DMCVR: Morphology-Guided Diffusion Model for 3D Cardiac Volume Reconstruction

Xiaoxiao He[1][0000−0003−4581−0712], Chaowei Tan[2], Ligong Han[1], Bo Liu[3], Leon Axel[4], Kang Li[5], and Dimitris N. Metaxas[1]⋆

[1] Department of Computer Science, Rutgers University
[2] FocusAI Inc.
[3] Walmart Global Tech
[4] School of Medicine, New York University
[5] West China Biomedical Big Data Center, Sichuan University West China Hospital

Abstract. Accurate 3D cardiac reconstruction from cine magnetic resonance imaging (cMRI) is crucial for improved cardiovascular disease diagnosis and understanding of the heart's motion. However, current cardiac MRI-based reconstruction technology used in clinical settings is 2D with limited through-plane resolution, resulting in low-quality reconstructed cardiac volumes. To better reconstruct 3D cardiac volumes from sparse 2D image stacks, we propose a morphology-guided diffusion model for 3D cardiac volume reconstruction, DMCVR, that synthesizes high-resolution 2D images and corresponding 3D reconstructed volumes. Our method outperforms previous approaches by conditioning the cardiac morphology on the generative model, eliminating the time-consuming iterative optimization process of the latent code, and improving generation quality. The learned latent spaces provide global semantics, local cardiac morphology and details of each 2D cMRI slice with highly interpretable value to reconstruct 3D cardiac shape. Our experiments show that DMCVR is highly effective in several aspects, such as 2D generation and 3D reconstruction performance. With DMCVR, we can produce high-resolution 3D cardiac MRI reconstructions, surpassing current techniques. Our proposed framework has great potential for improving the accuracy of cardiac disease diagnosis and treatment planning. Code can be accessed at https://github.com/hexiaoxiao-cs/DMCVR.

Keywords: Diffusion model · 3D Reconstruction · Generative model

1 Introduction

Medical imaging technology has revolutionized the field of cardiac disease diagnosis, enabling the assessment of both cardiac anatomical structures and motion, including the creation of 3D models of the heart [5]. Cardiac cine magnetic resonance imaging (cMRI) [16,20] is widely used in clinical diagnosis [14], allowing for non-invasive visualization of the heart in motion with detailed information

---

⋆ Corresponding Author

Fig. 1. (a) demonstrates the limitations of cardiac cMRI. The white line in the short axis (SAX) image is the location of 2 chamber (2ch) long axis (LAX) image slice and vice versa. The grey images indicate the missing slices which are not captured during the MRI scan. (b) is an overview of our DMCVR architecture. The SAX images x_0 are first encoded to global semantic ℓ_{sem}, regional morphology ℓ_{mor} and stochastic latent codes x_T, followed by interpolation in their respective latent space. The reconstructed images are sampled from a forward denoising diffusion implicit model (DDIM) process conditioned on the three latent codes. Finally, the 3D cardiac model is reconstructed via stacking the labels. The red, green, and blue regions represent the left ventricle cavity (LVC), left ventricle myocardium (LVM), and right ventricle cavity (RVC), respectively.

on cardiac function and anatomy [17]. While cMRI has great potential in helping doctors understand and analyze cardiac function [9,15], the imaging technique has certain drawbacks including low through-plane resolution to accommodate for the limited scanning time, as visualized in Fig. 1. Recently, researchers have approached the problem of cardiac volume reconstruction with learning-based generative models [2]. However, most of the methods suffer from low generation quality, missing key cardiac structures and long generation times. This paper focuses on improving the cardiac model generation quality, while reducing the generation time, aiming to better reconstruct the missing structure of the cardiac model from low through-plane resolution cMRI.

