LOT-Bridge: A Latent Optimal Transport Bridge Based on Signed Distance Fields for Automatic Skull Defect Reconstruction

Yanghui Yan[1], Xingce Wang[1]✉, Zhongke Wu[1], Xiaodong Ju[2]✉, Jingyi Liu[3], Changkai Ji[4], Yicheng Zhu[5], and Wuyang Shui[6]

[1] School of Artificial Intelligence, Beijing Normal University, Beijing, China
[2] Institute of Sports Medicine, Peking University Third Hospital, Beijing, China
[3] Department of Radiology, Peking Union Medical College Hospital, Beijing, China
[4] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
[5] Department of Neurology, Peking Union Medical College Hospital, Beijing, China
[6] Alder Hey Children's Hospital, Liverpool, UK
✉Corresponding Authors: `wangxingce@bnu.edu.cn` `juxiaodong@puh3.net.cn`

Abstract. Due to the reliance on manually designed implants, traditional skull defect reconstruction is time-consuming. Although latent diffusion models based on Signed Distance Fields (SDFs) accelerate the implant design, Gaussian noise initialization and stochastic differential equations lead to many iterations and may introduce incorrect geometric details. Moreover, separately training the decoder and latent-space network can accumulate errors across stages. To address these limitations, a Latent Optimal Transport Bridge (LOT-Bridge) based on SDFs is proposed. We model automatic skull reconstruction as an Optimal Transport (OT) problem in the latent and data spaces, and design an efficient training strategy to achieve accurate generation with fewer iterations. Firstly, to accelerate inference and improve the determinism of generation, we introduce a latent bridge model, where the reverse sampling begins from the defective data and follows an ordinary differential equation. Secondly, a joint training strategy performs OT-based optimization in the latent and data spaces while ensuring high-quality implant SDF prediction. Finally, the adaptive morphological block refines SDF features through dilation and erosion, enhancing surface reconstruction. Extensive experiments on three datasets demonstrate that LOT-Bridge generates more accurate implants with faster inference than existing methods. Code: https://github.com/yanaynghui1998/LOT-Bridge.

Keywords: Automatic Skull Defect Reconstruction · Generative Model · Signed Distance Fields · Optimal Transport.

1 Introduction

The automatic skull defect reconstruction is crucial for patients with accidental injuries or brain surgery. However, due to the significant heterogeneity between cranial defects, personalized implant design is necessary [8]. This process is generally performed manually by software, which is costly and time-consuming [18].

To accelerate the implant design, Deep Learning (DL) methods represent shapes as binary volumes [13, 26, 24] or point clouds [25, 9] and train a 3D network [16, 20] for end-to-end shape completion. Recently, several studies [4, 3] combine Signed Distance Fields (SDFs) with diffusion models for shape completion. SDFs provide richer surface descriptions than other representations and can be directly converted into meshes [1, 27]. Diffusion models avoid overfitting and yield more accurate shapes than end-to-end DL models [4, 19]. Latent Diffusion Models (LDMs) [21, 2] perform reverse sampling in the latent space, reducing the high memory cost caused by voxelized SDFs and enabling fast generation. However, due to Gaussian noise initialization, LDMs still require lengthy iterations to fit the high-quality complete skull. Moreover, stochastic differential equations make LDMs less deterministic [5, 15], which may produce distorted geometric details. Finally, most LDMs directly adopt the decoder pre-trained at the first stage to achieve the final reconstruction. This accumulates approximation errors from different training stages, resulting in suboptimal completion performance.

To address these problems, we propose a Latent Optimal Transport Bridge (LOT-Bridge) based on SDFs for automatic skull defect reconstruction. Our core idea is to model skull reconstruction as an Optimal Transport (OT) problem in both latent and data spaces and solve it with an effective model and training strategy. Main contributions are as follows: (1) Based on an Ordinary Differential Equation (ODE) derived from the dynamic OT problem, we introduce a bridge model in the latent space. The bridge model learns the optimal vector field and builds OT paths between defective and complete samples, achieving fast and deterministic inference. (2) A joint training strategy is designed to simultaneously optimize the bridge model and the decoder. Specifically, the adversarial and reconstruction losses enforce OT between the reconstructed and complete skulls in data space. The implant loss enhances implant SDF prediction. Combining them with the latent-space loss, LOT-Bridge captures a high-quality generative ability. (3) The Adaptive Morphological Block (AMB) is proposed and is incorporated into the decoder. AMB can adaptively perform erosion and dilation at different upsampling stages, improving the final surface reconstruction.

