Anatomic-constrained Medical Image Synthesis via Physiological Density Sampling

Yuetan Chu[1], Changchun Yang[1], Gongning Luo[1(✉)], Zhaowen Qiu[2(✉)], and Xin Gao[1(✉)]

[1] Center of Excellence Applications (CoE) on Smart Health, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
{gongning.luo, xin.gao}@kaust.edu.sa
[2] College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China
qiuzw@nefu.edu.cn

Abstract. Despite substantial progress in utilizing deep learning methods for clinical diagnosis, their efficacy depends on sufficient annotated data, which is often limited available owing to the extensive manual efforts required for labeling. Although prevalent data synthesis techniques can mitigate such data scarcity, they risk generating outputs with distorted anatomy that poorly represent real-world data. We address this challenge through a novel integration of anatomically constrained synthesis with registration uncertainty-based refinement, termed **A**natomic-**C**onstrained medical **I**mage **S**ynthesis (ACIS). Specifically, we (1) generate the pseudo-mask via the physiological density estimation and Voronoi tessellation to represent the spatial anatomical information as the image synthesis prior; (2) synthesize diverse yet realistic image-annotation guided by the pseudo-masks, and (3) refine the outputs by registration uncertainty estimation to encourage the anatomical consistency between synthesized and real-world images. We validate ACIS for improving performance in both segmentation and image reconstruction tasks for few-shot learning. Experiments across diverse datasets demonstrate that ACIS outperforms state-of-the-art image synthesis techniques and enables models trained on only 10% or less of the total training data to achieve comparable or superior performance to that of models trained on complete datasets. The source code is publicly available at https://github.com/Arturia-Pendragon-Iris/VonoroiGeneration.

Keywords: Data augmentation· Few-shot learning · Segmentation · Image reconstruction

1 Introduction

In recent times, artificial intelligence and deep learning (DL) techniques have assumed an indispensable role in medical imaging analysis and computer-aided diagnosis (CAD) [1]. These approaches have exhibited remarkable efficacy across

diverse applications, including disease classification, automated organ reconstruction, lesion segmentation, image super-resolution, and denoising. However, the efficacy of a deep-learning model heavily relies on the availability of large-scale, high-quality datasets with corresponding meticulous annotations or ground truth [2]. Unfortunately, such datasets cannot be always easily accessible due to the substantial investments of time and labor for meticulous annotation, limited samples for rare diseases, and ethical and privacy considerations surrounding medical data [3]. This paucity of adequate training data imposes significant constraints on DL models, especially for data-intensive applications, consequently impeding clinical adoption of DL-based CAD [4].

To overcome these challenges, various data augmentation strategies have been proposed, which can be broadly classified into two overall categories [5]: original data transformation and artificial data synthesis. The former encompasses the application of one or more image-transforming techniques to existing samples, ranging from affine transformations to pixel-level gray-scale transformations [6]. While demonstrating some ability to enhance network robustness, these techniques remain fundamentally limited in their capacity to amplify generalization beyond the confines of the initial training data [7]. In contrast, techniques for image synthesis, such as generative adversarial networks (GANs) and diffusion models [8], hold promise for producing more diverse artificial samples [3]. However, only limited studies have reported the synthesis of 3D volumetric images with corresponding annotations [9]. In practice, numerous medical imaging modalities, including computed tomography (CT), and magnetic resonance imaging (MRI), involve 3D volumetric images, with segmentation frequently also performed in a 3D context [10]. Moreover, the synthesized samples can lack constraints enforcing realistic anatomy and the consistency between images and annotations [11], and may not accurately represent real-world data. Consequently, unconstrained data synthesis strategies do not reliably improve segmentation models or confer clinical utility.

To address these limitations, we propose a novel medical image synthesis approach termed **A**natomic-**C**onstrained medical **I**mage **S**ynthesis (ACIS) (Fig. 1), enabling the synthesis of new data with realistic anatomy to facilitate few-shot learning. The synthesis network is guided by our proposed pseudo-mask, which imparts anatomical priors to the synthesis process. To achieve this, we devise a physiological density estimation method and spatial sampling technique, allowing regions with abundant anatomical information to be represented by more sampling points. Voronoi tessellation [13] is then utilized to generate the pseudo-mask based on these points (Fig. 1.a). Next, we develop a multi-stage image-annotation synthesis workflow, in sequence synthesizing the pseudo-mask, annotation, and associated image (Fig. 1.b). To further improve anatomical realism, we devise a refinement module of the entire network (Fig. 1.c). The goal is that via registration between real-world and synthesized images, the correspondingly deformed real-world and synthesized annotations are maximally similar. We employ appearance uncertainty to quantify this similarity, using it as a reward for refinement.

