Function Deep implicit functions, or neural fields, have enabled the parameterization of physical properties and dynamics through simple neural networks [4, 31, 32, 37, 43, 47, 52]. DeepSDF [37] serves as an auto-decoder model, commonly used as a baseline for shape representation [1, 16, 45]. NeRF [37] presents a novel approach for synthesizing photorealistic 3D scenes from 2D images. Occupancy Network [31] constructs solid meshes through the classification of 3D points, while Occupancy Flow [35] extends this idea to 4D with a continuous vector field in time and space. Recent trends incorporate locally conditioned representations [1, 4, 16, 39, 45], utilizing small MLPs that are computationally and memory-efficient while capturing local details effectively. One such representation is the hybrid triplane [2, 6, 21, 28, 38], which represents features on axis-aligned planes and aggregates them using a lightweight implicit feature decoder. In our work, we adopt the expressive triplane representation. However, instead of decoding the 3D object itself, we utilize triplane features to decode complex diffeomorphic deformations, allowing us to represent new 3D objects by deforming the template shape using the encoded deformation.

Point Correspondence and Topology preservation Capturing dense correspondences between shapes remains a significant challenge and a critical area of interest in the 3D vision community [10, 16, 19, 20, 44, 49–53, 55, 56, 58]. Various approaches have been proposed to address point correspondence, including template learning, elementary representation, and deformation field-based methods. Among them, mesh-based methods [19, 20] face difficulties in handling topological changes, sensitivity to mesh connectivity, and challenges in capturing fine-grained details. Elementary-based methods [10, 16], on the other hand, may struggle with capturing high-level structural features due to the simplicity of the elements used. DIT [61] and NDF [44] exemplify deformation field-based methods, with DIT exhibiting smoother deformations using LSTM [15] and NDF employing NODE [3] for achieving diffeomorphic deformation. ImplicitAtlas [48] integrates multiple templates

to improve the shape representation capacity at a negligible computational cost. In our work, we follow the NDF framework but enhance the representation's capacity to capture accurate correspondences by leveraging more powerful triplane representation. Experimental results highlight the importance of incorporating triplane features with hybrid supervision, which prevents local optimization issues, provides significantly more accurate correspondences, and ensures the preservation of topology.

3D Shape Generation Generative models, such as GANs, autoregressive models, score matching models, and denoising diffusion probabilistic models, have been extensively studied for 3D shape generation. However, GAN-based methods [2,8,9,11,29,34,36,54,57] still outperform alternative approaches. Voxel-based GANs [8,13,46], for example, directly extend the use of CNN generators from 2D to 3D settings with high memory requirement and computational burden. In recent years, there has been a shift towards leveraging expressive 2D generator backbones, such as StyleGAN2 [17]. EG3D [2] combines a hybrid explicit-implicit triplane representation to improve computational efficiency while maintaining expressiveness. Get3D [9] incorporates the deformable tetrahedral grid for explicit surface extraction and triplane representation for differentiable rendering to generate textured 3D shapes.

Compared to the existing GAN-based approaches for 3D generation, the development of 3D diffusion models is still in its early stages. Several notable works have explored the application of diffusion models in generating 3D shapes. PVD [62] proposed the use of a point-voxel representation combined with PVConv [23] to generate 3D shapes through diffusion. DPM [26] introduced a shape latent code to guide the Markov chain in the reverse diffusion process. MeshDiffusion [22] utilized the deformable tetrahedral grid parametrization for unconditionally generating 3D meshes. 3D-LDM [33] integrated DeepSDF [37] into diffusion-based shape generation, leveraging diffusion to generate a global latent code and improve the conditioning of the neural field. NFD [42] extended the use of 2D diffusion into 3D shape generation, exploring the potential of diffusion models in capturing and generating complex 3D shapes with Occupancy Network [31].

While existing approaches in shape generation focus on directly generating 3D shapes, they often neglect the preservation of underlying topology. This oversight can lead to artifacts in the generated shapes and limit their applicability in scenarios where topology is important. In our work, we introduce a baseline diffusion-based method that deforms a template to generate new shape. The diffeomorphic deformation is encoded by the generated triplane features. Our approach focuses on producing visually coherent and realistic shapes while preserving point correspondence and underlying topology.

3. Preliminaries

Diffeomorphic Flow is a continuous and smooth mapping that transforms a given manifold or space while preserving its differentiable structure. In the context of 3D geometry, diffeomorphic flow plays a crucial role in establishing dense point correspondences between 3D shapes and ensuring the preservation of their underlying topology during deformation. Mathematically, the forward diffeomorphic flow $\Phi(p,t) : \mathbb{R}^3 \times [0,1] \rightarrow \mathbb{R}^3$ describes the trajectory of a 3D point p over the interval $[0,1]$, where the starting point p is located in the space of instance shape S and the destination point corresponds to the target shape T. The velocity field $\mathbf{v}(p,t) : \mathbb{R}^3 \times [0,1] \rightarrow \mathbb{R}^3$ represents the derivative of deformation of 3D points. The diffeomorphic flow Φ is obtained by solving the initial value problem (IVP) of an ordinary differential equation (ODE),

$$\frac{\partial \Phi}{\partial t}(p,t) = \mathbf{v}\left(\Phi(p,t),t\right) \quad \text{s.t.} \quad \Phi(p,0) = p \tag{1}$$

Similarly, the inverse flow Ψ can be calculated by solving a corresponding ODE with negative velocity field $-\mathbf{v}$, allowing for the transformation from the template space to the instance space

$$\frac{\partial \Psi}{\partial t}(p,t) = -\mathbf{v}\left(\Psi(p,t),t\right) \quad \text{s.t.} \quad \Psi(p,0) = p \tag{2}$$

where p is the starting point on the target shape. The property of topology preservation is achieved through the Lipschitz continuity of the velocity field. The forward and backward diffeomorphic deformation can be calculated by the integration of the velocity field by solving the equation 1 2, respectively.

