Turning brain MRI into diagnostic PET: ^{15}O-water PET CBF synthesis from multi-contrast MRI via attention-based encoder–decoder networks

Ramy Hussein [a,*], David Shin [b], Moss Y. Zhao [a,c], Jia Guo [d], Guido Davidzon [e], Gary Steinberg [f], Michael Moseley [a], Greg Zaharchuk [a]

[a] Radiological Sciences Laboratory, Department of Radiology, Stanford University, Stanford, CA 94305, USA
[b] Global MR Applications & Workflow, GE Healthcare, Menlo Park, CA 94025, USA
[c] Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, USA
[d] Department of Bioengineering, University of California, Riverside, CA 92521, USA
[e] Division of Nuclear Medicine, Department of Radiology, Stanford University, Stanford, CA 94305, USA
[f] Department of Neurosurgery, Stanford University, Stanford, CA 94304, USA

ARTICLE INFO

ABSTRACT

Accurate quantification of cerebral blood flow (CBF) is essential for the diagnosis and assessment of a wide range of neurological diseases. Positron emission tomography (PET) with radiolabeled water (^{15}O-water) is the gold-standard for the measurement of CBF in humans, however, it is not widely available due to its prohibitive costs and the use of short-lived radiopharmaceutical tracers that require onsite cyclotron production. Magnetic resonance imaging (MRI), in contrast, is more accessible and does not involve ionizing radiation. This study presents a convolutional encoder–decoder network with attention mechanisms to predict the gold-standard ^{15}O-water PET CBF from multi-contrast MRI scans, thus eliminating the need for radioactive tracers. The model was trained and validated using 5-fold cross-validation in a group of 126 subjects consisting of healthy controls and cerebrovascular disease patients, all of whom underwent simultaneous ^{15}O-water PET/MRI. The results demonstrate that the model can successfully synthesize high-quality PET CBF measurements (with an average SSIM of 0.924 and PSNR of 38.8 dB) and is more accurate compared to concurrent and previous PET synthesis methods. We also demonstrate the clinical significance of the proposed algorithm by evaluating the agreement for identifying the vascular territories with impaired CBF. Such methods may enable more widespread and accurate CBF evaluation in larger cohorts who cannot undergo PET imaging due to radiation concerns, lack of access, or logistic challenges.

1. Introduction

Cerebrovascular diseases are a worldwide public health issue, impacting all racial and ethnic groups (Yusuf et al., 2001). Stroke alone affects 15 million individuals annually, resulting in five million deaths and five million permanent disabilities, placing a strain on family, community, and health care systems (Mukherjee and Patil, 2011). Early diagnosis and proper evaluation of cerebrovascular diseases can reduce damage to the brain and facilitate faster treatment. Moreover, abnormalities in cerebral blood flow (CBF) are often associated with multiple neurological conditions, including vascular malformations, seizure disorders, and neurodegenerative disorders such as Alzheimer's disease (Iturria-Medina et al., 2016; Leijenaar et al., 2017). Accordingly, accurate CBF quantification is essential for the diagnosis and evaluation of cerebrovascular diseases.

Positron emission tomography (PET) with radiolabeled water (^{15}O-water) is widely considered the gold-standard imaging technique for measuring CBF in humans (Ito et al., 2004). However, due to its prohibitive costs, difficult logistics, and use of ionizing radiation, PET is not widely available, with only around 20 centers in the world offering ^{15}O-water PET CBF imaging, mostly in a research setting. Magnetic resonance imaging (MRI) is a more accessible and cost-effective alternative, with arterial spin labeling (ASL) perfusion MRI and dynamic susceptibility contrast (DSC) perfusion MRI being the two most common exams used to quantify CBF (Detre et al., 1992; Villringer et al., 1988). Despite its widespread use, MRI-derived CBF maps can be inaccurate in the presence of global or focal CBF reductions, as is often seen in patients with cerebrovascular diseases (Grade et al., 2015). This has led to the development of image-to-image translation

* Corresponding author.
E-mail address: ramyh@stanford.edu (R. Hussein).

methods to synthesize PET-like CBF maps from MRI scans, which can potentially improve the quantitative and qualitative assessment of CBF when compared to perfusion MRI-derived CBF measurements, and would be applicable to a wider range of patients and indications than is feasible with PET imaging.