Conventional 3D cardiac modeling [12] consists of 2D cardiac image segmentation followed by 3D cardiac volume reconstruction. Recent advances in deep learning methods have shown great success in medical image segmentation [4,6,11,23]. After obtaining 2D labels, the neighboring labels are stacked to reconstruct the 3D model. Nevertheless, due to the low inter-slice spatial cMRI resolution, a significant amount of structural information is lost in the resulting

3D volume. Thus, the interpolation between cMRI slices is necessary. Traditional intensity-based interpolation methods often yield blurring effects and unrealistic results. Conventional deformable model-based method [13] does not need consistency across images of the corresponding cardiac structures, but requires image-based structure segmentation which is nontrivial and hinders their ability to generalize. To overcome these limitations, an end-to-end pipeline based on generative adversarial networks (GANs), DeepRecon, was recently proposed in [2] that utilizes the latent space to interpolate the missing information between adjacent 2D slices. The generative network is first trained and a semantic image embedding in the $\mathcal{W}^+$ space [1] is computed. Evidently, the acquired semantic latent code is not optimal and needs iterative optimization with segmentation information for improving image qualities. However, even with the optimization step, the generated images still miss details in the cardiac region, which indicates the $\mathcal{W}^+$ space DeepRecon found does not represent the heart accurately.

In order to eliminate the step for optimizing the latent code and improve the image generation quality, we propose a morphology-guided diffusion-based 3D cardiac volume reconstruction method that improves the axial resolution of 2D cMRIs through global semantic and regional morphology latent code interpolation as indicated in Fig. 1. Inspired by [19], we utilize the global semantic latent code to encode the image into a high-level meaningful representation of the image. To improve the cardiac volume reconstruction, our approach needs to focus on the cardiac region. Therefore, we introduce the regional morphology latent code which represents the shapes and locations of LVC, LVM and RVC, which will help generating the cardiac region. The method consists of three parts: an implicit diffusion model, a global semantic encoder and a segmentation network that encodes an image to regional morphology embeddings. The proposed method does not require iteratively fine-tuning the latent codes. Our contributions are: 1) the first diffusion-based method for 3D cardiac volume reconstruction, 2) introducing the local morphology-based latent code for improved conditioning on the image generation process, 3) 8% improvement of left ventricle myocardium (LVM) segmentation accuracy and 35% improvement of structural similarity index compared to previous methods, and 4) improved efficiency by eliminating the iterative step for optimizing the latent code.

2   Methods

Fig. 2 demonstrates the structure of our DMCVR approach that learns the global semantic, regional morphology, and stochastic latent spaces from MR images to yield a broad range of outcomes, including generation of high-quality 2D image and high-resolution 3D reconstructed volume. In this section, we will first describe the architecture of our DMCVR method and then elaborate on the latent space-based 3D volume generation which enables 3D volume reconstruction.

Fig. 2. On the left side, we demonstrates the network structure of the DMCVR, which consists of an global semantic encoder, a regional morphology encoder/decoder and a conditional DDIM. The right side shows the visualization of the stochastic latent space sampled from a high-dimensional Gaussian distribution $\mathcal{N}(0, I)$.

2.1   DMCVR Architecture

Our DMCVR is composed of a global semantic encoder E_{sem}, a regional morprphology network (E_{mor}, D_{mor}) and a diffusion-based generator G. The generating process G is defined as follows: given input $x_T, \ell_{sem}, \ell_{mor}$, which are the stochastic, global semantic and regional morphology latent codes, we want to reconstruct the image x_0 recursively as follows:

$$x_{t-1} = \sqrt{\alpha_{t-1}} f_\theta(x_t, t, \ell_{sem}, \ell_{mor}) + \sqrt{1 - \alpha_{t-1}} \epsilon_\theta(x_t, t, \ell_{sem}, \ell_{mor}), \quad (1)$$

where $\epsilon_\theta(x_t, t, \ell_{sem}, \ell_{mor})$ is the noise prediction network and f_θ is defined as removing the noise from x_t or Tweedie's formula [3]:

$$f_\theta(x_t, t, \ell_{sem}, \ell_{mor}) = \frac{1}{\sqrt{\alpha_t}}(x_t - \sqrt{1 - \alpha_t} \epsilon_\theta(x_t, t, \ell_{sem}, \ell_{mor})) \quad (2)$$

Here, the term α_t is a function of t affecting the sampling quality.

