Meta-Learning-Driven CT Morphology Disentangled Diffusion Model for Multi-Region SPECT Attenuation Correction

Haoran Yang[1], Jiansong Fan[1], Lihua Li[3], and Xiang Pan[1,2,*]

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China
[2] The PRC Ministry of Education Engineering Research Center of Intelligent Technology for Healthcare, Wuxi, Jiangsu 214122, China
[3] Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, China
[*]Corresponding author: xiangpan@jiangnan.edu.cn

Abstract. SPECT imaging faces persistent challenges from soft-tissue attenuation artifacts in clinical practice. While CT-based correction remains the clinical reference standard, associated radiation risks and infrastructure requirements limit its widespread adoption. To address this, we propose a Meta-Learning-Driven CT Morphology Disentangled Diffusion Model (MetaMorph-Diff), which achieves CT-independent attenuation correction. First, we design a Morphological Structure-Attentive Fusion module that explicitly guides the diffusion process using CT-derived anatomical priors. During training, its Morpho-Attentive Alignment submodule establishes voxel-level physical constraints between SPECT features and attenuation distributions by leveraging CT anatomical priors. During inference, its Morpho-Disentangling Gate achieves complete disentangling from CT dependencies through learned morphological embeddings. Crucially, the model uses only SPECT images during inference to achieve accurate attenuation correction without relying on CT data. Second, we propose a multi-region adaptive meta-learning strategy, which enhances cross-anatomical generalization capability by optimizing model initialization parameters, enabling a single model to achieve consistent and accurate correction across diverse anatomical regions. Our method surpasses existing approaches with higher-precision attenuation distribution prediction and stronger multi-region correction adaptability. The code is available at https://github.com/yhr1020/MetaMorph-Diff.

Keywords: Attenuation Correction · Diffusion Model · Meta-Learning · SPECT/CT

1 Introduction

Single photon emission computed tomography (SPECT) is widely used for diagnostic purposes across multiple anatomical regions, such as diagnosing brain tumors, epilepsy, and stroke, diagnosing and detecting recurrences of differentiated thyroid cancer, as well as assessing coronary artery disease and myocardial injury [14]. In clinical applications, SPECT imaging is susceptible to soft tissue attenuation artifacts, leading to false positives or false negatives, thus reducing diagnostic accuracy [20]. To address this issue, the most commonly used method is attenuation correction using computed tomography (CT) [1, 7]. Although CT plays a crucial role in attenuation correction, the X-ray radiation it uses poses a certain radiation risk to patients. Additionally, the high cost of SPECT/CT scanners limits the widespread use of SPECT in certain clinical settings, especially in resource-limited areas.

Previous works have attempted to address this issue using deep learning-based generative networks. Sakaguchi et al. [18] employed a CNN-based AutoEncoder for brain SPECT, while Chen et al. [2] developed a 3D Dual Squeeze Excitation Residual Network for cardiac applications. Shi et al. [19] utilized 3D cGANs to estimate myocardial attenuation maps. However, these approaches remain confined to single anatomical regions and crucially neglect CT morphological priors during training, resulting in physically inconsistent reconstructions with compromised anatomical fidelity.

Recently, diffusion models have been shown to generate better sample quality than state-of-the-art GANs [4], with applications spanning style transfer [15], image super-resolution [12], among others. In the medical field, they have been utilized in scenarios such as anomaly detection, medical image segmentation, denoising, registration, and generation [10]. However, traditional diffusion models are typically unconditional or based on simple class-conditioned controls, making it difficult to effectively extract multi-modal features from input images and achieve accurate multi-region medical image translation.