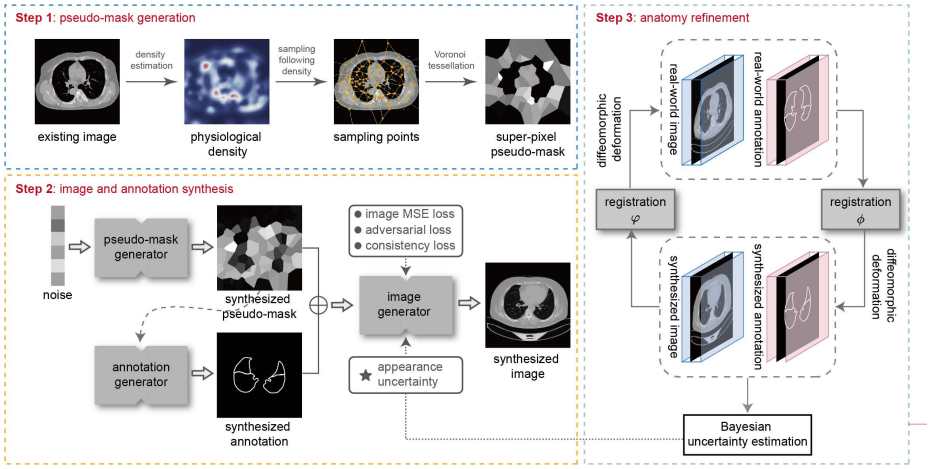

Fig. 1. The overview of our proposed ACIS framework, which consists of a pseudo-mask generation via anatomical-aware sampling (Step 1), a multi-stage image-annotation synthesis workflow (Step 2), and registration uncertainty estimation for anatomical improvement (Step 3).

We assess the efficiency of ACIS on both high-level segmentation tasks and low-level image reconstruction tasks. The validation of the latter has been limited reported thus far [14]. We select chest CT as the imaging modality for this study and utilize lung lobe segmentation to validate the benefits of employing ACIS for segmentation. Chest CT contains lots of anatomical structures, including heart, airways, vascular system, and lung fissures [12], and an optimal lung lobe segmentation relies heavily on global anatomy rather than local features [15], thus enabling an evaluation of the anatomical realism of the synthesized results. In this study, we validate the efficacy of ACIS across nine clinical tasks using two distinct datasets. Few-shot learning with ACIS significantly improves performance compared to other data synthesis techniques, approaching or exceeding models trained on full datasets. Moreover, these performance gains enhanced as the amount of available training data increased. Collectively, the experiments demonstrate that ACIS can effectively retain realistic anatomy to synthesize medical images for improving few-shot learning.

2 Method

The proposed ACIS methodology, as illustrated in Fig.1, encompasses three integral components: pseudo-mask generation via anatomical-aware sampling (Sec 2.1), a multi-stage image-annotation synthesis workflow (Sec 2.2), and registration-driven refinement (Sec 2.3). The pseudo-mask will be derived from real-world data, whereupon the synthesis workflow will first learn to generate similar pseudo-masks to provide anatomical guidance for subsequent image-annotation synthesis. After the initial training of the synthesis workflow, reg-

istration between real-world and synthesized image-annotation pairs will then be employed further to refine the consistency and realism of the synthesized anatomy. ACIS framework can be deployed on both 2D and 3D images without technical differences.

