Multiscale Wave-Shaping Neural Encoding with Cross-Attention Networks for 3D Organ Reconstruction and Generation

Diego A. Flores, Panagiotis G. Kalozoumis, and Dimitris K. Iakovidis(✉)

Department of Computer Science and Biomedical Engineering, University of Thessaly, Greece
{diakovidis}@uth.gr

Abstract. Accurate synthetic 3D shape generation and patient-specific reconstruction of complex anatomical structures is critical for ML-driven medical analysis, simulation, and digital twin creation. Signed distance field (SDF) implicit neural representations have recently demonstrated their effectiveness in the context of generating synthetic models of the human heart. However, current methods for the generation of shapes of organs with more complex geometries, such as the geometry of the colon, fail to accurately reproduce details, such as the folds throughout the colon. In this paper we propose a framework to enhance the robustness of the state-of-the-art SDF-based approach for synthetic heart generation for higher-fidelity organ modeling, and we evaluate its effectiveness in the generation of 3D colon shape representations. The proposed framework is based on SDF implicit neural representations, and effectively combines: a) multiscale latent grids, b) a wave-shaping neural activation encoding function, and c) cross-attention networks, for improved patient specific organ reconstruction and generation. An ablation study is provided to justify the contribution of each of the introduced components for heart modeling. Subsequently, a first proof-of-concept experimental investigation on public benchmark datasets is performed to assess the performance of the proposed framework for the generation and reconstruction of a human colon, which has a more complex shape as a longitudinal structure with puckered pouches and sharp turns. The results indicate that the proposed framework can provide improved 3D reconstruction and generation over the state-of-the-art methods for both heart and colon models.

Keywords: Signed distance field, implicit neural representations, multiscale modeling, cross-attention, synthetic data generation, 3D shapes, digital twins.

1 Introduction

Accurate modeling of three-dimensional (3D) anatomical structures from medical imaging data is foundational for clinical analysis, simulation, and the creation of patient-specific digital twins. Traditional explicit 3D representations, such as meshes and voxel grids, lack flexibility and continuous detail required for accurate reconstruction and generation of complex anatomical structures, particularly when

patient-specific variability or topological complexity is present. In contrast, implicit neural representations (INRs), especially neural signed distance functions (SDFs), offer a continuous, differentiable, and compact formulation by modeling surfaces as the zero level set of a learned function that maps 3D coordinates to continuous distance values. Early work in visual computing demonstrated that neural SDFs can represent entire classes of shapes with compact latent codes and enable interpolation, reconstruction, and completion of complex 3D geometry [1]. A recent survey in 3D generative deep learning highlights that INRs have extended beyond simple reconstruction to generative tasks that require a nuanced understanding of global structure and fine geometric details. Continuous implicit fields accommodate topology variation and surface complexity more naturally than explicit voxel or mesh representations, making them suitable for high-fidelity modeling tasks across scales [2].

Generative modeling via deep learning has shown that hierarchical latent structures and attention mechanisms [3] can significantly improve synthesis quality by capturing both global context and local detail [4]. Transformer-based and cross-attention models used in latent generative frameworks (*e.g.*, latent diffusion models) make it possible to condition generation processes on structured latent embeddings, allowing richer expressiveness and better capture of multi-scale patterns in data [5]. These insights indicate that multi-scale latent encodings and cross-region dependencies can be valuable elements for advancing implicit 3D generative modeling.

While the current studies demonstrate the utility of implicit neural fields for 3D modeling tasks, there remains a gap in architectural development for generative 3D modeling of complex anatomical structures. In the context of medical generative modeling, SDF4CHD (Signed Distance Fields for Congenital Heart Disease) applied SDF-based implicit representations to congenital heart disease, demonstrating that INR models can reconstruct and generate clinically plausible cardiac geometries [6]. To the best of our knowledge, architectural enhancements, such as multi-scale latent grids and cross-attention networks have not yet been fully integrated into INR-based medical generative methods. Moreover, the extension of these implicit generative frameworks beyond cardiac anatomy to other clinically significant organs, particularly hollow tubular structures such as the gastrointestinal tract, remains underexplored.

The main contributions of this study are:

1. It introduces Multiscale Latent Encoding with Cross-Attention Networks (MLE-CAN), an enhanced implicit neural representation framework that combines multi-scale latent grids, wave-shaping neural activation (WNA) encoding, and cross-attention networks to improve both reconstruction and generative modeling of complex anatomical shapes.
2. It performs a comprehensive evaluation of the proposed framework on public benchmark datasets, investigating its effectiveness for the 3D reconstruction and generation of simpler and more complex anatomical structures, such as the human heart and colon, respectively.
3. It extends the SDF4CHD state-of-the-art approach to establish a baseline in generating and reconstructing complex anatomies with limited labeling, such as the colon.

2 Related Work

Neural implicit shape representations, such as DeepSDF and follow up methods have enabled continuous modeling of 3D geometry from latent codes [1]. Neural Diffeomorphic Flow (NDF) extends this paradigm by learning topology-preserving diffeomorphic mappings in the implicit domain, facilitating accurate 3D shape reconstruction and registration on organ datasets [7]. Recent advances introduce structured latent embeddings and hierarchical architectures. For instance, Guillard *et al.* [8] designed a single network whose early layers produce coarse geometry and later layers add fine detail, effectively yielding multiple levels of detail from one latent code. Also, attention mechanisms have been leveraged in neural fields: transformer-style cross-attention layers can inject context vectors (*e.g.* shape priors or modality features) into the implicit decoding process, analogous to their success in 2D latent diffusion models. These multi-scale and attention-based architectures markedly improve the ability of implicit models to represent both global structure and local detail in generated shapes [5].

Neural implicit fields have begun to impact medical imaging by enabling continuous, differentiable models of anatomy. Early works in medical shape modeling often used PCA or mesh-based VAEs on fixed topologies [9], but these struggle with the topological variability common in pathology. Implicit SDF models overcome this by naturally representing diverse topologies. For example, Sander *et al*. [10] trained a multilayer perceptron-based framework that learns a continuous SDF for the left ventricle (LV) that reconstructs high-resolution geometry from sparse anisotropic MRI, effectively performing implicit super-resolution of cardiac shapes. Recent works explicitly integrate multi-scale and conditioning mechanisms into implicit models for medical shapes. Yang *et al.* [11] proposed Implicit Atlas, learning a shared template field plus deformation grids to represent organ shapes, improving shape representation capacity on certain medical datasets. More recently, De Wilde et al. [12] proposed a steerable generative model based on implicit neural representations for thyroid synthesis and reconstruction. Currently, the SDF4CHD framework proposed by Kong *et al.* [6] is the framework with the widest range of capabilities in the field of 3D congenital heart disease characterization and modeling. However, the framework's abilities can be further enhanced with strategic architectural enhancements, and further applications can be explored towards the 3D modeling of other anatomical structures.

Other works that attempt to reconstruct or generate synthetic 3D organs or 3D medical images include diffusion models like MedSDF [13], $TrIND$ [14], MeshHeart [15], Denoising Diffusion Probabilistic Models (DDPMs) [16] and traditional CNNs with Voxels as input like Voxel2Mesh [17]. Additionally, GAN based methods for point cloud 3D shape reconstruction with promising applications for synthetic 3D shape generation have been proposed [18]. Overall, the field of 3D medical shape characterization is progressing, but several models limit themselves to generating or reconstructing organs with simple geometries and most avoid organs or structures with complex geometries unless they have been carefully labeled in subsections.