In this paper, we propose a meta-learning-driven CT morphology disentangled diffusion model for multi-region attenuation correction. Firstly, we design a Morphological Structure-Attentive Fusion (MSAF) module that explicitly guides the diffusion process using CT-derived anatomical priors. During training, its Morpho-Attentive Alignment (MA) submodule hierarchically extracts CT image features and establishes voxel-level physical constraints between SPECT features and attenuation distributions by leveraging CT anatomical priors. During inference, the Morpho-Disentangling Gate (MDG) achieves complete disentangling from CT dependencies through learned morphological structure embeddings. Secondly, we develop a Multi-region Adaptive Meta-Learning strategy (MAML) to enable the model to adapt to attenuation correction tasks across diverse anatomical regions, thereby enhancing its generalization capability across multi-anatomical sites. Finally, we design and fine-tune a hybrid loss function tailored to the structural information requirements and voxel-level accuracy demands of SPECT attenuation correction. Experiments show our method achieves superior quantitative/qualitative performance and adaptability in multi-region tasks compared to existing approaches.

2 Methodology

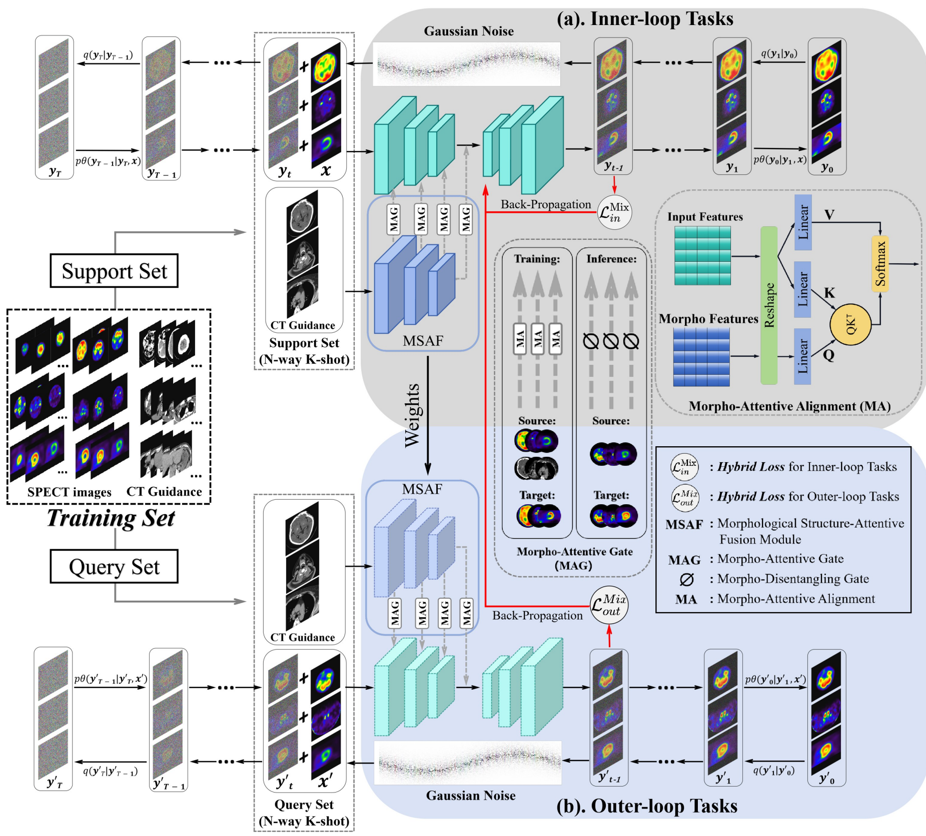

Fig. 1. The MetaMorph-Diff framework. The training set is divided into multiple support sets and query sets, with k-shot samples in each batch, where y_i/y_i' represents the attenuation-corrected SPECT images in support/query sets, and x/x' represents the conditional input of non-corrected SPECT images in support/query sets.

Our model consists of two main components: the Multi-Region Adaptive Meta-Learning-Driven Diffusion Model and the Morphological Structure-Attentive Fusion (MSAF) module. To more clearly illustrate the proposed architecture, Figure 1 depicts the overall framework of MetaMorph-Diff.

2.1 Morphological Structure-Attentive Fusion Module

Our study proposes an MSAF module to achieve CT morphological feature fusion while disentangling CT morphological inputs during inference. The MSAF module comprises two core components: a morphological encoder and a Morpho-Attentive Gate module (MAG).