2 Methodology

2.1 Problem Formulation

Given paired complete and defective SDF volumes $x_0 \sim \mathcal{P}_0$ and $x_1 \sim \mathcal{P}_1(\cdot|x_0)$, and their latent codes $z_0 \sim \mathcal{Q}_0(\cdot|x_0)$ and $z_1 \sim \mathcal{Q}_1(\cdot|x_1)$, our goal is to solve the following OT problem:

$$\inf_{\{T_t(z_0)\}_{t\in[0,1]}} \frac{1}{2}\int_0^1 \left\| \frac{d}{dt} T_t(z_0) \right\|_2^2 dt + \inf_{\Gamma\in\Pi} \int_{\mathcal{X}\times\mathcal{X}} \left\| x_0 - x_{0|1} \right\|_2^2 d\Gamma(x_0, x_{0|1}) \quad (1)$$

The first term represents the dynamic OT problem in the latent space and is constrained by $T_0(z_0) = z_0$ and $T_1(z_0) = z_1$, where $\{T_t(z_0)\}_{t\in[0,1]}$ is the OT path

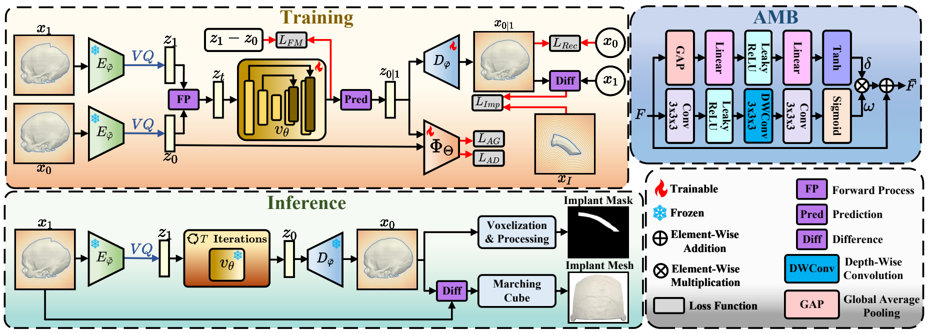

Fig. 1. Overview of LOT-Bridge and AMB. $\tilde{\varphi}$, $\bar{\varphi}$, θ and Θ are network parameters

from z_0 to z_1, and $dT_t(z_0)/dt$ is the optimal vector field. We consider this OT problem only with fixed endpoints. The second term is the OT problem in the data space $\mathcal{X}$, where $x_{0|1} \sim \mathcal{P}_{0|1}(\cdot \mid z_{0|1})$ is the final reconstructed SDF volume, $z_{0|1}$ is the complete latent code generated by the latent-space model, and Π is the set of all joint probability distributions with marginals $\mathcal{P}_0$ and $\mathcal{P}_{0|1}$. We refer to $z_{0|1}$ and $x_{0|1}$ as the fake complete latent code and SDF volume.

2.2 Latent Optimal Transport Bridge

The overview of LOT-Bridge is shown in Fig. 1. **During training**, the encoder $E_{\tilde{\varphi}}$, the quantizer VQ, and the decoder $D_{\bar{\varphi}}$ are initialized from a pre-trained Vector Quantized Variational Autoencoder (VQ-VAE). The complete and defective SDF volumes x_0 and x_1 are encoded and quantized into latent codes z_0 and z_1. The 3D U-Net v_θ receives the forward result z_t (Sec. 2.2.2) and approximates $z_1 - z_0$ with the flow matching loss L_{FM} (3). v_θ predicts the fake complete latent code $z_{0|1}$ (Sec. 2.2.3), which is decoded to the fake complete SDF volume $x_{0|1}$ for computing the reconstruction loss L_{Rec} (4). By the difference operation, the implant SDF is obtained and is enforced to match the ground-truth implant x_I using the implant loss L_{Imp} (7). Finally, z_0 and $z_{0|1}$ are fed into the discriminator Φ_Θ to calculate the adversarial losses L_{AG} (5) and L_{AD} (6). L_{FM}, L_{Rec}, L_{AG}, and L_{Imp} jointly optimize v_θ and $D_{\bar{\varphi}}$. Subsequently, L_{AD} updates Φ_Θ.