Recent advances in computer vision and the increasing size of medical imaging databases have enabled the development of image-to-image translation methods using deep learning, which can transform one medical imaging modality to another. Examples include predicting computed tomography (CT) images from MRI (Kearney et al., 2020), MRI from CT (Jin et al., 2019), CT from PET (Armanious et al., 2019), PET from CT (Ben-Cohen et al., 2019), MRI from PET (Bazangani et al., 2022), and different MRI contrasts from one another (Dar et al., 2019). However, the clinical utility of such cross-modality translations has yet to be established. In recent years, several image synthesis methods have been proposed to transform multi-parametric brain MRI images into ^{15}O-water PET CBF maps (Guo et al., 2020; Yousefi et al., 2021). These methods could potentially extend the ability to quantitatively characterize cerebrovascular disorders to sites without the capability to perform the gold-standard PET imaging. For instance, Guo et al. (2020) used a deep convolutional neural network (CNN) to predict ^{15}O-water PET scans from multi-contrast MRI inputs, achieving an average structural similarity index (SSIM) of 0.85 in both normal subjects and patients with cerebrovascular disease. Similarly, Yousefi et al. (2021) introduced an attention-guided CNN for translating T1-weighted and ASL data to PET-like images, achieving a comparable SSIM of 0.85 in normal subjects.

Although existing MRI-to-PET translation models could synthesize PET CBF maps of acceptable quality, yet their accuracy and clinical applicability can be further improved. This work introduces a novel multimodal encoder–decoder attention network for synthesizing ^{15}O-water PET CBF maps from structural and perfusion MRI scans. The inputs to the network include structural (T1w and T2-FLAIR) and perfusion images (single-delay [SD] and multi-delay [MD] ASL), as well as CBF estimates derived from the ASL sequences. Experiments were conducted to evaluate the model performance with different loss functions, network settings, and subsets of input MRI scans. The results demonstrate that the proposed method outperforms the state-of-the-art methods in terms of PET prediction performance. The main contributions of this work are as follows:

- A 3D convolutional encoder–decoder network incorporating attention mechanisms is proposed, along with a custom-designed loss function.
- Quantitative and qualitative analyses are conducted to study whether the integration of anatomical and tissue perfusion information from structural and ASL MRI exams can improve the prediction of PET CBF for both healthy controls and cerebrovascular disease patients.
- Ablation studies are performed to evaluate the impact and contributions of various loss functions, attention mechanisms, and different input MRI scans on the overall quality of the synthesized PET images.
- The diagnostic accuracy of the predicted PET CBF maps is evaluated using the Receiver Operating Characteristic (ROC) analysis, which assesses the ability to identify brain regions with impaired CBF. The classification accuracy, sensitivity, and specificity are also calculated to evaluate the clinical value of synthetic PET in diagnosing cerebrovascular diseases in MRI-only facilities.

2. Related work

This section initially reviews the concurrent and prior deep learning models that are commonly used in cross-modality brain image synthesis applications. Following this, we review the recent MRI-to-PET translation networks related to our study.

2.1. Cross-modality brain image-to-image translation

Recently, image-to-image translation networks have been utilized to address various image prediction problems in the medical field. Wolterink et al. (2017) took a pioneering role in applying deep learning for cross-modality medical image synthesis. They applied a generative adversarial network (GAN) to convert 2D brain MRI images into 2D brain CT images and vice versa. Results from a separate test set of six patients showed that GANs can generate CT images that closely resemble actual CT images, with an average PSNR of 32.3 dB. Subsequently, Dar et al. (2019) employed conditional generative adversarial networks for multi-contrast MRI synthesis. The combination of adversarial loss with pixel-wise and perceptual losses improved the synthesis performance of both T1- and T2-weighted images. Additionally, Yang et al. (2020b) developed a cross-modality MRI image generation method for multimodal registration and segmentation using conditional generative adversarial networks. This method achieved comparable results on five brain MRI datasets while using a single modality image as an input.

