Topology-Preserving Computed Tomography Super-Resolution Based on Dual-Stream Diffusion Model

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

[1] Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
{gongning.luo,xin.gao}@kaust.edu.sa
[2] Institute of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China
qiuzw@nefu.edu.cn

Abstract. X-ray computed tomography (CT) is indispensable for modern medical diagnosis, but the degradation of spatial resolution and image quality can adversely affect analysis and diagnosis. Although super-resolution (SR) techniques can help restore lost spatial information and improve imaging resolution for low-resolution CT (LRCT), they are always criticized for topology distortions and secondary artifacts. To address this challenge, we propose a dual-stream diffusion model for super-resolution with topology preservation and structure fidelity. The diffusion model employs a dual-stream structure-preserving network and an imaging enhancement operator in the denoising process for image information and structural feature recovery. The imaging enhancement operator can achieve simultaneous enhancement of vascular and blob structures in CT scans, providing the structure priors in the super-resolution process. The final super-resolved CT is optimized in both the convolutional imaging domain and the proposed vascular structure domain. Furthermore, for the first time, we constructed an ultra-high resolution CT scan dataset with a spatial resolution of 0.34×0.34 mm^2 and an image size of 1024×1024 as a super-resolution training set. Quantitative and qualitative evaluations show that our proposed model can achieve comparable information recovery and much better structure fidelity compared to the other state-of-the-art methods. The performance of high-level tasks, including vascular segmentation and lesion detection on super-resolved CT scans, is comparable to or even better than that of raw HRCT. The source code is publicly available at https://github.com/Arturia-Pendragon-Iris/UHRCT_SR.

Keywords: Computed tomography · Super resolution · Diffusion model · Image enhancement

1   Introduction

Computed tomography (CT) is a prevalent imaging modality with applications in biology, disease diagnosis, interventional imaging, and other areas. High-resolution CT (HRCT) is beneficial for clinical diagnosis and surgical planning because it can provide detailed spatial information and specific features, usually employed in advanced clinical routines [1]. HRCT usually requires high-precision CT machines to scan for a long time with high radiation doses to capture the internal structures, which is expensive and can impose the risk of radiation exposure [2]. These factors make HRCT relatively less available, especially in towns and villages, compared to low-resolution CT (LRCT). However, degradation in spatial resolution and imaging quality brought by LRCT can interfere with the original physiological and pathological information, adversely affecting the diagnosis [3]. Consequently, how to produce high-resolution CT scans at a smaller radiation dose level with lower scanning costs is a holy grail of the medical imaging field (Fig. 1).
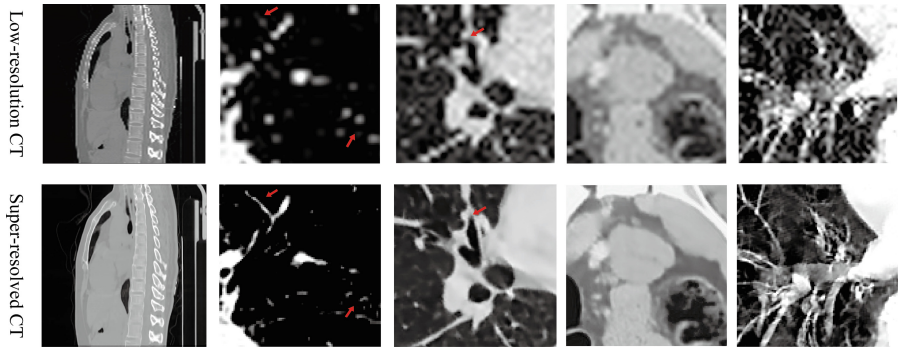

Fig. 1. Low-resolution CT scans can adversely affect the diagnosis. Restoring the original structural information while improving the spatial resolution is a great challenge.