3 Multi-Scale Cross-Attention Framework

The proposed framework is composed of a decomposition and flow/correction networks similar to the SDF4CHD framework proposed by Kong *et al*. [6], but enhanced with three key architectural subsystems (Fig. 1): a) A pair of multi-resolution shape and correction codes (latent grids) $\{Z_S^{(\ell)}\}^{K}{}_{\ell=1}$, $\{Z_C^{(\ell)}\}^{K}{}_{\ell=1}$, b) wave-shaping neural activation (WNA) [19] coordinate encoding, and c) the Type-token cross-attention injecting the type code latent vector (Z_T) as a learned token memory into the type SDF decoder, the deformation network and the correction network. The "Type" branch is the part of the framework that encodes the multi-hot binary (also called a multi-label binary encoding) Input Diagnosis Vector z_T, which is the vector that specifies if an organ (in this case a heart) possess none or several types of CHD, e.g. Tetralogy of Fallot (TOF), Double Outlet Right Ventricle (DORV), Transposition of the Great Arteries (TGA), etc. By default a healthy heart's input diagnosis vector (lower left corner at Figure 1) is set up to (0,0,0,0,0,0), while a heart with ToF is set up to (0,0,0,0,1,0) and a heart with all three TGA, DORV and ToF is set up to (0,0,1,0,1,1). This section will break down the theoretical aspects of why each of the three enhancements mentioned make the framework better at patient-specific organ reconstruction and synthetic generation.

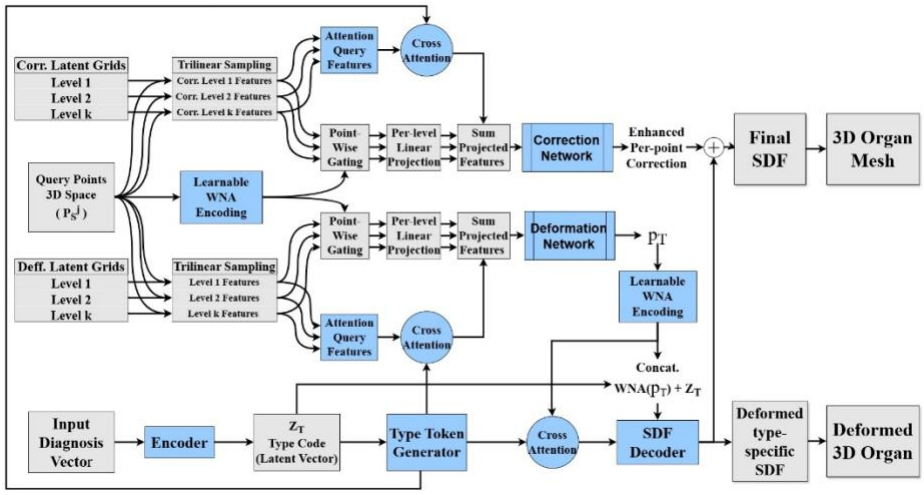

Fig. 1. Graph of our Multi-Scale Cross Attention Framework displaying a total of *k*-levels in the Correction and Deformation Latent Grids.

In Fig. 1, each of the *k* levels is a grid/tensor of dimension $2^{\ell} \times 2^{\ell} \times 2^{\ell} \times 27$ where ℓ and k are positive integers and $2 \leq \ell \leq k+1$. A single grid must represent both global and local anatomy variations. Multi-resolution grids allow a coarse level to represent global morphology and finer levels to represent boundary detail, which is analogous to multi-resolution implicit representations [20] [21]. For each resolution level, we build deformation and correction grids (located at the left section of Figure 1) defined as:

$$Z_D^{(\ell)} \in \mathbb{R}^{2^\ell \times 2^\ell \times 2^\ell \times C}, Z_C^{(\ell)} \in \mathbb{R}^{2^\ell \times 2^\ell \times 2^\ell \times C}, \quad (1)$$

where 2^ℓ is the spatial resolution and C is the channel dimension (in Fig. 1, $C = 27$). Consequently, for any point $\mathbf{p}$, trilinear sampling $\mathrm{Tri}()$ produces per-point features:

$$\mathbf{z}_D^{(\ell)}(\mathbf{p}) = \mathrm{Tri}\left(Z_D^{(\ell)}, \mathbf{p}\right) \in \mathbb{R}^C, \mathbf{z}_C^{(\ell)}(\mathbf{p}) = \mathrm{Tri}\left(Z_C^{(\ell)}, \mathbf{p}\right) \in \mathbb{R}^C, \quad (2)$$

where $\mathrm{Tri}()$ interpolates a coordinate inside a 3D grid cell using the values at its 8 corner points. When encoding sample points from a 3D space, we encode coordinates with a wave-shaped periodic mapping. For frequency bands $\{f_i\}_{i=1}^F$ (*e.g.*, $f_i = 2^{i-1}$) and learnable spectral scales $\{\omega_i\}_{i=1}^F$, define the wave-shaping primitive (applied element-wise)

$$\mathrm{WS}(x; \omega) = \tanh(\sin(\omega x)) \quad (3)$$

The WNA encoding of a 3D point $\mathbf{p} = [p_x, p_y, p_z]$ is:

$$\gamma_{\mathrm{WNA}}(\mathbf{p}) = [\mathbf{p}, \tanh(\sin(\omega_i f_i \mathbf{p})), \tanh(\cos(\omega_i f_i \mathbf{p}))]_{i=1}^F \in \mathbb{R}^C, \quad (4)$$

where the scalar $(\omega_i f_i)$ scales each coordinate, and $\sin()$, $\cos()$, $\tanh()$ are applied component-wise to a vector and $C = 3 + 6F$ (For $F = 4$, $C = 27$ as in Fig. 1). WNA excels in two critical aspects: (i) Introduces a learnable spectral scale, as the frequency ω_i is trainable, so the encoding adapts frequency emphasis to organ anatomy scale variability; (ii) The wave shaping functions $\tanh(\sin(\omega_i x))$ and $\tanh(\cos(\omega_i x))$ compress extremes (bounded nonlinearity), effectively altering the harmonic content and stabilizing gradients near peaks. This is conceptually aligned with the broader line of work using periodic structure for implicit fields (*e.g.*, sine activations in SIREN [22]) while specifically adopting the wave-shaping neural encoding approach of Triantafyllou *et al*. [19] to encode the query points in the 3D voxel space. Additionally, we convert the input diagnosis vector z_T into a small set of learnable tokens (a "memory") and let each spatial point query that memory (lower section of Figure 1). This makes type conditioning content-based and point-dependent: points in different anatomical regions can retrieve different type information. The latent vector is mapped into T tokens of width d_f with W_{tok} being the linear map from z_T to a flattened token set and b_{tok} the bias for token generation:

$$M = \mathrm{reshape}(W_{\mathrm{tok}} \cdot z_T + b_{\mathrm{tok}}) \in \mathbb{R}^{T \times d_f} \quad (5)$$

For each point p, we build a query feature $q(p)$ from its local point/geometry features (*e.g.*, WNA-encoded coordinates and/or multi-level grid features). Cross-attention returns a context vector $c(p)$ that we add to the point's hidden feature stream:

$$\mathrm{Attn}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_h}}\right)V, \quad (6)$$

$$c(p) = W_o \cdot \mathrm{Attn}\left(W_q q(p), W_k M, W_v M\right), \quad (7)$$

$$h(p) \leftarrow h(p) + c(p) \quad (8)$$

where *Q, K* and *V* are the stacked projected queries, keys and values from *N* points and T tokens respectively. *Wq*, *Wk*, *Wv* and *Wo* are matrices corresponding to query projection, key projection, value projection and output projection respectively. This approach directly mimics the transformer principle [3] [23].