The morphological encoder $f_{\text{Morph}}(\cdot)$ hierarchically extracts structural features from CT images, generating multi-scale representations $\{F^l_{\text{CT}}\}_{l=1}^L$, where L denotes the depth of the encoder. These features are fused with SPECT functional features $\{F^l_{\text{SPECT}}\}_{l=1}^L$ from the diffusion network's encoder $f_{\text{SPECT}}(\cdot)$ through MAG. The MAG operates in two distinct modes:

Morpho-Attentive Alignment. The MA establishes structural correspondence during training phase through cross-modal attention [16]:

$$F^l_{\text{fusion}} = \text{Softmax}\left(\frac{Q^l_{\text{SPECT}}(K^l_{\text{CT}})^\top}{\sqrt{d}}\right) V^l_{\text{CT}} \tag{1}$$

Where the query matrix is derived from SPECT features ($Q^l_{\text{SPECT}} = W_Q F^l_{\text{SPECT}}$, with $W_Q \in \mathbb{R}^{d\times d}$ being a learnable projection matrix), while key and value matrices derive from CT morphological features ($K^l_{\text{CT}} = W_K F^l_{\text{CT}}$, $V^l_{\text{CT}} = W_V F^l_{\text{CT}}$). Here, d denotes the feature dimension for scaling the dot-product attention to prevent gradient saturation. This cross-modal interaction enables the network to learn physically constrained attenuation mappings during the training phase.

Morpho-Disentangling Gate. During the inference phase, the MDG achieves complete CT dependency disentangling through a single-modal attention mechanism:

$$F^l_{\text{fusion}} = \text{Softmax}\left(\frac{Q^l_{\text{SPECT}}(K^l_{\text{SPECT}})^\top}{\sqrt{d}}\right) V^l_{\text{SPECT}} \tag{2}$$

All attention components (Q^l_{SPECT}, K^l_{SPECT}, V^l_{SPECT}) are endogenously derived from SPECT features. Leveraging morphological priors learned during the training phase, the model maintains anatomical consistency without requiring CT input and rapidly adapts to CT-free attenuation correction tasks through a meta-learning strategy.

2.2 Multi-Region Adaptive Meta-Learning-Driven Diffusion Model

Diffusion Model. The proposed method builds upon the standard diffusion framework [8], which defines forward and reverse Markov chains over T timesteps. The forward process gradually adds Gaussian noise to attenuation-corrected images y_0 through:

$$q(y_t|y_{t-1}) = \mathcal{N}(y_t; \sqrt{1-\beta_t}y_{t-1}, \beta_t\mathbf{I}) \tag{3}$$

where β_t controls the noise schedule.

For conditional generation, we adopt SR3's architecture [17] that concatenates non-corrected SPECT images x with noisy samples y_t as input. The reverse process learns to iteratively denoise y_T via:

$$p_\theta(y_{t-1}|y_t, x) = \mathcal{N}\left(y_{t-1} \mid \mu_\theta(x, y_t, t), \Sigma_\theta(x, y_t, t)\right) \tag{4}$$

Multi-Region Adaptive Meta-Learning. To improve the model's stability for attenuation correction without CT across different anatomical regions, we employ the MAML for optimization. Each anatomical region's correction is defined as a task with k-shot samples [6]. Initializing parameters θ_0, we perform G-step inner-loop updates on support set $S^j_{AC_i}$ from training tasks T_{train} (Fig. 1(a)):

$$\theta_j^G = \theta_{j-1}^G - \alpha \nabla_\theta \mathcal{L}_{S^j_{AC}}(F_{\theta^G_{j-1}}) \tag{5}$$

where α denotes the inner-loop learning rate. The outer-loop optimization evaluates θ_0 over J-scale tasks using query sets Q^j_{AC} (Fig. 1(b)):

$$\theta_0 = \theta_0 - \beta \nabla_\theta \sum_{j=1}^{J} \mathcal{L}_{Q^j_{AC}}(F_{\theta^G_j}(\theta_0)) \tag{6}$$

Where β is the outer-loop learning rate, and $\mathcal{L}_{Q^j_{AC}}$ is the loss function.

