

A Review on Low-Dose Emission Tomography Post-Reconstruction Denoising with Neural Network Approaches

Alexandre Bousse, *Member, IEEE*, Venkata Sai Sundar Kandarpa, Kuangyu Shi, *Member, IEEE*, Kuang Gong, *Member, IEEE*, Jae Sung Lee, *Senior Member, IEEE*, Chi Liu, *Senior Member, IEEE*, Dimitris Visvikis, *Fellow, IEEE*

Abstract—Low-dose emission tomography (ET) plays a crucial role in medical imaging, enabling the acquisition of functional information for various biological processes while minimizing the patient dose. However, the inherent randomness in the photon counting process is a source of noise which is amplified in low-dose ET. This review article provides an overview of existing post-processing techniques, with an emphasis on deep neural network (NN) approaches. Furthermore, we explore future directions in the field of NN-based low-dose ET. This comprehensive examination sheds light on the potential of deep learning in enhancing the quality and resolution of low-dose ET images, ultimately advancing the field of medical imaging.

Index Terms—Low-Dose, PET, SPECT, Deep Learning

ACRONYMS

3-D 3-dimensional
AI artificial intelligence
CDM conditional diffusion model
cGAN conditional GAN
CNN convolutional neural network
CT computed tomography
CycleGAN cycle-consistent GAN
DAE deep denoising autoencoder
DIP deep image prior
DM diffusion model
EM expectation-maximization
ET emission tomography
FBP filtered-backprojection
FCN fully convolutional network
GAN generative adversarial network

LOR line of response
MBIR model-based iterative reconstruction
MP myocardial perfusion
MR magnetic resonance
MRI MR imaging
MSA multi-head self-attention
MSE mean squared error
N2N Noise2Noise
NLM nonlocal means
NN neural network
OOD out-of-distribution
OSEM ordered-subset EM
PDF probability distribution function
PET positron emission tomography
PML penalized maximum log-likelihood
PMT photomultiplier tube
PSF point spread function
PVC partial volume correction
RED regularization by denoising
ResNet residual neural network
SA self-attention
SPECT single-photon emission computed tomography
SSIM structural similarity index measure
SSL self-supervised learning
ViT vision transformer
W-GAN Wasserstein GAN
WB whole-body
WT wavelet transform
XCAT extended cardiac-torso

This work did not involve human subjects or animals in its research.

This work was supported by the French National Research Agency (ANR) under grant No ANR-20-CE45-0020, the National Institutes of Health (NIH) under grant No R01EB025468, the Germaine de Stael Program and the Swiss National Science Foundation under grant number 188350

A. Bousse and V. S. S. Kandarpa equally contributed to this work.

A. Bousse, V. S. S. Kandarpa and D. Visvikis are with Univ. Brest, LATIM, INSERM UMR 1101, 29238 Brest, France.

K. Shi is with Lab for Artificial Intelligence & Translational Theranostics, Dept. Nuclear Medicine, Inselspital, University of Bern, 3010 Bern, Switzerland.

K. Gong is with The Center for Advanced Medical Computing and Analysis, Massachusetts General Hospital/Harvard Medical School, Boston, MA 02114, USA

C. Liu is with Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT, USA

J. S. Lee is with Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul 03080, Korea

Corresponding authors: A. Bousse, bousse@univ-brest.fr

I. INTRODUCTION

THE main components of ET are positron emission tomography (PET) and single-photon emission computed tomography (SPECT). They measure the radio-tracer distribution administered to the patient via gamma radiation arising from radioactive decay and have multiple use cases including oncology, cardiology, neurology, etc. The ability to get functional information on the various biological processes distinguish them from other imaging modalities such as MR imaging (MRI) and computed tomography (CT).

Radioactive decay is a random process which entails the difficulty in precise production of the images in ET. Noise or the speckled variation in ET images is caused by the inherent

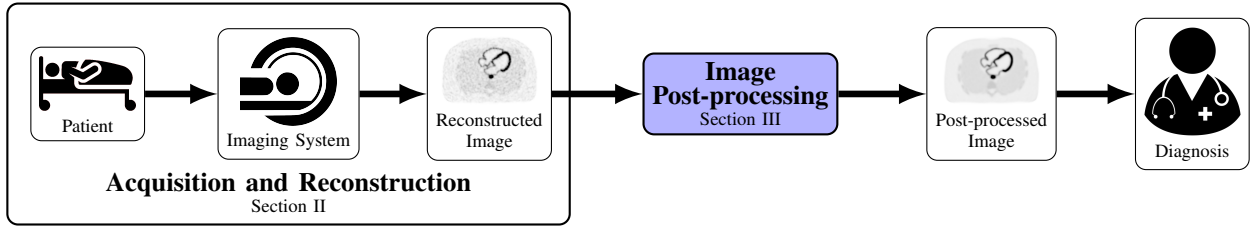


Fig. 1. Main steps in clinical ET. The patient is initially administered with a radioactive tracer. The emission is captured by an imaging system. The mapping of the raw detector data to an image is performed by an image reconstruction method. Finally, after post-processing, the image is used for diagnosis.

randomness of the photon counting process. In order to ensure patient safety, research has been extensively conducted in the regime of low-dose ET imaging [1], [2]. The reduction in the dose administered to the patient further adds to the challenge of obtaining a clear image.

ET images are reconstructed from the measured gamma rays, which is an ill-posed inverse problem subject to noise amplification. The images suffer from partial volume effects due to the low intrinsic resolution of the imaging systems, as well as positron range for PET. The resolution of the reconstructed images can be improved with model-based iterative reconstruction (MBIR) that incorporate the point spread function (PSF) in the system matrix. However, this further contributes to the ill-posedness of the inverse problem, which results in more noise. Hence, a number of post-processing techniques have been proposed in this regard. Artificial intelligence (AI) and more specifically deep learning-based methods (i.e., deep NNs) have been very effective in denoising and super resolution and hence have been explored at length in ET. The focus of this article is to discuss at length the NN approaches that have been proposed for image denoising in low-dose ET.

The main steps in the clinical application of ET, are highlighted in Fig. 1. The first step is the production of the radiopharmaceutical followed by its administration to the patient. Following this the emission data is acquired by a detector setup. Using the physical forward model, image reconstruction methods are utilized to map the raw detector data to an image. These steps are summarized in Section II. Following reconstruction, the image can be post-processed to reduce the noise and improve resolution. Section III gives an overview of existing post-processing techniques for low-dose ET image post-processing, with an emphasis on NN-based approaches, and is the main contribution of this paper. Finally, Section IV covers future directions of NN-based low-dose ET.

II. DATA ACQUISITION AND IMAGE RECONSTRUCTION

A. Data Acquisition

SPECT imaging is based on the emission of a single gamma photon per each radioactive decay event. These gamma photons are usually collected by a gamma camera [3], which typically consists of a collimator that selects relevant gamma photons to be detected; the gamma photons are converted into light in the visible spectra by the scintillation crystal. The optical-wavelength photons are sent to a photomultiplier tube (PMT) that converts light into electrons, generating a

detectable current. This current is then measured by an electronics setup, noting the occurrence of the event. The relative spatial coordinates of the event are determined by measurements from the point of contact in the PMTs. These events are stored in histograms based on their position, resulting in discrete (or vectorized) projection measurement data which is then utilized for image reconstruction.

PET on the other hand uses positron emitting radioisotope. The positron interacts with an electron resulting in an annihilation event that produces two gamma photons moving in nearly opposite direction. These gamma photons are simultaneously detected (coincidence event) by circularly arranged detector elements. Due to this inherent feature of PET, collimators are not present in the detector system. The coordinates of each decay event are recorded, through the detection of the corresponding coinciding pair of gamma photons. The binning of these detected coincidence events results in the projection data, used for reconstruction.

ET image reconstruction and post-processing requires a formalism that we briefly describe here. The radiotracer distribution takes the form of an image vector $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{R}^m$, ‘ T ’ denoting the matrix transposition, where each entry x_j denotes the radiotracer concentration at voxel j (in Bq per voxel). In both SPECT and PET, the imaging system is modeled by a system matrix $\mathbf{P} \in \mathbb{R}^{n \times m}$ where for all (i, j) each entry $[\mathbf{P}]_{i,j}$ is the probability that an emission in voxel j is detected along the i th line of response (LOR), taking into account the geometry of the system, the linear attenuation, the sensitivity of the detectors and the intrinsic resolution. For each LOR $i = 1, \dots, n$, the expected number of detections given a radiotracer distribution \mathbf{x} is

$$\bar{y}_i(\mathbf{x}) = \tau[\mathbf{P}\mathbf{x}]_i + r_i \quad (1)$$

where τ is the scan duration and r_i is a background term comprising expected scatter as well as randoms (for PET), and the number of detection is a random variable y_i that follows a Poisson distribution centered in $\bar{y}_i(\mathbf{x})$, i.e.,

$$y_i \sim \text{Poisson}(\bar{y}_i(\mathbf{x})) . \quad (2)$$

In the following we denote $\mathbf{y} = [y_1, \dots, y_n]^T$ and $\bar{\mathbf{y}}(\mathbf{x}) = [\bar{y}_1(\mathbf{x}), \dots, \bar{y}_n(\mathbf{x})]^T$ the measured and expected data respectively, and $\mathbf{r} = [r_1, \dots, r_n]^T$ the background events vector.

B. Reconstruction

Reconstructing an image \mathbf{x}^{rec} corresponds to solving the following inverse problem:

$$\text{finding } \mathbf{x}^{\text{rec}} \text{ s.t. } \mathbf{y} \approx \bar{\mathbf{y}}(\mathbf{x}^{\text{rec}}), \quad (3)$$

This can be achieved by analytical inversion of \mathbf{P} applied to $\frac{1}{\tau}(\mathbf{y} - \mathbf{r})$, also known as filtered-backprojection (FBP) (see for example [4]). Unfortunately, the inverse problem (3) is ill-posed and direct inversion leads to noise amplification which is impractical for low-dose imaging. Moreover, the inversion relies on an idealized model that does not incorporate resolution modeling. Finally, solving (3) does not guarantee positivity of the solution.

Another approach consists in finding an estimate \mathbf{x}^{rec} by penalized maximum log-likelihood (PML)

$$\mathbf{x}^{\text{rec}} \in \arg \min_{\mathbf{x} \in \mathbb{R}_+^m} \ell(\mathbf{y}, \bar{\mathbf{y}}(\mathbf{x})) + \beta R(\mathbf{x}) \quad (4)$$

where $\ell(\mathbf{y}, \bar{\mathbf{y}})$ is the Poisson negative log-likelihood of the expectation $\bar{\mathbf{y}}$ given the measurement \mathbf{y} , R is a penalty, or *prior*, that enforces image smoothness and $\beta > 0$ is a weight. Solving (4) can be achieved by MBIR such as the expectation-maximization (EM) algorithm [5] or ordered-subset EM (OSEM) algorithm [6] in absence of penalty (i.e., $\beta = 0$), and modified EM [7] with a smooth convex penalty. Reconstructing the image \mathbf{x} by PML (4) reduces the noise as compared with analytical inversion (3) by (i) the incorporation of the stochastic model of the noise in L and (ii) by the presence of the penalty that controls the noise. Typical penalty term R includes quadratic smoothness penalty [8], the edge-preserving Huber penalty [9] or the relative difference prior [10]. Anatomical priors have been used to smooth the PET image while preserving image resolution by taking advantage of high-resolution anatomical images such as CT or MRI [11]–[15]. More recently, Sudarshan *et al.* [16], [17] introduced patch-based dictionary learning for joint PET/magnetic resonance (MR) image reconstruction.

