VQ-SCD: Vector Quantization Meets Unknown Scan Condition Self-supervised Low-Dose CT Denoising

Bo Su[1], Jiabo Xu[1], Xiangyun Hu[1,2(✉)], Kai Deng[1], Jiancheng Li[2], and Zhouxian Lu[1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430072, China
huxy@whu.edu.cn
[2] Hubei Luojia Laboratory, Wuhan, 430079, China

Abstract. Artifacts and noise in low-dose CT images can degrade image quality, potentially hindering accurate diagnosis. In recent years, image-domain post-processing denoising methods have gained flexibility by eliminating the need for raw data. However, clinical scanning conditions vary widely, with most existing studies focusing on CT denoising under fixed or known conditions. Moreover, obtaining paired CT data in clinical settings is challenging, limiting the practical applicability of supervised learning methods. To address these challenges, we propose the self-supervised VQ-SCD, capable of denoising low-dose CT (LDCT) images under varying unknown scanning conditions using only normal-dose CT (NDCT) training data. For the first time, VQ-SCD uses a discretized codebook to approximate the distribution of LDCT features across various scanning conditions, enabling uniform characterization and denoising of data from multiple scanning setups. Additionally, we design a miniature diffusion model that uses up-sampled features as guidance to enhance image details. Our method outperforms both supervised and state-of-the-art self-supervised methods in terms of both quantitative metrics and visual quality, with a test time of only 0.25 seconds per image. Furthermore, training the model using only animal and phantom data still results in excellent denoising performance on human data. The code will be available at https://github.com/WHUSU/VQSCD.

Keywords: Low-dose CT · Denoising· Vector Quantization· Diffusion model.

1 Introduction

High doses of ionizing radiation can damage tissue cells, potentially increasing the risk of cancer. Reducing tube current can lower the radiation dose to generate LDCT images[3]; however, this often introduces noise and artifacts that

Bo Su and Jiabo Xu contribute equally to this work.

degrade image quality, potentially affecting diagnostic accuracy. Image domain post-processing allows for direct processing in the reconstructed image, facilitating broader application and adoption. Common image-domain techniques include filtering[18], NLM[20], and BM3D[7]. Recently, deep learning-based low-dose CT denoising methods have shown notable success[11, 14, 17, 25]. Some supervised learning approaches employ specific model architectures, analyze differences between low-dose and normal-dose images, and adaptively tune hyperparameters and regularization terms to achieve optimal denoising outcomes for known scanning conditions [29, 9, 1]. However, such models often rely heavily on paired data, with simulated data frequently differing from real clinical data. From an ethical perspective, the secondary radiation dose from repeated scans can be detrimental to the patient. Additionally, unavoidable secondary respiratory motion causes discrepancies between consecutive images, complicating the acquisition of a large volume of paired data. This challenge is further exacerbated by the variability in clinical scanning conditions, such as dose, slice thickness, and equipment, leading to significant imaging variations. Addressing these issues requires minimizing data dependence while enhancing the model's applicability.

Self-supervised denoising methods[21, 12, 6, 30, 24] have been extensively researched for natural images, with training typically involving only noisy images, primarily focusing on improving visual quality or high-level semantic features. In contrast, the primary goal of low-dose CT image denoising is to reduce noise while preserving critical diagnostic information. Ethical constraints complicate the collection of LDCT data, but hospitals can readily collect a substantial number of retrospective NDCT images. Thus, using NDCT images for training is a more practical and suitable approach. Given the complexity of CT imaging noise characteristics, excessive noise suppression can lead to the loss of small lesions. Although diffusion models have notable advantages in preserving image details and textures [8, 13, 27], their high computational cost and prolonged inference time limit their clinical applicability. Moreover, complex clinical scenarios require the model to adapt to diverse scanning conditions, with significant variations in image quality posing additional challenges for model robustness.