2.1 Pseudo-mask Generation via Anatomical-awareness Sampling

To address the concerns of anatomical distortion in images synthesized directly from noise, we introduce the super-pixel pseudo-mask to provide an anatomical prior to guide image synthesis. The main idea involves point-wise sampling of physiological information to represent and impart general anatomy for directing synthesis. First, we develop a physiological density estimation technique, enabling higher sampling probability in high-density physiological regions, while avoiding invalid sampling of background. Given a real-world image x, we assume that regions with significant fluctuations in image intensity, such as the hilum, spine, and vasculature, likely contain abundant physiological information necessitating intensive sampling. In contrast, smoother regions with sparser fluctuations, such as background, can be represented by fewer points. Specifically, We employ the Canny filter [16] to identify the image regions with great gradient and extract the binary edges e, then we utilize the kernel density estimation [17] to calculate the spatial density to these extracted edges:

$$f_i = \frac{1}{n}\sum_{j=1}^{n} K\left(\frac{d\left(x_i, \hat{x}_j\right)}{h}\right), \tag{1}$$

where $\hat{x}_j, j = 1, 2 \cdots n$ are n points randomly selected from e uniformly, to avoid the great computation consumption for involving all points in e. f_i is the estimated density in the point $x_i \in x$. $K(\cdot)$ is the kernel function and we use the Gaussian function $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ in our study. $d\left(x_i, x_j\right)$ denotes the Euclidean distance between x_i and x_j, while h is the bandwidth. The hyperparameter selection is described in the Supplementary Materials. Fast kernel density estimation algorithm [18] can also be employed to accelerate the calculation.

Following f as the spatial probability distribution, we randomly sample N points $p_k, k = 1, 2 \cdots N$ from x. A Voronoi tessellation [19] of the domain is then constructed based on these sampling points. Each Voronoi cell is assigned the image value of its contained sampling point:

$$I^m\left(x_i | x_i \in C\left(p_k\right)\right) = I\left(p_k\right). \tag{2}$$

I is the raw image intensity and I^m is the image intensity of the pseudo-masks. $C\left(p_k\right)$ is the Voronoi cell for the point p_k. By the defining properties of a Voronoi tessellation, each cell has exactly one sample point, and all pixels in a cell take the value of their nearest sampled point [13]. This yields a pseudo-mask that can provide the essential anatomical prior for subsequent image synthesis, without imposing excessive constraints that would limit output diversity.

2.2 Image-to-annotation Synthesis from Unsupervised Pseudo-masks

Our synthetic images are generated conditional on the pseudo-masks. To enhance the diversity of synthetic images under limited training data, we develop a technique for pseudo-mask synthesis and employ a pseudo-mask generator with a discriminator [20] as the backbone. Pseudo-masks are first derived from real-world data via the method described in Section 2.1. Owing to the randomness of sampling, our approach can produce unlimited masks, enabling the mask synthesis model to learn to generate analogous pseudo-masks as anatomical guidance. Moreover, we also implement an annotation generator to synthesize the multi-label segmentation annotations [9]. The synthesized pseudo-masks as well as the annotations will be concatenated along the channel dimension and then sent to the image generator.

We then develop a conditional pix2pix generative model [21] with an image discriminator to synthesize images and corresponding annotations from the pseudo-masks. The training comprises two phases: real-world image-based initialization and fine-tuning on generated pseudo-masks. Initially, we utilize the pseudo-masks and annotations from real-world images to jointly optimize the generator and discriminator, with an objective including a reconstruction consistency loss and the adversarial loss. The reconstruction consistency loss is implemented by the mean square error between the synthesized and the real-world image. Subsequently, we fine-tune the generator on synthesized pseudo-masks. The discriminator judges whether the synthesized image appears realistic, and evaluates the consistency between the synthesized image, pseudo-mask, and the synthesized annotation.

2.3 Refinement via Registration-based Uncertainty Estimation

To further enhance the anatomical realism of the synthesized images, we propose an innovative refinement approach. Let x and y signify the real-world image and annotation correspondingly, and $\hat{x}$ and $\hat{y}$ the synthesized counterparts. The rationale underpinning this proposition is that upon registering x to $\hat{x}$ via a deformation field ϕ as $x \circ \phi \rightarrow \hat{x}$, we encourage for the deformed annotation $y \circ \phi$ to likewise approximate $\hat{y}$. The appearance uncertainty [22] is leveraged to quantify the divergences between the two annotations post-registration.

$$\hat{\Sigma}_{app}^2 = \frac{1}{T} \sum_{t=1}^{T} \left((x \circ \phi_t - \hat{x})^2 + (\hat{x} \circ \varphi_t - x)^2 \right) \tag{3}$$

Here ϕ and φ represent the deformation field in the registration from x to $\hat{x}$, and $\hat{x}$ to x, respectively. We utilize the diffeomorphic deformation with ϕ_t and φ_t, denoting the deformation field produced by t-th forward pass across T total steps [23]. Our objective is to minimize the appearance uncertainty for the synthesized images and annotations.