PML methods have played a pivotal role in ET image reconstruction. These methods have proven effective in managing noise levels and improving image quality. However, in the context of low-dose ET imaging, striking a suitable balance between noise reduction and preservation of essential image details necessitates careful tuning of the prior weight. Failure to do so may lead to undesired over-smoothing artifacts. Furthermore, anatomically-guided penalties, while valuable in certain cases, can introduce artifacts when there is misalignment between the activity and anatomical images. It is also worth noting that the applicability of PML techniques is contingent upon the availability of raw data, which may not always be obtainable.

III. DEEP LEARNING-BASED IMAGE POST-PROCESSING

When ET raw data are not available, images can be post-processed to improve their quality. Prior to the advent of deep learning, conventional image post-processing methods were employed for incorporating corrections in ET images. The first image post-processing task is to improve the image

resolution, namely, partial volume correction (PVC). PVC first achieved with deconvolution techniques, which consists of correcting for the image PSF using iterative techniques, such as the Van Cittert algorithm [18] or the Richardson-Lucy algorithm [19], [20]. However, deconvolution is an ill-posed inverse problem and leads to noise amplification, which is non-practical for low-dose imaging. Therefore, it is necessary to deploy adequate techniques to control the noise. This topic has been the subject to numerous works over the last decade [21]–[25].

The inherent relationship between resolution and noise poses a significant challenge. Enhancing image resolution typically leads to increased noise levels, while reducing noise tends to compromise resolution. Addressing both aspects simultaneously requires a paradigm shift towards AI-based approaches. NNs have revolutionized imaging ever since the performance of AlexNet in the ImageNet challenge [26]. In medical imaging, they have contributed to image segmentation, cancer detection, registration, reconstruction, etc. [27]. In the context of low-dose ET imaging, they have been widely implemented to bring about improvements in images reconstructed by traditional algorithms.

This section reviews the NN-based low-dose ET denoising methods and categorizing them based on their NN design. We have broadly categorized the methods into supervised methods, self-attention (SA) mechanisms, unsupervised methods, multi-modality (i.e., additional anatomical information from another modality) and diffusion models (DMs). These subsections have further been divided into subcategories to distinguish and highlight specific workings of the approaches. Note that, although SA mechanisms are typically categorized as supervised methods, recognizing their increase in popularity and effectiveness, we have chosen to dedicate a separate section to these mechanisms, with an emphasis on transformers.

A. Supervised Methods

Machine learning methods that require labeled data for training come under this category. Owing to the data revolution and the availability of annotated datasets for different tasks along with competitions, the most popular methods for denoising are supervised.

A reconstructed image \mathbf{x}^{rec} produced by traditional methods like FBP (by solving (3)) or PML (by solving (4)) is processed through an image-to-image NN \mathbf{F}_θ depending of a trained parameter θ^* as

$$\mathbf{x}^{\text{PP}} = \mathbf{F}_{\theta^*}(\mathbf{x}^{\text{rec}}) \quad (5)$$

where \mathbf{x}^{PP} is the post-processed image. Supervised training of θ^* is generally achieved using a training dataset of K noisy/clean image pairs $(\mathbf{x}_k^{\text{noisy}}, \mathbf{x}_k^{\text{clean}})$, $k = 1, \dots, K$, as

$$\theta^* = \arg \min_{\theta} \sum_{k=1}^K L(\mathbf{F}_\theta(\mathbf{x}_k^{\text{noisy}}), \mathbf{x}_k^{\text{clean}}) \quad (6)$$

where $L(\cdot, \cdot)$ is a loss function.

There are many possible variations of the NN \mathbf{F}_θ and its training. This subsection will cover fully convolutional networks (FCNs) and generative adversarial networks (GANs).

1) *Fully Convolutional Networks*: Computer vision tasks like image segmentation, super-resolution, image enhancement, etc., typically utilize FCNs that are based on residual neural network (ResNet) [28] or U-Net [29]. One of the earliest works that used FCNs for denoising in PET was proposed in [30]. The proposed method predicted full-dose PET images from PET images with dosage reduced by 200 times. The authors used a convolutional encoder-decoder-styled architecture with three convolutional layers on the encoder and decoder part of the network. The encoder consisted of convolutions and max-pooling layers while the decoder consisted of upsampling through bilinear interpolation. In order to tackle the resolution loss experienced in the encoder-decoder type of structure the authors also employed concatenations between the encoder and decoder similar to a U-Net. In addition, a residual connection from the input to the output image was used. This enabled the network to learn the difference between the full-dose and low-dose PET images. The network was also trained with multi-slice input so as to help the network distinguish between noise and finer structural details. The loss function of choice in this work was the L^1 -norm. This approach is shown in Fig. 2. Gong *et al.* [31] proposed an architecture based on ResNet consisting of five residual blocks. Owing to the limited amount of real data, the authors initially trained the network on simulated data created using BrainWeb [32] and the extended cardiac-torso (XCAT) phantom [33]. The network was then fine-tuned on real data. Another aspect of this work is the use of perceptual loss to further improve the quality of predicted images. The authors found their proposed method to perform better than traditional Gaussian filtering. Dilated convolution [34] replaced the convolution operation in the U-Net in the work proposed by Spuhler *et al.* [35]. The network, called dNet, outperformed U-Net for PET image denoising. The advantage of dilated convolutions, which were first introduced for image segmentation, is that they remove the requirement of pooling and upsampling operators. The dataset used in this study was from a psychiatric study consisting of 35 patients. Deep denoising autoencoder (DAE) was proposed in [36] for dynamic PET denoising. The DAE was trained on noisy and noiseless spatiotemporal patches of simulated images. Although a promising voxel-level denoising method was proposed, the proposed DAE struggled to generalize to test data different from the training data. A 3-dimensional (3-D) version of the U-Net was used in [37] for mapping from noisy $64 \times 64 \times 64$ patches to noise-reduced patches of the same dimension. The authors trained the network on a lung cancer real dataset. The method was evaluated by three physicians through lesion detection tasks. The proposed method performed better than Gaussian smoothing, but its improvements were limited when it came to the count levels typically observed in a clinical setting. The effect on noise levels for denoising PET images was studied in [38]. A personalized weighting strategy for specific noise levels was proposed through the linear blending of results from different models. The authors trained five 3-D U-Nets each with a different noise level in the training data. Along with these, a separate network with all the varied noise levels also was trained. The one-network-for-all model did not generalize well on the testing data with multiple noise

levels. The networks trained on noisier images performed better at denoising but introduced more spatial blurring. The final method fused the deep image prior and regularization by denoising (RED) approach to obtain a final denoised image. DeepRED optimization was done using alternating direction method of multipliers [39].

Low-dose myocardial perfusion (MP) SPECT with deep learning was proposed in [40], where the authors used a 3-D convolutional neural network (CNN). The 3-D network consisted of autoencoders where the encoder and the decoder parts of the network were stacked with convolutions. The network was trained to map from various levels of dosage ($1/2, 1/4, \dots, 1/16$) to full-dose images. The authors found their method to perform better than the conventional spatial-post filtering method. The network was trained on real patient data reconstructed with both FBP and OSEM algorithms. A four-layer U-Net was utilized in [41] for SPECT denoising. The U-Net in this method was trained on simulated XCAT phantoms. Low-dose imaging in SPECT through the reduction in acquisition time and projection angles was explored in [42]. The authors used a ResNet for mapping from low-dose to full-dose images. The dataset consisted of MP SPECT images of 363 patients reconstructed with OSEM algorithm. A study on weighted loss functions for varying levels of statistics based on inter-subject changes for deep learning-based SPECT denoising was proposed in [43].

2) *Generative Adversarial Networks*: GANs [44] are a special type of NN model consisting of two units, with the generator unit synthesizing candidates while the discriminator unit attempts to decipher whether the candidate's images are synthetic or real. The development of GANs has strengthened the capability of NNs in this regard, allowing them to capture complex probability distributions. Lu *et al.* [45] investigated the accuracy of deep learning-based denoising methods including GANs for PET imaging of small lung nodules, focusing on quantitative accuracy and visual image quality. Ouyang *et al.* [46] explored GANs with feature matching and task-specific perceptual loss in the restoration of amyloid PET. Jeong *et al.* [47] demonstrated that GAN-based restoration of amyloid PET scans did not affect physician interpretation, indicating that the restored images were consistent with the original scans and preserved their diagnostic value.

Conditional GAN (cGAN) was introduced by Wang *et al.* [48] to recover full-dose brain [^{18}F]FDG PET images from low-dose measurements. Xue *et al.* [49] confirmed the cross-scanner and cross-tracer capability of customized cGAN, where the training was done from [^{18}F]FDG PET on one scanner and the test was performed on [^{18}F]FET and [^{18}F]Florbetapir PET imaging of different scanners.

Cycle-consistent GAN (CycleGAN) was applied to recover full-dose whole-body (WB) PET from low-dose measurements [50]. Zhou *et al.* [51] confirmed that their CycleGAN preserves edges and standardized uptake values from the restored low-dose dataset with biopsy-proven primary lung cancer or suspicious radiological abnormalities. A supervised GAN with the cycle-consistency loss, Wasserstein distance loss, and an additional supervised learning loss, named as S-CycleGAN was demonstrated to outperformed 3-D-cGAN in the recovery

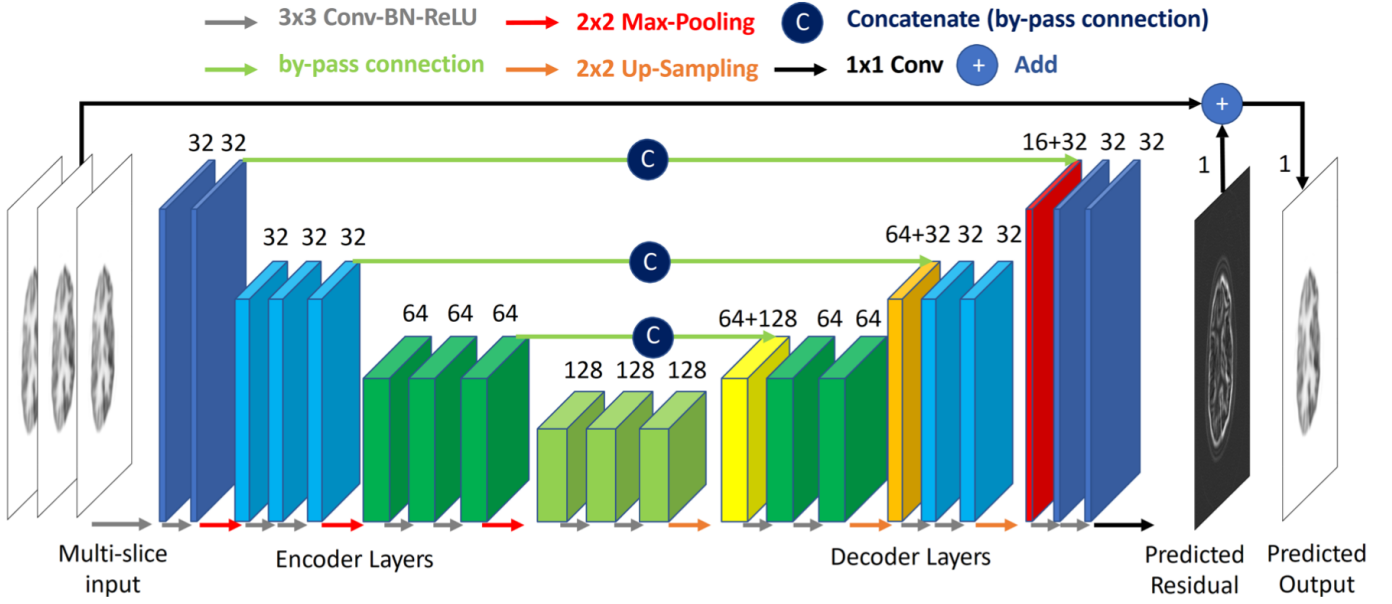


Fig. 2. Representation of supervised deep learning-based method from Xu *et al.* [30].

of low-dose brain PET [52]. The CycleGAN has demonstrated the advantage to train non-synthetic low-dose WB [^{18}F]FDG PET scans together with separate full-dose WB [^{18}F]FDG PET scans in a study by Sanaat *et al.* [53].