To address these challenges, we propose a self-supervised denoising framework incorporating Vector Quantization (VQ) [23]. The method exclusively utilizes NDCT images during training to achieve LDCT denoising under multi-scanning conditions, with the VQ-SCD overview presented in Fig. 1. The VQ-SCD framework comprises three principal components designed to enhance denoising performance. First, diverse noise patterns are generated through random noise injection and geometric flipping of input images, simulating heterogeneous scanning condition distributions. Subsequently, a Vision Transformer (ViT) model [2] extracts hierarchical features from both LDCT and NDCT images. These feature representations are constrained through Mean Squared Error (MSE) loss minimization to ensure cross-condition feature consistency. The VQ discretizes continuous features into codebook entries. Each encoded value corresponds to a codebook vector, enabling the replacement of original features with their nearest codebook representations. By representing various domains

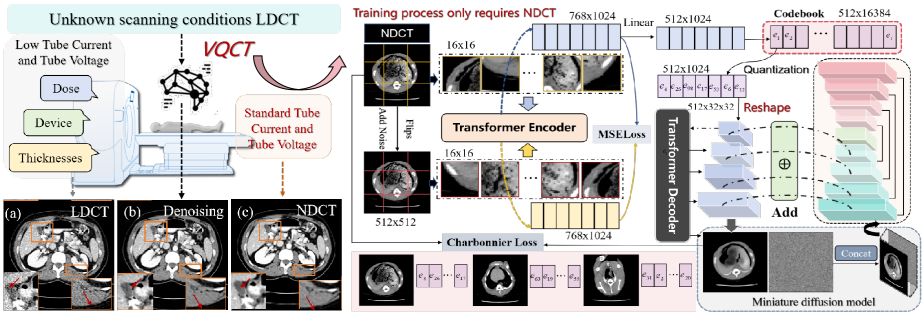

Fig. 1. Overview of the proposed VQ-SCD, including the encoder, vector quantization, decoder, and miniature diffusion model, which requires only NDCT for training to achieve LDCT denoising for unknown scan conditions.

as consistent distributions through the combination of codebook vectors, our model is able to describe and process heterogeneous data uniformly. Finally, a miniature diffusion model enhances texture preservation through multilevel upsampling feature conditioning. This module iteratively solves maximum a posteriori (MAP) estimation [34] by integrating diffusion model priors, achieving detail-optimized reconstruction within 0.25 seconds per image. The model outperforms both supervised and unsupervised methods in terms of quantitative metrics and visual quality. The proposed method demonstrates adaptability to complex multi-scanning scenarios while circumventing conventional data acquisition limitations, thereby offering an efficient and generalizable solution for LDCT denoising.

2 Method

2.1 Encoder architecture

We utilize the ViT to encode features of LDCT images. As shown in Fig. 1, the ViT encoder segments the input image into non-overlapping 16×16 pixel blocks, which are then linearly mapped to embedding vectors. These vectors are combined with positional encoding to construct a serialized representation of the image. The input image, with size of 512×512, is divided into 1024 blocks using a 32×32 grid. Each block is processed by the ViT encoder and converted into a feature vector of dimension 768. To enhance robustness against variations in dose and slice thickness, we employ a multilevel Gaussian noise injection strategy for data augmentation in the training phase. Additionally, horizontal and vertical flip operations are introduced to improve the adaptability of the codebook to spatial transformations. Adding Gaussian noise is intended to enhance the encoder's robustness to perturbations and promote the utilization of the codebook, which serves a different purpose compared to adding noise in the projection domain to simulate LDCT images [16, 31] for model training.

2.2 Vector Quantizer module

The VQ module learns a codebook comprising N vectors, each encoding a d-dimensional feature of the encoder's input. The codebook is represented as a purple vector in Fig. 2, $e_1, e_2, e_3..., e_k$. The original input image is processed by the encoder to obtain a feature vector, denoted $z_e(x)$ in a blue vector. The d-dimensional vectors of $z_e(x)$ are discretized to find the nearest e_i in the codebook. Discretization involves quantizing continuous high-dimensional features into their nearest-neighbor codebook vectors, which effectively reduces data dimensionality and minimizes redundant information. Finally, $z_e(x)$ is replaced by the nearest e_i in the codebook and outputs $z_q(x)$.