However, directly utilizing the appearance uncertainty as the loss function may be impractical given both 3D generation and registration networks occupying substantial video memory and computing resources [24], rendering end-to-end training on commonly available GPUs impossible (e.g. NVIDIA RTX A6000 with 48GB memory). Therefore, we employ a reinforcement learning-driven training approach, treating the appearance uncertainty as the reward and leveraging policy gradient techniques [25] for parameter updates. We first train the framework using the methods in Sections 2.1 and 2.2; subsequently, this reinforcement learning strategy is utilized to refine the entire network.

3 Results

3.1 Datasets and Experiment Settings

To evaluate the ACIS strategy on computer-vision tasks, we leveraged two distinct datasets TLobes and PENet chest CT [26] for 3D segmentation and image reconstruction evaluation, respectively. The TLobes is an in-house dataset, providing annotations of the five lung lobes and heart; the PENet dataset is a public dataset consisting of CT pulmonary angiography. The voxel intensity of all CT underwent truncation within the specified Hounsfield Unit (HU) window of $[-1000, 600]$ and subsequently normalization to $[0, 1]$. The size of CT volume resolution was normalized to $192 \times 192 \times 160$ for 3D synthesis and segmentation evaluation, while the size of CT slices was normalized to 256×256 for 2D synthesis and reconstruction evaluation. We compared ACIS against StyleGAN [27], MedGen3D [9], and V2UM2I [11]. Additionally, we implemented ACIS- without the proposed anatomical refinement strategy to validate its efficacy. We employed image augmentation techniques including random flipping, random rotation, random affine transformations, and random histogram shifting during the training process.

Segmentation: We implemented two models, 3DUNet [28] and MedNext [29], for image segmentation tasks. The 3DUNet comprised 4 encoder and 4 decoder layers, with the feature channels of $[16, 32, 64, 128]$ in the encoder layers. For the MedNext, we followed the architecture proposed by [29], using the small model with $3 \times 3 \times 3$ convolutional kernel size. Both models were trained using the dice loss as the objective function. The number of training data was set as 4, 8, 16, and 80, respectively to evaluate the efficiency of different synthesis approaches for few-shot learning. The dice similarity coefficient (DSC) and mean surface distance (MSD) [30] were utilized to evaluate the segmentation performance

Reconstruction: For the image reconstruction experiment, we performed three sub-tasks, including super-resolution, denoising, and deblurring. The synthesized images were further degraded via downsampling, Gaussian blurring, or adding Gaussian noise. For super-resolution, we performed $2\times$ and $4\times$ upsampling, with downsampled images obtained by bilinear interpolation [35]. For deblurring, we employed Gaussian filters with $\sigma \in \{1, 2, 5\}$. For denoising, we used noise levels $\sigma \in \{0.1, 0.2, 0.5\}$ with $\mu = 0$. We utilized two neural network architectures: ResNet [31] and Residual Encoder-Decoder (RED) [32] for general

image reconstruction and supplemented Efficient Super-Resolution Transformer (ESRT) [33] specifically for super-resolution. All the models were trained using the L2 loss as the objective function. We utilized 8 data to train the reconstruction tasks and established a baseline model without integrated synthesis methods. The model trained on the full 80 data was treated as the upper bound. Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [34] were utilized to evaluate the reconstruction performance. We also conducted a significance test between ACIS and the other-best performing models using a one-sided Wilcoxon signed-rank test, considering $p < 0.05$ as the threshold for statistical significance.