Gong *et al.* [54] proposed a parameter-transferred Wasserstein GAN (W-GAN), namely, PT-GAN, with a task-specific initialization for low-dose PET image denoising without compromising structural details. A representation of PT-GAN is shown in Fig. 3. Du *et al.* [55] developed a cascaded data consistency GAN to recover high-quality PET images from FBP-reconstructed PET images with streaking artifacts and high noise. Geng *et al.* [56] developed a content-noise complementary learning pipeline using GAN to reduce noise in medical images including CT, MR, and PET, which outperformed state-of-the-art denoising algorithms in terms of visual quality, quantitative metrics and robust generalization capability.

Additional modality such as MRI can be added into the training of GAN. Wang *et al.* [57] developed a 3-D auto-context-based locality-adaptive multi-modality GAN (LA-GAN) for estimating full-dose [^{18}F]FDG PET images from low-dose counterparts together with MR images and demonstrated the superiority over traditional multi-modality fusion methods for PET restoration. Zhou *et al.* [58] introduced a unified motion correction and denoising GAN for generating motion-compensated low-noise images from low-dose gated PET data.

GANs can be also applied on denoising directly during the reconstruction from low-count sinogram data. Xue *et al.* [59] use a LCPR-Net to enforced a cyclic consistency constraint on the least-squares loss of a GAN framework to establish a nonlinear end-to-end mapping process from low-count sinograms to full-count PET images. Similarly, an improved W-GAN framework was employed as a direct PET image reconstruction network (DPIR-Net) to enhance image

speed and quality of PET reconstruction [60].

GANs were also applied on the denoising of SPECT images. Sun *et al.* [61] developed a method based on a 3-D cGAN for denoising of dual-gated MP images. The same group also investigated the denoising performance of cGAN in projection-domain and compared it with the denoising in reconstruction-domain for low-dose MP SPECT imaging [62]. Sun *et al.* [63] introduced Pix2Pix GAN for denoising low-dose MP images and found it superior than other denoising methods. They also introduced attention mechanisms in GANs for the denoising of fast MP images [64]. Aghakhan Olia *et al.* [65] developed a GAN to predict non-gated standard-dose SPECT images in the projection space. Their finding revealed that recovery of underlying signals/information in low-dose images beyond a quarter of the standard dose would not be feasible and adversely affect the clinical interpretation of the resulting images.

B. Self-attention Mechanisms

The SA mechanism was first proposed for natural language processing by Vaswani *et al.* [66]. The generated attention map is similar to the weight elements utilized in the nonlocal means (NLM) denoising [67]. This connection was further explained in the work of nonlocal NNs[68].

During SA calculation, features are first extracted from the input to construct the *Query*, *Key*, and *Value* components. The *Query* and *Key* components are then utilized to generate the attention map through a matrix multiplication, scaling, and SoftMax operations. The calculated attention map is multiplied with the *Value* component to obtain the final output. Compared to the widely used convolution operation, the SA module has a spatially-variant filter defined by the attention map, which is calculated from the input itself. The SA module can be embedded into the popular U-Net and GAN structures to

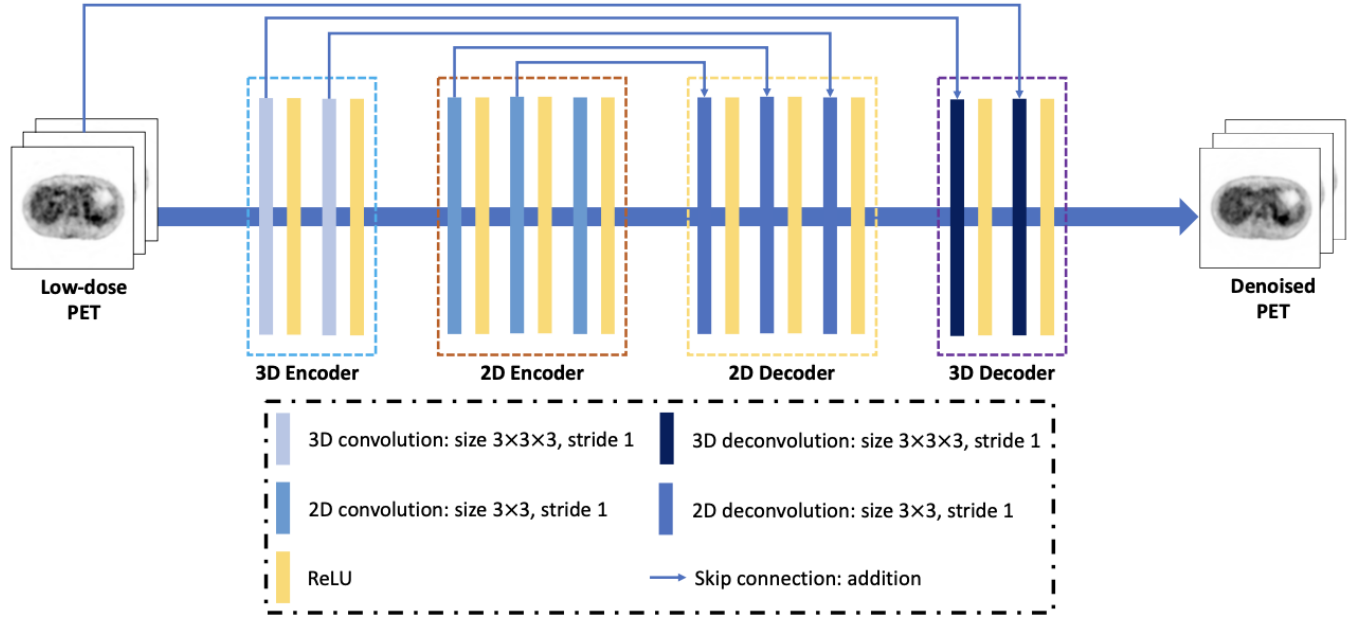


Fig. 3. Structure of PT-WGAN from Gong *et al.* [54].

further improve the performance as demonstrated in various computer vision tasks [69], [70].

For PET image denoising, Xue *et al.* [71] proposed embedding the SA block into the widely used U-Net structure. The SA can also be embedded into other network structures for PET image denoising, such as GAN [72] and CycleGAN [73], [74]. Apart from utilizing only low-dose PET images as the input, Onishi *et al.* [75] proposed an unsupervised PET image denoising framework that incorporated anatomical information into the network architecture via the SA mechanism. The attention gates employed in this work aimed to remove PET image noise by better utilizing the multi-scale semantic features extracted from the MR prior image. For dynamic PET imaging, Li *et al.* [76] proposed to directly generate high-quality Patlak images from five-frame sinograms by a network with SA blocks, potentially reducing the acquisition time and avoiding the input function needed for parametric imaging. For PET image reconstruction, Xie *et al.* [77] proposed to utilize the U-Net with a SA block for image representation and further employ this network representation into the PET MBIR framework. They also employed additional high-resolution anatomical images as the network input to further improve the reconstruction performance [78].

Though CNNs achieved great success in various medical imaging tasks, the network specifically focused on local spatial information, and the receptive field was also limited. The transformer networks extensively employed SA blocks (described in the previous subsection) as the network building blocks, which had the ability to capture long-range information. Vision transformer (ViT) [79] was the first application of the transformer network to computer vision. The input images were firstly divided into patches, linearly embedded along with the position information, and then fed to the transformer

network for image classification. One issue of ViT was the quadratically growing computational complexity along the spatial dimension, which limited the receptive field achievable.

To address this issue, the Swin Transformer [80] was proposed to efficiently calculate local multi-head self-attention (MSA) using shifted windows that offered linear computational complexity. Compared to ViT, the Swin Transformer divided the whole image into windows. The SA was computed based on a shifted windowing scheme, which achieved greater efficiency by limiting SA calculation to non-overlapping local windows while also considering cross-window connections. This method is represented in Fig. 4. Instead of SA calculation along the spatial domain, Restormer [81] was recently proposed to efficiently compute global MSA along the channel dimension. It had linear computational complexity for image restoration tasks.

The transformer networks were recently applied for PET image quality improvement. Luo *et al.* [82] proposed a GAN embedded with a transformer to perform low-dose PET image denoising. The transformer was inserted between the encoder and decoder paths of the generator network, and the training function was based on both voxel-wise estimation error and the adversarial loss. Jang *et al.* [83] proposed a transformer network that can leverage both spatial and channel information based on local and global MSAs. Quantitative evaluations based on datasets of different PET tracers, i.e., [^{18}F]FDG, [^{18}F]ACBC, [^{18}F]DCFPyL, and [^{68}Ga]DOTATATE, showed that the proposed transformer structure achieved better performance than other reference methods. When utilizing both low-dose PET and high-resolution MR prior images as the input, Zhang *et al.* [84] designed a network structure with two paths to extract PET and MR features and a transformer block to fuse the PET and MR features. Wang *et al.* [85] compared

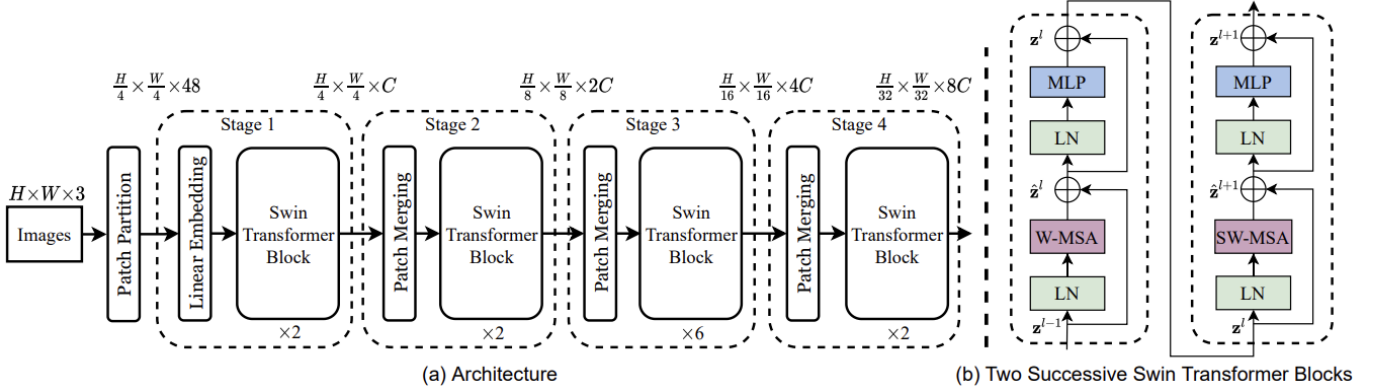


Fig. 4. (a) The architecture of the Swin transformer network and (b) consecutive blocks of the Swin transformer. Reprint from Liu *et al.* [80].

five deep learning-based denoising methods, including Swin Transformer and ViT, under different PET dose levels. Results showed that the Swin Transformer achieved better performance than other reference methods in most evaluation tasks. Apart from PET image denoising, Hu *et al.* [86] utilized the transformer network for PET image reconstruction following the unrolled-NN framework [87]–[89], and the results showed better performance than other unrolled networks where CNNs were adopted.

C. Unsupervised Methods

The advent of unsupervised methods was a result of the limitations that supervised learning methods posed namely large datasets and the effort that goes into annotating them. Two such methods that have been developed in the context of denoising are deep image prior (DIP) and Noise2Noise (N2N) methods.