In Armanious et al. (2020), a GAN-based framework with a novel generator architecture and style-transfer losses was proposed to address three medical image-to-image translation problems: PET-to-CT translation, MRI motion artifacts correction, and PET image denoising. The quantitative results and radiologists' evaluations demonstrated the superiority of the proposed GAN architecture compared to existing translation methods. Yang et al. (2020a) proposed a structure-constrained cycleGAN for unsupervised MRI-to-CT synthesis. This approach incorporated an additional structure-consistency loss function and a self-attention module to generate high-quality synthetic brain and abdomen CT images. Liu et al. (2022) proposed a transformer-based MRI synthesis approach, named multi-contrast multi-scale transformer (MMT), for missing MRI sequence imputation. Experiments on two multi-contrast MRI datasets showed that MMT can outperform state-of-the-art MRI synthesis methods both quantitatively and qualitatively. This suggests that vision transformers can be used not only for medical image recognition problems, but also for more challenging image-to-image translation problems.

Image-to-image translation networks were also used to reduce or even eliminate the need for gadolinium-based contrast agents (GBCAs) in MRI studies. In Gong et al. (2018), a convolutional encoder–decoder network was utilized in to synthesize high-quality contrast-enhanced MRI (CE-MRI) images from images taken with a reduced dose of the GBCA. Quantitative results and qualitative assessments demonstrated that the synthesized CE-MRI images had a significant improvement in image quality and contrast enhancement compared to the acquired low-dose images. Moreover, a 3D high-resolution fully convolutional network was proposed in Chen et al. (2021) to map a set of pre-contrast MRI scans to CE-MRI. The pre-contrast MRI sequences of T1-weighted (T1w), T2-weighted (T2w), and apparent diffusion coefficient (ADC) map were used as inputs to the network and the post-contrast T1w being the target output. Results showed high-quality synthetic CE-MRI images, potentially allowing deep learning to substitute for GBCAs and reduce gadolinium deposition.

Others have explored the same synthesis problem by using T1w, T2w, and fluid-attenuated inversion recovery (FLAIR) MRI sequences as inputs for two deep convolutional neural networks (dCNN) (Preetha et al., 2021). Results indicated that the quantification of the synthetic CE-MRI images could effectively gauge the patient's response to treatment with minimal discrepancy compared to true CE-MRI images obtained through gadolinium administration. Furthermore, Xie et al. developed a more advanced cascade neural network architecture that incorporated the contour information of the input unenhanced MRI images to enhance the quality of the synthetic CE-MRI images (Xie et al., 2022). Quantitative and qualitative assessments on a test set of 169 patients revealed that, due to the contour information, no intensity differences were seen in both tumor and non-tumor brain regions.

2.2. MRI-to-PET translation

Researchers have investigated the possibility of converting MRI data into PET-like images. Li et al. (2014) utilized a 3D CNN model to generate FDG-PET patterns from MRI data of the ADNI database. This model was equipped with a large number of parameters that could capture the nonlinear relationship between MRI and PET data. The trained network was then used to predict PET patterns in subjects with only MRI data. This technique was shown to be effective in diagnosing Alzheimer's disease (AD). In Sikka et al. (2018), Sikka et al. proposed a 3D U-Net architecture to synthesize FDG-PET images from MRI scans, with the aim of improving the diagnosis of Alzheimer's disease. Their model yielded an average increase in the classification accuracy of 4.25%. Similarly, in Pan et al. (2018), Pan et al. developed a two-stage deep learning framework using incomplete multi-modal imaging data, incorporating a cycleGAN model to impute the missing PET data from MRI scans. The resulting synthetic PET images were found to enhance the accuracy of Alzheimer's disease classification.