The forward diffusion process takes the noise x_T as input and produces x_0 the target image. Since the change in x_T will affect the details of the output images, we can treat x_T as the stochastic latent code. Therefore, finding the correct stochastic latent code is crucial for generating image details. Thanks to DDIM proposed by Song *et al.* [21], it is possible to get x_T in a deterministic fashion by running the generative process backwards to obtain the stochastic latent code x_T for a given image x_0. This process is viewed as a stochastic encoder $x_T = E_{sto}(x_0, \ell_{sem}, \ell_{mor})$, which is conditioned on ℓ_{sem} and ℓ_{mor}. This conditioning helps us to remove the iterative optimization step used by previous method. We formulate the inversion process from x_0 to x_T as follows:

$$x_{t+1} = \sqrt{\alpha_{t+1}} f_\theta(x_t, t, \ell_{sem}, \ell_{mor}) + \sqrt{1 - \alpha_{t+1}} \epsilon_\theta(x_t, t, \ell_{sem}, \ell_{mor}) \quad (3)$$

Although using the stochastic latent variables we are able to reconstruct the image accurately, the stochastic latent space does not contain interpolatable high-level semantics. Here we utilize a semantic encoder proposed by Preechakul *et al.* [19] to encode the global high-level semantics into a descriptive vector for conditioning the diffusion process, similar to the style vector in StyleGAN [10]. The global semantic encoder utilizes the first half of the UNet, and is trained end-to-end with the conditional diffusion model.

One drawback of the global semantic encoder is that it encodes the general high-level features, but tends to pay little attention to the cardiac region. This is due to the relatively small area of LVC, LVM and RVC in the cMRI slice. However, the generation accuracy of the cardiac region is crucial for the cardiac reconstruction task. For this reason, we introduce the regional morphology encoder E_{mor} that embeds the image into the latent space containing necessary information to produce the segmentation map of the target cardiac tissues. With this extra morphology information, we are able to guide the generative model to focus on the boundary of the ventricular cavity and myocardium region, which will produce increased image accuracy in the cardiac region and the downstream segmentation task. Here, we do not assume any particular architecture for the segmentation network. However, in our experiments, we utilize the segmentation network MedFormer proposed by Gao *et al.* [4] for its excellent performance.

The training of DMCVR contains the training of the segmentation network and the training of the generative model. We first train the segmentation model with summation of focal loss and dice loss [4]. We utilize the simple loss introduced in [7] for training the conditional diffusion implicit model, where

$$L_{gen}(x) = \mathbb{E}_{t \sim \text{Unif}(1,T), \epsilon \sim \mathcal{N}(0,I)} ||\epsilon_\theta(x_t, t, E_{sem}(x_0), E_{mor}(x_0)) - \epsilon||_2^2. \quad (4)$$

2.2 3D volume reconstruction and latent-space-based interpolation

Due to various limitations, the gap between consecutive cardiac slices in cMRI is large, which results in an under-sampled 3D model. In order to output a smooth super-resolution cine image volume, we generate the missing slices by using the interpolated global semantic, regional morphology and stochastic latent codes. For global semantic and regional morphology latent code ℓ, since it is similar to the idea of latent code in StyleGAN, we utilize the same interpolation strategies as in the original paper between adjacent slices. Assume that $k < j - i, i < j$,

$$\ell^{i+k} = (1 - \frac{k}{j-i})\ell^i + \frac{k}{j-i}\ell^j. \quad (5)$$

For interpolating the stochastic latent variable, it is important to consider that the distribution of stochastic noise is high-dimensional Gaussian, as shown in Eq. (4). Thus, our stochastic embedding is positioned on a sphere shown in Fig. 2. Using linear interpolation on the stochastic noise deviates from the underlying distribution assumption and causes the diffusion model to generate unrealistic images. Hence, to preserve the Gaussian property of the stochastic

Table 1. Quantitative comparison among the segmentation results of the original image (Original), DeepRecon with 1k optimization steps (DeepRecon$_{1k}$), Diffusion AutoEncoder [19] (DiffAE) and our DMCVR. We use a pretrained segmentation model on images generated by different methods. All metrics are evaluated against the ground truth based on 3D SAX images.