1) *Deep Image Prior*: The DIP [90] is a learning-free method that demonstrated that a randomly initialized NN could be used as a prior for inverse problems like denoising, inpainting, and super-resolution. The DIP changes the paradigm of the standard deep denoising approach (5) in the sense that instead of training the parameter θ from a dataset as described in (6), the post-processed image x^{pp} is obtained from the reconstructed image x^{rec} as

$$\theta^* = \arg \min_{\theta} L(F_{\theta}(z), x^{\text{rec}}), \quad x^{\text{pp}} = F_{\theta^*}(z), \quad (7)$$

where z is a random image input. This approach relies on the implicit regularization imposed by the architecture of the NN F_{θ} which prevents over-fitting, the structure of a NN being sufficient to capture the low-level information of the image. This section describes the denoising methods in ET that utilize DIP.

A 3-D U-Net was trained in [91], [92] to predict denoised PET images using CT/MRI images as the high-quality prior input (i.e., z in (7)) and noisy PET images as the label. It was observed that using a prior image further improved the results rather than training the NN on a random noise input (as done in [90]). The network was trained on a single set of MR images and noisy PET volumes and then evaluated on

simulated phantoms as well as real datasets. The loss function used was the mean squared error (MSE) combined with the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm [93]. The DIP method was extended to dynamic PET denoising in [94]. The 3-D U-Net in this work was trained on static PET images as inputs to the network and noisy dynamic PET images as training labels. DeepRED [95] is a method that utilized DIP with RED [96]. A U-Net-like architecture was trained on noisy dynamic PET images with MSE. The input to the network was a random noise vector. Hashimoto *et al.* extended their previous work [94] in [97] where they proposed two modular approaches to dynamic PET image denoising. The first module consisting of a 3-D U-Net extracted features from static PET images. The extracted features were then fed to a reconstruction module consisting of a typical CNN with convolution layers that predicted denoised images while being trained on noisy dynamic PET volumes. Simultaneous denoising of dynamic PET images was proposed by Yang *et al.* [98]. Their network consisted of multiple convolutional layers, taking as input time averaged PET images with the training labels being the noisy dynamic PET volume. A variation of this network called double DIP was also proposed which additionally generates the time averaged PET images.

2) *Noise2noise*: N2N is a self-supervised learning (SSL) technique, i.e., a machine learning technique that trains models using input data only, without labeled data and explicit supervision. Labeling data takes considerable time and effort and in some situation obtaining gold standard-labeled data is impossible. SSL has the advantage of significantly increasing the number of datasets for model training as it does not require labeled data. In medical imaging techniques that employ ionizing radiation, such as X-ray and gamma-ray, obtaining a clean target (label) with a high radiation dose can elevate the potential health risks associated with radiation exposure. Although it is possible to obtain a clean target by increasing the scanning time, it may cause image blurring or distortion by increasing the likelihood of patient movement.

The N2N approach, as proposed by Lehtinen *et al.* [99], trains a deep NN F_{θ} parametrized with θ for image denoising using K noisy image pairs, i.e., two noise instances of the

same image, as input and target, as follows:

$$\theta^* = \arg \min_{\theta} \sum_{k=1}^K \left\| \mathbf{F}_{\theta} (\mathbf{x}_k^{\text{clean}} + \epsilon_{k,1}) - (\mathbf{x}_k^{\text{clean}} + \epsilon_{k,2}) \right\|^2 \quad (8)$$

where $\epsilon_{k,1}, \epsilon_{k,2}$ are independent noise realizations for each image $\mathbf{x}_k^{\text{clean}}, k = 1, \dots, K$. This SSL technique, which only requires the input and target noise distribution to be identical and independent, has demonstrated its efficacy in reducing various types of noise, including Gaussian and Poisson noise.

There has been a growing interest in applying the N2N to medical images, such as CT and MRI [100]–[102]. In these studies, the denoising efficacy of N2N was compared to that of conventional denoising techniques. CT images reconstructed using N2N prior images yielded better root MSE and structural similarity index measure (SSIM), as well as improved texture preservation as compared with conventional methods using total variation, NLM, and convolutional sparse coding [100]. Fang *et al.* [101] showed that using a N2N-denoised image as a prior within MBIR showed promising results in suppressing noise while preserving subtle structures when applied to spectral CT data for material decomposition. Jung *et al.* [102] evaluated the performance of N2N for image quality improvement in sub-millimeter resolution 3-D MR images. In this study, the K-space data of 3-D MR images were split into two separate sets with independent noise realizations, which was achieved by undersampling data alternatively along the k - z -axis and estimating missing data using a GRAPPA kernel. Volumetric accuracy, as well as image quality, was improved by N2N, which utilized only a single fully sampled K-space data.

N2N and SSL technology are also potentially useful for improving the image quality and diagnostic performance in ET. List-mode data recorded in nuclear medicine image scans and containing detail information about the location, energy, and time of each detected event allows for flexible data binning. Its flexibility has provided excellent platform for investigating and implementing N2N technologies that require identical and independent dataset as input and target for network training. A N2N application study on [^{18}F]FDG brain PET utilized short time-bin images (10s–40s) generated from list-mode data and showed equivalent peak signal-to-noise ratio of N2N outcomes compared to supervised learning with 300s images as target, at all tested noise levels [103]. The usefulness of N2N for noise reduction in [^{15}O]water dynamic PET and [$^{99\text{m}}\text{Tc}$]MDP/DPD WB bone scan studies was also reported [104], [105]. In addition, Chan *et al.* [106] proposed a technique to improve the accuracy and robustness of N2N network trained for WB PET image denoising by mitigating the high variance in N2N denoising outcomes. The variability is mainly caused by the spatially non-stationary nature of PET image noise distribution. In this study, instead of training the network with pairs of individual noisy realizations, the number of training samples was increased by pairing a single noisy realization with an ensemble of noisy realizations at the same count level. When applied to low-count WB PET images, the original N2N produced speckle and clustered noise artifacts. Nevertheless, the proposed method was effective in reducing the noise while

preserving natural noise texture. An endeavor has been also undertaken to enhance the generalization ability of N2N-based noise reduction algorithm through the incorporation of wavelet transforms (WTs) into the N2N network [107]. The proposed method entails the utilization of the forward WT to decompose a given noisy image into its low-pass and high-pass frequency components. These data are then fed into an N2N network, followed by an inverse WT. The final output was compared with another noisy image as in typical N2N framework. The forward and backward WT coefficients were also determined through training, thereby enabling the proposed method to outperform the original N2N method in suppressing artifacts and preserving abnormal uptakes. This method is shown in Fig. 5 and Fig. 6.

Multiple variants of N2N have been developed to mitigate different types of noise and enhance the efficacy of self-supervised denoising algorithms. These variants include Noiser2Noise, Noise2Void, and Noise2Self [108]–[110]. When applied to low-count [^{18}F]FDG brain PET images, Noiser2Noise, which requires only a single noisy realization of each training sample, was more effective than N2N in preserving the noise texture of the input images [103]. Noise2Void, another SSL technique that does not require paired training samples, outperformed traditional denoising methods for PET when pre-trained through transfer learning and guided by anatomical images [111].

D. Multi-modality

In hybrid PET/CT, PET/MR, and SPECT/CT systems, the anatomical information derived from CT and MRI can also be used to enhance the denoising performance of PET and SPECT. Numerous studies using CNNs in brain PET/MR datasets achieved substantial dose reduction, sometimes up to 99%, and incremental improvements in image quality and resolution have been demonstrated by combining PET and MR images as multiple channels into the network compared to using PET images alone as input [57], [112]–[116]. While CNNs requirement on training datasets is high, a relatively low-complexity CNN (micro-net) that is more robust to very limited amounts of training data was proposed for MR-guided PET denoising, and demonstrated to have robust performance [117]. Other networks structures, such as a spatial adaptive and transformer fusion network [84], have also been proposed to improve PET denoising by more effectively incorporating MRI information.

Not only helping noise reduction, anatomical CT and/or MRI information can also improve the PET image resolution. For example, anatomically-guided PET reconstruction using the Bowsher prior [118] can be generated by a CNN in the image, thus space allowing the generation of anatomically-guided high-resolution PET images without the need to access raw data and reconstruction console [119]. When PET raw data are available, incorporating sinogram-based physics into the loss function of PET/MR networks has been shown to further improve the denoising performance that is more robust to out-of-distribution (OOD) data [120].

In addition to the strategy of incorporating anatomical images as multi-channel inputs, such information can also be

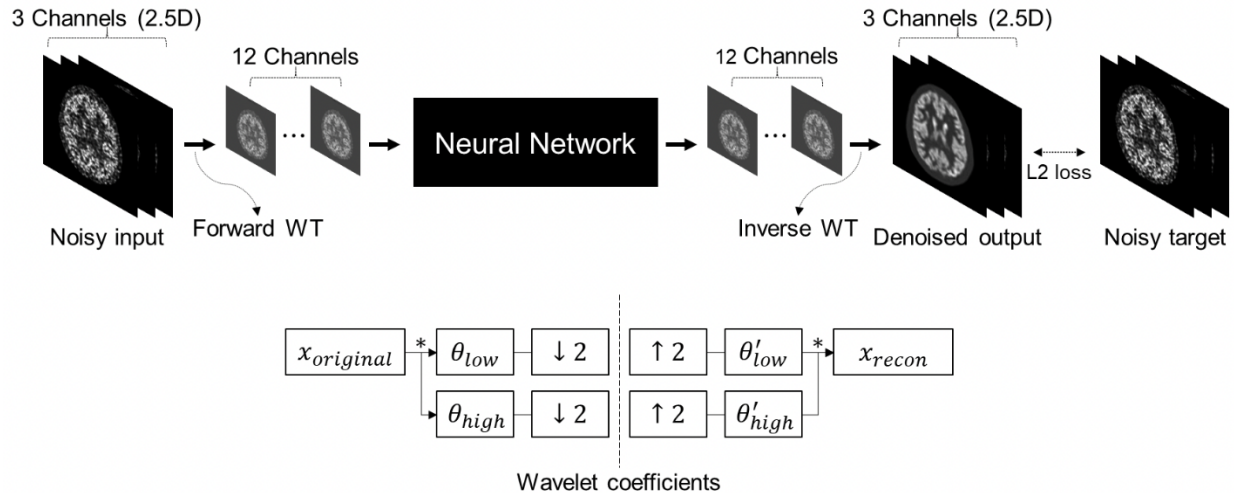


Fig. 5. Schematic of N2N network model improved by the incorporation of WTs. Reprint from Kang *et al.* [107].