Diffusion Probabilistic Model (DPM) [14] is a parameterized Markov chain designed to learn the underlying data distribution $p(X)$.

During the Forward Diffusion Process (FDP), the diffused data point X_t is obtained at each time step t by sampling from the conditional distribution:

$$q\left(X_t \mid X_{t-1}\right) = \mathcal{N}\left(X_t; \sqrt{1-\beta_t} X_{t-1}, \beta_t I\right) \tag{3}$$

where X_0 is sampled from the initial distribution $q(X_0)$, and X_T follows a Gaussian distribution $N(X_T; 0, I)$. The parameter $\beta_t \in (0,1)$ represents a variance schedule that gradually introduces Gaussian noise to the data. By defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_t)$, X_t can be sampled conditionally on X_0 as $q\left(X_t \mid X_0\right) = \mathcal{N}\left(X_t; \sqrt{\bar{\alpha}_t} X_0, \left(1-\bar{\alpha}_t\right) I\right)$, providing a distribution for sampling X_t from the initial data X_0.

In contrast, the Reverse Diffusion Process aims to approximate the posterior distribution $p(X_{t-1}|X_t)$ to recreate a realistic X_0 starting from random noise X_T. The Reverse Diffusion Process is formulated as a trajectory of posterior distributions starting from X_T:

$$p\left(X_{0:T}\right) = p\left(X_T\right) \prod_{t=1}^{T} p_\theta\left(X_{t-1} \mid X_t\right) \tag{4}$$

The conditional distribution $p_\theta(X_{t-1}|X_t)$ is approximated by a neural network with parameters θ:

$$p_\theta\left(X_{t-1} \mid X_t\right)=\mathcal{N}\left(X_t ; \mu_\theta\left(X_t, t\right), \Sigma_\theta\left(X_t, t\right)\right) \quad (5)$$

4. Method

In this section, we present our Hybrid Neural Diffeomorphic Flow (HNDF) for shape representation and generation. Section 4.1 reviews our baseline method [44]. In Section 4.2, we introduce the utilization of triplane features, and the hybrid supervision for capturing local and global correspondences. Finally, in Section 4.3, we describe our proposed method for generating topology-preserving shapes.

4.1. Review of NDF

NDF [44], similar to DeepSDF [37], represents a 3D shape S_i using a continuous signed distance field (SDF) $\mathcal{F}$. Given a random 3D point p and a latent code c_i of length k, $\mathcal{F}$ outputs the distance from the point p to the closest surface of shape S_i. However, unlike DeepSDF, which directly represents 3D shapes, NDF uses a deform code c_i to control the deformation of each instance shape from the template shape. As a result, the conditional continuous SDF $\mathcal{F}$ can be decomposed into $\mathcal{T} \circ \mathcal{D}$, where $\mathcal{D} : \mathbb{R}^3 \times \mathbb{R}^k \mapsto \mathbb{R}^3$ provides the deformation mapping from the coordinates of p in the instance space of S_i to a canonical position p' in the template space. The function $\mathcal{T}$ represents a single shape DeepSDF that models the implicit template shape.

4.2. Hybrid Shape Representation via Triplane

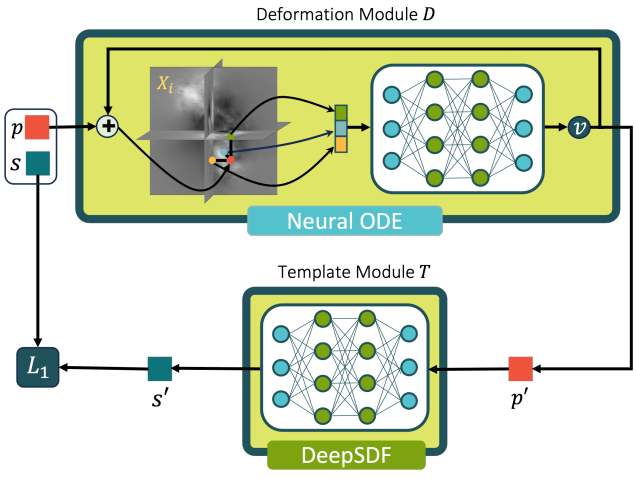

Figure 2. **Shape Representation** framework consists of a deformation module $\mathcal{D}$, a template module $\mathcal{T}$, and per-object triplane features X_i. Given a point p in the instance space, we compute its corresponding destination point p' in the template space using Eq. 1. The template module then provides the sign distance value s' for this point. During training, we optimize the framework by minimizing the L_1 loss between the represented s' and the ground truth s, while incorporating regularization terms.