After the trilinear sampling, the attention query features are not raw trilinear samples; they are the gated sum of sampled grid features of the correction and deformation latent grids. In combination with WNA, the cross-attention (CA) pathway injects type semantics not only into the SDF decoder but also into the deformation and correction branches, improving topology-consistent synthesis. At each level k, multiscale latent-grid features are obtained by trilinear sampling Tri() and combined with WNA-encoded query points $\gamma_{\mathrm{WNA}}(\mathbf{p})$ through point-wise gating, per-level linear projection, and summation of the gated projected features. The resulting feature vector is further enriched by a type-token context obtained via cross-attention and then passed to the deformation and correction networks (MLPs with 6 hidden layers each located at the center of Figure 1), which predict the point velocity field and the SDF correction residual, respectively. If k=1, the sum projected features becomes a one gated projection and the attention query from gated multi-level summary becomes a gated single-level sample.

4 Experimental Setup and Results

To assess the proposed framework, it was trained and tested on the congenital heart disease (CHD) dataset obtained from Kong *et al*. [6] and the HQColon dataset provided by Finocchiaro *et al*. [24]. The CHD dataset provides 3D labels of the 7 major regions of the heart as well as providing labels according to the CHD a given heart suffers. The HQColon dataset does not provide segmentation labels or disease classification. While the same data preprocessing algorithm is used to preprocess the CHD and HQColon datasets, the CHD heart 3D CT scans are stored as 7-channel SDFs depicting the 7 major regions of the heart while the CT colonography HQColon data will be stored as a single channel SDF depicting the full colon with no subregions. Hence, the HQColon proves to be a harder challenge for both the reconstruction and the generation task.

During training, the full network and the per-sample latent grids are optimized under a compound loss: MSE on occupancy for both reconstructions (without and with correction), a Gaussian prior on the latent grids, a light total-variation term on those grids, and a Lipschitz penalty on the type decoder MLP of 6 hidden layers. For both datasets, the generation task consists of the latent grids being initiated as random gaussian values and then for them to serve as input for the generation framework to later output conditional shapes according to the input diagnosis vector (if labeling is provided). For the reconstruction task, network weights are frozen and the latent grids are samples from random gaussian values to later be optimized for 200 iterations using the Adam optimizer with a learning rate of 0.01 per SDF of a given structure meant to be reconstructed under a compound loss composed of MSE loss on sampled points of the type/deformation output without correction and the final output with correction (with the final output being upweighted on interior and surface samples), plus a small Gaussian penalty on the grids. A reduce on plateau learning rate scheduler for the 200 Adam optimization operations is set up with a patience of 15 and reduction factor of 0.6 on plateau. Finally, reconstructions are rasterized on a 128^3 or 256^3 occupancy grid using a coarse-to-fine query. Occupancy above 0.5 is treated as inside the organ; each

voxel is assigned the highest-scoring anatomical class (for multi-channel SDFs), or background if all scores fall below 0.5. Dice score reconstruction is computed on this labeled volume and marching cubes at the same 0.5 threshold is used to export the surfaces.

4.1 Experiments on the CHD dataset

This study explores the effects of the WNA, cross-attention and multilevel latent grids against the SDF4CHD framework. For a fair comparison, the same train-test split used by Kong *et al*. [6] for evaluating SDF4CHD and the rest of INR-based models was used to train and test our framework. The training set consisted of 67 patients spanning 6 CHD types and 14 combinations of CHD types and the testing set consisted of 34 additional patients with CHD combinations not present in the training set. Also, we trained and tested the SDF4CHD on additional single (k=1) $Z_D^{(4)}, Z_C^{(4)}$ and $Z_D^{(5)}, Z_C^{(5)}$ latent grids to compare their framework's reconstruction capabilities with our multiscale cross attention framework with $1 \leq k \leq 4$ on whole heart (WH) reconstruction and individual cardiac anatomies: left ventricle (LV), right ventricle (RV), left atrium (LA), right atrium (RA), myocardium (Myo), aorta (Ao) and pulmonary artery (PA). The same train and test split were used as provided by the CHD dataset (Table 1 and Figure 2 and 3). The reason each framework/model must be able to accurately reconstruct a given heart is that by varying the input diagnosis vector on a learned heart morphology's SDF, the model can further deform the heart to later resemble or depict different sets of combinations of congenital heart disease on a patient specific heart. We preprocessed the CHD CT scan dataset into SDFs and ground truth heart meshes with the same method as Kong *et al*. [6] which delimits the INR in a 3D voxel space of 128^3.

Our training approach is similar to the one of Kong *et al*. [6]. However, for the CHD dataset, a healthy (no CHD) sample of the dataset is chosen, and the type CHD network encoder (which encodes the multi-hot binary input diagnosis vector), WNA encoding (which encodes the 3D space), the type token generator, the attention mechanism and SDF decoder are all fitted under MSE loss between predicted and the ground truth occupancy of the healthy sample for 500 iterations of gradient descent updates with learning rate $1\text{x}10^{-4}$. At the start of the training both the correction and deformation networks were trained in an alternating approach for the first 100 epochs in which one network remains frozen, and the other one gets trained for 5 epochs and vice versa. Afterwards, the full network weights are trained for 1000 epochs. The initial learning rate for the network is $1\text{x}10^{-4}$ while for the latent grids is $1\text{x}10^{-3}$. A reduce-on-plateau learning rate scheduler is set, where the learning rate of both the network and latent grids is halved if there is no reduction in the reconstruction loss for 20 consecutive epochs. The minimum learning rate for both the network and its multilevel latent grids the training is set up to reach is $1\text{x}10^{-6}$.

For our multiscale WNA Cross-Attention-based framework we also provided an ablation study to further justify the presence of each of its main components. For the WNA encoding method, the ablation study also includes two variants with Fourier Positional Encoding (FPE) replacing WNA as well as the framework having or not

having the cross-attention (CA) mechanism. It can be observed in Figure 2 and 3 that the dice score results (Y axis) of our model (Red Lines) increase as $1 \leq k \leq 4$ increases (4 graphs in total) and outperforming most of its ablation variants in whole heart and individual cardiac structure (X axis) reconstruction measured in dice score when $k = 4$ and $2 \leq \ell \leq 5$. This experiment sets no level of the latent grids to have equal dimensions to another level for all $1 \leq k \leq 4$. This is to have an ℓ value that always increases as k increases, hence creating a full set of latent grid levels of strictly ascending complexity $\{Z_D^{(2)}, Z_C^{(2)}, Z_D^{(3)}, Z_C^{(3)}, Z_D^{(4)}, Z_C^{(4)}, Z_D^{(5)}, Z_C^{(5)}\}$ when $k = 4$.