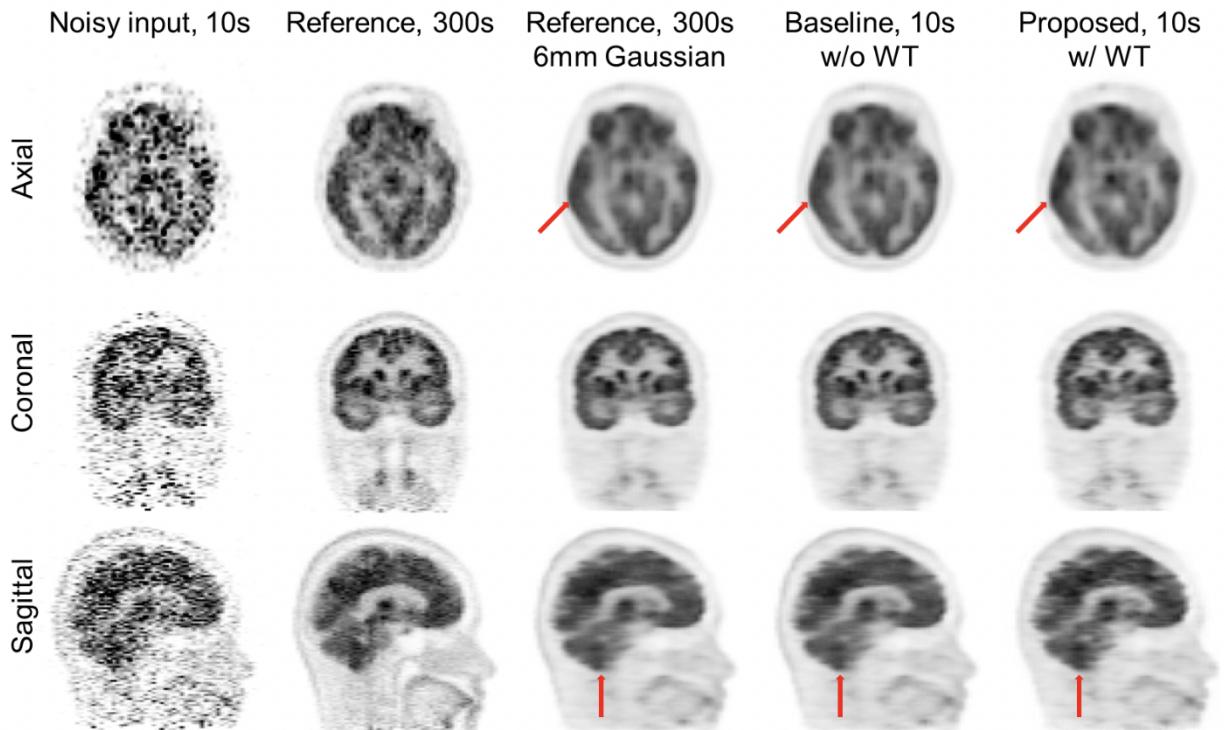


Fig. 6. Full count image, noisy input, Gaussian filtered and denoised images using the N2N without and with incorporating the trainable WT for clinical data. Reprint from Kang *et al.* [107].

input to the network through a SA mechanism, for unsupervised PET denoising for example [75], [78].

In another strategy, investigations using MR and CT images as prior information to guide DIP (cf. Section III-C1) have been performed for brain and body PET/CT datasets [91], [92], [121]. The group further extended the anatomical-guided DIP approach to direct parametric reconstruction framework, where CT and MR images were incorporated as the network input to provide a manifold constraint, and also utilized to construct a kernel layer to perform non-local feature denoising [122].

One requirement of incorporating anatomical information into PET denoising is the integrated or simultaneous scanners such as PET/MR, which are still limited to widespread use. A study suggested that accurate full-dose amyloid PET images can be generated from low-dose PET and either simultaneous or non-simultaneous (acquired up to 42 days apart) MR images, broadening the utility of low-dose amyloid PET imaging [123].

While a large number of hybrid PET denoising work were performed for brain PET/MR datasets, investigations using

CT and/or MRI information for denoising body PET images were also performed. In an application to cardiac viability [^{18}F]FDG imaging in patients with ischemic heart disease, both attenuation correction CT and low-count PET were input into networks through two channels, leading to effective denoising [124]. In a study of [^{68}Ga]PSMA prostate PET/MR, 50% dose reduction could be achieved by using a discrete-WT CNN with MRI priors [125]. Rather than concatenating the MR and PET at the input level, a study combined them in the feature space with attention-weighted loss, and applied the methods to WB PET for children and young adults lymphoma patients [126].

While brain MR images are usually well registered with PET images, for WB applications, mismatch between PET and CT/MR due to motion could complicate PET denoising methods that incorporate anatomical information. A study showed that when CT and PET are aligned, incorporating CT as additional channels improves the quantitative accuracy of lung lesions derived from denoised low-count PET images. However, when CT and PET are misaligned, incorporating CT information resulted in additional lesion quantification bias as compared with using PET data only [45]. The results suggest that motion correction and image registration are important pre-processing steps when incorporating anatomical information into PET denoising.

Similar approaches described above can also be applied to hybrid SPECT/CT scanners. In a SPECT bone imaging application, a lesion-attention weighted U²-Net incorporating both 1/7-count SPECT and resampled attenuation correction CT can lead to the successful generation of synthetic standard-count SPECT images [127].

E. Diffusion Models

DMs have arisen as an alternative to GANs for image generation and other tasks [128], and have been widely explored in medical imaging [129]. In the following paragraph we briefly summarize score-based DMs for image generation as described in Ho *et al.* [130].

Assume that the (clean) feasible images \mathbf{x} are distributed according to some unknown probability distribution function (PDF) $p_0(\mathbf{x})$ that we wish to use to randomly generate images. The forward diffusion is achieved by generating a collection of images $(\mathbf{x}_t)_{t=1,\dots,T}$, starting from a clean image $\mathbf{x} = \mathbf{x}_0$ drawn from p_0 (i.e., by randomly choosing an image from the training dataset), as a Markov chain defined as

$$\mathbf{x}_t \mid \mathbf{x}_{t-1} \sim \mathcal{N}(\sqrt{\alpha_t}\mathbf{x}_{t-1}, (1 - \alpha_t)\mathbf{I}) \quad (9)$$

where \mathbf{I} is the identity operator in the image space, and $(\alpha_t)_{t=1,\dots,T}$ is a collection of parameters in $[0, 1]$ defined such that \mathbf{x}_T is (approximately) a white noise. Sampling an image according to the initial distribution p_0 can therefore be achieved by reversing the diffusion process, i.e., starting from a white noise \mathbf{x}_T from which \mathbf{x}_{T-1} is generated, and so on until \mathbf{x}_0 . Unfortunately, the reverse conditional PDFs $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ are unknown. Instead, the approximate model is used:

$$\mathbf{x}_{t-1} \mid \mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}_t, t), \sigma^2(t)\mathbf{I}) \quad (10)$$

where $\sigma^2(t)$ is a known function of α_t and $\boldsymbol{\mu}(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\nabla \log p_t(\mathbf{x}_t)$ where p_t is the PDF of \mathbf{x}_t . The score function $\mathbf{s}_t(\mathbf{x}_t) = \nabla \log p_t(\mathbf{x}_t)$ is untractable and is therefore replaced by a NN $\mathbf{S}_\theta(\mathbf{x}_t, t)$, where the parameter θ is trained unsupervisingly via score-matching from several instances of $(\mathbf{x}_t)_{t=1,\dots,T}$ generated following (9) and by sampling \mathbf{x}_0 from a training dataset of clean images [131].

In addition to generating images, DMs can also be used for inverse problems such as denoising and reconstruction. One type of approach, namely conditional diffusion models (CDMs), consists in generating a noise-free image \mathbf{x} according to the a-posteriori PDF $p_0(\cdot \mid \mathbf{y})$, where \mathbf{y} is the measurement (in our case, $\mathbf{y} = \mathbf{x}^{\text{noisy}}$), by successive sampling of $\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y}$. This can be achieved using the same above-mentioned methodology using the conditional score $\mathbf{s}_t(\mathbf{x}_t, \mathbf{y}) = \nabla \log p_t(\cdot \mid \mathbf{y})(\mathbf{x}_t)$. The conditional score can be approximated score-matching in two ways: (i) by conditional score-matching of a NN $\mathbf{S}_\theta(\mathbf{x}_t, \mathbf{y}, t)$ that takes as input both the image \mathbf{x}_t for all t and the measurement \mathbf{y} [128], [132], [133] (supervised), and (ii) unconditional score-matching of $\mathbf{S}_\theta(\mathbf{x}_t, t)$ (cf. previous paragraph) combined with the Bayes rule and an approximation of the posterior distribution $p(\mathbf{y} \mid \mathbf{x}_t)$ [134], [135] (unsupervised). Note that approach (ii) requires the knowledge of the noise model of \mathbf{y} . With a different formulation but in the same spirit, Mardani *et al.* [136] proposed to approximate the a-posteriori PDF $p_0(\cdot \mid \mathbf{y})$ by minimization of the Kullback-Leibler divergence between a standard Gaussian model and the posterior, which yields to a penalized least-squares optimization problem with a score-matching penalty.

Recently, Gong *et al.* [137] used a CDM for brain PET image denoising with incorporation of the MR image as prior information \mathbf{y} to the approximated conditional score $\mathbf{S}_\theta(\mathbf{x}_t, \mathbf{y}, t)$ trained from a clean PET/MRI dataset. Pan *et al.* [138] used a CDM reinforced with a consistency model [139] to improve the efficiency. Shen *et al.* [140] proposed to utilize diffusion models for low-count PET image denoising through a bidirectional condition diffusion probabilistic model, which was validated on WB PET datasets. Jiang *et al.* [141] proposed an unsupervised PET enhancement framework through the latent diffusion model, which can be trained only on standard-count PET data. Instead of a Gaussian noise, a Poisson noise was inserted in the diffusion process to better accommodate PET imaging. Also, a CT-guided cross-attention was proposed to incorporate additional CT images into the inverse process. Han *et al.* [142] proposes to generate high-quality PET based on the diffusion models through a coarse-to-fine PET reconstruction framework that consists of a coarse prediction module and an iterative refinement module.

Apart from PET image denoising, Xie *et al.* [143] proposed to generate synthetic PET image from MR images based on the diffusion models. Singh *et al.* [144] proposed a diffusion model-based PET image reconstruction framework, where the PET forward model (eq. (1) and eq. (2)) was utilized together with the score function (i.e., approach (ii) above with the PET raw data \mathbf{y}) for high-quality PET image generation. Their method was further improved by adding MR anatomical prior.

Despite their potential in image denoising for ET, it is important to note that DMs are relatively novel in this field.

TABLE I
SUMMARY OF THE REVIEWED ARTICLES

Application	Supervised	Method/Architecture	Anatomical Information	References
Section III-A: Supervised Methods				
Brain PET	yes	CNN, U-Net	no	[30], [35]
Brain & Chest PET	yes	ResNet	no	[31]
Brain PET	yes	DAE	no	[36]
Chest PET	yes	U-Net	no	[37]
Whole-body PET	yes	U-Net	no	[38]
MP SPECT	yes	CNN, U-Net, ResNet	no	[40]–[43]
Chest PET	yes	GAN, cGAN, CycleGAN, W-GAN	no	[49], [51], [54], [56], [58], [59]
Chest PET	yes	CNN, U-Net, GAN	CT	[45]
Brain PET	yes	GAN	no	[46], [47]
Brain PET	yes	GAN, cGAN, CycleGAN	no	[48], [52], [55]
Brain PET	yes	cGAN	MRI	[57]
Whole-body PET	yes	CycleGAN	no	[50], [53], [60]
MP SPECT	yes	cGAN, attention-based GAN	no	[61]–[65]
Section III-B: SA Mechanisms				
Whole-body PET	yes	U-Net + SA block	no	[71]
Chest PET	yes	U-Net + SA block	no	[77]
Chest PET	yes	U-Net + SA block	CT, MRI	[78]
Brain PET	yes	GAN + SA block	no	[72]
WB PET	yes	CNN, CycleGAN + SA block	no	[73], [76]
Brain PET	no	CNN + SA block	MRI	[75]
WB PET	yes	Transformer	no	[82], [83]
Brain, WB PET	yes	U-Net, GAN, Transformer	MRI	[84], [85]
Brain PET	yes	Transformer	no	[86]
Section III-C: Unsupervised Methods				
WB PET	no	DIP	CT, MRI	[91], [92]
Brain PET	no	DIP	no	[94], [95], [97], [98]
Brain PET	no	N2N	no	[103], [107], [111]
Bone SPECT	no	N2N	no	[105]
¹⁵ O water PET	no	N2N	no	[104]
WB PET	no	N2N	no	[106]
Section III-D: Multi-modality				
Brain PET	yes	CNN, U-Net	MRI	[112]–[117], [119], [120], [123]
Brain PET	no	DIP	MRI	[122]
⁶⁸ Ga PET (prostate)	yes	CNN	MRI	[125]
MP PET	yes	U-Net	CT	[124]
WB PET	yes	CNN	MRI	[126]
Bone SPECT	yes	U-Net	CT	[127]
Section III-E: DMs				
Brain PET	no	DM	MRI	[137]
WB PET	no	DM	no	[138], [140]
WB PET	no	DM	CT	[141]
Brain PET	yes	DM	no	[142]
Brain PET (image synthesis)	yes	DM	MRI	[143]
Brain PET (image reconstruction)	no	DM	MRI	[142]

As of the current literature review, there is a limited number of research papers that have explored the application of DMs to denoise (or reconstruct) ET images. This scarcity of studies underlines the emerging nature of DMs in ET image processing, and as a result, it represents an exciting and promising avenue for further research and development.