In Gao et al. (2021), Gao et al. proposed a task-induced pyramid and attention generative adversarial network for FDG-PET imputation from MRI. Experiments conducted on the ADNI database showed the model to be effective in FDG-PET synthesis, achieving an average SSIM of 0.915 and PSNR of 29 dB. Lan et al. (2021) utilized a 3D self-attention conditional GAN (SC-GAN) that incorporated attention modules to establish connections between different neuroimaging modalities. The SC-GAN was evaluated on the ADNI database, where it was used to generate various downstream image contrasts, such as amyloid PET, fractional anisotropy (FA) and mean diffusivity (MD) maps. However, the PET synthesis error was found to be relatively high in brain regions with high amyloid-β load, suggesting that synthetic PET cannot replace amyloid PET imaging for clinical purposes. Chen et al. (2019) developed a 2D U-Net model to generate full-dose amyloid PET images from a combination of extremely low-dose amyloid PET and multi-contrast MR images. Quantitative analyses revealed that this form of imaging integration can successfully produce synthetic PET images with standardized uptake value ratio (SUVR) values that are comparable to those of the true full-dose PET images.

In Wei et al. (2019), Wei et al. introduced a sketcher-refiner GAN to predict the PET-derived myelin content map from multimodal MRI. The predictions showed results similar to the gold-standard PET-derived maps, indicating its potential for clinical applications in the management of Multiple Sclerosis patients. Similarly, Yaakub et al. presented a two-stage deep learning framework to support the clinical assessment of patients with focal epilepsy by identifying potential areas of hypometabolism in FDG-PET scans (Yaakub et al., 2019). This framework first trained a GAN to learn the mapping between healthy FDG-PET and T1-weighted (T1w) MRI data. Then, pseudo-normal PET images were synthesized from T1w MRI scans of patients with epilepsy for comparison to the real PET scans. With the synthetic PET data, an average sensitivity of 92.9% and 74.8% was observed in seven cases with MR-visible epileptogenic lesions and 13 cases with non-contributory MR, respectively.

Guo et al. demonstrated the potential of utilizing multi-contrast MRI scans to synthesize ^{15}O-water PET CBF images in a study published in 2020 (Guo et al., 2020). A two-dimensional U-Net model was employed, which took structural MRI and single-delay and multi-delay ASL exams as inputs to generate a PET CBF map, allowing for more precise CBF measurements in MRI-only sites. Yousefi et al. (2021) developed a residual CNN-based synthesis method to transform ASL data into PET CBF maps. The approach also incorporated T1w MRI for anatomical information to increase the accuracy of PET synthesis. Quantitative measures and a blind reader study revealed a high level of similarity between the true and synthetic PET images, with an average SSIM of 0.85. All participants in the study were healthy controls without any pathology. In Chen et al. (2020), Chen et al. employed a similar deep learning architecture to predict cerebrovascular reserve

(CVR, defined as the percent CBF increase from a baseline value after acetazolamide administration) images from structural and perfusion-based MRI scans as well as the baseline ^{15}O-water PET CBF measured before acetazolamide administration. Quantitative and comparative analyses showed a high diagnostic performance in identifying vascular territories with impaired CVR.

Hu et al. (2020) utilized a bidirectional GAN to transform MRI into PET images while preserving the various brain structures of different individuals. They achieved satisfactory quantitative results, though the quality of the generated PET images was somewhat limited. Subsequently, Shin et al. (2020) applied the BERT algorithm (Devlin et al., 2018) to generate synthetic amyloid and FDG-PET images from T1w MRI data with minimal pre- and post-processing. Nevertheless, the quantitative and qualitative PET prediction results were limited, making it unsuitable for clinical applications. Also, a conditional GAN (c-GAN)-based approach was introduced in Wang et al. (2018) to generate full-dose PET images from low-dose ones. By testing the model on brain data of both healthy subjects and mild cognitive impairment patients, it was demonstrated to outperform baseline methods in terms of both quantitative and qualitative results. Zhang et al. (2022) further presented a 3D end-to-end generative adversarial network, which was designed to learn a mapping function to transform MRI scans into their underlying PET scans. The 3D multiple convolution U-Net (MCU) generator architecture has been implemented in order to enhance the quality of the synthesis results while preserving the diverse brain structures of different subjects. By combining MRI and synthetic PET scans, the accuracy of multi-class AD diagnosis has been increased by approximately 1% when compared to using MRI alone.