As shown in [1, 4, 16, 38, 39, 45], previous methods [37, 44, 48, 61] utilizing a single latent vector to control the entire shape or deformation space could not capture the details of the complex 3D shape or the deformation. Motivated by recent advancements in hybrid representation [2], we propose to encode complex diffeomorphic deformations as a set of three axis-aligned 2D feature planes, as shown in Fig. 2. This enables us to capture fine-grained details and variations in the shape space more effectively.

The triplane representation is a hybrid architecture for neural fields that combines explicit and implicit components [2]. For each instance shape S_i, it employs three axis-aligned orthogonal feature planes ($X_i = [F^i_{xy}, F^i_{xz}, F^i_{yz}]$), each with a resolution of $L \times L \times C$. These planes serve as the encoded representations of the deformation. To query a deformation, the position of given point p_i is projected onto each of the feature planes, and the corresponding feature vectors are retrieved using bilinear interpolation. Subsequently, a lightweight multilayer perceptron (MLP) decoder is employed to interpret the aggregated features as corresponding velocity vector v_i. The diffeomorphic deformation d_i for point p_i can be calculated by integrating the velocity vector using an explicit Runge-Kutta solver [3], as defined in Eq. 1. In contrast to the approach in [2], where feature aggregation is performed through summation, we have found that concatenating the interpolated features from the triplane yields better results.

4.2.1 Training

In our method, we represent the instance shape S_i as a deformed template shape ($\mathcal{T} \circ \mathcal{D}_i$). To capture the continuous shape of S_i, we employ two modules: a continuous diffeomorphic deformation module $\mathcal{D}$ and a template shape representation $\mathcal{T}$. As discussed in Sec. 4.2, the diffeomorphic deformation d_i of a point p_i is obtained by integrating the velocity field. The signed distance field (SDF) value of p_i is determined by evaluating the implicit template shape module $\mathcal{T}$ at the transformed point p'_i, where $p'_i = p_i + d_i$.

During training, our method jointly optimizes the deformation module $\mathcal{D}$, template DeepSDF shape $\mathcal{T}$, and per-object triplane features X_i to represent a training set of S objects. The triplane representation provides an expressive representation power, allowing us to achieve accurate deformation and correspondence. Unlike NDF [44], which requires multiple deformation modules, our method only requires one deformation module. This not only enables more accurate deformation representation but also reduces the memory and computation requirements.

The training objective function includes a reconstruction loss and a regularization loss:

$$\mathcal{L}_{train} = \mathcal{L}_{\text{rec}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} \quad (6)$$

where $\mathcal{L}_{\text{rec}}$ shows the reconstruction loss between the ground truth SDF value s_i and the represented SDF value s'_i,

and $\mathcal{L}_{\text{reg}}$ includes a series of regularization terms. Specifically, reconstruction loss $\mathcal{L}_{\text{rec}}$ can be written as

$$\mathcal{L}_{\text{rec}} = \sum_{i=1}^{S} \sum_{j=1}^{N} L_1(\mathcal{T} \circ \mathcal{D}_i(p_{i,j}), s_{i,j}) \tag{7}$$

where S is the number of instance shapes in the training set, N is the number of sampling points for each shape, $p_{i,j}$ is the j-th point on the i-th shape and $s_{i,j}$ is the corresponding ground truth SDF value.

In addition to the point-wise deformation regularization ($\sum_{i,j} \|\mathcal{T} \circ \mathcal{D}_i(p_{i,j}) - s_{i,j}\|_2$) and the L_2 norm feature regularization ($\left\|F^i_{xy}\right\|_2 + \left\|F^i_{yz}\right\|_2 + \left\|F^i_{xz}\right\|_2$), the inclusion of total variation (TV) regularization [40] is crucial for simplifying the triplane representation and ensuring smooth deformations. The overall regularization term in the training objective is defined as:

$$\mathcal{L}_{\text{reg}} = \lambda_{\text{PW}} \mathcal{L}_{\text{PW}} + \lambda_{\text{L2}} \mathcal{L}_{\text{L2}} + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}} \tag{8}$$

4.2.2 Hybrid Supervision for Inference Time Reconstruction

In contrast to previous methods [37, 44] that utilize a single latent vector for shape reconstruction, the incorporation of triplane representation in our work introduces specific challenges when reconstructing new shapes. Specifically, during the optimization process, the features interpolated from the triplane representation for different positions p_i are optimized locally. Since the final diffeomorphic deformation is the integration of velocity vectors along the trajectory in the entire space, the optimized deformation can become trapped in local optima, leading to incorrect global correspondence, as shown in Fig. 3. As a consequence, the reconstructed shape and deformation may exhibit artifacts, and the overall correspondence may be compromised.

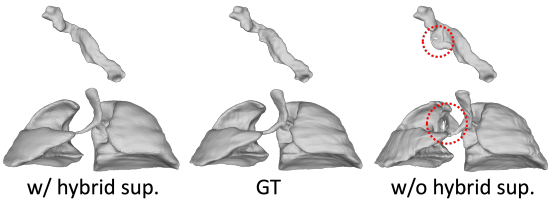

Figure 3. Left is the reconstruction results with proposed hybrid supervision. Middle is the ground truth. Right is the result from purely local supervision, which failed to capture the global correspondence.

Therefore, we introduce a hybrid supervision strategy that incorporates both global and local correspondence. In addition to randomly sampled points that provide local supervision, we downsample the entire $N \times N \times N$ coordinate grid with predefined step size and include these regularly sampled points for global supervision during optimization.