2.3 Decoder module

The decoder progressively reconstructs the low-dimensional feature map to a high-resolution 512×512. We design a miniature diffusion model to restore textures. Let y and x denote the output of the VQ model and the original normal-dose CT, respectively. To preserve the original details of the LDCT images while effectively removing noise, we expect x_{t-1} to obey the diffusion prior p_θ and to be guided by y_{t-1} in its content. p_θ can be calculated concerning Eq. (1).

$$p_\theta\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \mu_\theta\left(\boldsymbol{x}_t, t\right), \Sigma_\theta\left(\boldsymbol{x}_t, t\right)\right), \tag{1}$$

$p(x_{t-1}|x_t)$ represents the reverse denoising process. μ_θ and $\sum_\theta$ are the mean and variance [10]. We use the up-sampled feature maps as conditions to guide the inverse process of the diffusion model, combining it with a priori iterations to solve the maximum a posteriori distribution problem. x_{t-1} satisfies the following conditions:

$$arg\min_{x_{t-1}}\left\{\left\|x_{t-1}-y_{t-1}\right\|^2-\lambda_t \log p_\theta(\hat{x}_{t-1}|x_t)\right\}, \tag{2}$$

We introduce an adaptive trade-off coefficient $\lambda_t = \lambda_0 \cdot \sqrt{\bar{\alpha}_t}$ [15] (initial $\lambda_0 = 0.08$) to balance noise suppression and detail retention across diverse scanning conditions and clinical requirements. Larger λ_t values enhance denoising efficacy at the expense of potential detail loss. Our sampling strategy employs the DDIM method [26], implementing differential step intervals: 20-step sampling for the initial 400 iterations followed by 200-step intervals. This configuration optimizes likelihood term influence through finer temporal sampling while effectively capturing low-dose CT texture characteristics through interval adjustment.

2.4 Traning Losses

The proposed method employs a composite loss function composed of four components: L_{MSE}, which constrains the coded features of the input and output images in the encoder; L_{commit}, with weight set to 0.1 [23], which constrains the alignment of encoder outputs with the codebook in the VQ layer; L_{diff}, which constrains the update of the predicted noise relative to the real noise in the

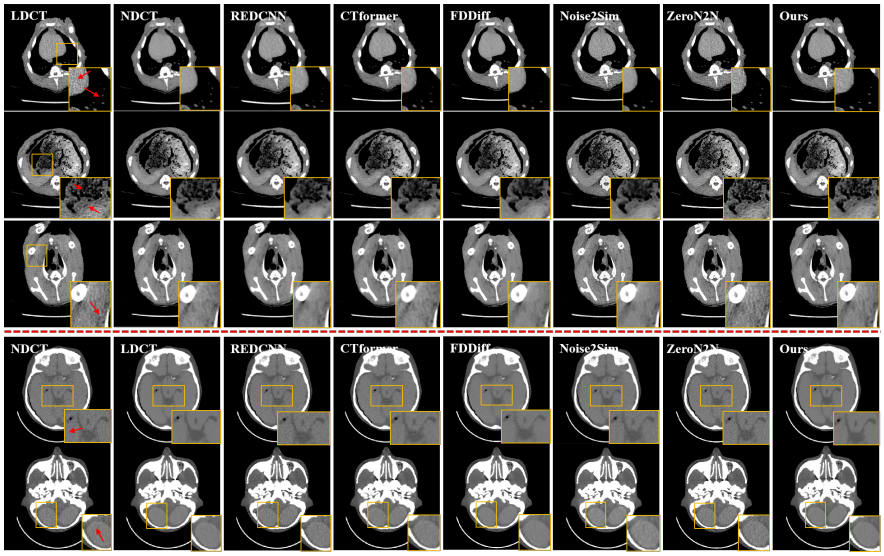

Fig. 2. Comparison of results in different methods with [-100, 200] HU. Above the red dotted line, the piglet representative slices are 10%, 25%, and 50% doses, listed from top to bottom. Below the line, the phantom results are 10% dose, 5% slice, 25% dose, and 0.625 mm, listed from top to bottom.