3. Materials and data preprocessing

This study was approved by the Institutional Review Board of Stanford University in accordance with the ethical standards of the Helsinki Declaration for medical research involving human subjects, and is HIPAA compliant. Written informed consent was obtained from all participants prior to the study. Our dataset was acquired between April 2017 and March 2022. Data were acquired from 131 subjects (72 healthy controls [HC] and 59 cerebrovascular disease patients [PT]) on a 3T PET/MR hybrid system (SIGNA PET/MR, GE Healthcare, Waukesha, WI, USA) using an eight-channel head coil. Participants were instructed to refrain from food or beverage containing caffeine at least six hours before imaging.

Our dataset was composed of two cohorts, as illustrated in Fig. 1. The first cohort included PET/MRI data from 30 healthy controls (HCs) and 40 patients (PTs), with 4 cases excluded due to missing MRI or PET scans. Each participant had a single visit, during which three simultaneous PET/ASL acquisitions were acquired (two before and one 15 min after the intravenous administration of the vasodilator [acetazolamide, ACZ] at a dose of 15 mg/kg up to a maximum of 1 g). The second cohort included PET/MRI data from 42 HCs and 19 PTs, with 1 HC participant excluded due to missing PET scans. Of the 41 HCs, 31 underwent two identical sessions on different days. During each session, two simultaneous PET/ASL acquisitions were acquired from the participants (one before and one 15 min after ACZ administration). The demographic information of the 126 participants involved in our study is presented in Table 1.