Table 1. SDF4CHD model dice score results for whole heart (WH) and individual cardiac structures in reconstructed geometries compared to different baselines for unseen cases (test set) with their respective latent grid dimension ℓ and k = 1. "Corr" denotes the SDF4CHD model variant with correction network added to improve the reconstruction results.

Single Latent Grid ℓ	WH	LV	RV	LA	RA	Myo	Ao	PA
2	86.3±2.7	91.4±2.7	88.3±8.0	82.6±5.3	88.8±4.0	88.9±5.5	76.3±12.8	72.1±16.0
2+Corr.	88.8±2.0	91.4±2.7	88.3±7.9	85.5±4.7	91.9±2.1	90.0±3.8	83.0±8.6	76.7±14.3
3	92.9±1.3	95.5±0.8	94.1±3.2	90.4±3.9	95.1±1.2	93.9±2.2	88.1±7.0	81.6±15.0
3+Corr.	94.2±1.0	95.5±0.8	94.3±2.3	92.7±3.2	96.5±0.8	94.3±1.9	91.2±5.0	85.5±12.3
4	95.4±0.7	96.9±0.5	96.3±1.1	94.5±1.9	96.9±0.7	95.4±1.4	93.7±2.6	90.5±5.7
4+Corr.	96.2±0.4	96.9±0.5	96.3±1.1	96.2±1.3	97.3±0.5	95.6±1.3	95.5±1.1	93.2±3.4
5	95.8±0.9	97.1±0.6	96.7±0.7	95.5±2.0	97.3±0.7	95.6±1.6	93.5±3.3	90.6±7.9
5+Corr.	**96.9±0.4**	**97.2±0.5**	**96.7±0.7**	**97.4±0.8**	**97.9±0.3**	**96.0±1.3**	**96.5±0.7**	**95.2±2.5**

Fig. 2. Ablation study of the Multiscale Cross Attention SDF4CHD model presenting the dice score results (Y axis) for whole heart (WH) and individual cardiac structures (X axis) in reconstructed geometries compared to different baselines for unseen cases (test set) with their respective latent grid levels and up to 4 levels (depicting $1 \leq k \leq 4$), stating $2 \leq \ell \leq 5$. The overall standard deviation was of the order of 10^{-1}.

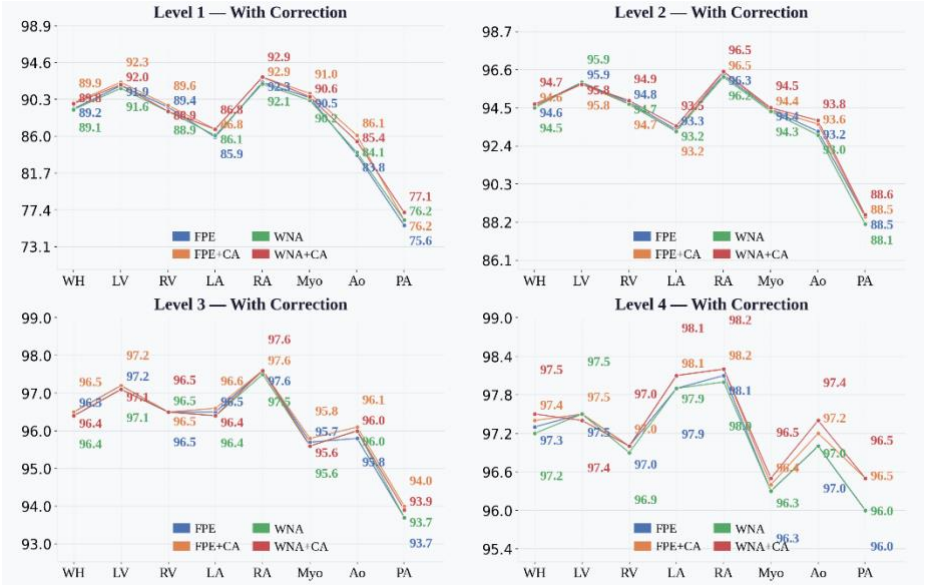

Fig. 3. Ablation study of the Multiscale Cross Attention SDF4CHD model presenting the dice score results (Y axis) for whole heart (WH) and individual cardiac structures (X axis) in reconstructed geometries with the correction module compared to different baselines for unseen cases (test set) with their respective latent grid levels and up to 4 levels (depicting $1 \leq k \leq 4$), stating $2 \leq \ell \leq 5$. The overall standard deviation was of the order of 10^{-1}.

Regarding synthetic 3D heart generation, all the evaluated models in Table 2 have the property of relying on single or dual latent grids. Kong *et al*. [6] already determined that in the present scenario and dataset that optimal generation is achieved with single level latent grids of dimensions $Z_D^{(2)}, Z_C^{(2)}$. For a fair comparison we have used the provided pretrained model of Kong *et al*. [6] to generate 110 synthetic hearts of varying CHD combinations as the optimal training for their framework is different than the optimal training for our framework. In our framework, to achieve optimal generation all shape code dimensions in all levels must be kept at $Z_D^{(2)}, Z_C^{(2)}$. We observed that optimal generation is achieved with 5 levels of latent grids all having $Z_D^{(2)}, Z_C^{(2)}$. If a single latent grid is elevated at $Z_D^{(3)}, Z_C^{(3)}$, the reconstruction ability increases but high irregularities appear in the surface of the synthetic generated hearts.

The quantitative and qualitative evaluation of the synthetic 3D heart meshes is shown in Table 2 and Figure 4 respectively. For each model, each generated mesh is represented by 32,000 points drawn on the triangles with probability proportional to face area, so the samples follow anatomical surface area rather than local vertex density

and remain comparable across meshes with different tessellations. The pairwise similarity is the mean squared Chamfer Distance [25]. For a given mesh every sample point on a given surface is matched to its nearest neighbor sampled point on the surface of another mesh and every generated mesh is compared to every ground truth mesh (and conversely) so that the provided set metrics assess two collections of shapes rather than patient-wise overlay. Each synthetic mesh is compared to every ground truth mesh and each ground truth mesh is compared to every generated mesh. The metrics provided are Symmetric Minimum Matching Distance (MMD-CD) [26], Set Precision and Recall [27] [28], Root Mean Square Roughness (RMSR) [29], Dihedral Angle Standard Deviation (DA-STD) [30] and Mean Aspect Ratio (MAR) using the edge ratio method [31]. Lower MMD-CD implies synthetic heart distribution is closer to real heart distribution. Higher set precision (Set-P) means more of the generated organs look like the real ones and higher set recall (Set-R) implies the generated set covers more of the real anatomical variety. Reduced RMSR after high-pass filtering suggests smoother surfaces with fewer high-frequency artifacts. DA-STD, inspired by dihedral-angle-based mesh quality assessment by Váša & Rus (2012) [30], quantifies local surface regularity as the standard deviation of the angles between normals of adjacent triangular faces, computed over all internal mesh edges. Additionally, MAR values closer to 1 (ideal equilateral triangles) demonstrate that our method generates more regular usable meshes suitable for Finite Element-based simulation and analysis.

Table 2. Evaluation and comparison of synthetic heart generation among several baselines and state-of-the-art-models and our multiscale cross-attention model with 5 equal levels of latent grids of $Z_D^{(2)}, Z_C^{(2)}$. 110 heart meshes with varying pathologies were generated by each model and then compared to the full cohort of the CHD dataset composed of 110 ground truth heart meshes. "+Corr" denotes the model plus the correction network.