During inference, the trained v_θ transforms z_1 to z_0 over T iterations, and the trained $D_{\bar{\varphi}}$ upsamples z_0 to x_0. Through voxelization and post-processing, the binary implant mask is produced. Specifically, x_0 is first voxelized into a binary mask by setting SDF values > 0 to 0 and ≤ 0 to 1. The Boolean difference operation is then performed between this mask and the defective mask to obtain the implant mask. Finally, the opening operation and the connected component analysis remove noise and retain the largest connected component. Moreover, the implant SDF can be computed directly by the difference operation (7) between x_0 and x_1, from which the implant mesh is extracted using the marching cube.

VQ-VAE with Adoptive Morphological Block A variant of VQ-VAE is designed, where the encoder and the quantizer follow [2]. For the decoder, we add AMB after each upsampling layer to enhances the final surface reconstruction. As shown in Fig. 1, AMB consists of two branches. The upper branch selects erosion or dilation, while the lower branch controls the intensity. Given an upsampled feature $F \in \mathbb{R}^{B\times C\times D\times H\times W}$, the upper branch compresses F into a factor $\delta \in \mathbb{R}^{B\times 1\times 1\times 1\times 1}$, which is constrained within $[-1, 1]$ by a Tanh activation. As δ changes, the boundary of the SDF features shrinks or expands, performing erosion or dilation. The lower branch reduces the channel number of F to 1 and employs a Sigmoid activation to obtain a positive scaling map $\omega \in \mathbb{R}^{B\times 1\times D\times H\times W}$. The morphological operation of AMB can be defined as:

$$\bar{F} = F + \delta\omega \quad \delta \begin{cases} > 0 & \text{Erosion} \\ < 0 & \text{Dilation} \end{cases} \tag{2}$$

where $\bar{F}$ is the output of AMB. Following [2], we train VQ-VAE only on complete SDF volumes with the reconstruction (L_2), commitment, and VQ losses.

Bridge Model in the Latent Space To achieve fast and deterministic generation, we introduce an ODE-based bridge model. Firstly, we define the forward process, which can transport the complete sample z_0 to the defective sample z_1 along the optimal vector field $dT_t(z_0)/dt$ in (1). According to the Euler–Lagrange equations [11], the constrained first term of (1) leads to an ODE: $dT_t(z_0)/dt = (z_1 - z_0)$, where $t \in [0, 1]$. The solution to the ODE is $T_t(z_0) = z_t = tz_1 + (1-t)z_0$, which is the forward process (**FP** in Fig. 1). A U-Net v_θ learns the optimal vector field with fixed direction $(z_1 - z_0)$ using L_{FM}:

$$L_{FM} = \mathbb{E}_{t\sim\mathcal{U}([0,1]), z_0\sim\mathcal{Q}_0, z_1\sim\mathcal{Q}_1}\left[\|(z_1 - z_0) - v_\theta(z_t, t)\|_2^2 \right] \tag{3}$$

Finally, the deterministic DDPM posterior [14] enables the reverse sampling, which is: $z_s = \frac{s}{t}z_t + (1 - \frac{s}{t})z_0$, where s and t are two consecutive time steps $(s < t)$. Since v_θ learns the instantaneous velocity, there is a direct estimation from z_t to z_0: $z_{0,t} = z_t - t\, v_\theta(z_t, t)$. By replacing z_0 with $z_{0,t}$ and sampling time steps at equal intervals between $[0, 1]$, the reverse sampling starts from z_1 and progressively generates z_0 in the latent space (**Inference** in Fig. 1).

Joint Training Strategy A joint training strategy is designed to enhance the generative ability of LOT-Bridge (**Training** in Fig. 1). We first realize OT in the data space by solving the second term of (1). Generating the fake complete latent code $z_{0|1}$ (Sec. 2.1) via reverse sampling leads to a lengthy training time. Therefore, based on (3), we propose a one-step prediction from z_1 to z_0 (**Pred** in Fig. 1) : $z_{0|1} = z_1 - v_\theta(z_t, t)$. Since $z_0 \sim \mathcal{Q}_0(\cdot \mid x_0)$ and $z_1 \sim \mathcal{Q}_1(\cdot \mid x_1)$, $z_{0|1}$ follows the conditional distribution $\mathcal{Q}_{0|1}(\cdot \mid x_1, x_0, t; \theta)$. According to [22], instead of optimizing over couplings Γ in (1), we minimize the transport cost between the real complete skull $x_0 \sim \mathcal{P}_0$ and the fake complete skull $x_{0|1} \sim \mathcal{P}_{0|1}(\cdot \mid z_{0|1}; \bar{\varphi})$ in the data space via L_{Rec}, while enforcing $\mathcal{Q}_{0|1}(\cdot \mid x_0, x_1, t; \theta)$ to match $\mathcal{Q}_0(\cdot \mid x_0)$ in the latent space using L_{AG} and L_{AD}. These losses are defined as follows:

$$L_{Rec} = \mathbb{E}_{x_0 \sim \mathcal{P}_0, x_{0|1} \sim \mathcal{P}_{0|1}} \left[\| x_0 - x_{0|1} \|_2^2 \right] \tag{4}$$

$$L_{\mathrm{AG}} = -\mathbb{E}_{z_{0|1} \sim \mathcal{Q}_{0|1}} \left[\log \Phi_\Theta (z_{0|1}) \right] \tag{5}$$

$$L_{AD} = \mathbb{E}_{z_0 \sim \mathcal{Q}_0} \left[\log \Phi_\Theta (z_0) \right] + \mathbb{E}_{z_{0|1} \sim \mathcal{Q}_{0|1}} \left[\log (1 - \Phi_\Theta (z_{0|1})) \right] \tag{6}$$

where $x_{0|1} = D_{\bar{\varphi}}(z_{0|1})$. LOT-Bridge cannot directly produce the implant SDF. Therefore, we apply a difference operation (**Diff** in Fig. 1) [17] to extract a pseudo implant x_{PI}, where $x_{PI} = max(x_{0|1}, -x_1)$, and max compares the SDF voxels in $x_{0|1}$ and $-x_1$ and retains the larger value. Then, L_{Imp} reduces the difference between $x_{PI} \sim \mathcal{P}_{PI}(\cdot \mid x_{0|1}, x_1)$ and the reference implant $x_I \sim \mathcal{P}_I$:

$$L_{Imp} = \mathbb{E}_{x_I \sim \mathcal{P}_I, x_{PI} \sim \mathcal{P}_{PI}} \left[\| x_I - x_{PI} \|_2^2 \right] \tag{7}$$

The total loss function L_{Total} for joint training v_θ and $D_{\bar{\varphi}}$ is defined as follows:

$$L_{Total} = L_{FM} + L_{Rec} + L_{AG} + \lambda L_{Imp} \tag{8}$$

where λ is a hyperparameter. After each update of v_θ and $D_{\bar{\varphi}}$, L_{AD} optimizes Φ_Θ. Note that all loss functions are computed using the same $t \sim \mathcal{U}([0, 1])$.

3 Experiments

Datasets and Implementation Details LOT-Bridge is evaluated on the SkullBreak, SkullFix [10], and MUG500 datasets [12]. The SkullBreak and SkullFix datasets contain paired 3D complete and defective binary masks. The SkullBreak dataset is downsampled to $256 \times 256 \times 256$ with a voxel size of $0.8 \times 0.8 \times 0.8\text{mm}^3$ and is split into 430 training and 140 testing masks. The SkullFix dataset is resampled, downsampled, and zero-padded to $256 \times 256 \times 256$ with a voxel size of $0.9 \times 0.9 \times 0.9\text{mm}^3$, and is divided into 75 training and 25 testing masks. The MUG500 dataset provides the paired defective and implant meshes. which are merged and converted to $256 \times 256 \times 256$ binary masks with a voxel size of $1.0 \times 1.0 \times 1.0\text{mm}^3$ using trimesh. The MUG500 dataset is split into 19 training and 10 testing masks. Data augmentation in [7] expands the training samples to 1290 for SkullBreak, 825 for SkullFix, and 265 for MUG500. The SDF volumes with the same size are extracted from binary masks using SimpleITK [28] and are scaled to fit $[-1, 1]^3$, and then clipped to the truncated range $[-0.3, 0.3]$.