MRI perfusion scans included two ASL acquisitions: a single-delay (SD) and a multi-delay (MD) pseudo-continuous ASL (pCASL). Additionally, dynamic susceptibility contrast (DSC) perfusion MRI was performed following ACZ administration. MRI-based CBF maps were generated from both SD-ASL and MD-ASL using a general kinetic model (Buxton et al., 1998; Alsop et al., 2015). Non-iterative methods were used to derive arterial transit time (ATT) from MD-ASL (Dai et al., 2012). Proton density (PD) images were also collected as part of the ASL acquisitions for quantitation. Magnetic resonance angiography (MRA), gradient-echo (GRE), T1-weighted, and T2-weighted fluid-attenuated inversion recovery (T2- FLAIR) were also acquired from all
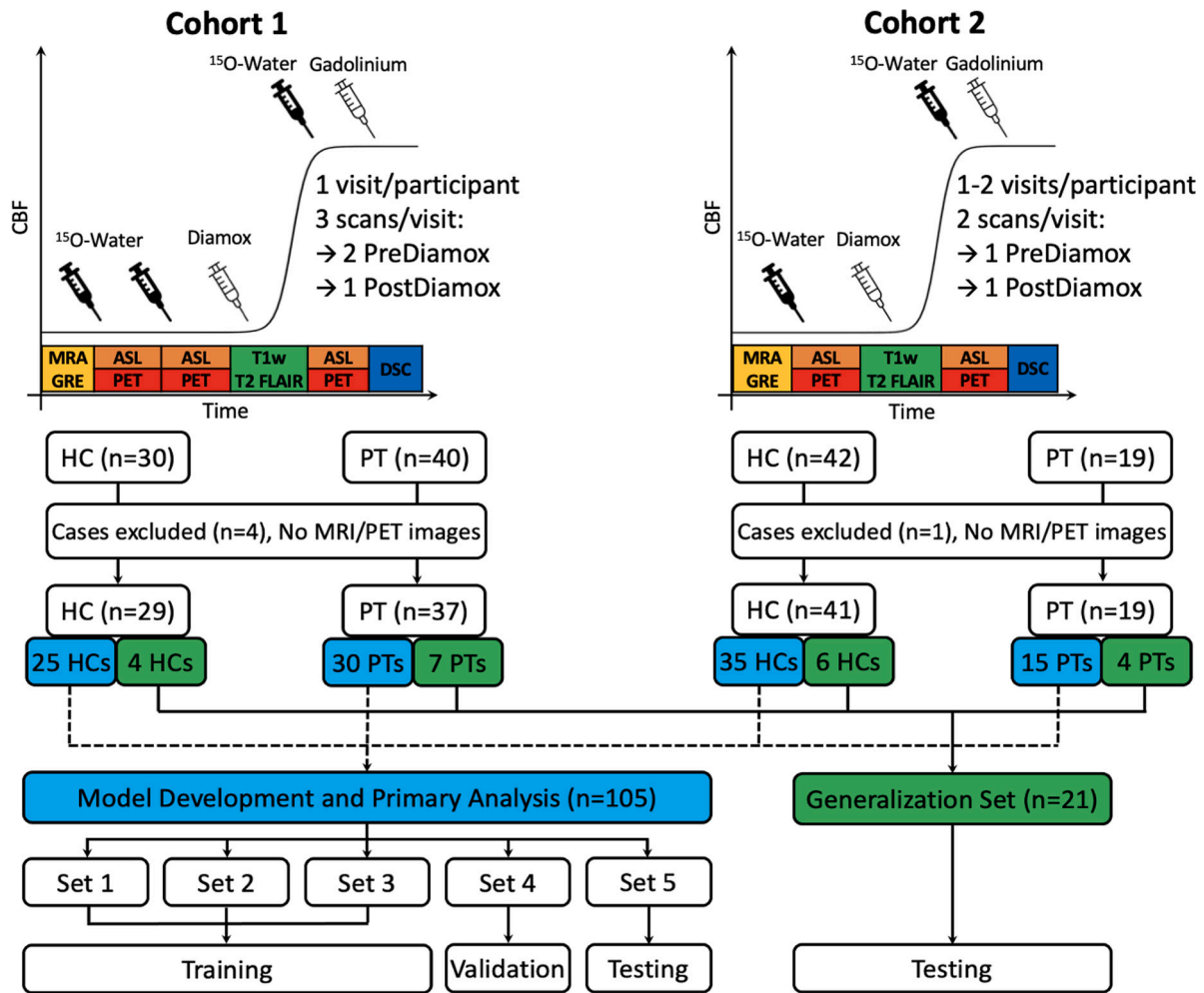
Fig. 1. Experimental design for measuring CBF using PET/MRI in two cohorts. In cohort 1, three simultaneous PET/ASL acquisitions were acquired from the participants in a single visit (two scans before and one scan 15 min after the administration of the vasodilator [acetazolamide, ACZ]). In cohort 2, two simultaneous PET/ASL acquisitions were acquired from the participants in each visit (one before and one 15 min after ACZ administration); Of the 41 HCs in cohort 2, 31 had two separate imaging sessions on different days.

Table 1
Demographic information of the dataset. Age is presented as mean ± standard deviation.

Group	Controls		Patients	
Gender	Male	Female	Male	Female
Number	27	43	25	31
Age	36.0 (±9.6)	37.5 (±13.0)	48.0 (±17.0)	42.0 (±13.2)
Ethnicity				
Hispanic or Latino		Not Hispanic or Latino		Unknown
12		68		46
Race					
White	Asian	African-American	One+ Race	Other	Unknown
60	42	8	6	6	4

participants. A summary of the parameters used for ASL and structural MR imaging is presented in Table A.1.

In both cohorts, PET attenuation correction was performed using a two-point Dixon MRI acquisition and an atlas-based algorithm (Zhao et al., 2021). The quantitative PET CBF was measured using ^{15}O-water injection and an image-derived arterial input function (AIF) kinetic model (Khalighi et al., 2018), along with a 1-compartment model (Zhou et al., 2001) using PMOD software. The PET images were reconstructed with a resolution of $1.56 \times 1.56 \times 2.78$ mm^3 and were corrected for signal decay and attenuation. All images were co-registered to the T1w images

and normalized to the Montreal Neurological Institute (MNI) brain template (Mazziotta et al., 2001) with 2 mm isotropic resolution using Advanced Normalization Tools (ANTs) software (Tustison et al., 2014). The ANTs was employed with the standard settings, which conducted rigid and affine registrations. The brain tissue segmentation was performed using FSL (Smith et al., 2004) and all 3D PET/MRI images were cropped to $96 \times 96 \times 64$ voxels for faster computations.