Model	MMD-CD↓	Set-P↑	Set-R↑	RMSR↓	DA-STD↓	MAR↓
DeepSDF [1]	0.001544	0.0818	0.0636	0.002799	6.585876°	6.924780
CDeepSDF [6]	0.001278	0.4272	0.2364	0.002783	6.069758°	6.574505
NDF [7]	0.001440	0.1909	0.0364	0.002794	6.163693°	6.257644
SDF4CHD [6]	0.001214	0.8545	0.2364	0.002807	6.522208°	6.454072
SDF4CHD+Corr	0.001209	0.8454	**0.2727**	0.002806	6.519051°	6.928455
Our Framework	0.001199	**0.9364**	0.2182	**0.000859**	4.038681°	1.233010
Our Framework+Corr	**0.001192**	0.9273	0.2182	0.000861	**4.010311°**	**1.232529**

In this experiment, the advantage of our framework over the rest is that the implementation of multiscale latent grids allows it to maneuver between the two endpoints of the SDF4CHD. As Kong et al. [6] stated in their single level experiments that better reconstruction is achieved with latent grids (k=1) $Z_D^{(3)}, Z_C^{(3)}$ for all SDF4CHD, DeepSDF, CDeepSDF and NDF but with decrease in synthetic generation capacity (hearts with surface irregularities and artifacts). Kong et al. used (k=1) $Z_D^{(2)}, Z_C^{(2)}$ for SDF4CHD, DeepSDF, CDeepSDF and NDF to generate conditional or unconditional synthetic hearts. Given the fact that our framework has the multiscale latent grids, it can increase the dimensional complexity of the latent grids by varying

$1 \leq k \leq 5$ but leaving all five latent grid levels all equal to $Z_D^{(2)}, Z_C^{(2)}$ $(\ell = 2)$. This is done with the intention of finding a midpoint of complexity higher than $Z_D^{(2)}, Z_C^{(2)}$ but lower than $Z_D^{(3)}, Z_C^{(3)}$ which is when generation capacity begins to decrease. Hence, Table 2 demonstrates that our framework produces quantitatively more accurate hearts compared to the ground truth ones with better quality meshes for computational modeling downstream tasks.

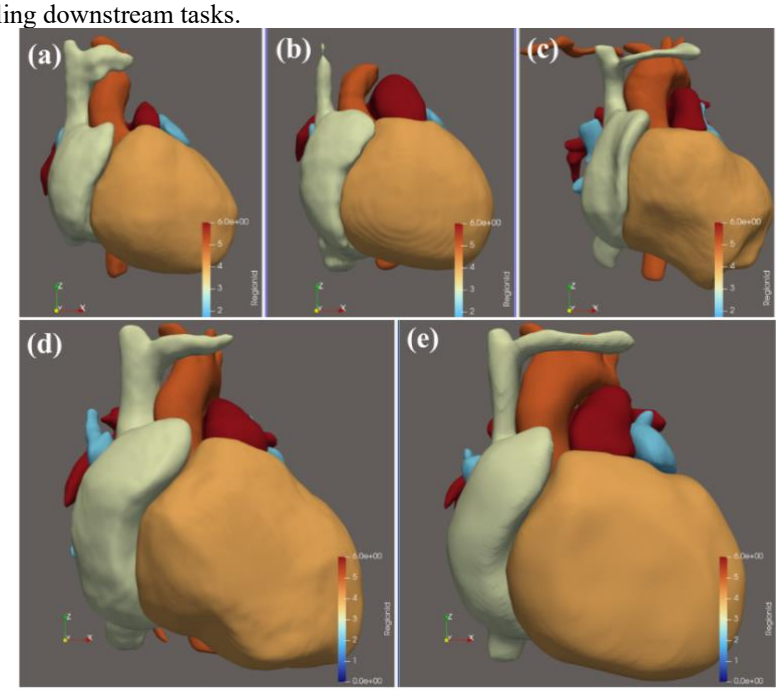

Fig. 4. 3D synthetic hearts generated by (a) DeepSDF, (b) CDeepSDF, (c) NDF, (d) SDF4CHD+Corr and (e) our Multiscale Cross-Attention SDF4CHD+Corr model with 5 equal levels of latent grids of $Z_D^{(2)}, Z_C^{(2)}$.

4.2 Experiments on the HQColon dataset

In this section, experiments with two types of data preprocessing approaches were conducted using the HQColon dataset. In the first case, data with the same voxel space dimensions are preprocessed as the CHD experiments (i.e., 128^3 voxel space). With this setup, our multiscale cross attention model had a different behavior in comparison to the CHD dataset experiments. In our multiscale model, the advantage provided by the multi-level latent grids in the heart dataset's generation task is allowing the complexity to maneuver in between $Z_D^{(2)}, Z_C^{(2)}$ and $Z_D^{(3)}, Z_C^{(3)}$. However, given the higher degree of anatomical complexity and lack of labels partitioning the colons, higher complexity is needed for the latent grids in both reconstruction and generation. Hence, we set the deformation and correction latent grids to $Z_D^{(5)}, Z_C^{(5)}$ to achieve top reconstruction and synthetic 3D colon generation. For all evaluated models, low latent grid dimensions

yielded poor reconstruction and generation quality. Hence, for a fair evaluation we set the latent grids to $Z_D^{(5)}, Z_C^{(5)}$ in all the models and used the same train – test split (Table 3, Fig. 5). The HQColon dataset was separated into a training set composed of the first 400 colons and a testing set composed of the remaining 35 colons. For the training of our framework and SDF4CHD, a sample of the dataset is chosen and the type network encoder, WNA encoding, the type token generator, the attention mechanism and SDF decoder are all fitted under MSE loss between predicted and the ground truth occupancy of the fixed sample for 1000 iterations of gradient descent updates with learning rate $1x10^{-4}$. At the start of the training both the correction and deformation networks are trained in an alternating approach for the first 200 epochs in which one network remains frozen, and the other one gets trained for 5 epochs and vice versa. Afterwards, the full network weights are trained for 1000 epochs. The initial learning rate for the network is $1x10^{-4}$ while for the latent grids is $1x10^{-3}$. A reduce on plateau scheduler is set, under which the learning rate of both the network and latent grids is halved if there is no reduction on the reconstruction loss for 20 consecutive epochs. The minimum learning rate the training is set up to reach is $1x10^{-6}$. Additionally, unlike the CHD dataset, the HQColon dataset does not have labels for one or more conditions a colon may have (blockage, diverticulitis, cancer, polyps, etc) and it also does not provide labels for specific segments of the colon (e.g., ascending, transversal, descending colon and rectum). Hence, the evaluation of the generation task in this dataset is unconditional (the generated 3D colons are not meant to depict or resemble a specific disease, Fig. 6) and the models' generation and reconstruction ability is measured regarding their ability to generate or reconstruct the overall geometry of a colon.

Table 3. Dice score results per model for whole colon in reconstructed geometries in a voxel space of 128^3 compared to different baselines for unseen cases (35 test samples) with latent grids $Z_D^{(5)}, Z_C^{(5)}$. For GANs only the generator network is considered for the parameter and GFLOPs count. For the rest of the models the latent grids' 200 Adam optimization operations and number of parameters are considered. "+Corr" denotes the model plus the correction network.