With the advancement of artificial intelligence, super-resolution (SR) techniques based on neural networks indicate new approaches to this problem. By inferring detailed high-frequency features from LRCT, super-resolution can introduce additional knowledge and restore lost information due to low-resolution scanning. Deep-learning (DL) based methods, compared to traditional methods, can incorporate hierarchical features and representations from prior knowledge, resulting in improved results in SR tasks [4]. According to different neural-network frameworks, these SR methods can be broadly categorized into two classes: 1) convolutional neural network (CNN) based model [5–7], and 2) generative adversarial network (GAN) based model [2,8,9]. Very recently, the diffusion model is emerging as the most promising deep generative model [11], which usually consists of two stages: a forward stage to add noises and a reverse

stage to separate noises and recover the original images. The diffusion model shows impressive generative capabilities for many tasks, including image generation, inpainting, translation, and super-resolution [10,12,13].

While DL-based methods can generate promising results, there can still be geometric distortions and artifacts along with structural edges in the super-resolved results [15,16]. These structural features always represent essential physiological structures, including vasculature, fibrosis, tumor, and other lesions. The distortion and infidelity of these features can lead to potential misjudgment for diagnosis, which is unacceptable for clinical application. Moreover, the target image size and spatial resolution of HRCT for most existing SR methods is about 512×512 and 0.8×0.8 mm^2. With the progress in hardware settings, ultra-high-resolution CT (UHRCT) with an image size of 1024×1024 and spatial resolution of 0.3×0.3 mm^2 can be available very recently [17]. Though UHRCT can provide much more detailed information, to our best knowledge, SR tasks targeting UHRCT have rarely been discussed and reported.

In this paper, we propose a novel dual-stream conditional diffusion model for CT scan super-resolution to generate UHRCT results with high image quality and structure fidelity. The conditional diffusion model takes the form $p(y|x)$, where x is the LRCT, and y is the targeted UHRCT [14]. The novel diffusion model incorporates a dual-stream structure-preserving network and a novel imaging enhancement operator in the denoising process. The imaging enhancement operator can simultaneously extract the vascular and blob structures in the CT scans and provide structure prior to the dual-stream network. The dual-stream network can fully exploit the prior information with two branches. One branch optimizes the SR results in the image domain, and the other branch optimizes the results in the structure domain. In practice, we use a convolution-based lightweight module to simulate the filtering operations, which enables faster and easier back-propagation in the training process. Furthermore, we constructed a new ultra-high resolution CT scan dataset obtained with the most advanced CT machines. The dataset contained 87 UHRCT scans with a spatial resolution of 0.34×0.34 mm^2 and an image size of 1024×1024. Extensive experiments, including qualitative and quantitative comparisons in both image consistency, structure fidelity, and high-level tasks, demonstrated the superiority of our method. Our contributions can be summarized as follows:

1) We proposed a novel dual-stream diffusion model framework for CT super-resolution. The framework incorporates a dual-stream structure-preserving network in the denoising process to realize better physiological structure restoration.

2) We designed a new image enhancement operator to model the vascular and blob structures in medical images. To avoid non-derivative operations in image enhancement, we proposed a novel enhancement module consisting of lightweight convolutional layers to replace the filtering operation for faster and easier back-propagation in structural domain optimization.

3) We established an ultra-high-resolution CT scan dataset with a spatial resolution of 0.34×0.34 mm^2 and an image size of 1024×1024 for training and testing the SR task.