2.3 Hybrid Loss Function

Due to the high demand for image detail preservation in the attenuation correction task, we designed and fine-tuned a hybrid loss function. Specifically, we combined Multi-Scale Structural Similarity Index (MS-SSIM) and L2 loss, mixing them with appropriate weight ratios to effectively balance the retention of image structural information with the control of smoothness. While preserving high-frequency details of the image, this method effectively suppresses low-frequency noise, thereby improving the image quality and reconstruction accuracy. The loss function is expressed as:

$$\mathcal{L}^{\text{Mix}} = \alpha \cdot \mathcal{L}^{\text{MS-SSIM}} + (1-\alpha) \cdot G_{\sigma^G_M} \cdot \mathcal{L}^{l_2} \tag{7}$$

Where $\mathcal{L}^{\text{MS-SSIM}}$ is the multi-scale structural similarity loss [22], and $G_{\sigma^G_M}$ approximates the MS-SSIM pyramid structure by applying different Gaussian smoothing parameters σ_G on the full-resolution image, thereby reducing computational overhead and capturing multi-scale information. $\mathcal{L}^{l_2}$ is the standard l_2 loss. We experimented with and fine-tuned the parameter α to control the balance between the two loss functions.

3 Experiments and Results

3.1 Dataset and Implementation Details

Dataset. In this study, we used an in-house dataset from a hospital in China. The dataset comprises 821 brain perfusion cases, 810 thyroid cases, and 814 myocardial perfusion cases. All images were acquired using the Philips Precedence 16 SPECT/CT system (Philips, Netherlands), equipped with a low-energy general-purpose parallel-hole collimator with a peak energy of 140 keV and a window width of 20%. The radiopharmaceuticals used for brain perfusion, thyroid,

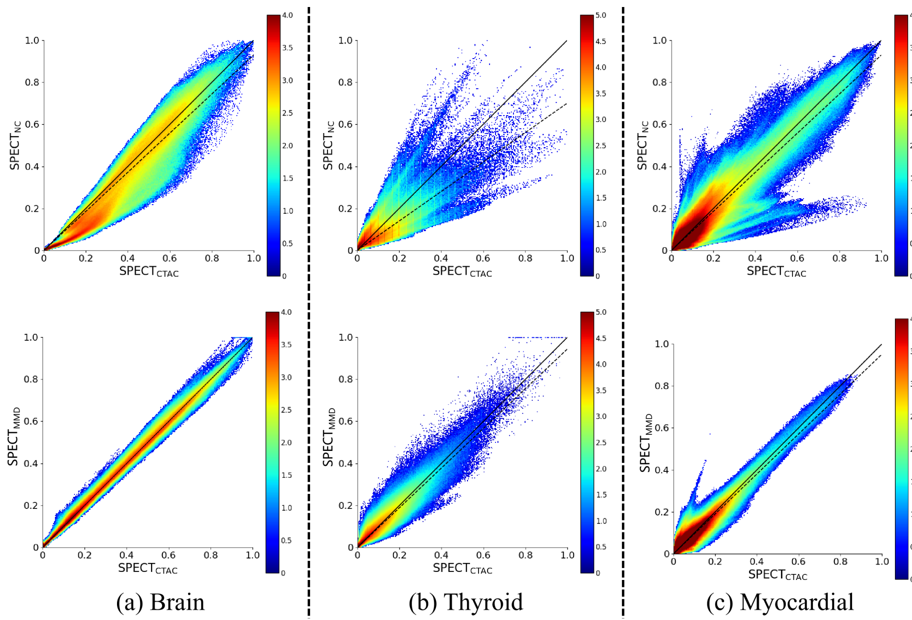

Fig. 2. The Joint Histogram of Voxels presents the voxel value distributions between the non-corrected image ($\mathrm{SPECT_{NC}}$), the image generated by our MetaMorph-Diff ($\mathrm{SPECT_{MMD}}$), and the CT-based attenuation-corrected image ($\mathrm{SPECT_{CTAC}}$). The voxel values are normalized to the range [0, 1] and undergo a logarithmic transformation.

and myocardial perfusion imaging were $^{99}\mathrm{Tc^m}$-ECD 925∼1110 MBq, $^{99}\mathrm{Tc^m O_4}$ 555∼740 MBq, and $^{99}\mathrm{Tc^m}$-MIBI 925∼1110 MBq, respectively. Each sample includes a set of paired images, specifically the original non-corrected SPECT image, the corresponding CT image, and the attenuation-corrected SPECT image.