The reconstruction loss during inference is defined as:

$$\mathcal{L}_{\text{rec}} = \mathcal{L}^{\text{grid}}_{\text{rec}} + \lambda_{\text{random}} \mathcal{L}^{\text{random}}_{\text{rec}} \tag{9}$$

where λ_{random} is initialized as 0 and gets increased as the optimization continues.

After we get the grid-structure deformation Φ, we utilize two additional regularization terms to ensure the diffeomorphism of the deformation field and maintain structural integrity. The first term, selective Jacobian determinant regularization ($\mathcal{L}_{\text{Jdet}}$), enforces local orientation consistency.

$$\mathcal{L}_{Jdet} = \frac{1}{N} \sum_{p} relu\left(-|J_\Phi(p)|\right) \tag{10}$$

where the Jacobian matrix J_Φ is defined as:

$$J_\Phi(p) = \begin{bmatrix} \frac{\partial \Phi_x(p)}{\partial x} & \frac{\partial \Phi_x(p)}{\partial y} & \frac{\partial \Phi_x(p)}{\partial z} \\ \frac{\partial \Phi_y(p)}{\partial x} & \frac{\partial \Phi_y(p)}{\partial y} & \frac{\partial \Phi_y(p)}{\partial z} \\ \frac{\partial \Phi_z(p)}{\partial x} & \frac{\partial \Phi_z(p)}{\partial y} & \frac{\partial \Phi_z(p)}{\partial z} \end{bmatrix} \tag{11}$$

The second term, deformation regularization ($\mathcal{L}_{\text{def}}$), discourages excessively skewed deformations that may lead to unnatural shapes.

$$\mathcal{L}_{def} = \sum_{p} \|\nabla \Phi(p)\|^2 \tag{12}$$

The combination of global and local supervision provides comprehensive guidance during optimization, enabling the model to capture both fine-grained details and global structural consistency.

4.2.3 Point Correspondence and Shape Registration

During inference, our method utilizes the learned template shape from training and the diffeomorphic deformation encoded by the triplane feature to establish point correspondence and shape registration between different instance shapes. For each point p_t on the template shape, we apply the inverse diffeomorphic flow Ψ, as defined in Eq. 2, to obtain the corresponding points p_i and p_j on instance shapes S_i and S_j respectively, based on their respective triplane features X_i and X_j. This process allows us to accurately capture point correspondence and establish registration between the instances, facilitating tasks such as shape comparison, shape synthesis, and texture transfer.

4.3. Topology-preserving Shape Generation

In this section, we present our proposed method for topology-preserving shape generation. Rather than directly generating shapes from scratch, our approach focuses on generating new shapes by deforming a template shape using synthesized diffeomorphic deformations.

4.3.1 Training a Diffusion Model

After the training of the diffeomorphic deformation module $\mathcal{D}$ and the template shape representation $\mathcal{T}$, as described in Section 4.2.1, we can leverage the hybrid supervision introduced in Section 4.2.2 to obtain the corresponding per-shape triplane features for the dataset. These optimized sets

of triplane features, denoted as $X \in \mathbb{R}^{N\times(L\times L\times 3C)}$, will be utilized to train our generative model, where N denotes the number of shapes in the dataset, L is the dimension of triplane features and C is the number of channels for each 2D plane $(F^i_{xy}, F^i_{xz}, F^i_{yz})$.

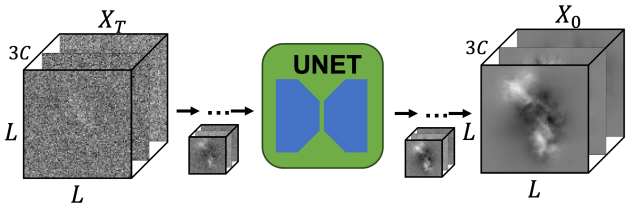

Figure 4. The triplane feature can be represented as multi-channel images. In our work, we adopt the 2D diffusion model as our shape generation model. The generated triplane feature encodes the diffeomorphic deformation that deforms the template to produce the new shapes.

In our framework, the triplane feature is composed of three 2D plane features. We concatenate these feature planes and takes advantage of the strong generative capability of existing 2D diffusion models. Following Sec. 3, we train a diffusion model to learn the reverse diffusion process and predict the added noise from its noisy input by minimizing the following loss function:

$$Loss(\theta) = \mathbb{E}_{X_0\sim q(X),\epsilon\sim\mathcal{N}(0,I),t} \left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}X_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right] \quad (13)$$

where ϵ_θ is predicted noise and θ represents the model parameters.

4.3.2 New Shape Generation

During the inference phase, the generation of a new shape involves deforming the template shape based on the diffeomorphic deformation encoded by the sampled triplane features. Following [14], we initiate the process by sampling a random Gaussian noise $X_T \sim \mathcal{N}(0, I) \in \mathbb{R}^{L\times L\times 3C}$. Subsequently, we perform iterative denoising for a total of T steps as:

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(X_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta\left(X_t, t\right)\right) + \sigma_t\boldsymbol{\epsilon} \quad (14)$$

where $\epsilon \sim \mathcal{N}(0, I)$ if $t > 1$, else, $\epsilon = 0$.