IV. DISCUSSION AND CONCLUSION

Reduction in the dosage of the radio-pharmaceutical administered to a patient undergoing a functional imaging scan is essential to reduce the risk and radiation the patient is being exposed to. Low-dose imaging aims to produce images with quality on par with regular-dose imaging while operating on a fraction of the radio-pharmaceutical. Getting a clear image in ET imaging is a difficult task due to attenuation, scatter, and the ill-posed nature of the physics model used to characterize PET and SPECT. Low-dose ET further complicates an already challenging problem. Images are produced by converting the detector data into readable images through the process of reconstruction. Low-dose ET reconstructed images suffer from noise. One way to tackle this noise is through the post-processing of the reconstructed images.

Image processing methods like Gaussian filtering have been implemented over the years to tackle the reconstructed noisy images. Recently deep learning-based methods have been found to surpass traditional algorithms in imaging tasks. In this article, various post-processing approaches that utilize deep NNs have been discussed at length. These approaches have been appropriately classified into various sub-sections to facilitate better readability and to distinguish them from one another. U-Net was the earliest NN adapted for denoising. Adding a discriminator network to a U-Net resulted in GANs. These networks were discussed in the Section III-A. Recently developed supervised methods include a self-attention mechanism to U-Net and GANs were highlighted in Section III-B. Deep image prior and N2N methods are the most popular unsupervised NN approaches, and were discussed in Section III-C. Section III-D presents methods that utilize multi-modality input data for training NNs. Finally, Section III-E covers recent trends in DMs. The primary focus of this article has been to discuss the diverse set of NN-based approaches while keeping up with the latest trends in the denoising regime. The presented papers are summarized in Table I.

One of the main challenges in training a NN is the avail-

ability of data. There is an increasing number of publicly available datasets such as Brainweb [32]. However, these are limited to the use cases they have been compiled for. There is a concern about using NNs trained on one specific dataset to be generalized to a larger data pool. This challenge of dealing with OOD data is addressed to an extent by unsupervised methods, which do not require large labeled datasets. However, unsupervised methods are yet to perform convincingly better than supervised methods. Efforts are being made in this regard to develop NNs that produce robust results independent of labeled data.

Evaluation of the denoised images predicted is another aspect that all NN methods need to be held accountable for. Typically, MSE and SSIM are used for quantitative comparison and analysis. Plot profiles and region of interest analysis for tumors are also utilized for a more thorough evaluation. Some articles have further employed a scoring system used by radiologists for the qualitative assessment of the images. Such efforts to check the images produced by NN-based approaches are essential to build trust and translate these methods to clinical cases.

NN-based denoising has indeed revolutionized low-dose ET imaging. This article through the emphasis on both PET and SPECT has highlighted the plethora of ways deep learning has improved image quality through post-processing. We discussed the earliest methods as well as the most promising methods in recent times that have the potential to be translated into clinical usage. The advantage of denoising is that methods applicable to one modality can easily be implemented for other modalities also, and this review on low-dose ET could be useful to image denoising and enhancement in other modalities too.

The necessity to compare the multitude of methods presented in this review poses a formidable challenge, primarily due to their application-specific nature, spanning cardiology, neurology, and oncology, as well as the utilization of diverse training datasets. In an ideal scenario, method performances would be evaluated using dedicated benchmark datasets specifically designed for comparative purposes, accompanied by standardized metrics akin to those employed in renowned challenges within the field. Drawing inspiration from successful model evaluation frameworks, such as the Ultra-low Dose PET Imaging Challenge [145], it is paramount for the community to consider the creation of specialized datasets and standardized metrics aimed at comprehensively evaluating these methods. Such initiatives hold the potential to significantly enhance our ability to objectively compare and contrast the efficacy of different techniques across diverse medical applications, thus enabling more informed decisions and driving progress within the field.

ACKNOWLEDGMENT

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper.

REFERENCES

- [1] S. Mattsson and M. Söderberg, "Radiation dose management in CT, SPECT/CT and PET/CT techniques," *Radiation protection dosimetry*, vol. 147, no. 1-2, pp. 13–21, 2011.
- [2] A. Gimelli, S. Achenbach, R. R. Buechel, T. Edvardsen, M. Franccone, O. Gaemperli, M. Hacker, F. Hyafil, P. A. Kaufmann, P. Lancellotti, *et al.*, "Strategies for radiation dose reduction in nuclear cardiology and cardiac computed tomography imaging: A report from the European Association of Cardiovascular Imaging (EACVI), the Cardiovascular Committee of European Association of Nuclear Medicine (EANM), and the European Society of Cardiovascular Radiology (ESCR)," *European heart journal*, vol. 39, no. 4, pp. 286–296, 2018.
- [3] H. O. Anger, "Scintillation camera," *Review of scientific instruments*, vol. 29, no. 1, pp. 27–33, 1958.
- [4] F. Natterer, *The Mathematics of Computerized Tomography* (Classics in Applied Mathematics). SIAM, 2001.
- [5] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE transactions on medical imaging*, vol. 1, no. 2, pp. 113–122, 1982.
- [6] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994.
- [7] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE transactions on medical imaging*, vol. 14, no. 1, pp. 132–137, 1995.
- [8] J. Qi, R. M. Leahy, C. Hsu, T. H. Farquhar, and S. R. Cherry, "Fully 3D bayesian image reconstruction for the ECAT EXACT HR+," *IEEE Transactions on Nuclear Science*, vol. 45, no. 3, pp. 1096–1103, 1998.
- [9] P. J. Huber, *Robust statistics*. John Wiley & Sons, 2004, vol. 523.
- [10] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Transactions on nuclear science*, vol. 49, no. 1, pp. 56–60, 2002.
- [11] J. E. Bowsher, H. Yuan, L. W. Hedlund, T. G. Turkington, G. Akabani, A. Badea, W. C. Kurylo, C. T. Wheeler, G. P. Cofer, M. W. Dewhurst, *et al.*, "Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors," in *IEEE Symposium Conference Record Nuclear Science 2004.*, IEEE, vol. 4, 2004, pp. 2488–2492.
- [12] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy, "PET image reconstruction using information theoretic anatomical priors," *IEEE transactions on medical imaging*, vol. 30, no. 3, pp. 537–549, 2010.
- [13] D. Kazantsev, S. Arridge, S. Pedemonte, A. Bousse, K. Erlandsson, B. F. Hutton, and S. Ourselin, "An anatomically driven anisotropic diffusion filtering method for 3D SPECT reconstruction," *Physics in Medicine & Biology*, vol. 57, no. 12, p. 3793, 2012. DOI: 10.1088/0031-9155/57/12/3793.
- [14] S. Pedemonte, A. Bousse, B. F. Hutton, S. Arridge, and S. Ourselin, "4-D generative model for PET/MRI reconstruction," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011: 14th International Conference, Toronto, Canada, September 18–22, 2011, Proceedings, Part 14*, Springer, 2011, pp. 581–588.
- [15] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, *et al.*, "PET reconstruction with an anatomical MRI prior using parallel level sets," *IEEE transactions on medical imaging*, vol. 35, no. 9, pp. 2189–2199, 2016.
- [16] V. P. Sudarshan, Z. Chen, and S. P. Awate, "Joint PET+MRI patch-based dictionary for bayesian random field PET reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 338–346.
- [17] V. P. Sudarshan, G. F. Egan, Z. Chen, and S. P. Awate, "Joint PET-MRI image reconstruction using a patch-based joint-dictionary prior," *Medical image analysis*, vol. 62, p. 101669, 2020.
- [18] P. Van Cittert, "Zum einfluss der spaltbreite auf die intensitätsverteilung in spektrallinien. ii," *Zeitschrift für Physik*, vol. 69, no. 5-6, pp. 298–308, 1931.
- [19] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JoSA*, vol. 62, no. 1, pp. 55–59, 1972.
- [20] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, vol. 79, p. 745, 1974.