miniature diffusion model; and $L_{charbonnier}$, which constrains that the decoder's output is aligned with the NDCT of the input, thus balancing denoising with structure preservation. The final loss L_{total} is expressed as:

$$L = L_{MSE} + 0.1L_{commit} + L_{diff} + L_{charbonnier}. \tag{3}$$

3 Experiments

Datasets (1) Piglet Dataset [32]. CT data at 50% dose, 25% dose, and 10% dose are applied to assess the denoising efficacy of various methods. (2) Phantom Dataset. Human anatomy is replicated using phantoms that simulate the head, manufactured by General Electric (GE), USA. This replication ensures that these body phantoms exhibit x-ray attenuation properties during CT scans that are very similar to those of the human body. GE equipment is used to acquire head CT images at 10%, 25%, and standard doses. The GE scanning protocol is as follows: head CT axial scan with a probe range of 40 mm and a scanning time of 1 s; slice thicknesses are configured to be 5 mm and 0.625 mm. Images are scanned at a standard resolution of 512x512. (3) 2016 Mayo Dataset. The dataset described in [4] contains data from ten paired 25% low-dose and normal-dose abdominal CT scans. To demonstrate the capacity for zero-sample migration of our method, we select 3 mm low-dose slices for testing. Scanning conditions are differed from those used in the training dataset.

Table 1. Quantitative results of different methods on the Piglet dataset.

Method	10% dose		25% dose		50% dose		Time(s)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
Low-dose	33.4423	0.8579	34.7077	0.8657	35.2169	0.9092	-
REDCNN	35.8637	0.9198	36.4999	0.9216	36.3088	0.9381	0.1
CTformer	35.7722	0.9054	36.3938	0.9154	36.8441	0.9354	0.2
FDDiff	36.2923	0.9115	36.4673	0.9285	36.7844	0.9345	2.5
Noise2Sim	36.7517	0.9143	36.6136	0.9229	36.8953	0.9391	0.1
ZeroN2N	35.3556	0.8669	35.7389	0.8782	36.1493	0.9012	20
Ours	**36.9639**	**0.9290**	**36.7116**	**0.9337**	**36.9883**	**0.9469**	0.25

Table 2. Quantitative results of different methods on the Phantom dataset.

Method	5mm, 10%dose		1mm, 25%dose		Time(s)
	PSNR	SSIM	PSNR	SSIM	
Low-dose	32.3618	0.8338	33.6624	0.8701	-
REDCNN	36.6297	0.9166	35.8066	0.9027	0.1
CTformer	37.2995	0.9216	35.2483	0.9164	0.2
FDDiff	37.3418	0.9231	35.8423	0.9166	2.5
Noise2Sim	37.0151	0.9176	35.1765	0.9078	0.1
ZeroN2N	34.7649	0.8368	35.1308	0.8807	20
Ours	**37.9611**	**0.9262**	**36.1617**	**0.9173**	0.25

Parameter settings All models were implemented in PyTorch and trained and tested on an NVIDIA 4090 GPU. The dimensions of the codebook were set to 512×16384 [33]. The diffusion model's noise scheduling increased incrementally from 1e-6 to 2e-2, with an initial step count of 1,000. For the experiments with piglets and phantoms, our model training process utilizes only the corresponding NDCT images. For the experiments with Mayo data, the training dataset includes NDCT images from piglets and phantoms but does not contain any human body images. The compared supervised models include REDCNN[5], CTformer[28] and diffusion-based method FDDiff[27]. All supervised methods used fully paired normal-dose and low-dose training data. Self-supervised methods, such as Noise2Sim[22], and ZeroN2N[19], are also included, with training conducted solely on low-dose data. Hyperparameters of the deep learning methods followed the configurations in the original publications or official open-source code. Standard CT quality evaluation metrics, such as SSIM and PSNR, are applied to quantitatively assess model performance.