After sampling, the concatenated triplane feature is split into three axis-aligned 2D planes $(F^i_{xy}, F^i_{xz}, F^i_{yz})$. This generated triplane feature can be interpreted as the diffeomorphic deformation. By following the trajectory defined by the ODE function in Eq. 2, each point on the template shape is displaced towards its corresponding destination point in the instance space. Consequently, the new generated shape, known as the deformed template, retains the same underlying topology as the template shape, ensuring consistent connectivity.

5. Experiments

In this section, we present the experiments conducted to evaluate our proposed Hybrid Neural Diffeomorphic Flow (HNDF) for shape **representation** and **generation** tasks.

Datasets: To assess the effectiveness of our shape representation, we utilize the same medical datasets as [44]: Pancreas CT, Inhouse Liver, Inhouse Lung and MultiModality Whole Heart Segmentation, as these datasets exhibit clear common topology while demonstrating shape variation, making them suitable for our evaluation. For shape generation evaluation, we employ liver and pancreas from the Abdomen1k dataset [27], and heart and lung from [44]. Please refer to the supplementary material for detailed data sources and preprocessing information.

Shape Representation Evaluation: We evaluate HNDF for shape representation through two experiments. First, we demonstrate the expressive power of triplane representation and the importance of our hybrid supervision. Evaluation metrics include Chamfer distance (CD) and normal consistency (NC). Second, we evaluate point correspondence and shape registration accuracy, incorporating self-intersection (SI) as an additional metric for geometrical fidelity.

Shape Generation Evaluation: For shape generation evaluation, following [42], we adopt an adapted version of Frechet inception distance (FID). This metric considers rendered shading images of our generated meshes, taking human perception into account. As discussed in [60], shading-image FID overcomes limitations of other mesh-based evaluation metrics. FID is computed across 20 views and averaged to obtain a final score

$$\text{FID} = \frac{1}{20}\left[\sum_{i=1}^{20}\left\|\mu^i_g - \mu^i_r\right\|^2 + \text{Tr}\left(\Sigma^i_g + \Sigma^i_r - 2\left(\Sigma^i_r\Sigma^i_g\right)^{\frac{1}{2}}\right)\right] \quad (15)$$

Additionally, precision and recall scores are reported using the method proposed by [41]. Precision reflects the quality of the rendered images, while recall measures the diversity of the generative model.

Baseline Methods We compare our proposed Hybrid Neural Diffeomorphic Flow (HNDF) with several baselines for the shape representation task. This includes DIT [61], DIF-Net [5], and NDF [44], which share the same representation formula as ours, where the shape is represented as a deformed template. We also include AtlasNet [10], which uses explicit mesh parameterization for shape reconstruction. Additionally, we compare with DeepSDF [37] and NFD [42], which directly represent 3D shapes from scratch.

For the shape generation task, we explore different sampling strategies and generative models. We compare against DeepSDF [37] and NDF [44], which assume a Gaussian distribution for the global latent vector. We sample new shapes by randomly sampling global vectors from a Gaussian distribution or performing PCA analysis on optimized

	Reconstruction								Registration											
	CD Mean(↓)				NC Mean(↑)				CD Mean(↓)				NC Mean(↑)				SI Mean(↓)			
Model/Data	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung
DeepSDF [37]	0.711	0.539	0.951	0.669	0.898	0.866	0.913	0.928	-	-	-	-	-	-	-	-	-	-	-	-
NFD [42]	**0.080**	0.118	0.287	**0.255**	**0.982**	**0.898**	0.947	**0.947**	-	-	-	-	-	-	-	-	-	-	-	-
AtlasNet [10]	8.08	3.46	7.55	5.01	0.703	0.823	0.808	0.824	8.08	3.46	7.55	5.01	0.703	0.823	0.808	0.824	5860	29.5	**0**	13.8
DIT [61]	0.63	0.509	1.05	0.712	0.903	0.87	0.919	0.934	0.677	0.528	1.07	0.736	0.893	0.868	0.918	0.931	346	11.8	**0**	2.06
DIF-Net [5]	4.18	1.58	2.23	1.86	0.756	0.832	0.838	0.882	10.5	2.06	2.42	1.94	0.694	0.832	0.838	0.881	2560	4.61	1090	786
NDF [44]	0.512	0.476	0.993	0.643	0.917	0.873	0.923	0.937	0.518	0.49	1.02	0.67	0.916	0.873	0.923	0.936	**0**	**2**	**0**	**0**
Ours	0.082	**0.116**	**0.277**	**0.255**	0.961	0.885	**0.948**	0.945	**0.099**	**0.125**	**0.306**	**0.304**	**0.946**	**0.882**	**0.936**	**0.939**	15	8	6	**0**

Table 1. **Shape Reconstruction** and **Shape Registration** results on Unseen Shapes. The chamfer distance results shown above are multiplied by 10^3. **DeepSDF** and **NFD** cannot model the deformation between shapes, therefore, they are unable to conduct the shape registration task.

global latent vectors. We also compare with recent generative models such as point-cloud-based PVD [62], and neural-field-based 3D-LDM [33] and NFD [42]. However, it's important to note that these models do not consider the preservation of underlying topology.