4) We have conducted extensive experiments and demonstrated the excellent performance of the proposed SR methods in both the image and structure domains. In addition, we have evaluated our proposed method on high-level tasks, including vascular-system segmentation and lesion detection on the SRCT, indicating the reliability of our SR results.
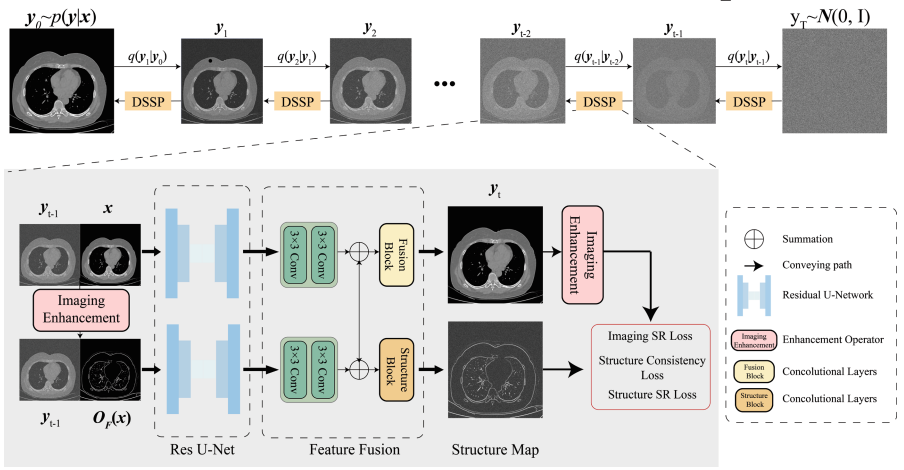

Fig. 2. Overview of the proposed Dual-stream Diffusion Model. The forward diffusion process q gradually adds Gaussian noises to the target HRCT. The reverse process p iteratively denoises the target image, conditioned on the source LRCT x, which is realized with the dual-stream structure-preserving network.

2 Method

2.1 Dual-Stream Diffusion Model for Structure-Preserving Super-Resolution

To preserve the structure and topology relationship in the denoising progress, we designed a novel Dual-Stream Diffusion Model (DSDM) for better super-resolution and topology restoration (Fig. 2). In the DSDM framework, given a HRCT slice y, we generate a noisy version $\tilde{y}$, and train the network G_{DSSP} to denoise $\tilde{y}$ with the corresponding LRCT slice x and a noise level indicator γ. The optimization is defined as

$$E_{(x,y)}E_{\epsilon\ \mathcal{N}(0,I)}E_{\gamma}\left\|G_{\text{DSSP}}\left(x,\sqrt{\gamma}y+\sqrt{1-\gamma}\epsilon,\gamma\right)-\epsilon\right\|^{\text{P}} \tag{1}$$

In the denoising process, we used a dual-stream structure-preserving (DSSP) network for supervised structure restoration. The DSSP network optimizes the denoised results in the image domain and the structure domain, respectively. The structural domain, obtained with the image enhancement operator, is concatenated with the LRCT slice as the input of the structure branch. The final SR results are obtained after the feature fusion model between the image map and the structure map.

2.2   Imaging Enhancement Operator for Vascular and Blob Structure

We introduced an enhancement operator in the DSSP network to model the vascular and blob structures, which can represent important physiological information according to clinical experience, and provide the prior structural information for the DSSP network. For one pixel $x = [x_1, x_2]^T$ in the CT slice, let $I(x)$ denote the imaging intensity at this point. The 2×2 Hessian matrix at the scale s is defined as [18]

$$H_{i,j}(x) = s^2 I(x) \frac{\partial^2 G(x,s)}{\partial x_i \partial x_j} \tag{2}$$

where $G(x,s)$ is the 2D Gaussian kernel. The two eigenvalues of the Hessian matrix are denoted as $\lambda = (\lambda_1, \lambda_2)$ and here we agree that $|\lambda_1| <= |\lambda_2|$. The eigenvalues of the Hessian matrix can reflect the geometric shape, curvature, and brightness of the local images. For the blob-like structures, the three eigenvalues are about the same, $\lambda_1 \approx \lambda_2$; for the vascular-like structures, λ_2 can be much larger than the absolute value of λ_1, $|\lambda_2| >> |\lambda_1|$ [19]. The eigenvalue relations at scale s can be indicated by several different functions. Here we proposed a novel structure kernel function, which is defined as

$$\mathcal{V}_C = \frac{\lambda_2 (\kappa_1 \lambda_1 + \lambda_\tau)(\kappa_2 \lambda_2 + \lambda_\tau)}{(\lambda_1 + \lambda_2 + \lambda_\tau)^3} \tag{3}$$