- [21] O. Rousset, A. Rahmim, A. Alavi, and H. Zaidi, "Partial volume correction strategies in PET," *PET clinics*, vol. 2, no. 2, pp. 235–249, 2007.
- [22] K. Erlandsson, I. Buvat, P. H. Pretorius, B. A. Thomas, and B. F. Hutton, "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology," *Physics in Medicine & Biology*, vol. 57, no. 21, R119, 2012.
- [23] B. F. Hutton, B. A. Thomas, K. Erlandsson, A. Bousse, A. Reilhac-Laborde, D. Kazantsev, S. Pedemonte, K. Vunckx, S. Arridge, and S. Ourselin, "What approach to brain partial volume correction is best for PET/MRI?" *Nuclear Instruments and Methods in Physics Research Section A*, vol. 702, pp. 29–33, 2013. DOI: 10.1016/j.nima.2012.07.059. [Online]. Available: <https://doi.org/10.1016/j.nima.2012.07.059>.
- [24] K. Erlandsson, J. Dickson, S. Arridge, D. Atkinson, S. Ourselin, and B. F. Hutton, "MR imaging-guided partial volume correction of PET data in PET/MR imaging," *PET clinics*, vol. 11, no. 2, pp. 161–177, 2016.
- [25] B. A. Thomas, V. Cuplov, A. Bousse, A. Mendes, K. Thielemans, B. H. Hutton, and K. Erlandsson, "PETPVC: a toolbox for performing partial volume correction techniques in positron emission tomography," *Physics in Medicine & Biology*, vol. 61, no. 22, pp. 7975–7993, 2016. DOI: 10.1088/0031-9155/61/22/7975. [Online]. Available: <https://discovery.ucl.ac.uk/id/eprint/1523346>.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [27] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciampi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.
- [30] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk, "200x low-dose PET reconstruction using deep learning," *arXiv preprint arXiv:1712.04119*, 2017.
- [31] K. Gong, J. Guan, C.-C. Liu, and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 153–161, 2018.
- [32] C. A. Cocosco, "Brainweb: Online interface to a 3D MRI simulated brain database," *NeuroImage*, vol. 5, 1997.
- [33] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical physics*, vol. 35, no. 8, pp. 3800–3808, 2008.
- [34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
- [35] K. Spuhler, M. Serrano-Sosa, R. Cattell, C. DeLorenzo, and C. Huang, "Full-count PET recovery from low-count image using a dilated convolutional neural network," *Medical Physics*, vol. 47, no. 10, pp. 4928–4938, 2020.
- [36] I. S. Klyuzhin, J.-C. Cheng, C. Bevington, and V. Sossi, "Use of a tracer-specific deep artificial neural net to denoise dynamic PET images," *IEEE transactions on medical imaging*, vol. 39, no. 2, pp. 366–376, 2019.
- [37] J. Schaefferkoetter, J. Yan, C. Ortega, A. Sertic, E. Lechtman, Y. Eshet, U. Metser, and P. Veit-Haibach, "Convolutional neural networks for improving image quality with noisy PET data," *EJNMMI research*, vol. 10, pp. 1–11, 2020.
- [38] Q. Liu, H. Liu, N. Mirian, S. Ren, V. Viswanath, J. Karp, S. Surti, and C. Liu, "A personalized deep learning denoising strategy for low-count PET images," *Physics in Medicine & Biology*, vol. 67, no. 14, p. 145014, 2022.
- [39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [40] A. J. Ramon, Y. Yang, P. H. Pretorius, K. L. Johnson, M. A. King, and M. N. Wernick, "Improving diagnostic accuracy in low-dose SPECT myocardial perfusion imaging with convolutional denoising networks," *IEEE transactions on medical imaging*, vol. 39, no. 9, pp. 2893–2903, 2020.
- [41] M. P. Reymann, T. Würfl, P. Ritt, B. Stimpel, M. Cachovan, A. H. Vija, and A. Maier, "U-net for SPECT image denoising," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2019, pp. 1–2.
- [42] I. Shiri, K. AmirMozafari Sabet, H. Arabi, M. Pourkeshavarz, B. Teimourian, M. R. Ay, and H. Zaidi, "Standard SPECT myocardial perfusion estimation from half-time acquisitions using deep convolutional residual neural networks," *Journal of Nuclear Cardiology*, pp. 1–19, 2020.
- [43] J. Liu, Y. Yang, M. N. Wernick, P. H. Pretorius, and M. A. King, "Accounting for inter-subject variations in deep learning for reduced-dose studies in cardiac SPECT," in *2021 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2021, pp. 2913–2917.
- [44] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Advances in neural information processing systems," *Curran Associates, Inc*, vol. 27, pp. 2672–2680, 2014.
- [45] W. Lu, J. A. Onofrey, Y. Lu, L. Shi, T. Ma, Y. Liu, and C. Liu, "An investigation of quantitative accuracy for deep learning based denoising in oncological PET," *Physics in Medicine & Biology*, vol. 64, no. 16, p. 165019, 2019.
- [46] J. Ouyang, K. T. Chen, E. Gong, J. Pauly, and G. Zaharchuk, "Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss," *Medical physics*, vol. 46, no. 8, pp. 3555–3564, 2019.
- [47] Y. J. Jeong, H. S. Park, J. E. Jeong, H. J. Yoon, K. Jeon, K. Cho, and D.-Y. Kang, "Restoration of amyloid PET images obtained with short-time data using a generative adversarial networks framework," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.
- [48] Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *Neuroimage*, vol. 174, pp. 550–562, 2018.
- [49] S. Xue, R. Guo, K. P. Bohn, J. Matzke, M. Viscione, I. Alberts, H. Meng, C. Sun, M. Zhang, M. Zhang, *et al.*, "A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose PET," *European journal of nuclear medicine and molecular imaging*, pp. 1–14, 2021.
- [50] Y. Lei, X. Dong, T. Wang, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, "Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks," *Physics in Medicine & Biology*, vol. 64, no. 21, p. 215017, 2019.
- [51] L. Zhou, J. D. Schaefferkoetter, I. W. Tham, G. Huang, and J. Yan, "Supervised learning with cycleGAN for low-dose FDG PET image denoising," *Medical image analysis*, vol. 65, p. 101770, 2020.
- [52] K. Zhao, L. Zhou, S. Gao, X. Wang, Y. Wang, X. Zhao, H. Wang, K. Liu, Y. Zhu, and H. Ye, "Study of low-dose PET image recovery using supervised learning with CycleGAN," *Plos one*, vol. 15, no. 9, e0238455, 2020.
- [53] A. Sanaat, I. Shiri, H. Arabi, I. Mainta, R. Nkoulou, and H. Zaidi, "Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging," *European journal of nuclear medicine and molecular imaging*, vol. 48, pp. 2405–2415, 2021.
- [54] Y. Gong, H. Shan, Y. Teng, N. Tu, M. Li, G. Liang, G. Wang, and S. Wang, "Parameter-transferred wasserstein generative adversarial network (PT-WGAN) for low-dose PET image denoising," *IEEE transactions on radiation and plasma medical sciences*, vol. 5, no. 2, pp. 213–223, 2020.
- [55] Q. Du, X. Ren, J. Wang, Y. Qiang, X. Yang, and N. G.-F. Kazihise, "Iterative PET image reconstruction using cascaded data consistency generative adversarial network," *IET Image Processing*, vol. 14, no. 15, pp. 3989–3999, 2020.
- [56] M. Geng, X. Meng, J. Yu, L. Zhu, L. Jin, Z. Jiang, B. Qiu, H. Li, H. Kong, J. Yuan, *et al.*, "Content-noise complementary learning for medical image denoising," *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 407–419, 2021.
- [57] Y. Wang, L. Zhou, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, and D. Shen, "3D auto-context-based locality adaptive multi-modality GANs for PET synthesis," *IEEE transactions on medical imaging*, vol. 38, no. 6, pp. 1328–1339, 2018.

- [58] B. Zhou, Y.-J. Tsai, X. Chen, J. S. Duncan, and C. Liu, "MDPET: A unified motion correction and denoising adversarial network for low-dose gated PET," *IEEE transactions on medical imaging*, vol. 40, no. 11, pp. 3154–3164, 2021.
- [59] H. Xue, Q. Zhang, S. Zou, W. Zhang, C. Zhou, C. Tie, Q. Wan, Y. Teng, Y. Li, D. Liang, *et al.*, "LCPR-Net: Low-count PET image reconstruction using the domain transform and cycle-consistent generative adversarial networks," *Quantitative imaging in medicine and surgery*, vol. 11, no. 2, p. 749, 2021.
- [60] Z. Hu, H. Xue, Q. Zhang, J. Gao, N. Zhang, S. Zou, Y. Teng, X. Liu, Y. Yang, D. Liang, *et al.*, "DPIR-Net: Direct PET image reconstruction based on the wasserstein generative adversarial network," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 35–43, 2020.
- [61] J. Sun, Q. Zhang, Y. Du, D. Zhang, P. H. Pretorius, M. A. King, and G. S. Mok, "Dual gating myocardial perfusion SPECT denoising using a conditional generative adversarial network," *Medical Physics*, vol. 49, no. 8, pp. 5093–5106, 2022.
- [62] J. Sun, H. Jiang, Y. Du, C.-Y. Li, T.-H. Wu, Y.-H. Liu, B.-H. Yang, and G. S. Mok, "Deep learning-based denoising in projection-domain and reconstruction-domain for low-dose myocardial perfusion SPECT," *Journal of Nuclear Cardiology*, pp. 1–16, 2022.
- [63] J. Sun, Y. Du, C. Li, T.-H. Wu, B. Yang, and G. S. Mok, "Pix2Pix generative adversarial network for low dose myocardial perfusion SPECT denoising," *Quantitative Imaging in Medicine and Surgery*, vol. 12, no. 7, p. 3539, 2022.
- [64] J. Sun, B.-H. Yang, C.-Y. Li, Y. Du, Y.-H. Liu, T.-H. Wu, and G. S. Mok, "Fast myocardial perfusion SPECT denoising using an attention-guided generative adversarial network," *Frontiers in Medicine*, vol. 10, 2023.
- [65] N. Aghakhan Olia, A. Kamali-Asl, S. Hariri Tabrizi, P. Geramifar, P. Sheikhzadeh, S. Farzanefar, H. Arabi, and H. Zaidi, "Deep learning-based denoising of low-dose SPECT myocardial perfusion images: Quantitative assessment and clinical performance," *European journal of nuclear medicine and molecular imaging*, pp. 1–15, 2022.
- [66] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [67] C. Chan, R. Fulton, R. Barnett, D. D. Feng, and S. Meikle, "Postreconstruction nonlocal means filtering of whole-body PET with an anatomical prior," *IEEE Transactions on medical imaging*, vol. 33, no. 3, pp. 636–650, 2013.
- [68] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.
- [69] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [70] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*, PMLR, 2019, pp. 7354–7363.
- [71] H. Xue, Y. Teng, C. Tie, Q. Wan, J. Wu, M. Li, G. Liang, D. Liang, X. Liu, H. Zheng, *et al.*, "A 3D attention residual encoder–decoder least-square GAN for low-count PET denoising," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 983, p. 164638, 2020.
- [72] Y. Fu, S. Dong, M. Niu, L. Xue, H. Guo, Y. Huang, Y. Xu, T. Yu, K. Shi, Q. Yang, *et al.*, "AIGAN: Attention–encoding Integrated Generative Adversarial Network for the reconstruction of low-dose CT and low-dose PET images," *Medical Image Analysis*, vol. 86, p. 102787, 2023.
- [73] C. Shang, G. Zhao, Y. Li, J. Yuan, M. Wang, Y. Wu, and Y. Lin, "Short-Axis PET Image Quality Improvement by Attention CycleGAN Using Total-Body PET," *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [74] Y. Lei, T. Wang, X. Dong, K. Higgins, T. Liu, W. J. Curran, H. Mao, J. A. Nye, and X. Yang, "Low dose PET imaging with CT-aided cycle-consistent adversarial networks," in *Medical Imaging 2020: Physics of Medical Imaging*, SPIE, vol. 11312, 2020, pp. 1043–1049.
- [75] Y. Onishi, F. Hashimoto, K. Ote, H. Ohba, R. Ota, E. Yoshikawa, and Y. Ouchi, "Anatomical-guided attention enhances unsupervised PET image denoising performance," *Medical Image Analysis*, vol. 74, p. 102226, 2021.
- [76] Y. Li, J. Hu, H. Sari, S. Xue, R. Ma, S. Kandarpa, D. Visvikis, A. Rominger, H. Liu, and K. Shi, "A deep neural network for parametric image reconstruction on a large axial field-of-view PET," *European journal of nuclear medicine and molecular imaging*, vol. 50, no. 3, pp. 701–714, 2023.
- [77] Z. Xie, R. Baikejiang, T. Li, X. Zhang, K. Gong, M. Zhang, W. Qi, E. Asma, and J. Qi, "Generative adversarial network based regularized image reconstruction for PET," *Physics in Medicine & Biology*, vol. 65, no. 12, p. 125016, 2020.
- [78] Z. Xie, T. Li, X. Zhang, W. Qi, E. Asma, and J. Qi, "Anatomically aided PET image reconstruction using deep neural networks," *Medical Physics*, vol. 48, no. 9, pp. 5244–5258, 2021.
- [79] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [80] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.
- [81] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5728–5739.
- [82] Y. Luo, Y. Wang, C. Zu, B. Zhan, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3D transformer-GAN for high-quality PET reconstruction," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, Springer, 2021, pp. 276–285.
- [83] S.-I. Jang, T. Pan, Y. Li, P. Heidari, J. Chen, Q. Li, and K. Gong, "Spach transformer: Spatial and channel-wise transformer based on local and global self-attentions for PET image denoising," *arXiv preprint arXiv:2209.03300*, 2022.
- [84] L. Zhang, Z. Xiao, C. Zhou, J. Yuan, Q. He, Y. Yang, X. Liu, D. Liang, H. Zheng, W. Fan, *et al.*, "Spatial adaptive and transformer fusion network (STFNet) for low-count PET blind denoising with MRI," *Medical Physics*, vol. 49, no. 1, pp. 343–356, 2022.
- [85] Y.-R. Wang, P. Wang, L. C. Adams, N. D. Sheybani, L. Qu, A. H. Sarrami, A. J. Theruvath, S. Gatidis, T. Ho, Q. Zhou, *et al.*, "Low-count whole-body PET/MRI restoration: An evaluation of dose reduction spectrum and five state-of-the-art artificial intelligence models," *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–14, 2023.
- [86] R. Hu and H. Liu, "TransEM: Residual Swin-Transformer based regularized PET image reconstruction," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, Springer, 2022, pp. 184–193.
- [87] K. Gong, D. Wu, K. Kim, J. Yang, T. Sun, G. El Fakhri, Y. Seo, and Q. Li, "MAPEM-Net: an unrolled neural network for Fully 3D PET image reconstruction," in *15th International meeting on fully three-dimensional image reconstruction in radiology and nuclear medicine*, SPIE, vol. 11072, 2019, pp. 109–113.
- [88] A. Mehranian and A. J. Reader, "Model-based deep learning PET image reconstruction using forward–backward splitting expectation–maximization," *IEEE transactions on radiation and plasma medical sciences*, vol. 5, no. 1, pp. 54–64, 2020.
- [89] H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler, "Improved low-count quantitative PET reconstruction with an iterative neural network," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3512–3522, 2020.
- [90] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.
- [91] J. Cui, K. Gong, N. Guo, X. Meng, K. Kim, H. Liu, and Q. Li, "CT-guided PET image denoising using deep neural network without prior training data," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, IEEE, 2018, pp. 1–3.
- [92] J. Cui, K. Gong, N. Guo, C. Wu, X. Meng, K. Kim, K. Zheng, Z. Wu, L. Fu, B. Xu, *et al.*, "PET image denoising using unsupervised deep learning," *European journal of nuclear medicine and molecular imaging*, vol. 46, pp. 2780–2789, 2019.
- [93] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical programming*, vol. 45, no. 1–3, pp. 503–528, 1989.