5.1. Shape Representation

We evaluate our shape representation through two evaluations: **representation** on training data and **reconstruction** on unseen data, following the setting of [44]. For each point p in the instance space, according to Eq. 1, we can get the corresponding destination point p' in the template space, and the trained template module will return the sign distance value for this point. After retrieving the sign distance value for all the grid points, we can then utilize the marching cube algorithm [24] to extract the mesh for each instance. In the representation comparison, we utilize the trained per-object latent feature to assess the effectiveness of different representation methods. In the reconstruction comparison, we independently optimize the per-object latent feature while keeping the network parameters fixed to evaluate the generability of the methods in shape reconstruction. Fig. 5 shows the reconstruction results of different methods. Due to the space limitation, we put the shape representation result on the supplementary.

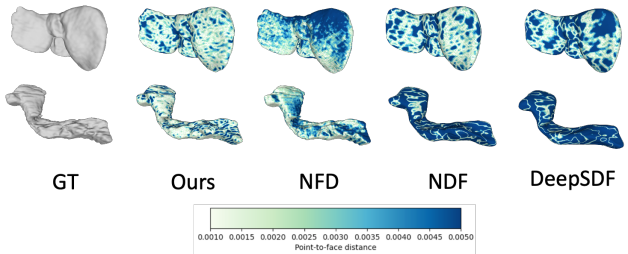

Figure 5. Reconstruction Result on unseen data.

DIF-Net achieves the best results on the training data representation but worse results on the shape reconstruction tasks, indicating the overfitting on the training data. Our method and NFD achieve similar overall performance, benefiting from the enhanced representation power of the triplane feature. Comparing with NDF, our method achieves superior performance even with a single deformation module, outperforming NDF with 4 consecutive deformation modules. The ablation study conducted on regularization, as shown in Tab. 3, demonstrates the significance of our proposed hybrid supervision in achieving accurate reconstruction for new shapes reconstruction.

5.2. Point Correspondence and Shape Registration

As the methods DeepSDF and NFD can only represent the shape without capturing point correspondence, we compare the remaining methods in Table 1 for shape registration evaluation and the instance shape is represented by deforming the template, as described in Sec. 4.2.3. Following the trajectory defined by the ODE function in Eq. 2, each point on the template shape moves towards the corresponding destination point on the instance space. As a result, the instance shape, defined as the deformed template, shares the same underlying topology as the template shape, ensuring consistent connectivity. The diffeomorphic deformation from the template towards instance shapes is shown in the left half of Fig. 1.

To evaluate the point correspondence and shape registration results, we compare the deformed template with the corresponding ground truth instance shape. We also utilize self-intersection as a metric to assess the preservation of topology and geometric fidelity during the deformation. To ensure a fair comparison, we remesh the template meshes to have the same number of vertices (5000), following the approach in [44]. Based on the comparison presented in Table 1, our proposed method achieves better registration accuracy and correct dense correspondence, with only slight self-intersection, which can be considered negligible given the large number of vertices and faces in the template shape.

5.3. Shape Generation

Table 2 presents the evaluation of shape generation across different methods. For DeepSDF and NDF, we sam-

Model/Data	FID Mean(↓)				Prec. Mean(↑)				Recall Mean(↑)			
	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung	Pancreas	Liver	Heart	Lung
DeepSDF △	99.46	93.74	85.76	83.63	0.810	0.858	0.773	0.581	0.078	0.089	0.501	0.630
DeepSDF ★	80.03	85.64	72.22	69.46	0.729	0.810	0.771	0.670	0.430	0.534	0.735	0.664
NDF △	69.66	60.50	65.63	60.28	0.797	0.714	0.762	0.765	0.508	0.593	0.615	0.695
NDF ★	69.66	66.45	64.21	58.27	0.844	0.821	0.804	0.813	0.505	0.571	0.842	0.794
PVD	89.26	86.32	82.44	87.12	0.760	0.821	0.764	0.732	0.420	0.466	0.671	0.712
3D-LDM	78.64	79.58	73.25	76.42	0.782	0.824	0.819	0.803	0.470	0.554	0.813	0.821
NFD	72.83	74.24	61.58	64.34	0.812	0.831	0.832	0.821	0.523	0.560	**0.972**	**0.937**
Ours	**52.01**	**48.54**	**54.5**	**41.3**	**0.992**	**0.994**	**0.984**	**0.951**	**0.661**	**0.613**	0.893	0.914

Table 2. **Shape generation** results. Our method achieve better performance according to the FID, precision and recall. △ denotes sampling from Gaussian distribution while ★ denotes sampling from PCA.

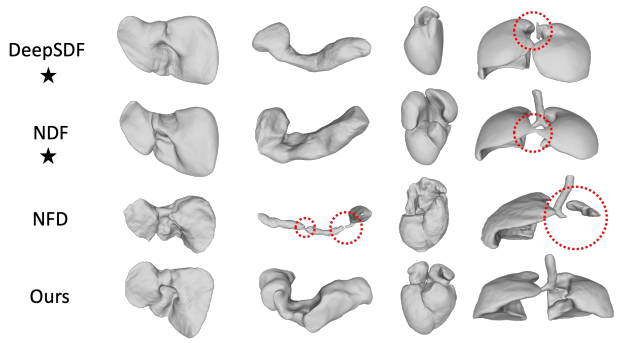

Figure 6. Visualization of generated 3D shape.

ple global latent vectors from a Gaussian distribution and perform PCA analysis, where the parameters are determined by grid search. However, similar to the results in previous experiments, the shapes sampled from DeepSDF and NDF tend to be smoother compared to real instance shapes. PVD is capable of generating variable shapes, but it is limited by its nature to generate only coarse object shapes. 3D-LDM attempts to capture the distribution of the global latent vectors of DeepSDF, but still faces the smoothing issue from the global latent vector. NFD can also generate variable shapes. However, compared to our methods, the shapes generated by NFD may not preserve topology, resulting in potentially separated components in the generated shapes, as shown in Fig. 6. In contrast, our method focuses on generating diffeomorphic deformations encoded by triplane features. The new shapes are generated by deforming the template, allowing us to achieve high fidelity and variability while preserving the underlying topology.