Performance Comparison on Piglet Dataset. As shown in Figure 2, supervised methods effectively remove noise; however, they result in the blurring of fine structures and loss of information, especially pronounced when noise levels are high. CTformer also shows flocculation artifacts and poor visualization quality. FDDiff shows effective denoising ability, but introduces more streaks and dot artifacts at the 10% dose level. ZeroN2N preserves more image detail

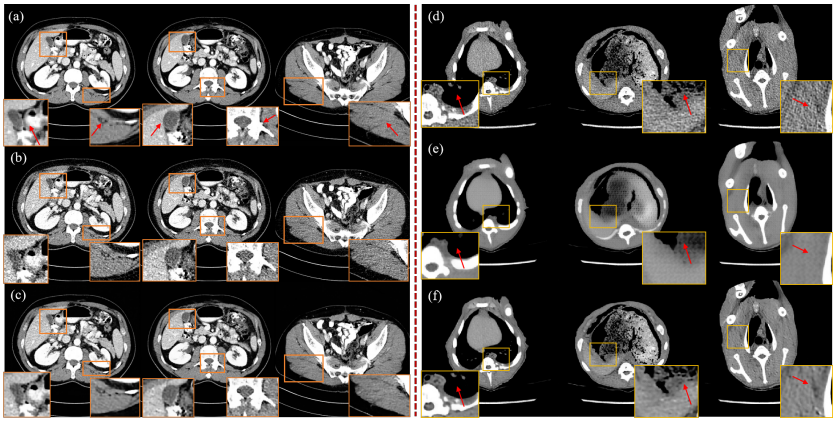

Fig. 3. The left side of the red dashed line displays 3mm Mayo2016 results [-100, 200] HU, arranged from top to bottom as follows: normal dose (a), low dose (b), and denoising results (c). The right side presents the miniature diffusion model ablation experiments: (d) LDCT, (e) without diffusion model and (f) with diffusion model.

and texture, but its denoising effectiveness is limited, with noticeable residual noise. Noise2Sim demonstrates improved denoising without excessive smoothing, showing finer details in the 10% and 50% slices; however, blurring in the central region remains prominent at 25% dose, with significant loss of detail and edge information. Our proposed method, trained exclusively on NDCT images, demonstrated effective denoising across dose levels, providing clear image details, minimal blurring, and preserved contrast. As shown in Tab. 1, our approach outperforms the compared models in the testing metrics under varying degrees of LDCT conditions. Our method is 10 times more efficient when compared to diffusion-based methods.

Performance Comparison on Phantom Dataset. As shown in Figure 2 below the red line, the supervised method exhibit some denoising on slices of varying thickness, but the CTformer still introduced sandy artifacts. FDDiff demonstrated better denoising and preserved details on the 25% dose, 0.625 mm slices; however, the overall visualization is blurrier on the 10% dose, 5 mm slices, with a reduction in contrast, potentially affecting diagnostic assessment of the head slice. Noise2Sim exhibited a little blurring, with poorer visualization on both slices and loss of edge features. ZeroN2N demonstrated limited denoising, though contrast is largely maintained. Our method achieved effective denoising under scanning conditions of unknown thickness and dose, with image texture most similar to that of normal-dose slices, clear contrast, and no blurring of edges or details. As shown in Tab. 2, our method outperforms the compared supervised and self-supervised methods on testing metrics under two different scanning strategies.

Zero-Shot Generalization to Human Datasets. Figure 3 illustrates the denoising results on the Mayo dataset, and the training process of our method uses only NDCT images of piglet and phantom data. This dataset has a slice thickness of 3 mm, and the scanning device differs from that used in our training data. Thus, this task involves unknown device specifications, dose levels, and slice thickness. Our method demonstrated consistently high-quality denoising results: zoomed-in regions retained clear texture, edges showed no blurring artifacts, and the overall images introduced no additional artifactual noise. However, we observed a loss of contrast in some features, including enhanced blood vessel regions. We hypothesize that this is due to the lack of enhanced images in the training data and significant imaging differences between devices, contributing to reduced performance in enhanced regions.