Model	Whole Colon	No. of Parameters	GFLOPs
DeepSDF [1]	31.7±5.3	2,601,632	7,185.666
NDF [7]	91.9±5.6	4,073,484	39,991.875
SDF-VAEGAN [32]	62.2±6.0	53,324,097	492.077
SDF-VAEGAN+Diff. [33]	60.3±6.1	54,194,369	492.081
PointCloud-GAN [18]	86.0±1.3	212,533,248	0.425
SDF4CHD [6]	93.1±1.3	8,158,765	40,060.876
SDF4CHD+Corr	94.3±1.0	8,158,765	46,718.080
Our Framework	93.4±1.1	11,222,519	47,343.676
Our Framework+Corr	**96.8±0.4**	11,222,519	55,188.953

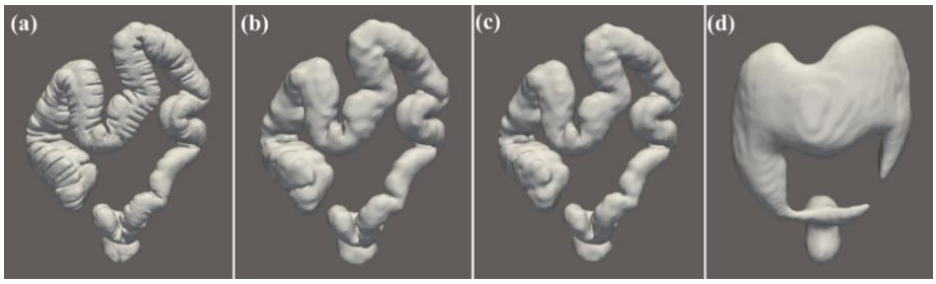

Fig. 5: Visual example of the reconstruction task of Colon #408 of the testing set in a voxel space of 128^3. (a) Ground truth testing set colon. (b) Proposed WNA Cross-Attention Model+Corr. (c) SDF4CHD+Corr. (d) SDF-VAEGAN.

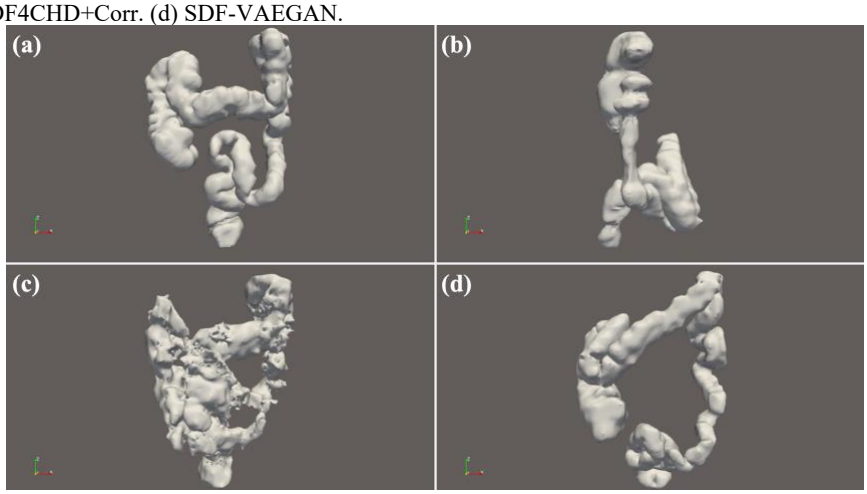

Fig. 6: Visual example of synthetically generated colon structures using different models in a voxel space of 128^3. (a) Proposed WNA Cross-Attention Model+Corr. (b) VAEGAN. (c) SDF4CHD+Corr. (d) NDF.

The GAN models did not provide good reconstruction and generation quality. Meanwhile some of the outputs of the SDF4CHD, the NDF model and our WNA Cross-attention model, despite resembling the shape of a human colon, are noisy, possess artifacts, irregularities and there is the memorization issue (generated colons are almost exact copies of the colons in the training set). Colon 3D modeling is challenging given that the colon's geometry presents strong connectivity and overall path, but also fine local detail. Although our model provides a lower incidence of artifacts and more instances of anatomically accurate colons, some of these generated colons are memorized, implying that generating morphologically complex and anatomically accurate organs with scarce labeling is still an open challenge.

On the second preprocessing approach, we preprocessed the data with the same algorithm as our CHD dataset experiments but in a voxel dimension of 256^3. This was made with the goal of the higher dimensionality of the preprocessing to translate into the preservation of most fine details (haustral folds, and other surface geometrical features) of the colons, that the lower resolution preprocessing might not capture. The

results were that increasing the dimensionality of the data helped better reconstruct the colons (Table 4, Fig. 7). For the synthetic generation quantitative evaluation, the same mesh preprocessing strategy and metrics were used as in the CHD dataset experiments (Table 5). The qualitative evaluation is shown in (Fig. 8). A nearest neighbor analysis of the generated colon meshes against the training set and unseen colon geometries of the test set was performed. Each synthetic or ground truth mesh was represented by 32,000 area-weighted surface points (Same as in Table 2 and 5). The pairwise similarity is the mean squared Chamfer Distance [25]. Every generated colon was compared to every training and test colon, with no assumed pairing. Two summaries are reported in Table 6: prefers training over test determines whether on average each generated colon, is closer to random training subset of the same size as the test set (35 vs 35) or to the test set itself. 50% implies no bias and above 50% a preference for training shapes. Unseen test colons recovered measures, for each of the 35 held out test colon geometries, whether some generated mesh is closer to it than any of the 400 training meshes.

Table 4. Dice score results per model for whole colon in reconstructed geometries in a voxel space of 256^3 compared to different baselines for unseen cases (35 test samples) with latent grids $Z_D^{(5)}, Z_C^{(5)}$. For GANs only the generator network is considered for the parameter and GFLOPs count. For the rest of the models the latent grids' 200 Adam optimization operations and number of parameters are considered. "+Corr" denotes the model plus the correction network.

Model	Whole Colon	No. of Parameters	GFLOPs
DeepSDF [1]	57.8±10.4	2,601,632	57,485.335
NDF [7]	78.2±5.9	4,073,484	319,935.002
PointCloud-GAN [18]	80.9±2.0	212,533,248	0.425
MedSDF [13]	90.6±1.7	85,984,625	3156.641
SDF4CHD [6]	93.1±1.6	8,158,765	320,487.006
SDF4CHD+Corr	97.3±0.5	8,158,765	373,744.641
Our Framework	92.0±5.2	11,222,519	378,749.412
Our Framework+Corr	**97.5±0.5**	11,222,519	441,511.624

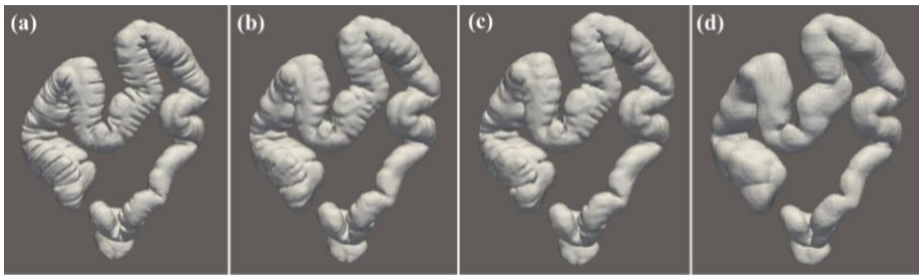

Fig. 7: Visual example of the reconstruction task of Colon #408 of the testing set in a voxel space of 256^3. (a) Ground truth testing set colon. (b) Proposed WNA Cross-Attention Model+Corr. (c) SDF4CHD+Corr. (d) MedSDF.