We adopt 3D VQ-VAE in [2], the discriminator Φ_Θ in [6], and the 3D U-Net v_θ in [14]. VQ-VAE is pre-trained for 500 epochs with a learning rate of 4×10^{-6} and batch size of 4. During training, the learning rate of v_θ and the decoder $D_{\bar{\varphi}}$ is 5×10^{-8}, the learning rate of Φ_Θ is 5×10^{-6}. The batch size is 4, the training epoch is 800, and $\lambda = 0.5$ in (8). During inference, we set $T = 5$ in Fig. 1. Experiments are conducted on an NVIDIA A800 80 GB GPU with the Adam optimizer. We chose the Dice Score (DSC), the Boundary Dice score (BDSC), the 95th percentile Hausdorff Distance (HD95), and the Mean Surface Distance (MSD) as quantitative metrics. Higher DSC and BDSC, along with lower HD95 and MSD, indicate better quality of the generated implant mask.

Table 1. Quantitative comparison on the SkullBreak, SkullFix and MUG500 datasets. **Bold** and Underlined values indicate the best and second-best results. All metrics are reported as mean values.

Method	SkullBreak				SkullFix				MUG500			
	DSC↑	BDSC↑	HD95↓	MSD↓	DSC↑	BDSC↑	HD95↓	MSD↓	DSC↑	BDSC↑	HD95↓	MSD↓
PBT	0.856	0.899	2.597	0.806	0.884	0.912	2.321	0.788	0.841	0.857	1.253	0.350
SegRegNet	0.928	0.936	1.412	0.483	0.936	0.945	1.331	0.453	0.920	0.934	0.974	0.193
PCTrans	0.887	0.912	1.858	0.656	0.902	0.927	1.650	0.569	0.916	0.940	0.983	0.201
PCDiff	0.912	0.914	1.768	0.615	0.917	0.927	1.618	0.552	0.923	0.942	0.924	0.185
SDFusion	0.917	0.930	1.528	0.515	0.921	0.931	1.524	0.507	0.935	0.939	0.941	0.188
OctGPT	0.920	0.935	1.468	0.493	0.934	0.944	1.374	0.463	0.941	0.945	0.911	0.174
LOT-Bridge	**0.938**	**0.945**	**1.358**	**0.459**	**0.948**	**0.960**	**1.313**	**0.414**	**0.949**	**0.955**	**0.813**	**0.144**

3.1 Results and Discussion

Comparison Experiments LOT-Bridge is compared with six methods: two binary volume-based skull reconstruction methods: a patch-based training method (PBT [13]) and a residual UNet with symmetry networks (SegRegNet [24]), two point cloud-based skull reconstruction methods: a transformer-based method (PCTrans [25]) and a diffusion-based method (PCDiff [9]), two SDF-based shape completion and generation methods: a LDM-based method (SDFusion [2]) and an autoregressive method for latent-space generation (OctGPT [23]).

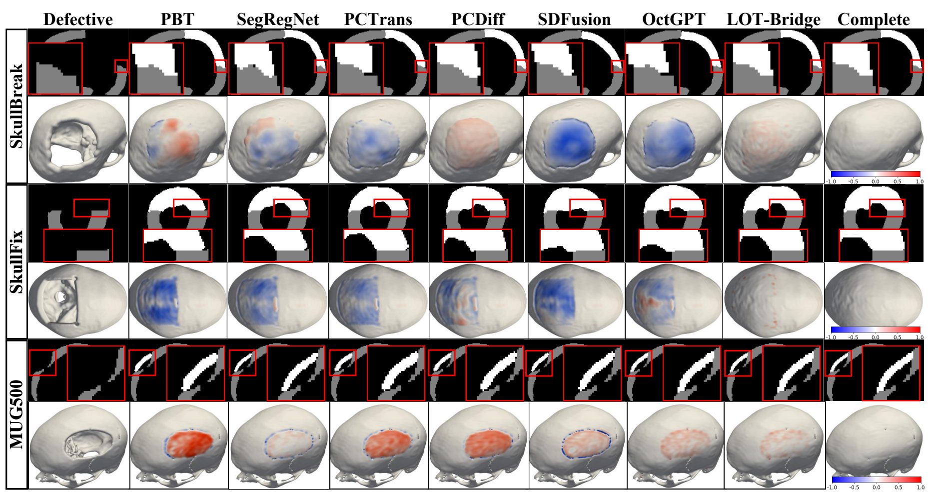

Fig. 2. Qualitative comparison on the SkullBreak, SkullFix, and MUG500 datasets. Rows 1, 3, and 5 are binary masks in the sagittal, axial, and coronal views. Rows 2, 4, and 6 are meshes. Surface distances between the predicted and ground-truth implants are shown by a blue–white–red colormap centered at zero. Please zoom in for details.