κ_1 and κ_2 are the parameters to control the sensitivity for the vascular-like structures and blob-like structures, respectively. λ_τ is the self-regulating factor.

$$\lambda_\tau = (\lambda_2 - \lambda_1) \frac{e^\gamma - e^{-\gamma}}{e^\gamma + e^{-\gamma}} + \lambda_1 \tag{4}$$

$$\gamma = \left| \frac{\lambda_2}{\lambda_1} \right| - 1 \tag{5}$$

When λ_1 is about the same with λ_2, λ_τ is closed to λ_1; and when λ_1 is much smaller to λ_2, λ_τ is closed to λ_2, which can achieve a balance between two conditions.

2.3   Objective Function

We designed a new loss function to ensure that the final SR result can be optimized in both the image domain and the structure domain. Denoting the reference image as y^t, the pixel-wise loss in the imaging domain is formulated as

$$\mathcal{L}_{SR}^{pixel} = \left\| G_{DSSP}^{image}\left(y^t\right) - y^t \right\|^2 + \lambda_1 \left\| G_{DSSP}^{image}\left(y^t\right) - y^t \right\|^1 \tag{6}$$

$G_{DSSP}^{image}\left(y^t\right)$ is the recovered image from the image-domain branch. L1 loss yields a significantly higher consistency and lower diversity, while L2 loss can better capture the outliers. Here we used a parameter λ_{L1} to balance these two losses.

In the meantime, a structure-constraint loss is also necessary to help the network achieve better performance in structure consistency. The loss function consists of two parts, which measure the consistency of the image-domain branch and the structure-domain branch. Denoting the structure-domain output as $G_{DSSP}^{struct}\left(y^t\right)$, the structure-constraint loss can be presented as

$$\mathcal{L}_{SR}^{struct} = \left\| F_c \circ G_{DSSP}^{image}\left(y^t\right) - F_c \circ y^t \right\|^1 + \left\| G_{DSSP}^{struct}\left(y^t\right) - F_c \circ y^t \right\|^1 \tag{7}$$

However, the image enhancement described above involves overly complex calculations, making back-propagation difficult in the training process. Here we utilized a convolution-based operator $O_{F_c}(\cdot)$ to simplify the calculation, which consists of several lightweight convolutional layers to simulate the operation of image enhancement. In this way, we transform the complex filtering operation into a simple convolution operation, thus back-propagation can be easily processed. The loss function is then modified as

$$\mathcal{L}_{SR}^{struct} = \left\| O_{F_c} \circ G_{DSSP}^{image}\left(y^t\right) - O_{F_c} \circ y^t \right\|^1 + \left\| G_{DSSP}^{struct}\left(y^t\right) - O_{F_c} \circ y^t \right\|^1 \tag{8}$$

The total objective function is the sum of two losses.

$$\mathcal{L}_{SR} = \mathcal{L}_{SR}^{pixel} + \lambda_2 \mathcal{L}_{SR}^{struct} \tag{9}$$

3   Experiments and Conclusion

3.1   Datasets and Evaluation

We constructed three datasets for framework training and evaluation; two of them were in-house data collected from two CT scanners(the ethics number is 20220359), and the other was the public Luna16 dataset [22]. More details about the two in-house datasets are described in the Supplementary Materials. We evaluated our SR model on three CT datasets:

- Dataset 1: 2D super-resolution from 256×256 to 1024×1024, with the spatial resolution from 1.36×1.36 mm^2 to 0.34×0.34 mm^2.

- Dataset 2: 3D super-resolution from $256 \times 256 \times 1X$ to $512 \times 512 \times 5X$, with the spatial resolution from $1.60 \times 1.60 \times 5.00$ mm^3 to $0.80 \times 0.80 \times 1.00$ mm^3.
- Dataset 3: 2D super-resolution from 256×256 to 512×512 on the Luna16 dataset.