Ablation Study. (1) Miniature diffusion model. As illustrated in Fig. 3, three CT slices with detailed zoom-ins are presented to the right of the red line: the first column shows the original low-dose CT slice, the second shows the denoising result without the model, and the third shows the result with the model. The results indicate that the VQ approach is unable to restore detailed features, and the overall image exhibits arrow-like artifacts. Upon incorporating the miniature diffusion model, the image detail improves significantly, and the arrowhead-like artifacts are eliminated. We posit that due to the rich high-frequency detail in the original slices, it is challenging to fully restore all details using only the VQ approach. Therefore, we incorporate the diffusion model in the subsequent up-sampling of the VQ results, using features from all levels of the up-sampled data as priming conditions to retain detailed features effectively. (2) The dimension of codebook. We investigated the impact of codebook dimensionality on performance by conducting ablation experiments using 1mm, 25% low-dose model data. When the codebook dimension was set to 1024, the PSNR and SSIM values were 36.1171 and 0.9276, respectively. In contrast, when the dimension was reduced to 512, the PSNR and SSIM values are 36.1617 and 0.9173, respectively. The 512-dimensional configuration required 662 MB fewer parameters than the 1024-dimensional configuration, which required 697 MB. These results suggest that the dimensionality of the codebook has a little impact on the experimental outcomes.

4 Conclusion

We propose the self-supervised vector quantization LDCT denoising model, while only utilizing NDCT for training. We employ vector quantization learning to approximate the diverse features of low-dose images under varying scanning conditions. The proposed miniature diffusion model, guided by the up-sampled feature map and prior diffusion model information, effectively restores image details and maintains denoising efficiency without reducing sampling steps, achieving a test speed of 0.25 seconds per image. We have conducted a comprehensive set of experiments using existing public data, our method outperforms supervised and

state-of-the-art self-supervised methods in quantitative metrics and visual quality. Training on phantom and animal data alone enables effective noise removal on human data across varied equipment, dose, and slice thicknesses.

Acknowledgments. This work was funded by the Special Fund of the Hubei Luojia Laboratory under Grant 230100001.

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

References

1. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE Transactions on Medical Imaging **37**(6), 1322–1332 (2018)
2. Alexey, D.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
3. Brenner, D.J., Hall, E.J.: Computed tomography—an increasing source of radiation exposure. New England journal of medicine **357**(22), 2277–2284 (2007)
4. Chen, B., Leng, S., Yu, L., Holmes III, D., Fletcher, J., McCollough, C.: An open library of ct patient projection data. In: Medical Imaging 2016: Physics of Medical Imaging. vol. 9783, pp. 330–335. SPIE (2016)
5. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging **36**(12), 2524–2535 (2017)
6. Chen, S., Zhang, J., Yu, Z., Huang, T.: Exploring efficient asymmetric blind-spots for self-supervised denoising in real-world scenarios. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2814–2823 (2024)
7. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007)
8. Gao, Q., Li, Z., Zhang, J., Zhang, Y., Shan, H.: Corediff: Contextual error-modulated generalized diffusion model for low-dose ct denoising and generalization. IEEE Transactions on Medical Imaging **43**(2), 745–759 (2023)
9. He, J., Yang, Y., Wang, Y., Zeng, D., Bian, Z., Zhang, H., Sun, J., Xu, Z., Ma, J.: Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction. IEEE Transactions on Medical Imaging **38**(2), 371–382 (2019)
10. Ho, J., Jain, A., Abbeel, P.: Advances in neural information processing systems. vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020)
11. Hu, D., Zhang, Y., Liu, J., Luo, S., Chen, Y.: Dior: Deep iterative optimization-based residual-learning for limited-angle ct reconstruction. IEEE Transactions on Medical Imaging **41**(7), 1778–1790 (2022)
12. Huang, H., Zhang, C., Zhao, L., Ding, S., Wang, H., Wu, H.: Self-supervised medical image denoising based on wista-net for human healthcare in metaverse. IEEE Journal of Biomedical and Health Informatics **28**(11), 6329–6337 (2024)
13. Karageorgos, G.M., Zhang, J., Peters, N., Xia, W., Niu, C., Paganetti, H., Wang, G., De Man, B.: A denoising diffusion probabilistic model for metal artifact reduction in ct. IEEE Transactions on Medical Imaging (2024)