The quantitative results are shown in Table 1. LOT-Bridge obtains the best metrics on all datasets. Moreover, SDF-based methods exhibit better performance than point cloud-based methods and PBT, highlighting the superiority of the SDF representation. Due to the SDE framework and the separate training of the latent-space network and decoder, SDFusion typically underperforms SegRegNet. Compared with SegRegNet, LOT-Bridge improves BDSC by 0.015 and reduces HD95 by 0.018 mm on the SkullFix dataset. This benefits from the ODE framework, the joint training strategy, and AMB. The complexity comparisons are reported in Table 2, where Time denotes the inference time, and Param denotes the number of trainable parameters. Due to operating on volumetric data, binary volume-based and SDF-based methods generally maintain a large number of network parameters. Meanwhile, data-space methods show a slower reconstruction speed than latent-space methods. Among latent-space models, LOT-Bridge achieves the fastest inference and relatively low MACs.

We present the qualitative results in Fig. 2. For implant mask prediction, the other methods suffer from overly thick boundaries, non-smooth curvatures, and poor connectivity with the defective skull. In comparison, LOT-Bridge reconstructs implant masks closely matching the ground truth and enables an accurate transition from the implant to the defect boundary. Similarly, LOT-Bridge generates implant meshes with the smallest surface errors among all methods.

Table 2. Complexity comparison on the SkullFix datasets. All methods are tested on an NVIDIA A800 GPU. (Number) denotes the number of sampling steps, s denotes seconds, M denotes million, and T denotes Tera, where 1Tera = 1000Giga.

Method	Time(s)	Param(M)	MACs(T)
PBT	15.195	**22.863**	10.791
SegRegNet	10.720	140.354	11.350
PCTrans	10.613	57.296	**3.121**
PCDiff(1000)	1096.647	<u>50.508</u>	250.095
SDFusion(100)	10.125	552.115	131.577
OctGPT	<u>7.364</u>	417.196	17.065
LOT-Bridge(5)	**1.812**	539.081	<u>10.353</u>

Table 3. Ablation study of the joint training strategy and AMB on the SkullFix dataset.

Method	DSC↑	BDSC↑	HD95↓	MSD↓
Base	0.931	0.940	1.502	0.529
w AMB	0.937	0.946	1.430	0.497
w L_{Adv}	0.933	0.942	1.476	0.519
w L_{Rec}	0.935	0.944	1.453	0.513
w L_{Imp}	0.934	0.944	1.457	0.515
w/o AMB	0.942	0.950	1.408	0.448
w/o L_{Adv}	0.943	0.952	1.354	0.430
w/o L_{Rec}	0.941	0.948	1.374	0.437
w/o L_{Imp}	<u>0.944</u>	<u>0.953</u>	<u>1.333</u>	<u>0.424</u>
Our	**0.948**	**0.960**	**1.313**	**0.414**

Ablation Study We conduct ablation studies on the SkullFix dataset to verify the effectiveness of the joint training strategy and AMB. The results are shown in Table 3, where Base indicates directly combining the pre-trained original VQ-VAE [2] with the bridge model trained using only L_{FM} (3), Our indicates the full method, w indicates adding the key component into Base, w/o indicates removing the key component from Our, and L_{Adv} indicates using both L_{AG} (5)

and L_{AD} (6). Note that Base, w AMB, and w L_{Adv} all use the separate training for the bridge model and the decoder. As shown in Table 3 and Fig. 3(a), AMB enhances the performance of LOT-Bridge and contributes to generating the implant masks and meshes with accurate boundaries and surfaces. Moreover, w/o AMB boosts the generative ability of Base, validating the effectiveness of the joint training strategy. Due to the powerful synergy between AMB and the joint training strategy, Our attains the highest scores across all metrics.

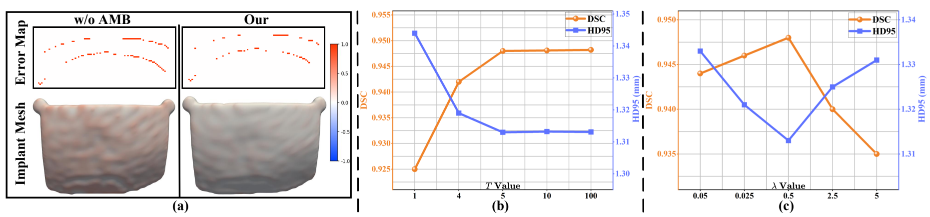

Fig. 3. (**a**) Row 1 shows the error maps between the predicted implant masks and the ground truth, and Row 2 presents the surface distances between the predicted implant meshes and the ground truth. (**b**) and (**c**) report DSC and HD95 values on the SkullFix dataset under different numbers of iterations (T) and hyperparameter settings (λ). In other experiments, we set $T = 5$ and $\lambda = 0.5$ by default. Please zoom in for details.