Implementation Details. Based on the NVIDIA RTX A6000 GPU, we implemented our network using the PyTorch framework. The Adam optimizer was employed with an initial learning rate set to 2×10^{-4} for updating the network parameters. The size of all input images were adjusted to 128×128 pixels.

3.2 Results

In this study, we used several quantitative metrics to evaluate the performance of the proposed model: Structural Similarity Index (SSIM) [21], Peak Signal-to-Noise Ratio (PSNR), Mean Absolute Error (MAE), Pearson Correlation Coefficient (PCC) [13], and Euclidean Distance (ED) [5]. Additionally, we present the voxel-level joint histogram (see Fig. 2). The joint histogram of voxels indicates

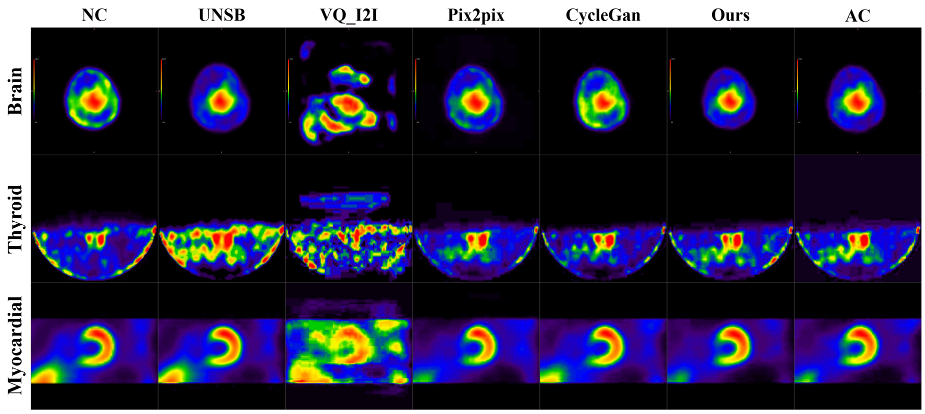

Fig. 3. Qualitative results of different methods. NC represents the non-corrected image, while AC refers to the attenuation-corrected image using CT.

that the voxel distribution of $\text{SPECT}_{\text{MMD}}$ closely matches the voxel distribution of the ground truth image, demonstrating its greater accuracy in voxel-level mapping.

We compared MetaMorph-Diff with four state-of-the-art image translation methods [3, 9, 11, 23] to evaluate its performance in multi-region attenuation correction. Additionally, we integrated a meta-learning framework [6] into two comparison models to assess the effectiveness of combining it with our base model. Finally, we analyzed MetaMorph-Diff's performance across different k-shot settings to identify the optimal k value. All experiments were performed with the same dataset and setup for consistency and fairness.

We present the quantitative results of all comparison experiments in Table 1. The experiments demonstrate that our MetaMorph-Diff ($k = 5$) consistently outperforms all baseline methods across all evaluation metrics. To further validate the statistical significance of these results, we conducted one-way ANOVA tests using IBM SPSS Statistics 27.0. The analysis revealed that all metrics yielded P-values less than 0.01, indicating statistically significant differences among the models ($P < 0.05$ was considered significant).

Fig. 3 illustrates the generation results of different methods. UNSB [11] and CycleGAN [23] show better generation results for certain specific sites, but their generation performance is unstable for other sites. VQ-I2I [3] shows suboptimal generation, struggling to achieve precise image translation and exhibiting noticeable distortion. While Pix2pix [9] performs similarly to our model in terms of structural similarity, its translation performance in the regions of interest remains less accurate, particularly in handling detailed areas where it lacks sufficient precision.