5.4. Ablation Study

Supervision Table 3 highlights the significance of our global supervision in shape reconstruction, mitigating the risk of local minima. While incorporating additional mesh supervision improved the results marginally, it also increased computational and memory demands. Thus, we opted to utilize global supervision in our approach.

Feature Representation We explored the use of 3D voxel-grid features as an alternative to triplane features, and found that they yielded similar results as shown in Table 4. However, voxel-grid features required more computation and memory resources for representation and generation tasks. In contrast, triplane feature representation achieved high reconstruction accuracy with improved memory and computation efficiency.

Model/Data	CD Mean(↓)		NC Mean(↑)	
	Pancreas	Liver	Pancreas	Liver
Ours	**0.082**	0.116	**0.961**	0.885
Ours - Global Sup.	0.264	0.368	0.932	0.877
Ours + Mesh Sup.	**0.082**	**0.112**	0.960	**0.886**

Table 3. Shape Reconstruction with various supervision.

Model/Data	CD Mean(↓)		NC Mean(↑)	
	Pancreas	Liver	Pancreas	Liver
Vector	0.512	0.476	0.917	0.873
Triplane	**0.082**	0.116	**0.961**	**0.885**
Voxel	0.146	**0.112**	0.957	**0.885**

Table 4. Shape Reconstruction with various feature representations.

6. Conclusion

In this paper, we introduce Hybrid Neural Diffeomorphic Flow (HNDF) as a novel approach for topology-preserving shape representation and generation. Our method leverages the expressive power of triplane representation, enabling accurate dense correspondence and high representation accuracy. The proposed hybrid supervision plays a crucial role in capturing both local and global correspondence. Unlike existing methods that primarily focus on directly generating shapes, we explore the concept of generating shapes using deformed templates to preserve the underpying topology. We present a baseline method for topology-preserving shape generation and will continue our exploration for more complex shapes and scenarios. By presenting our research, we aim to contribute to the 3D vision community and provide insights into the potential of topology-preserving shape representation and generation.

References

[1] 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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 608–625. Springer, 2020. 2, 4

[2] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16123–16133, 2022. 2, 3, 4

[3] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018. 2, 4

[4] 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. 2, 4

[5] 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. 2, 6, 7

[6] Terrance DeVries, Miguel Angel Bautista, Nitish Srivastava, Graham W Taylor, and Joshua M Susskind. Unconstrained scene generation with locally conditioned radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14304–14313, 2021. 2

[7] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346, 2001. 2

[8] Matheus Gadelha, Subhransu Maji, and Rui Wang. 3d shape induction from 2d views of multiple objects. In *2017 International Conference on 3D Vision (3DV)*, pages 402–411. IEEE, 2017. 3

[9] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. Get3d: A generative model of high quality 3d textured shapes learned from images. *Advances In Neural Information Processing Systems*, 35:31841–31854, 2022. 3

[10] 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, 6, 7

[11] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. Stylenerf: A style-based 3d-aware generator for high-resolution image synthesis. *arXiv preprint arXiv:2110.08985*, 2021. 3

[12] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009. 2

[13] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato's cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019. 3

[14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 3, 6

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 2

[16] 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, 4

[17] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 3

[18] Hans Lamecker, Martin Seebass, Hans-Christian Hege, and Peter Deuflhard. A 3d statistical shape model of the pelvic bone for segmentation. In *Medical Imaging 2004: Image Processing*, volume 5370, pages 1341–1351. SPIE, 2004. 2

[19] Isaak Lim, Alexander Dielen, Marcel Campen, and Leif Kobbelt. A simple approach to intrinsic correspondence learning on unstructured 3d meshes. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 2

[20] Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE international conference on computer vision*, pages 5659–5667, 2017. 2

[21] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *Advances in Neural Information Processing Systems*, 33:15651–15663, 2020. 2

[22] Zhen Liu, Yao Feng, Michael J Black, Derek Nowrouzezahrai, Liam Paull, and Weiyang Liu. Meshdiffusion: Score-based generative 3d mesh modeling. *arXiv preprint arXiv:2303.08133*, 2023. 3

[23] Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[24] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 7

[25] Cristian Lorenz and N Krahnstover. 3d statistical shape models for medical image segmentation. In *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No. PR00062)*, pages 414–423. IEEE, 1999. 2

[26] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021. 3

[27] Jun Ma, Yao Zhang, Song Gu, Cheng Zhu, Cheng Ge, Yichi Zhang, Xingle An, Congcong Wang, Qiyuan Wang, Xin Liu,

et al. Abdomenct-1k: Is abdominal organ segmentation a solved problem? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6695–6714, 2021. 6

[28] Julien NP Martel, David B Lindell, Connor Z Lin, Eric R Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788*, 2021. 2

[29] Quan Meng, Anpei Chen, Haimin Luo, Minye Wu, Hao Su, Lan Xu, Xuming He, and Jingyi Yu. Gnerf: Gan-based neural radiance field without posed camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6351–6361, 2021. 3