To assess the effect of the number of iterations, the reverse sampling is performed with iterations $T = [1, 4, 5, 10, 100]$. As shown in Fig. 3(b), LOT-Bridge reaches a performance bottleneck at $T = 5$. Therefore, by combining the bridge model, the joint training strategy, and AMB, LOT-Bridge yields high-quality skull reconstructions with fewer sampling steps. To explore the impact of the hyperparameter in (8), we set $\lambda = [0.05, 0.025, 0.5, 2.5, 5]$. As shown in Fig. 3(c), once λ exceeds 0.5, the generation accuracy of LOT-Bridge drops markedly. This suggests that L_{Imp} provides more valuable supervision when $\lambda = 0.5$.

4 Conclusion

We propose LOT-Bridge based on SDFs for automatic skull defect reconstruction. In the latent space, an ODE-based bridge model transports defective data to complete data along the deterministic OT path with fewer iterations. Meanwhile, a joint training strategy enables OT in the latent and data spaces, as well as the direct acquisition of the reliable implant SDF. To improve the surface, AMB adaptively applies erosion and dilation during upsampling, generating more accurate details on the final implants. Compared with other encoder-decoder generative models, our method is based on the OT framework. Extensive experiments have demonstrated that LOT-Bridge surpasses other methods in both reconstruction quality and inference speed, highlighting its potential for clinical implant design. Moreover, LOT-Bridge is highly scalable and can be applied to general shape completion and transformation tasks.

Acknowledgments. This work was funded in part by the National Natural Science Foundation of China (No. 62072045), in part by the Beijing Municipal Science and Technology Commission, and the Zhongguancun Science Park Management Committee (No. Z221100002722020), and in part by the Natural Science Foundation of Beijing (No. 7242167).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, Z., Tang, J., Dong, Y., Cao, Z., Hong, F., Lan, Y., Wang, T., Xie, H., Wu, T., Saito, S., et al.: 3DTopia-XL: Scaling high-quality 3D asset generation via primitive diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 26576–26586 (2025)
2. Cheng, Y.C., Lee, H.Y., Tulyakov, S., Schwing, A.G., Gui, L.Y.: SDFusion: Multimodal 3D shape completion, reconstruction, and generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4456–4465 (2023)
3. Chou, G., Bahat, Y., Heide, F.: Diffusion-sdf: Conditional generative modeling of signed distance functions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2262–2272 (2023)
4. Chu, R., Xie, E., Mo, S., Li, Z., Nießner, M., Fu, C.W., Jia, J.: Diffcomplete: Diffusion-based generative 3D shape completion. In: Advances in Neural Information Processing Systems. vol. 36, pp. 75951–75966 (2023)
5. Chung, H., Kim, J., Ye, J.C.: Direct diffusion bridge using data consistency for inverse problems. In: Advances in Neural Information Processing Systems. vol. 36, pp. 7158–7169 (2023)
6. Corona-Figueroa, A., Bond-Taylor, S., Bhowmik, N., Gaus, Y.F.A., Breckon, T.P., Shum, H.P., Willcocks, C.G.: Unaligned 2D to 3D translation with conditional vector-quantized code diffusion using transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14585–14594 (2023)
7. Ellis, D.G., Aizenberg, M.R.: Deep learning using augmentation via registration: 1st place solution to the autoimplant 2020 challenge. In: Cranial Implant Design Challenge, pp. 47–55. Springer (2020)
8. Fishman, Z., Mainprize, J.G., Edwards, G., Antonyshyn, O., Hardisty, M., Whyne, C.M.: Thickness and design features of clinical cranial implants—what should automated methods strive to replicate? International Journal of Computer Assisted Radiology and Surgery **19**(4), 747–756 (2024)
9. Friedrich, P., Wolleb, J., Bieder, F., Thieringer, F.M., Cattin, P.C.: Point cloud diffusion models for automatic implant generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 112–122. Springer (2023)
10. Kodym, O., Li, J., Pepe, A., Gsaxner, C., Chilamkurthy, S., Egger, J., Španěl, M.: SkullBreak/SkullFix–Dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks. Data in Brief **35**, 106902 (2021)
11. Kornilov, N., Mokrov, P., Gasnikov, A., Korotin, A.: Optimal flow matching: Learning straight trajectories in just one step. In: Advances in Neural Information Processing Systems. vol. 37, pp. 104180–104204 (2024)