Table 5. Evaluation and comparison of synthetic colon generation among several baselines and state-of-the-art-models and our multiscale cross-attention model with 1 level of latent grids of

$Z_D^{(5)}, Z_C^{(5)}$. 435 unconditional colon meshes were generated by each model and then compared to the full cohort of 435 ground truth colon meshes belonging to the HQColon dataset. "+Corr" denotes the model plus the correction network.

Model	MMD-CD↓	Set-P↑	Set-R↑	RMSR↓	DA-STD↓	MAR↓
DeepSDF [1]	0.002631	0.5678	0.5425	**0.000611**	**5.211964°**	**1.360222**
MedSDF [13]	0.002428	0.3310	0.3241	0.001472	8.219325°	7.924783
NDF [7]	0.003033	0.5333	0.5425	0.000696	11.710187°	1.555547
SDF4CHD [6]	0.001211	0.8597	0.6781	0.000644	11.729441°	1.545792
SDF4CHD+Corr	0.001206	0.8574	0.6850	0.000672	12.360881°	1.559860
Our Framework	**0.001031**	**0.8920**	0.7287	0.000633	11.143069°	1.530602
Our Framework+Corr	0.001045	**0.8920**	**0.7333**	0.000652	11.594596°	1.540349

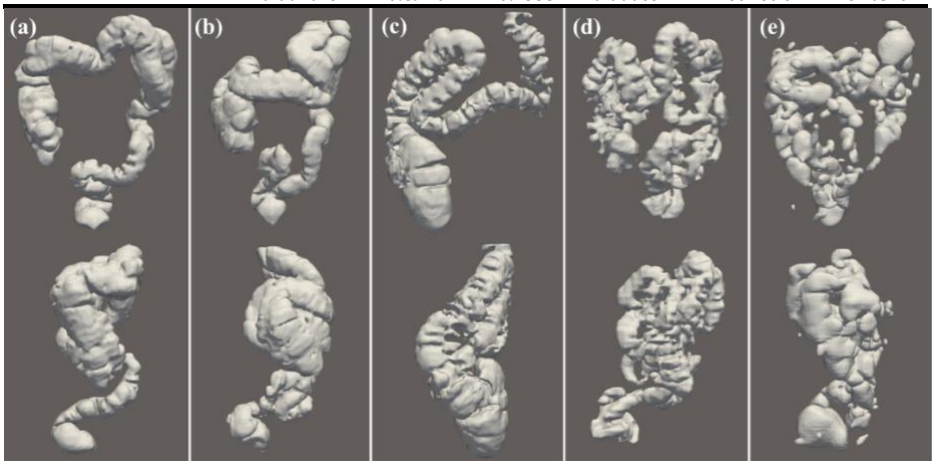

Fig. 8: Visual example of synthetically generated colon structures using different models in a voxel space of 256^3 with their respective frontal view (up) and lateral view (down). (a) Proposed WNA Cross-Attention Model+Corr. (b) SDF4CHD+Corr. (c) NDF. (d) DeepSDF. (e) MedSDF.

Table 6. Nearest Neighbor Analysis of the Generated Colons against the training set and the test set. "+Corr" denotes the model plus the correction network.

Model	Prefers Training over Test	Unseen test colons recovered
DeepSDF [1]	56.3%	8/35
MedSDF [13]	55.2%	4/35
NDF [7]	54.9%	7/35
SDF4CHD [6]	54.5%	8/35
SDF4CHD+Corr	58.6%	8/35
Our Framework	47.4%	13/35
Our Framework+Corr	**45.7%**	**14/35**

5 Conclusions and Future Directions

Our Multiscale Cross Attention Framework extends previous state-of-the-art SDF-based machine learning models by (i) introducing multi-level deformation/correction latent grids to allow model variant to maneuver in between several degrees of complexities, (ii) introducing a wave-shaping neural encoding featuring learnable spectral scaling, and (iii) injecting type information through a compact token memory queried via cross-attention across all three decoders. Altogether, this architectural setup yields state-of-the-art patient-specific reconstructions and high-fidelity organ generation under specific latent grids dimensional setups. For the heart, the present framework produced quantitatively more realistic hearts in comparison to the ground truth, less surface irregularities/artifacts and more suitable meshes for further biomechanical modeling and simulations (e.g. Finite Element Analysis), given a mean aspect ratio score of 1.23 on its mesh's triangular components. In terms of colon 3D modeling, qualitative evaluation of the outputs of our framework revealed less occurrence of artifacts, and higher anatomical fidelity in the generation task in comparison to other SDF-based machine learning as well as peak reconstruction of unseen colon geometries. Additionally, mesh quality metrics resulted to be acceptable for computational modeling downstream tasks. Finally, although all INR based models occasionally suffer from the memorization issue in the more complex 256^3 voxel space, our framework generated quantitatively more accurate colons compared to the rest of the models with a slight decrease on the mesh quality metrics compared DeepSDF and after the correction module is the least train-biased and generates the most unseen test anatomies according to the nearest neighbor analysis.

Future work will include noise and artifact reduction methods in 3D generation, investigation of model memorization and anti-memorization techniques in 3D space generation. Further labeling of the HQColon dataset by trained gastroenterologists, regarding both pathological labeling and morphological labeling, as well as specialized loss functions (not targeting only reconstruction) and the involvement of the colonic centerline data (taking advantage that the colon is a tubular structure with a very complex distribution in a 3D space) may also help INR-based methods better learn the overall healthy and pathological morphology of human colons. The colonic and overall gastrointestinal morphological data are very well documented in medical literature throughout decades. Machine learning models can also be given the ability to ingest these ground truth measurements to provide better outputs or employ post-processing steps to refine their outputs. Pathological labeling will also make the reconstruction task more meaningful as the proposed WNA Cross-attention framework and many conditional generation frameworks will not only be able to reconstruct the colons but also deform them to create several patient-specific virtual variations with one or more pathological conditions of the same specific colon. This ability would be highly regarded for the development of a patient specific digital twins of the colon or other organs. This paper concludes that the present INR-based methodologies are exceptional at characterizing, generating and deforming geometrically compact and carefully labeled organs, but struggle when characterizing complex morphologies when these morphologies have not been segmented in simpler subregions.

Code Availability. The code used in this study is available upon request from the authors.

Data Availability. This study relies exclusively on the publicly available CHD and HQColon datasets by Kong et al. (2024) and Finocchiaro et al. (2026) respectively. Such data are fully de-identified. As no new data involving human participants were collected, no additional ethical approval or participant consent was required.

Acknowledgments. This project has received funding from the European Union's HORIZON TMA MSCA Doctoral Networks programme (HORIZON-MSCA-2023-DN-01) under Grant Agreement No. 101169012 (INTELLI-INGEST Project).

Disclosure of Interests. The authors have no competing interests in the present paper.