Cardiac Region	Method	DICE ↑	VOE↓	ASD↓	HD↓	ASSD↓
All labels	Original	*0.943*	*10.730*	*0.229*	*4.056*	*0.229*
	DeepRecon$_{1k}$	0.914	15.179	0.367	5.879	0.397
	DiffAE	0.919	14.913	0.322	4.654	0.326
	DMCVR	**0.935**	**12.153**	**0.261**	**4.093**	**0.266**
LVC	Original	*0.937*	*11.579*	*0.221*	*3.156*	*0.224*
	DeepRecon$_{1k}$	0.928	12.955	0.336	4.299	0.328
	DiffAE	0.910	16.049	0.330	3.710	0.320
	DMCVR	**0.929**	**12.940**	**0.250**	**3.236**	**0.254**
LVM	Original	*0.875*	*22.082*	*0.226*	*3.140*	*0.237*
	DeepRecon$_{1k}$	0.796	33.382	0.390	5.730	0.389
	DiffAE	0.825	29.333	0.351	4.032	0.338
	DMCVR	**0.865**	**23.636**	**0.282**	**3.519**	**0.267**
RVC	Original	*0.898*	*18.187*	*0.273*	*4.458*	*0.267*
	DeepRecon$_{1k}$	0.858	23.662	0.381	6.304	0.473
	DiffAE	0.857	24.518	0.346	5.217	0.382
	DMCVR	**0.884**	**20.467**	**0.273**	**4.460**	**0.308**

latent space, we interpolate the stochastic latent codes over a unit sphere, which can be written as follows: Let $k < j - i, i < j$ and $x_T^i \cdot x_T^j = \cos\theta$,

$$x_T^{i+k} = \frac{\sin((1 - \frac{k}{j-i})\theta)}{\sin(\theta)}x_T^i + \frac{\sin(\frac{k}{j-i}\theta)}{\sin(\theta)}x_T^j. \tag{6}$$

3   Experiments

3.1   Experimental Settings

In this study we use data from the publicly available UK Biobank cardiac MRI data [18], which contains SAX and LAX cine CMR images of normal subjects. LVC, LVM and RVC are manually annotated on SAX images at the end-diastolic (ED) and end-systolic (ES) cardiac phases. We use 808 cases containing 484,800 2D SAX MR slices for training and 200 cases containing 120,000 2D images for testing. To evaluate the 3D volume reconstruction performance, we randomly choose 50 testing 2D LAX cases to evaluate the 3D volume reconstruction task. All models are implemented on PyTorch 1.13 and trained with 4×RTX8000.

3.2   Evaluation of the 2D slice generation quality

We provide peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [8] to evaluate the similarity between the generated images and

Fig. 3. 2D and 3D visualization results of the generated images and segmentation. (a,e) original image, (b,f) DeepRecon$_{1k}$, (c,g) DiffAE [19], (d,h) our proposed DMCVR.

the original images. In addition to image quality assessment, we want to consider the segmentation performance on the generated images by using a segmentation network trained on the real training data as the evaluator and segment the testing images generated by DeepRecon$_{1k}$, DiffAE which only uses the global semantic latent code as the condition on the DDIM model, and our DMCVR methods. The segmentation accuracy of the evaluator on the generated images can be viewed as a quantitative metric to represent the generation quality of the generated data compared to the cMRI data. We compare segmentation obtained based on three methods against ground truth on the SAX images in Tab. 1. The Dice coefficient (DICE), volumetric overlap error (VOE), average surface distance (ASD), Hausdorff distance (HD) and average symmetric surface distance (ASSD) [22] are reported for comparison.