We compare our model with other SOTA super-resolution methods, including bicubic interpolation, SRCNN [7], SRResNet [6], Cycle-GAN [2], and SR3 [12]. Performance is assessed qualitatively and quantitatively, using PSNR, SSIM [23], Visual Information Fidelity (VIF) [24], and Structure Mean Square Error (SMSE). VIF value is correlated well with the human perception of SR images, which can measure diagnostic acceptance and information maintenance. SMSE is proposed to evaluate the structure difference between the ground truth and SRCT. Specially, we obtained the structural features of the ground truth and SRCT with Frangi filtering and then calculated the pixel-wise difference [20].

$$SMSE = \|F_{\text{Frangi}}(HRCT) - F_{\text{Frangi}}(SRCT)\|^2 \qquad (10)$$
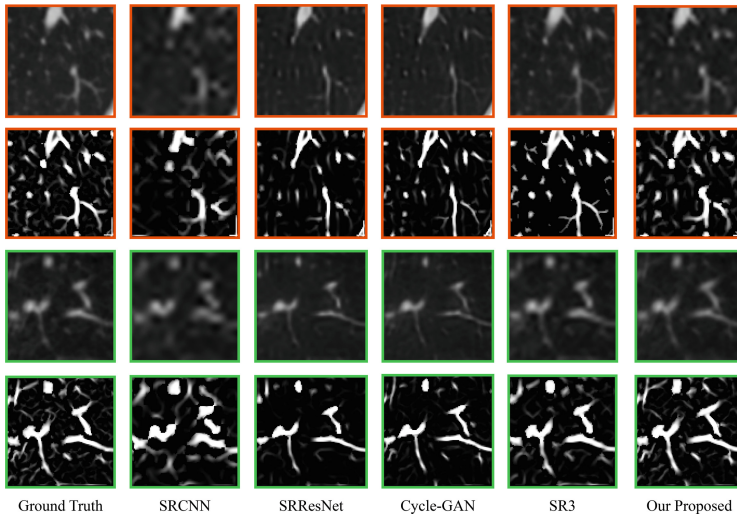

Fig. 3. Visual comparison of super-resolved CT from the ultra-high-resolution dataset. The display window is $[-600, 400]$ HU. The restored images are shown in the first and third lines, and the structural features are shown in the second and fourth lines.

3.2   Qualitative and Quantitative Results

Qualitative comparisons are shown in Fig. 3 and the quantitative results are shown in Table 1. The super-resolution results with our proposed methods achieve the highest scores in both image restoration and structure consistency for most indices, and there are no obvious secondary artifacts introduced in the SR

results. Although the GAN-based methods and SR3 can produce sharp details, they tend to generate artifacts for the vascular systems, which is more evident in the structure-enhanced figures. The problem of inconsistent structure is also reflected in the value of VIF and SMSE on both GAN-based methods and SR3.

Lesion Detection and Vessel Segmentation on Super-Resolved CT. To further evaluate the information maintenance of our SR methods, we conducted some high-level tasks, including lung nodules detection and pulmonary airway and blood vessel segmentation on the super-resolved CT scans. We compared the performance of different methods on SRCT and the ground truth. For nodule detection, these methods included U-Net, V-Net [25], ResNet [26], DCNN [27] and 3D-DCNN [28]. For the vessel segmentation, these methods included 3D U-Net, V-Net [25], nnUNet [31], Nardelli et al. [30] and Qin et al. [29]. Figure 4 shows the quantitative results of the performance comparison. The performance of these high-level tasks on the SR results is comparable to or even better than that on the ground-truth CT. Such results, to some extent, demonstrated that our SR method does not introduce artifacts or structural inconsistencies and cause misjudgment, while the improved spatial resolution and image quality generated by our proposed results shows great potential in improving the performance of high-level tasks.

Table 1. Quantitative evaluation of state-of-the-art super-resolution algorithms.