References

1. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 165–174 (2019). https://doi.org/10.1109/CVPR.2019.00025.
2. Sun, J.-M., Wu, T., Gao, L.: Recent advances in implicit representation-based 3D shape generation. Vis. Intell. 2, 9 (2024). https://doi.org/10.1007/s44267-024-00042-1.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. ukasz, Polosukhin, I.: Attention is All you Need. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. (2017).
4. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10674–10685 (2022). https://doi.org/10.1109/CVPR52688.2022.01042.
5. Chen, M., Mei, S., Fan, J., Wang, M.: Opportunities and challenges of diffusion models for generative AI. Natl Sci Rev. 11, nwae348 (2024). https://doi.org/10.1093/nsr/nwae348.
6. Kong, F., Stocker, S., Choi, P.S., Ma, M., Ennis, D.B., Marsden, A.L.: SDF4CHD: Generative modeling of cardiac anatomies with congenital heart defects. Medical Image Analysis. 97, 103293 (2024). https://doi.org/10.1016/j.media.2024.103293.
7. Sun, S., Han, K., Kong, D., Tang, H., Yan, X., Xie, X.: Topology-Preserving Shape Reconstruction and Registration via Neural Diffeomorphic Flow. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20813–20823 (2022). https://doi.org/10.1109/CVPR52688.2022.02018.
8. Guillard, B., Habermann, M., Theobalt, C., Fua, P.: A Latent Implicit 3D Shape Model for Multiple Levels of Detail. In: Cremers, D., Lähner, Z., Moeller, M., Nießner, M., Ommer, B., and Triebel, R. (eds.) Pattern Recognition. pp. 37–52. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-85187-2_3.

9. Biffi, C., Cerrolaza, J.J., Tarroni, G., Bai, W., de Marvao, A., Oktay, O., Ledig, C., Le Folgoc, L., Kamnitsas, K., Doumou, G., Duan, J., Prasad, S.K., Cook, S.A., O'Regan, D.P., Rueckert, D.: Explainable Anatomical Shape Analysis Through Deep Hierarchical Generative Models. IEEE Transactions on Medical Imaging. 39, 2088–2099 (2020). https://doi.org/10.1109/TMI.2020.2964499.
10. Sander, J., de Vos, B.D., Bruns, S., Planken, N., Viergever, M.A., Leiner, T., Išgum, I.: Reconstruction and completion of high-resolution 3D cardiac shapes using anisotropic CMRI segmentations and continuous implicit neural representations. Computers in Biology and Medicine. 164, 107266 (2023). https://doi.org/10.1016/j.compbiomed.2023.107266.
11. Yang, J., Wickramasinghe, U., Ni, B., Fua, P.: ImplicitAtlas: Learning Deformable Shape Templates in Medical Imaging. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15840–15850 (2022). https://doi.org/10.1109/CVPR52688.2022.01540.
12. de Wilde, B., Rietberg, M.T., Lajoinie, G., Wolterink, J.M.: Steerable Anatomical Shape Synthesis with Implicit Neural Representations. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., and Park, J. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 630–639. Springer Nature Switzerland, Cham (2026). https://doi.org/10.1007/978-3-032-04947-6_60.
13. Zhang, G., Yang, J., Chen, S., Zhang, A., Li, Y.: High-Fidelity Medical Shape Generation via Skeletal Latent Diffusion, http://arxiv.org/abs/2603.07504, (2026). https://doi.org/10.48550/arXiv.2603.07504.
14. Sinha, A., Hamarneh, G.: TrIND: Representing Anatomical Trees by Denoising Diffusion of Implicit Neural Fields. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., and Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 344–354. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72390-2_33.
15. Qiao, M., McGurk, K.A., Wang, S., Matthews, P.M., O'Regan, D.P., Bai, W.: A personalized time-resolved 3D mesh generative model for unveiling normal heart dynamics. Nat Mach Intell. 7, 800–811 (2025). https://doi.org/10.1038/s42256-025-01035-5.
16. Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S., Stegmaier, J., Kuhl, C., Nebelung, S., Kather, J.N., Truhn, D.: Denoising diffusion probabilistic models for 3D medical image generation. Sci Rep. 13, 7303 (2023). https://doi.org/10.1038/s41598-023-34341-2.
17. Wickramasinghe, U., Remelli, E., Knott, G., Fua, P.: Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., and Joskowicz, L. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 299–308. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_30.

18. Zhao, D., Mao, S., Shao, J., Huang, H.: Generative adversarial networks for high-fidelity 3D point cloud completion. Sci Rep. 16, 14076 (2026). https://doi.org/10.1038/s41598-026-44111-5.
19. Triantafyllou, G., Dimas, G., Kalozoumis, P.G., Iakovidis, D.K.: Wave-Shaping Neural Activation for Improved 3D Model Reconstruction from Sparse Point Clouds. In: Blanc-Talon, J., Delmas, P., Philips, W., and Scheunders, P. (eds.) Advanced Concepts for Intelligent Vision Systems. pp. 172–183. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-45382-3_15.
20. Chen, Z., Zhang, Y., Genova, K., Fanello, S., Bouaziz, S., Häne, C., Du, R., Keskin, C., Funkhouser, T., Tang, D.: Multiresolution Deep Implicit Functions for 3D Shape Representation. Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV) October 1 (2021). https://doi.org/10.1109/ICCV48922.2021.01284.
21. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. 41, 102:1-102:15 (2022). https://doi.org/10.1145/3528223.3530127.
22. Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. pp. 7462–7473. Curran Associates Inc., Red Hook, NY, USA (2020).
23. Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., Teh, Y.W.: Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning. pp. 3744–3753. PMLR (2019).
24. Finocchiaro, M., Stern, R., Vilhelmsborg, R., Smith, A.G., Petersen, J., Cold, K., Konge, L., Erleben, K., Ganz, M.: Clinically validated dataset of 435 human colons segmented from CT colonography. Sci Data. 13, 199 (2026). https://doi.org/10.1038/s41597-025-06518-z.
25. Fan, H., Su, H., Guibas, L.: A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2463–2471 (2017). https://doi.org/10.1109/CVPR.2017.264.
26. Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L.: Learning Representations and Generative Models for 3D Point Clouds. (2018).
27. Sajjadi, M.S.M., Bachem, O., Lucic, M., Bousquet, O., Gelly, S.: Assessing Generative Models via Precision and Recall. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. (2018).
28. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., Aila, T.: Improved Precision and Recall Metric for Assessing Generative Models. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. (2019).
29. Gadelmawla, E.S., Koura, M.M., Maksoud, T.M.A., Elewa, I.M., Soliman, H.H.: Roughness parameters. Journal of Materials Processing Technology. 123, 133–145 (2002). https://doi.org/10.1016/S0924-0136(02)00060-2.

30. Váša, L., Rus, J.: Dihedral Angle Mesh Error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. Computer Graphics Forum. 31, 1715–1724 (2012). https://doi.org/10.1111/j.1467-8659.2012.03176.x.
31. Zhang, Y., Bajaj, C., Sohn, B.-S.: 3D finite element meshing from imaging data. Computer Methods in Applied Mechanics and Engineering. 194, 5083–5106 (2005). https://doi.org/10.1016/j.cma.2004.11.026.
32. Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. pp. 1558–1566. JMLR.org, New York, NY, USA (2016).
33. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models. Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) June 1 (2022). https://doi.org/10.1109/CVPR52688.2022.01042.