Table 1. Quantitative results of different methods(Mean±Std)

Methods	**Setting**	**SSIM**$\uparrow$	**PSNR**$\uparrow$(dB)	**MAE**$\downarrow$ ($\times 10^{-2}$)	**PCC**$\uparrow$ ($\times 10^{-2}$)	**ED**$\downarrow$
UNSB	No MAML	0.85 ± 0.09	28.87 ± 4.38	3.37 ± 3.42	93.03 ± 4.43	7.82 ± 4.64
VQ_I2I	No MAML	0.60 ± 0.15	20.48 ± 5.39	9.38 ± 3.51	72.32 ± 8.96	9.58 ± 7.83
Pix2Pix	No MAML	0.92 ± 0.04	28.63 ± 3.85	2.24 ± 1.52	84.02 ± 3.75	3.93 ± 1.43
	3-way 5-shot	0.92 ± 0.04	29.43 ± 2.74	1.58 ± 0.63	96.27 ± 2.51	2.58 ± 1.68
CycleGan	No MAML	0.79 ± 0.08	27.61 ± 2.83	2.43 ± 1.14	86.45 ± 12.54	3.81 ± 1.81
	3-way 5-shot	0.91 ± 0.03	34.47 ± 0.39	2.07 ± 0.57	96.66 ± 1.86	2.08 ± 0.38
Ours	3-way 1-shot	0.60 ± 0.03	29.56 ± 1.70	2.80 ± 0.49	87.24 ± 7.61	5.02 ± 0.90
	3-way 2-shot	0.77 ± 0.02	32.55 ± 2.28	1.74 ± 0.48	94.28 ± 3.45	3.55 ± 0.93
	3-way 3-shot	0.90 ± 0.03	34.88 ± 3.34	1.33 ± 0.58	96.32 ± 2.57	2.93 ± 1.14
	3-way 4-shot	0.95 ± 0.02	39.51 ± 3.66	0.83 ± 0.43	98.69 ± 1.11	1.75 ± 0.84
	3-way 5-shot	**0.96 ± 0.02**	**41.42 ± 3.74**	**0.63 ± 0.32**	**98.84 ± 1.04**	**1.00 ± 0.58**
	3-way 6-shot	0.96 ± 0.02	40.92 ± 4.38	0.67 ± 0.54	98.60 ± 1.34	1.08 ± 0.61

Table 2. Results of ablation study(Mean±Std)

MSAF	**Hybrid Loss**	**MAML**	**SSIM**$\uparrow$	**PSNR**$\uparrow$**(dB)**	**MAE**$\downarrow$ ($\times 10^{-2}$)	**PCC**$\uparrow$ ($\times 10^{-2}$)	**ED**$\downarrow$
×	✓	✓	0.92 ± 0.01	39.38 ± 2.74	0.81 ± 0.23	98.34 ± 1.16	1.59 ± 0.51
✓	✓	×	0.92 ± 0.03	36.53 ± 4.11	1.16 ± 0.59	97.81 ± 1.86	2.59 ± 1.02
✓	×	✓	0.93 ± 0.02	38.64 ± 4.47	0.90 ± 0.54	98.51 ± 1.69	1.87 ± 1.03
✓	✓	✓	**0.96 ± 0.02**	**41.42 ± 3.74**	**0.63 ± 0.32**	**98.84 ± 1.04**	**1.00 ± 0.58**

3.3 Ablation Study

To further investigate the importance of the design components in MetaMorph-Diff, we systematically analyzed the impact of removing or modifying three modules: MSAF, MAML, and the Hybrid Loss Function.

For the MSAF, we removed the MSAF module from the network and replaced the inter-layer connections between convolutional outputs with self-attention mechanisms. For the MAML module, we removed the meta-learning framework and instead trained the base model separately for each site. For the Hybrid Loss module, we replaced the original hybrid loss calculation with mean squared error (MSE) loss function.

As shown in Table. 2, the ablation study results indicate that our modules play a crucial role in enhancing the model's performance.