	Dataset 1				Dataset 2				Dataset 3			
	PSNR	SSIM	VIF	SMSE	PSNR	SSIM	VIF	SMSE	PSNR	SSIM	VIF	SMSE
Interpolation	28.27	0.926	0.583	1.46	22.73	0.817	0.516	1.76	23.84	0.776	0.679	1.51
SRCNN [7]	31.71	0.957	0.743	1.43	28.39	0.875	0.559	1.27	31.62	0.842	0.738	1.43
SRResNet [6]	32.69	0.96	0.762	0.992	29.6	0.854	0.685	1.03	32.84	0.897	0.796	0.892
Cycle-GAN [2]	37.32	**0.993**	0.901	0.462	32.31	0.918	0.881	0.822	37.82	0.921	0.915	0.282
SR3 [12]	37.18	0.974	0.812	0.474	36.85	0.957	0.916	0.859	**39.57**	0.968	0.902	0.274
Our Proposed	**40.75**	0.992	**0.977**	**0.162**	**38.76**	**0.979**	**0.967**	**0.274**	38.91	**0.977**	**0.974**	**0.162**

Lung Nodule Detection   Lung Vascular Segmentation
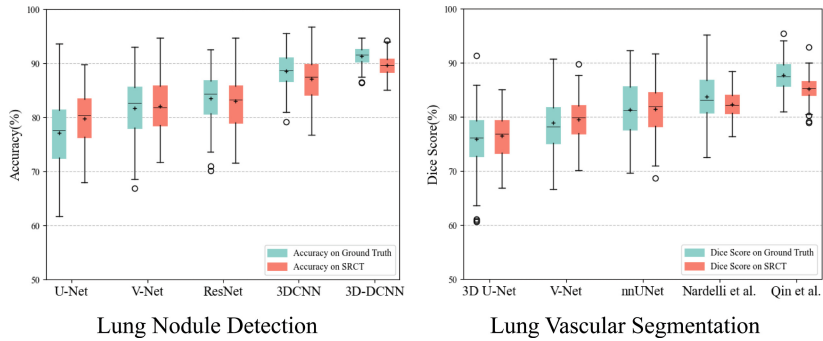
Fig. 4. Comparison of the detection and segmentation performances on HRCT and SRCT.

4   Conclusion

In this paper, we have established a dual-stream diffusion model framework to address the problem of topology distortion and artifact introduction that generally exists in the medical super-resolution results. We first propose a novel image enhancement operator to model the vessel and blob structures in the CT slice, which can provide a structure prior to the SR framework. Then, we design a dual-stream diffusion model that employs a dual-stream ream structure-preserving network in the denoising process. The final SR outputs are optimized not only by convolutional image-space losses but also by the proposed structure-space losses. Extensive experiments have shown that our SR methods can achieve high performance in both image restoration and structure fidelity, demonstrating the promising performance of information preservation and the potential of applying our SR results to downstream tasks.

References

1. Akagi, M., Nakamura, Y., Higaki, T., et al.: Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. Eur. Radiol. **29**, 6163–6171 (2019)
2. You, C., Li, G., Zhang, Y., et al.: CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). IEEE Trans. Med. Imaging **39**(1), 188–203 (2019)
3. Wolterink, J.M., Leiner, T., Viergever, M.A., et al.: Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans. Med. Imaging **36**(12), 2536–2545 (2017)
4. Wang, Z., Chen, J., Hoi, S.C.H.: Deep learning for image super-resolution: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **43**(10), 3365–3387 (2020)
5. Ren, H., El-Khamy, M., Lee, J.: CT-SRCNN: cascade trained and trimmed deep convolutional neural networks for image super resolution. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1423–1431. IEEE (2018)
6. Wu, Z., Shen, C., Van Den Hengel, A.: Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recogn. **90**, 119–133 (2019)
7. Georgescu, M.I., Ionescu, R.T., Verga, N.: Convolutional neural networks with intermediate loss for 3D super-resolution of CT and MRI scans. IEEE Access **8**, 49112–49124 (2020)
8. Xie, Y., Franz, E., Chu, M., et al.: tempoGAN: a temporally coherent, volumetric GAN for super-resolution fluid flow. ACM Trans. Graph. (TOG) **37**(4), 1–15 (2018)
9. Lyu, Q., You, C., Shan, H., et al.: Super-resolution MRI and CT through GAN-circle. In: Developments in X-ray tomography XII, vol. 11113, pp. 202–208. SPIE (2019)