- [94] F. Hashimoto, H. Ohba, K. Ote, A. Teramoto, and H. Tsukada, "Dynamic PET image denoising using deep convolutional neural networks without prior training datasets," *IEEE access*, vol. 7, pp. 96 594–96 603, 2019.
- [95] H. Sun, L. Peng, H. Zhang, Y. He, S. Cao, and L. Lu, "Dynamic PET image denoising using deep image prior combined with regularization by denoising," *IEEE Access*, vol. 9, pp. 52 378–52 392, 2021.
- [96] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by RED," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.
- [97] F. Hashimoto, H. Ohba, K. Ote, A. Kakimoto, H. Tsukada, and Y. Ouchi, "4D deep image prior: Dynamic PET image denoising using an unsupervised four-dimensional branch convolutional neural network," *Physics in Medicine & Biology*, vol. 66, no. 1, p. 015 006, 2021.
- [98] C.-H. Yang and H.-M. Huang, "Simultaneous denoising of dynamic PET images based on deep image prior," *Journal of Digital Imaging*, vol. 35, no. 4, pp. 834–845, 2022.
- [99] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.
- [100] D. Wu, K. Kim, and Q. Li, "Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning," *Medical Physics*, vol. 48, no. 12, pp. 7657–7672, 2021.
- [101] W. Fang, D. Wu, K. Kim, M. K. Kalra, R. Singh, L. Li, and Q. Li, "Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior," *Physics in Medicine & Biology*, vol. 66, no. 15, p. 155 013, 2021.
- [102] W. Jung, H.-S. Lee, M. Seo, Y. Nam, Y. Choi, N.-Y. Shin, K.-J. Ahn, B.-s. Kim, and J. Jang, "MR-self Noise2Noise: Self-supervised deep learning-based image quality improvement of submillimeter resolution 3D MR images," *European Radiology*, pp. 1–13, 2022.
- [103] S. Y. Yie, S. K. Kang, D. Hwang, and J. S. Lee, "Self-supervised PET denoising," *Nuclear Medicine and Molecular Imaging*, vol. 54, pp. 299–304, 2020.
- [104] D. Wu, K. Gong, K. Kim, X. Zhang, J. Ouyang, and Q. Li, *Deep denoising of O-15 water dynamic PET images without training data*, 2020.
- [105] S. Y. Yie, S. K. Kang, D. Hwang, H. Choi, and J. S. Lee, *Self-supervised bone scan denoising*, 2021.
- [106] C. Chan, J. Zhou, L. Yang, W. Qi, and E. Asma, "Noise to noise ensemble learning for PET image denoising," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2019, pp. 1–3.
- [107] S.-K. Kang, S.-Y. Yie, and J.-S. Lee, "Noise2Noise improved by trainable wavelet coefficients for PET denoising," *Electronics*, vol. 10, no. 13, p. 1529, 2021.
- [108] N. Moran, D. Schmidt, Y. Zhong, and P. Coady, "Noisier2noise: Learning to denoise from unpaired noisy data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 064–12 072.
- [109] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.
- [110] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *International Conference on Machine Learning*, PMLR, 2019, pp. 524–533.
- [111] T.-A. Song, F. Yang, and J. Dutta, "Noise2void: Unsupervised denoising of PET images," *Physics in Medicine & Biology*, vol. 66, no. 21, p. 214 002, 2021.
- [112] L. Xiang, Y. Qiao, D. Nie, L. An, W. Lin, Q. Wang, and D. Shen, "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/," *Neurocomputing*, vol. 267, pp. 406–416, 2017.
- [113] K. T. Chen, E. Gong, F. B. de Carvalho Macruz, J. Xu, A. Boumis, M. Khalighi, K. L. Poston, S. J. Sha, M. D. Greicius, E. Mormino, *et al.*, "Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs," *Radiology*, vol. 290, no. 3, pp. 649–656, 2019.
- [114] K. T. Chen, M. Schürer, J. Ouyang, M. E. I. Koran, G. Davidzon, E. Mormino, S. Tiepolt, K.-T. Hoffmann, O. Sabri, G. Zaharchuk, *et al.*, "Generalization of deep learning models for ultra-low-count amyloid PET/MRI using transfer learning," *European journal of nuclear medicine and molecular imaging*, vol. 47, pp. 2998–3007, 2020.
- [115] C.-C. Liu and J. Qi, "Higher SNR PET image prediction using a deep learning model and MRI image," *Physics in Medicine & Biology*, vol. 64, no. 11, p. 115 004, 2019.
- [116] K. T. Chen, T. N. Toueg, M. E. I. Koran, G. Davidzon, M. Zeineh, D. Holley, H. Gandhi, K. Halbert, A. Boumis, G. Kennedy, *et al.*, "True ultra-low-dose amyloid PET/MRI enhanced with deep learning for clinical interpretation," *European journal of nuclear medicine and molecular imaging*, vol. 48, pp. 2416–2425, 2021.
- [117] C. O. Da Costa-Luis and A. J. Reader, "Micro-networks for robust MR-guided low count PET imaging," *IEEE transactions on radiation and plasma medical sciences*, vol. 5, no. 2, pp. 202–212, 2020.
- [118] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczyk, C. Floyd, and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography," *IEEE transactions on medical imaging*, vol. 15, no. 5, pp. 673–686, 1996.
- [119] G. Schramm, D. Rigie, T. Vahle, A. Rezaei, K. Van Laere, T. Shepherd, J. Nuyts, and F. Boada, "Approximating anatomically-guided PET reconstruction in image space using a convolutional neural network," *Neuroimage*, vol. 224, p. 117 399, 2021.
- [120] V. P. Sudarshan, U. Upadhyay, G. F. Egan, Z. Chen, and S. P. Awate, "Towards lower-dose PET using physics-based uncertainty-aware multimodal learning with robustness to out-of-distribution data," *Medical Image Analysis*, vol. 73, p. 102 187, 2021.
- [121] K. Gong, E. Berg, S. R. Cherry, and J. Qi, "Machine learning in PET: From photon detection to quantitative image reconstruction," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 51–68, 2019.
- [122] K. Gong, C. Catana, J. Qi, and Q. Li, "Direct reconstruction of linear parametric images from dynamic PET using nonlocal deep image prior," *IEEE Transactions on Medical Imaging*, vol. 41, no. 3, pp. 680–689, 2021.
- [123] K. T. Chen, O. Adeyeri, T. N. Toueg, M. Zeineh, E. Mormino, M. Khalighi, and G. Zaharchuk, "Investigating simultaneity for deep learning-enhanced actual ultra-low-dose amyloid PET/MR imaging," *American Journal of Neuroradiology*, vol. 43, no. 3, pp. 354–360, 2022.
- [124] C. N. Ladefoged, P. Hasbak, C. Hornnes, L. Højgaard, and F. L. Andersen, "Low-dose PET image noise reduction using deep learning: Application to cardiac viability FDG imaging in patients with ischemic heart disease," *Physics in Medicine & Biology*, vol. 66, no. 5, p. 054 003, 2021.
- [125] F. Deng, X. Li, F. Yang, H. Sun, J. Yuan, Q. He, W. Xu, Y. Yang, D. Liang, X. Liu, *et al.*, "Low-dose 68 Ga-PSMA prostate PET/MRI imaging using deep learning based on MRI priors," *Frontiers in Oncology*, vol. 11, p. 5856, 2022.
- [126] Y.-R. Wang, L. Baratto, K. E. Hawk, A. J. Theruvath, A. Pribnow, A. S. Thakor, S. Gatidis, R. Lu, S. E. Gummidipundi, J. Garcia-Diaz, *et al.*, "Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure," *European journal of nuclear medicine and molecular imaging*, vol. 48, pp. 2771–2781, 2021.
- [127] B. Pan, N. Qi, Q. Meng, J. Wang, S. Peng, C. Qi, N.-J. Gong, and J. Zhao, "Ultra high speed SPECT bone imaging enabled by a deep learning enhancement method: A proof of concept," *EJNMMI physics*, vol. 9, no. 1, p. 43, 2022.
- [128] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," *Advances in Neural Information Processing Systems*, vol. 34, pp. 8780–8794, 2021.
- [129] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, "Diffusion models in medical imaging: A comprehensive survey," *Medical Image Analysis*, p. 102 846, 2023.
- [130] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.
- [131] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.
- [132] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.
- [133] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim, "Dolce: A model-based probabilistic diffusion framework for limited-angle CT reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 498–10 508.

- [134] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, “Diffusion posterior sampling for general noisy inverse problems,” *arXiv preprint arXiv:2209.14687*, 2022.
- [135] B. Kawar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 23 593–23 606, 2022.
- [136] M. Mardani, J. Song, J. Kautz, and A. Vahdat, “A variational perspective on solving inverse problems with diffusion models,” *arXiv preprint arXiv:2305.04391*, 2023.
- [137] K. Gong, K. Johnson, G. El Fakhri, Q. Li, and T. Pan, “PET image denoising based on denoising diffusion probabilistic model,” *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–11, 2023.
- [138] S. Pan, E. Abouei, J. Peng, J. Qian, J. F. Wynne, T. Wang, C.-W. Chang, J. Roper, J. A. Nye, H. Mao, *et al.*, “Full-dose PET synthesis from low-dose PET using high-efficiency diffusion denoising probabilistic model,” *arXiv preprint arXiv:2308.13072*, 2023.
- [139] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, “Consistency models,” 2023.
- [140] C. Shen, Z. Yang, and Y. Zhang, “PET image denoising with score-based diffusion probabilistic models,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2023, pp. 270–278.
- [141] C. Jiang, Y. Pan, M. Liu, L. Ma, X. Zhang, J. Liu, X. Xiong, and D. Shen, “PET-diffusion: Unsupervised PET enhancement based on the latent diffusion model,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2023, pp. 3–12.
- [142] Z. Han, Y. Wang, L. Zhou, P. Wang, B. Yan, J. Zhou, Y. Wang, and D. Shen, “Contrastive diffusion model with auxiliary guidance for coarse-to-fine PET reconstruction,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2023, pp. 239–249.
- [143] T. Xie, C. Cao, Z. Cui, F. Li, Z. Wei, Y. Zhu, Y. Li, D. Liang, Q. jin, G. Chen, and H. Wang, “Brain PET synthesis from MRI using joint probability distribution of diffusion model at ultrahigh fields,” *arXiv preprint arXiv:2211.08901*, 2022.
- [144] I. R. D. Singh, A. Denker, R. Barbano, Ž. Kereta, B. Jin, K. Thielemans, P. Maass, and S. Arridge, “Score-based generative models for PET image reconstruction,” *arXiv preprint arXiv:2308.14190*, 2023.
- [145] “Ultra-low dose PET imaging challenge.” (2023), [Online]. Available: <https://ultra-low-dose-pet.grand-challenge.org/> (visited on 10/10/2023).