3.2 Evaluation of the Improvement on Segmentation Task

		Baseline		StyleGAN		MedGen3D		V2UM2I		ACIS-		ACIS	
		DCS (%)	NSD	DCS (%)	NSD	DCS (%)	NSD	DCS (%)	NSD	DCS (%)	NSD	DCS (%)	NSD
n=4	3DUnet	71.82	18.10	82.90	7.18	84.99	8.43	84.21	7.29	84.16	6.40	**87.61***	**5.73***
	MedNext	72.72	19.61	85.90	6.39	88.06	4.26	88.07	4.81	87.91	4.42	**91.10***	**4.27***
n=8	3DUnet	87.09	5.07	91.16	5.13	93.42	**2.31**	92.64	2.82	92.36	2.72	**94.05***	2.34
	MedNext	88.39	5.68	90.96	4.86	93.62	2.18	92.46	2.63	92.36	2.63	**94.83***	**2.17***
n=16	3DUnet	90.50	3.38	91.98	3.41	93.33	2.05	92.40	3.18	92.85	2.05	**94.05***	**1.89**
	MedNext	92.52	3.08	93.03	3.75	93.78	2.15	93.18	2.39	93.98	2.15	**94.34**	**1.95***
n=80	3DUnet	93.06	1.81	92.88	1.49	94.30	1.20	94.46	1.27	93.97	1.27	**94.47**	**1.13**
	MedNext	94.40	1.58	93.89	1.28	**95.42**	1.20	94.91	1.14	95.03	1.12	95.24	**1.06**

Table 1. A comparative assessment of the efficacy of various data synthesis techniques for pulmonary anatomical segmentation."DCS"=dice score, "MSD"=mean surface distance. A higher value of DCS and a lower value of MSD means a better segmentation result. The synthesis method yielding the best performance is denoted in **bold** font. "*" means the statistical significance.

The anatomical segmentation performance on the TLobe dataset, as presented in Table 1, demonstrated the generalization capability of our proposed ACIS synthesis method. The empirical findings revealed that the improvements afforded by ACIS were broadly applicable: when training data was scarce, ACIS conferred considerable performance boosts per both dice score and mean surface distance metrics. Even with sufficient training data, ACIS continued to enhance segmentation performance and model robustness. These results implied that ACIS represented an efficient few-shot learning strategy that mitigates dependence on training set size and a general means of improving neural network performance and generalizability.

3.3 Evaluation of the Improvement on Reconstruction Task

Table 2 delineated the quantitative outcomes of the image reconstruction tasks on the PENet dataset. Our analyses revealed that the proposed ACIS approach surpassed other data synthesis techniques in most metrics, with performance

			Baseline		StyleGAN		MedGen3D		V2UM2I		ACIS-		ACIS		Upper Bound	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Super resolution	SR×2	ResNet	29.43	0.911	30.76	0.918	**31.40**	0.923	30.49	0.921	31.24	0.922	31.20	**0.927***	32.44	0.932
		RED	29.56	0.910	30.87	0.922	32.19	0.941	31.30	0.932	32.12	0.937	**32.90***	**0.947***	33.80	0.935
		ESRT	30.39	0.929	32.51	0.940	33.45	**0.951**	33.38	0.942	34.01	0.942	**34.49**	0.950	34.53	0.957
	SR×4	ResNet	26.80	0.904	27.91	0.913	**28.90**	0.915	28.66	0.914	28.50	0.915	28.76	**0.922**	28.68	0.923
		RED	27.79	0.916	28.81	0.922	29.34	0.918	**29.66**	**0.927**	28.98	0.923	29.22	0.924	29.15	0.924
		ESRT	28.02	0.917	29.11	0.919	29.05	0.922	**30.13**	0.916	29.28	0.926	29.36	**0.926**	29.53	0.927
Denoising	σ=0.1	ResNet	30.34	0.849	31.58	0.855	31.75	0.876	32.41	0.885	**32.62**	**0.906**	32.04	0.904	32.61	0.914
		RED	30.98	0.894	32.03	0.905	32.97	0.910	33.06	0.906	33.26	0.911	**33.84***	**0.914**	34.29	0.920
	σ=0.2	ResNet	27.55	0.871	28.83	0.899	30.76	0.892	29.98	0.905	**30.99**	**0.913**	30.83	0.912	31.24	0.918
		RED	28.88	0.897	30.24	0.906	**31.66**	0.910	31.52	0.918	30.32	0.919	30.34	**0.925***	31.55	0.928
	σ=0.5	ResNet	24.19	0.805	25.38	0.820	26.02	0.844	**26.49**	0.845	25.91	0.852	26.25	**0.856**	26.95	0.863
		RED	27.15	0.837	28.33	0.844	28.39	0.849	**29.39**	0.846	29.35	0.845	29.22	**0.853**	29.90	0.860
Deblurring	σ=2	ResNet	30.09	0.889	31.31	0.904	32.74	0.938	32.69	0.944	**34.35**	0.951	34.22	**0.960**	34.52	0.956
		RED	34.56	0.964	35.71	0.974	34.84	0.958	**36.51**	0.964	35.24	0.960	35.22	**0.970***	35.52	0.973
	σ=3	ResNet	29.44	0.876	30.98	0.890	31.34	0.906	31.69	0.911	31.23	0.916	**31.87**	**0.927**	32.59	0.930
		RED	33.28	0.946	34.78	0.959	35.20	0.948	**35.50**	0.955	34.61	0.959	34.91	**0.965***	35.95	0.968
	σ=5	ResNet	27.02	0.790	28.10	0.799	28.25	0.825	**29.04**	**0.863**	28.80	0.861	28.77	0.863	28.93	0.857
		RED	27.73	0.851	29.03	0.870	29.07	0.875	29.96	0.864	29.99	0.892	**30.24***	**0.898**	29.89	0.897