Our method achieves a PSNR score of **30.504** and SSIM score of **0.982**, which is a significant improvement (35% increase in SSIM) compared to Deep-Recon (PSNR: **27.684**, SSIM: **0.724**) with 1k optimization steps. This indicates that our method generates more realistic image compared to DeepRecon. The segmentation results on the original images in Tab. 1 provide an upper bound for other results. DMCVR outperforms all other methods in every metric with an 8% increase in LVM segmentation compared to DiffRecon$_{1k}$Moreover, by comparing the DiffAE and DMCVR, the introduction of the regional morphology latent code drastically improves the generation results due to the extra information on the shape of LVC, LVM, and RVC. Fig. 3 demonstrates the original image and corresponding synthetic images. The white arrow points towards the presence of cardiac papillary muscles. As indicated in the images, DeepRecon$_{1k}$ (b) cannot effectively recover the information of the papillary muscles from the latent space. However, both diffusion-based (c,d) methods accurately synthesize the information. Our method (d) generates a cleaner image with less artifacts than (c), especially around the LV and RV regions. By comparing the yellow circled area, our method produces image closer to the ground truth compared to

Table 2. Evaluation of 3D volumetric reconstruction from the DICE score of the intersection on each LAX plane against ground truth based on 2D LAX sampled images: mean (standard deviation). Nearest Neighbor, Image-based Linear Interpolation, DeepRecon$_{1k}$ and our DMCVR method are compared.

Method	Average DICE	2ch DICE	3ch DICE	4ch DICE
Nearest Neighbor	0.780 (0.111)	0.787 (0.091)	0.793 (0.105)	0.766 (0.128)
Linear Interpolation	0.781 (0.080)	0.797 (0.051)	0.773 (0.070)	0.768 (0.102)
DeepRecon$_{1k}$	0.817 (0.097)	**0.848** (0.056)	0.802 (0.141)	0.797 (0.091)
DMCVR	**0.836 (0.052)**	0.841 **(0.042)**	**0.809 (0.069)**	**0.854 (0.043)**

Fig. 4. Visual comparison of 3D volumetric reconstruction from SAX images to LAX. Each row from top to bottom are 2ch, 3ch and 4ch images. The column from left to right represents: resampled original images using nearest neighbour (NN), resampled original labels using NN, resampled DMCVR images, resampled DMCVR labels and the corresponding LAX images.

DeepRecon$_{1k}$. Also, the white circle in Fig. 3 demonstrates the benefits of incorporating regional morphology information. Besides, the generative model used in DeepRecon$_{1k}$ needs to be trained for 14 days with additional time to iteratively optimize the latent code for each slice. Our method uses 4.8 days for training. Since DDIM inversion does not have test-time optimization as DeepRecon does, DMCVR generates images faster than DeepRecon.

3.3 Evaluation of the 3D volume reconstruction quality through latent space interpolation

In this section, we exploit the relationship between SAX and LAX images and leverage the LAX label to evaluate the volume reconstruction quality. In cardiac MRI, long axis (LAX) slices typically comprise 2-chamber (2ch), 3-chamber (3ch), and 4-chamber (4ch) views. To evaluate the performance of different interpolation methods on LAX slices, we conducted the following experiments: 1)

Nearest Neighbor resampling of short-axis (SAX) volume to each LAX view, 2) Image-based Linear Interpolation, 3) DeepRecon$_{1k}$, and 4) our DMCVR. Tab. 2 shows the computed 2D DICE score between the annotation of different LAX views and the intersection between the corresponding LAX plane and 3D reconstructed volume. Our method outperforms other methods in three categories and has only less than 1% performance degradation compared to DeepRecon$_{1k}$ but with more stable performance. Fig. 4 presents three examples for each LAX view, showing better reconstructed LAX results compared to the original images.

4   Conclusion

Integrating analysis of cMRI holds significant clinical importance in understanding and evaluating cardiac function. We propose a diffusion-model-based volume reconstruction method. Our finding shows that through an interpolatable latent space, we are able to improve the spatial resolution and produce meaningful MR images. In the future, we will consider incorporating LAX slices as part of the generation process to help refine the latent space.