[30] Tom Mertens, Jan Kautz, Jiawen Chen, Philippe Bekaert, and Frédo Durand. Texture transfer using geometry correlation. *Rendering Techniques*, 273(10.2312):273–284, 2006. 2

[31] 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. 2, 3

[32] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2

[33] Gimin Nam, Mariem Khlifi, Andrew Rodriguez, Alberto Tono, Linqi Zhou, and Paul Guerrero. 3d-ldm: Neural implicit 3d shape generation with latent diffusion models. *arXiv preprint arXiv:2212.00842*, 2022. 2, 3, 7

[34] Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. Hologan: Unsupervised learning of 3d representations from natural images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7588–7597, 2019. 3

[35] 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. 2

[36] Roy Or-El, Xuan Luo, Mengyi Shan, Eli Shechtman, Jeong Joon Park, and Ira Kemelmacher-Shlizerman. Stylesdf: High-resolution 3d-consistent image and geometry generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13503–13513, 2022. 2, 3

[37] 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. 2, 3, 4, 5, 6, 7

[38] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020. 2, 4

[39] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14335–14345, 2021. 2, 4

[40] Leonid I Rudin and Stanley Osher. Total variation based image restoration with free local constraints. In *Proceedings of 1st international conference on image processing*, volume 1, pages 31–35. IEEE, 1994. 5

[41] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. *Advances in neural information processing systems*, 31, 2018. 6

[42] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu, and Gordon Wetzstein. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20875–20886, 2023. 2, 3, 6, 7

[43] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[44] Shanlin Sun, Kun Han, Deying Kong, Hao Tang, Xiangyi Yan, and Xiaohui Xie. Topology-preserving shape reconstruction and registration via neural diffeomorphic flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20845–20855, 2022. 2, 4, 5, 6, 7

[45] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Carsten Stoll, and Christian Theobalt. Patchnets: Patch-based generalizable deep implicit 3d shape representations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, pages 293–309. Springer, 2020. 2, 4

[46] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. *Advances in neural information processing systems*, 29, 2016. 3

[47] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, pages 641–676. Wiley Online Library, 2022. 2

[48] Jiancheng Yang, Udaranga Wickramasinghe, Bingbing Ni, and Pascal Fua. Implicitatlas: learning deformable shape templates in medical imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15861–15871, 2022. 2, 4

[49] Chenyu You, Weicheng Dai, Fenglin Liu, Yifei Min, Haoran Su, Xiaoran Zhang, Xiaoxiao Li, David A Clifton, Lawrence Staib, and James S Duncan. Mine your own anatomy: Revisiting medical image segmentation with extremely limited labels. *arXiv preprint arXiv:2209.13476*, 2022. 2

[50] Chenyu You, Weicheng Dai, Yifei Min, Fenglin Liu, Xiaoran Zhang, Chen Feng, David A Clifton, S Kevin Zhou, Lawrence Hamilton Staib, and James S Duncan. Rethinking

semi-supervised medical image segmentation: A variance-reduction perspective. *Advances in Neural Information Processing Systems*, 2023. 2

[51] Chenyu You, Weicheng Dai, Yifei Min, Lawrence Staib, and James S Duncan. Bootstrapping semi-supervised medical image segmentation with anatomical-aware contrastive distillation. In *International Conference on Information Processing in Medical Imaging*, 2023. 2

[52] Chenyu You, Weicheng Dai, Yifei Min, Lawrence Staib, and James S Duncan. Implicit anatomical rendering for medical image segmentation with stochastic experts. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023. 2

[53] Chenyu You, Weicheng Dai, Yifei Min, Lawrence Staib, Jas Sekhon, and James S Duncan. Action++: Improving semi-supervised medical image segmentation with adaptive anatomical contrast. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023. 2

[54] 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*, 2019. 3

[55] Chenyu You, Jinlin Xiang, Kun Su, Xiaoran Zhang, Siyuan Dong, John Onofrey, Lawrence Staib, and James S Duncan. Incremental learning meets transfer learning: Application to multi-site prostate mri segmentation. In *International Workshop on Distributed, Collaborative, and Federated Learning*. Springer, 2022. 2

[56] Chenyu You, Junlin Yang, Julius Chapiro, and James S Duncan. Unsupervised wasserstein distance guided domain adaptation for 3d multi-domain liver segmentation. In *International Workshop on Interpretable and Annotation-Efficient Learning for Medical Image Computing*. Springer, 2020. 2

[57] Chenyu You, Ruihan Zhao, Fenglin Liu, Siyuan Dong, Sandeep Chinchali, Ufuk Topcu, Lawrence Staib, and James Duncan. Class-aware adversarial transformers for medical image segmentation. *Advances in Neural Information Processing Systems*, 2022. 3

[58] Chenyu You, Ruihan Zhao, Lawrence H Staib, and James S Duncan. Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2022. 2

[59] 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. *IEEE Transactions on Medical Imaging*, 2022. 2

[60] X Zheng, Yang Liu, P Wang, and Xin Tong. Sdf-stylegan: Implicit sdf-based stylegan for 3d shape generation. In *Computer Graphics Forum*, volume 41, pages 52–63. Wiley Online Library, 2022. 2, 6

[61] 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, 4, 6, 7

[62] Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5826–5835, 2021. 3, 7