10. Su, X., Song, J., Meng, C., et al.: Dual diffusion implicit bridges for image-to-image translation. International Conference on Learning Representations (2022)
11. Nair, N.G., Mei, K., Patel, V.M.: At-ddpm: restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3434–3443 (2023)
12. Saharia, C., Ho, J., Chan, W., et al.: Image super-resolution via iterative refinement. IEEE Trans. Pattern Anal. Mach. Intell. (2022)
13. Saharia, C., Chan, W., Chang, H., et al.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH on Conference Proceedings 2022, pp. 1–10 (2022)
14. Lyu, Q., Wang, G.: Conversion Between CT and MRI Images Using Diffusion and Score-Matching Models. arXiv preprint arXiv:2209.12104 (2022)
15. Shi, Y., Wang, K., Chen, C., et al.: Structure-preserving image super-resolution via contextualized multitask learning. IEEE Trans. Multimedia **19**(12), 2804–2815 (2017)
16. Ma, C., Rao, Y., Lu, J., et al.: Structure-preserving image super-resolution. IEEE Trans. Pattern Anal. Mach. Intell. **44**(11), 7898–7911 (2021)
17. Oostveen, L.J., Boedeker, K.L., Brink, M., et al.: Physical evaluation of an ultra-high-resolution CT scanner. Eur. Radiol. **30**, 2552–2560 (2020)
18. Jerman, T., Pernuš, F., Likar, B., et al.: Blob enhancement and visualization for improved intracranial aneurysm detection. IEEE Trans. Visual Comput. Graph. **22**(6), 1705–1717 (2015)
19. Jerman, T., Pernuš, F., Likar, B., et al.: Enhancement of vascular structures in 3D and 2D angiographic images. IEEE Trans. Med. Imaging **35**(9), 2107–2118 (2016)
20. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0056195
21. Jiang, M., Wang, G., Skinner, M.W., Rubinstein, J.T., Vannier, M.W.: Blind deblurring of spiral CT images. IEEE Trans. Med. Imaging **22**(7), 837–845 (2003). https://doi.org/10.1109/TMI.2003.815075
22. Armato, S.G., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. **38**, 915–931 (2011)
23. Horé, A., Ziou, D.: Image Quality Metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 2366–2369 (2010). https://doi.org/10.1109/ICPR.2010.579
24. Mahmoudpour, S., Kim, M.: A study on the relationship between depth map quality and stereoscopic image quality using upsampled depth maps. In: Emerging Trends in Image Processing, Computer Vision and Pattern Recognition, pp. 149–160. Morgan Kaufmann (2015)
25. Abdollahi, A., Pradhan, B., Alamri, A.: VNet: an end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. IEEE Access **8**, 179424–179436 (2020)
26. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
27. Jin, H., Li, Z., Tong, R., et al.: A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. Med. Phys. **45**(5), 2097–2107 (2018)
28. Naseer, I., Akram, S., Masood, T., et al.: Performance analysis of state-of-the-art CNN architectures for luna16. Sensors **22**(12), 4426 (2022)

29. Qin, Y., Zheng, H., Gu, Y., et al.: Learning tubule-sensitive cnns for pulmonary airway and artery-vein segmentation in ct. IEEE Trans. Med. Imaging **40**(6), 1603–1617 (2021)
30. Nardelli, P., Jimenez-Carretero, D., Bermejo-Pelaez, D., et al.: Pulmonary artery-vein classification in CT images using deep learning. IEEE Trans. Med. Imaging **37**(11), 2428–2440 (2018)
31. Isensee, F., Jaeger, P.F., Kohl, S.A.A., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
32. Trinh, D.H., Luong, M., Dibos, F., et al.: Novel example-based method for super-resolution and denoising of medical images. IEEE Trans. Image Process. **23**(4), 1882–1895 (2014)