Table 2. A comparative assessment of the efficacy of various image synthesis techniques for image super-resolution, denoising, and deblurring. "PSNR"=signal-to-noise ratio, "SSIM"=structural similarity index measure. A higher value of PSNR and SSIM means a better result. The synthesis techniques yielding the optimal performance are denoted in **bold** font. "*" means the statistical significance.

comparable to the upper bounds in some cases. This conclusion persisted across diverse reconstruction tasks and degrees of image degradation. The experiments evinced that the ACIS engendered substantive improvements in reconstruction performance, yielding PSNR enhancements of approximately 2 to 4 dB and SSIM increases of 0.01 to 0.03. Relative to the upper bound trained on 80 scans, the performance with ACIS augmentation exhibited only a 1% decline, despite reducing the training data by over 90%.

4 Conclusion

In this study, we propose a novel anatomic-constrained medical image synthesis framework termed ACIS that introduces anatomically diverse training data to downstream tasks, enhancing the fidelity and variability of synthesized images while preserving anatomical realism. We devise pseudo-masks derived from real-world data as anatomy priors coupled with a multi-stage synthesis pipeline to enable the generation of diverse yet realistic visual data. Additionally, we employ registration uncertainty to refine outputs and promote anatomical congruity between synthesized and real-world images. Experiments on high-level segmentation and low-level image reconstruction tasks demonstrate that our proposed synthesis approach efficaciously augments limited training data for few-shot learning, attaining superior performance over state-of-the-art synthesis techniques across diverse datasets. Remarkably, ACIS approaches or surpasses models trained on full datasets even when utilizing only 10% or less of the total training data. Overall, ACIS signifies promising advancements in medical im-

age synthesis, paving the way for enhanced diagnostic performance and clinical applications under limited supervision.

Acknowledgments. This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Research Administration (ORA) under Award No REI/1/5234-01-01, REI/1/5414-01-01, RGC/3/4816-01-01, REI/1/5289-01-01, REI/1/5404-01-01, REI/1/5992-01-01, and URF/1/4663-01-01.

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

References

1. Chan H P, Hadjiiski L M, Samala R K. Computer-aided diagnosis in the era of deep learning[J]. Medical physics, 2020, 47(5): e218-e227.
2. Najafabadi M M, Villanustre F, Khoshgoftaar T M, et al. Deep learning applications and challenges in big data analytics[J]. Journal of big data, 2015, 2(1): 1-21.
3. Han C, Kitamura Y, Kudo A, et al. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection[C]//2019 International Conference on 3D Vision (3DV). IEEE, 2019: 729-737.
4. Bhandarkar A, Naik P, Vakkund K, et al. Deep learning based computer aided diagnosis of Alzheimer's disease: a snapshot of last 5 years, gaps, and future directions[J]. Artificial Intelligence Review, 2024, 57(2): 1-62.
5. Nalepa J, Marcinkiewicz M, Kawulok M. Data augmentation for brain-tumor segmentation: a review[J]. Frontiers in computational neuroscience, 2019, 13: 83.
6. Vincent L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms[J]. IEEE transactions on image processing, 1993, 2(2): 176-201.
7. Garcea F, Serra A, Lamberti F, et al. Data augmentation for medical imaging: A systematic literature review[J]. Computers in Biology and Medicine, 2023, 152: 106391.
8. Croitoru F A, Hondru V, Ionescu R T, et al. Diffusion models in vision: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
9. Han K, Xiong Y, You C, et al. Medgen3d: A deep generative framework for paired 3d image and mask generation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 759-769.
10. Du Y, Bai F, Huang T, et al. Segvol: Universal and interactive volumetric medical image segmentation[J]. arXiv preprint arXiv:2311.13385, 2023.
11. Xing X, Papanastasiou G, Walsh S, et al. Less is More: Unsupervised Mask-guided Annotated CT Image Synthesis with Minimum Manual Segmentations[J]. IEEE Transactions on Medical Imaging, 2023
12. Chu Y, Luo G, Zhou L, et al. Deep learning-driven pulmonary arteries and veins segmentation reveals demography-associated pulmonary vasculature anatomy[J]. arXiv preprint arXiv:2404.07671, 2024.
13. Polianskii V, Marchetti G L, Kravberg A, et al. Voronoi density estimator for high-dimensional data: Computation, compactification and convergence[C]//Uncertainty in Artificial Intelligence. PMLR, 2022: 1644-1653.