12. Li, J., Krall, M., Trummer, F., Memon, A.R., Pepe, A., Gsaxner, C., Jin, Y., Chen, X., Deutschmann, H., Zefferer, U., et al.: Mug500+: Database of 500 high-resolution healthy human skulls and 29 craniotomy skulls and implants. Data in Brief **39**, 107524 (2021)
13. Li, J., Von Campe, G., Pepe, A., Gsaxner, C., Wang, E., Chen, X., Zefferer, U., Tödtling, M., Krall, M., Deutschmann, H., et al.: Automatic skull defect restoration and cranial implant generation for cranioplasty. Medical Image Analysis **73**, 102171 (2021)
14. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E., Nie, W., Anandkumar, A.: I2SB: Image-to-Image Schrödinger Bridge. In: International Conference on Machine Learning. pp. 22042–22062 (2023)
15. Liu, X., Gong, C., et al.: Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In: International Conference on Learning Representations. pp. 23666–23698 (2023)
16. Mahdi, H., Clement, A., Kim, E., Fishman, Z., Whyne, C.M., Mainprize, J.G., Hardisty, M.R.: A u-net based system for cranial implant design with pre-processing and learned implant filtering. In: Cranial Implant Design Challenge, pp. 63–79. Springer (2021)
17. Marschner, Z., Sellán, S., Liu, H.T.D., Jacobson, A.: Constructive solid geometry on neural signed distance fields. In: ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia. pp. 1–12 (2023)
18. Memon, A.R., Shi, H., Memon, T.R., Egger, J., Chen, X.: Deep learning-based automatic cranial implant design through direct defect shape prediction and its comparison study. Medical & Biological Engineering & Computing **63**(9), 2815–2826 (2025)
19. Nunes, L., Marcuzzi, R., Mersch, B., Behley, J., Stachniss, C.: Scaling diffusion models to real-world 3D lidar scene completion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14770–14780 (2024)
20. Pathak, S., Sindhura, C., Gorthi, R.K.S.S., Kiran, D.V., Gorthi, S.: Cranial implant design using V-Net based region of interest reconstruction. In: Cranial Implant Design Challenge, pp. 116–128. Springer (2021)
21. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
22. Tolstikhin, I., Bousquet, O., Gelly, S., Schölkopf, B.: Wasserstein Auto-Encoders. In: International Conference on Learning Representations. pp. 197–212 (2018)
23. Wei, S.T., Wang, R.H., Zhou, C.Z., Chen, B., Wang, P.S.: Octgpt: Octree-based multiscale autoregressive models for 3D shape generation. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers. pp. 1–11 (2025)
24. Wodzinski, M., Daniol, M., Hemmerling, D.: Automatic skull reconstruction by deep learnable symmetry enforcement. Computer Methods and Programs in Biomedicine **263**, 108670 (2025)
25. Wodzinski, M., Daniol, M., Hemmerling, D., Socha, M.: High-resolution cranial defect reconstruction by iterative, low-resolution, point cloud completion transformers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 333–343. Springer (2023)
26. Wodzinski, M., Kwarciak, K., Daniol, M., Hemmerling, D.: Improving deep learning-based automatic cranial defect reconstruction by heavy data augmentation: from image registration to latent diffusion models. Computers in Biology and Medicine **182**, 109129 (2024)

27. Xiang, H., Li, X., Cheng, K., Lai, X., Zhang, W., Liao, Z., Zeng, L., Liu, X.: Gaussianroom: Improving 3D gaussian splatting with sdf guidance and monocular cues for indoor scene reconstruction. In: 2025 IEEE International Conference on Robotics and Automation. pp. 2686–2693. IEEE (2025)
28. Yaniv, Z., Lowekamp, B.C., Johnson, H.J., Beare, R.: Simpleitk image-analysis notebooks: a collaborative environment for education and reproducible research. Journal of Digital Imaging **31**(3), 290–303 (2018)