14. Li, H., Yang, X., Yang, S., Wang, D., Jeon, G.: Transformer with double enhancement for low-dose ct denoising. IEEE Journal of Biomedical and Health Informatics **27**(10), 4660–4671 (2023)
15. Liu, X., Xie, Y., Liu, C., Cheng, J., Diao, S., Tan, S., Liang, X.: Diffusion probabilistic priors for zero-shot low-dose ct image denoising. Medical Physics **52**(1), 329–345 (2025)
16. Liu, Y., Chen, G., Pang, S., Zeng, D., Ding, Y., Xie, G., Ma, J., He, J.: Cross-domain unpaired learning for low-dose ct imaging. IEEE Journal of Biomedical and Health Informatics **27**(11), 5471–5482 (2023)
17. Lu, Y., Xu, Z., Hyung Choi, M., Kim, J., Jung, S.W.: Cross-domain denoising for low-dose multi-frame spiral computed tomography. IEEE Transactions on Medical Imaging **43**(11), 3949–3963 (2024)
18. Ma, J., Huang, J., Feng, Q., Zhang, H., Lu, H., Liang, Z., Chen, W.: Low-dose computed tomography image restoration using previous normal-dose scan. Medical physics **38**(10), 5713–5731 (2011)
19. Mansour, Y., Heckel, R.: Zero-shot noise2noise: Efficient image denoising without any data. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14018–14027 (2023)
20. Mendrik, A.M., Vonken, E.J., Rutten, A., Viergever, M.A., van Ginneken, B.: Noise reduction in computed tomography scans using 3-d anisotropic hybrid diffusion with continuous switch. IEEE transactions on medical imaging **28**(10), 1585–1594 (2009)
21. Niu, C., Li, M., Fan, F., Wu, W., Guo, X., Lyu, Q., Wang, G.: Noise suppression with similarity-based self-supervised deep learning. IEEE Transactions on Medical Imaging **42**(6), 1590–1602 (2023)
22. Niu, C., Li, M., Fan, F., Wu, W., Guo, X., Lyu, Q., Wang, G.: Noise suppression with similarity-based self-supervised deep learning. IEEE Transactions on Medical Imaging **42**(6), 1590–1602 (2023)
23. van den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning. In: International Conference on Neural Information Processing Systems. p. 6309–6318. NIPS'17 (2017)
24. Pan, Y., Liu, X., Liao, X., Cao, Y., Ren, C.: Random sub-samples generation for self-supervised real image denoising. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12116–12125 (2023)
25. Saidulu, N., Muduli, P.R., Dasgupta, A.: Rhlnet: Robust hybrid loss-based network for low-dose ct image denoising. IEEE Transactions on Instrumentation and Measurement pp. 1–1 (2024)
26. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021)
27. Su, B., Dong, P., Hu, X., Wang, B., Zha, Y., Wu, Z., Wan, J.: Fast and detail-preserving low-dose ct denoising with diffusion model. Biomedical Signal Processing and Control **105**, 107580 (2025)
28. Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., Yu, H.: Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising. Physics in Medicine & Biology **68**(6), 065012 (2023)
29. Wu, D., Kim, K., El Fakhri, G., Li, Q.: Iterative low-dose ct reconstruction with priors trained by artificial neural network. IEEE Transactions on Medical Imaging **36**(12), 2479–2486 (2017)
30. Xiang, W., Yang, H., Huang, D., Wang, Y.: Denoising diffusion autoencoders are unified self-supervised learners. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 15756–15766 (2023)

31. Yang, L., Li, Z., Ge, R., Zhao, J., Si, H., Zhang, D.: Low-dose ct denoising via sinogram inner-structure transformer. IEEE Transactions on Medical Imaging **42**(4), 910–921 (2022)
32. Yi, X., Babyn, P.: Sharpness-aware low-dose ct denoising using conditional generative adversarial network. Journal of digital imaging **31**, 655–669 (2018)
33. Yu, J., Li, X., Koh, J.Y., Zhang, H., Pang, R., Qin, J., Ku, A., Xu, Y., Baldridge, J., Wu, Y.: Vector-quantized image modeling with improved vqgan. In: International Conference on Learning Representations (2022)
34. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 International Conference on Computer Vision. pp. 479–486 (2011)