14. Yoo J, Ahn N, Sohn K A. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 8375-8384.
15. Xie W, Jacobs C, Charbonnier J P, et al. Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans[J]. IEEE transactions on medical imaging, 2020, 39(8): 2664-2675.
16. McIlhagga W. The Canny edge detector revisited[J]. International Journal of Computer Vision, 2011, 91: 251-261.
17. Terrell G R, Scott D W. Variable kernel density estimation[J]. The Annals of Statistics, 1992: 1236-1265.
18. Gray A G, Moore A W. Nonparametric density estimation: Toward computational tractability[C]//Proceedings of the 2003 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2003: 203-211.
19. Du Q, Faber V, Gunzburger M. Centroidal Voronoi tessellations: Applications and algorithms[J]. SIAM review, 1999, 41(4): 637-676.
20. Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[J]. Advances in neural information processing systems, 2017, 30.
21. Singh N K, Raza K. Medical image generation using generative adversarial networks: A review[J]. Health informatics: A computational perspective in healthcare, 2021: 77-96.
22. Chen J, Frey E C, He Y, et al. Transmorph: Transformer for unsupervised medical image registration[J]. Medical image analysis, 2022, 82: 102615.
23. Balakrishnan G, Zhao A, Sabuncu M R, et al. VoxelMorph: a learning framework for deformable medical image registration[J]. IEEE transactions on medical imaging, 2019, 38(8): 1788-1800.
24. Wang D, Pan Y, Durumeric O C, et al. PLOSL: Population learning followed by one shot learning pulmonary image registration using tissue volume preserving and vesselness constraints[J]. Medical image analysis, 2022, 79: 102434.
25. Paria B, Lahiri A, Biswas P K. PolicyGAN: Training generative adversarial networks using policy gradient[C]//2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR). IEEE, 2017: 1-6.
26. Huang S C, Kothari T, Banerjee I, et al. PENet—a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging[J]. NPJ digital medicine, 2020, 3(1): 61.
27. Endo Y, Kanamori Y. Few-shot semantic image synthesis using stylegan prior[J]. arXiv preprint arXiv:2103.14877, 2021.
28. Qamar S, Jin H, Zheng R, et al. A variant form of 3D-UNet for infant brain segmentation[J]. Future Generation Computer Systems, 2020, 108: 613-623.
29. Roy S, Koehler G, Ulrich C, et al. Mednext: transformer-driven scaling of convnets for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 405-415.
30. Park B, Park H, Lee S M, et al. Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks[J]. Journal of digital imaging, 2019, 32: 1019-1026.
31. Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136-144.

32. Chen H, Zhang Y, Kalra M K, et al. Low-dose CT with a residual encoder-decoder convolutional neural network[J]. IEEE transactions on medical imaging, 2017, 36(12): 2524-2535.
33. Lu Z, Li J, Liu H, et al. Transformer for single image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 457-466.
34. Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th international conference on pattern recognition. IEEE, 2010: 2366-2369.
35. Chu Y, Zhou L, Luo G, et al. Topology-Preserving Computed Tomography Super-Resolution Based on Dual-Stream Diffusion Model[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 260-270.