4 Conclusion and Future Work

We propose a meta-learning-driven CT morphology disentangled diffusion model for multi-region SPECT attenuation correction, overcoming the reliance on CT scans. Our experimental results show superior performance in both quantitative and qualitative assessments, with generated images demonstrating enhanced quality, stability, and voxel consistency. This approach offers a promising solution for attenuation correction in resource-limited settings, reducing the need for costly and radiation-prone CT scans while maintaining diagnostic accuracy. Future work will focus on developing a foundational model with improved generalization capability for attenuation correction across diverse anatomical regions.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China under grants W2411054, U21A20521 and 62271178, the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX23_2524, National Foreign Expert Project of China under Grant G2023144009L, Zhejiang Provincial Natural Science Foundation of China (LR23F010002), Wuxi Health Commission Precision Medicine Project (J202106), Jiangsu Provincial Six Talent Peaks Project (YY-124), and Major Projects of Wuxi Health Commission (Z202324).

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

References

1. Blankespoor, S., Xu, X., Kaiki, K., Brown, J., Tang, H., Cann, C., Hasegawa, B.: Attenuation correction of spect using x-ray ct on an emission-transmission ct system: myocardial perfusion assessment. IEEE Transactions on Nuclear Science **43**(4), 2263–2274 (1996)
2. Chen, X., Zhou, B., Xie, H., Shi, L., Liu, H., Holler, W., Lin, M., Liu, Y.H., Miller, E.J., Sinusas, A.J., et al.: Direct and indirect strategies of deep-learning-based attenuation correction for general purpose and dedicated cardiac spect. European journal of nuclear medicine and molecular imaging **49**(9), 3046–3060 (2022)
3. Chen, Y.J., Cheng, S.I., Chiu, W.C., Tseng, H.Y., Lee, H.Y.: Vector quantized image-to-image translation. In: European Conference on Computer Vision. pp. 440–456. Springer (2022)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
5. Dokmanic, I., Parhizkar, R., Ranieri, J., Vetterli, M.: Euclidean distance matrices: essential theory, algorithms, and applications. IEEE Signal Processing Magazine **32**(6), 12–30 (2015)
6. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017)
7. Hayashi, M., Deguchi, J., Utsunomiya, K., Yamada, M., Komori, T., Takeuchi, M., Kanna, K., Narabayashi, I.: Comparison of methods of attenuation and scatter correction in brain perfusion spect. Journal of Nuclear Medicine Technology **33**(4), 224–229 (2005)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
9. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
10. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D.: Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis **88**, 102846 (2023)
11. Kim, B., Kwon, G., Kim, K., Ye, J.C.: Unpaired image-to-image translation via neural schr\" odinger bridge. arXiv preprint arXiv:2305.15086 (2023)
12. Kim, J., Kim, T.K.: Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9202–9211 (2024)

13. Lee Rodgers, J., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. The American Statistician **42**(1), 59–66 (1988)
14. Mariani, G., Bruselli, L., Kuwert, T., Kim, E.E., Flotats, A., Israel, O., Dondi, M., Watanabe, N.: A review on the clinical uses of spect/ct. European journal of nuclear medicine and molecular imaging **37**, 1959–1985 (2010)
15. Qi, T., Fang, S., Wu, Y., Xie, H., Liu, J., Chen, L., He, Q., Zhang, Y.: Deadiff: An efficient stylization diffusion model with disentangled representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8693–8702 (2024)
16. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
17. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE transactions on pattern analysis and machine intelligence **45**(4), 4713–4726 (2022)
18. Sakaguchi, K., Kaida, H., Yoshida, S., Ishii, K.: Attenuation correction using deep learning for brain perfusion spect images. Annals of Nuclear Medicine **35**, 589–599 (2021)
19. Shi, L., Onofrey, J.A., Liu, H., Liu, Y.H., Liu, C.: Deep learning-based attenuation map generation for myocardial perfusion spect. European Journal of Nuclear Medicine and Molecular Imaging **47**, 2383–2395 (2020)
20. Singh, B., Bateman, T.M., Case, J.A., Heller, G.: Attenuation artifact, attenuation correction, and the future of myocardial perfusion spect. Journal of Nuclear Cardiology **14**, 153–164 (2007)
21. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
22. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. IEEE Transactions on computational imaging **3**(1), 47–57 (2016)
23. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)