References

1. Abdal, R., Qin, Y., Wonka, P.: Image2stylegan: How to embed images into the stylegan latent space? In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4432–4441 (2019)
2. Chang, Q., Yan, Z., Zhou, M., Liu, D., Sawalha, K., Ye, M., Zhangli, Q., Kanski, M., Al'Aref, S., Axel, L., et al.: Deeprecon: Joint 2d cardiac segmentation and 3d volume reconstruction via a structure-specific generative method. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV. pp. 567–577. Springer (2022)
3. Efron, B.: Tweedie's formula and selection bias. Journal of the American Statistical Association **106**(496), 1602–1614 (2011)
4. Gao, Y., Zhou, M., Liu, D., Yan, Z., Zhang, S., Metaxas, D.N.: A data-scalable transformer for medical image segmentation: architecture, model efficiency, and benchmark. arXiv preprint arXiv:2203.00131 (2022)
5. van der Geest, R.J., Reiber, J.H.: Quantification in cardiac mri. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **10**(5), 602–608 (1999)
6. He, X., Tan, C., Qiao, Y., Tan, V., Metaxas, D., Li, K.: Effective 3d humerus and scapula extraction using low-contrast and high-shape-variability mr data. In: Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging. vol. 10953, pp. 118–124. SPIE (2019)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)

8. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)

9. Isensee, F., Jaeger, P.F., Full, P.M., Wolf, I., Engelhardt, S., Maier-Hein, K.H.: Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. In: Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8. pp. 120–129. Springer (2018)

10. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)

11. Liu, D., Gao, Y., Zhangli, Q., Han, L., He, X., Xia, Z., Wen, S., Chang, Q., Yan, Z., Zhou, M., et al.: Transfusion: multi-view divergent fusion for medical image segmentation with transformers. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. pp. 485–495. Springer Nature Switzerland Cham (2022)

12. Lopez-Perez, A., Sebastian, R., Ferrero, J.M.: Three-dimensional cardiac computational modelling: methods, features and applications. Biomedical engineering online **14**, 1–31 (2015)

13. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE transactions on pattern analysis and machine intelligence **32**(12), 2262–2275 (2010)

14. Patel, R., Lim, R.P., Saric, M., Nayar, A., Babb, J., Ettel, M., Axel, L., Srichai, M.B.: Diagnostic performance of cardiac magnetic resonance imaging and echocardiography in evaluation of cardiac and paracardiac masses. The American Journal of Cardiology **117**(1), 135–140 (2016)

15. Pattynama, P.M., De Roos, A., Van der Wall, E.E., Van Voorthuisen, A.E.: Evaluation of cardiac function with magnetic resonance imaging. American heart journal **128**(3), 595–607 (1994)

16. Pelc, N.J., Herfkens, R.J., Shimakawa, A., Enzmann, D.R., et al.: Phase contrast cine magnetic resonance imaging. Magnetic resonance quarterly **7**(4), 229–254 (1991)

17. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S.E., Frangi, A.F.: A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. Magnetic Resonance Materials in Physics, Biology and Medicine **29**, 155–195 (2016)

18. Petersen, S.E., Matthews, P.M., Francis, J.M., Robson, M.D., Zemrak, F., Boubertakh, R., Young, A.A., Hudson, S., Weale, P., Garratt, S., et al.: Uk biobank's cardiovascular magnetic resonance protocol. Journal of cardiovascular magnetic resonance **18**(1), 1–7 (2015)

19. Preechakul, K., Chatthee, N., Wizadwongsa, S., Suwajanakorn, S.: Diffusion autoencoders: Toward a meaningful and decodable representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10619–10629 (2022)

20. Sechtem, U., Pflugfelder, P., Higgins, C.B.: Quantification of cardiac function by conventional and cine magnetic resonance imaging. Cardiovascular and interventional radiology **10**, 365–373 (1987)

21. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=St1giarCHLP

22. Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. BMC medical imaging **15**(1), 1–28 (2015)
23. Zhangli, Q., Yi, J., Liu, D., He, X., Xia, Z., Chang, Q., Han, L., Gao, Y., Wen, S., Tang, H., et al.: Region proposal rectification towards robust instance segmentation of biological images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV. pp. 129–139. Springer Nature Switzerland Cham (2022)