In order to facilitate a comprehensive evaluation of our proposed MRI-to-PET translation model's performance, the entire dataset comprising PET/MRI data from Cohorts 1 and 2 was divided into two principal subsets, i.e. (n = 105 vs. 21). The first subset, referred to as the "model development and primary analysis" set, served as the primary data for conducting extensive grid search, as well as for model training, validation, and testing. Rigorous experimentation, optimization, and analysis were performed on this set to ensure the development of a robust and effective model. Conversely, the second subset, known as the "generalization set", was deliberately set aside and remained entirely separate from the model training process. This distinct dataset was reserved for the purpose of emulating real-life scenarios and assessing the model's ability to generalize and perform on unseen data. By employing the "generalization set", we aimed to evaluate the model's performance beyond the confines of the training data, thus increasing its reliability and applicability in practical settings.
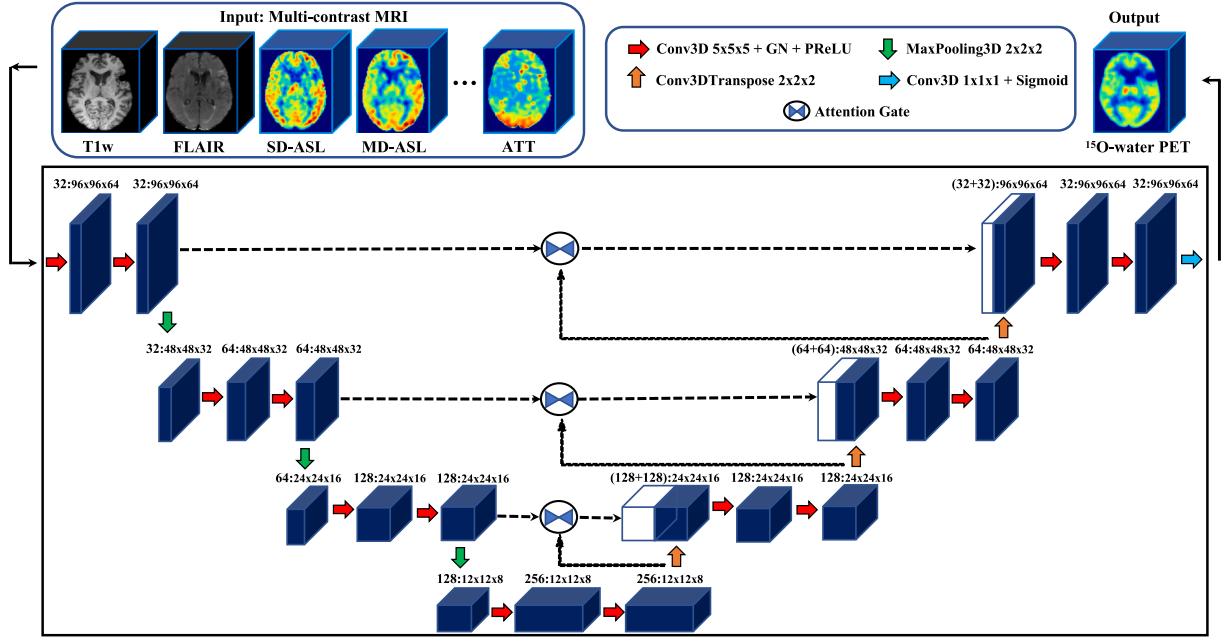
Fig. 2. Attention-based encoder–decoder network architecture for predicting PET CBF maps from multi-contrast MRI. The input to the network is an 8-channel tensor $\mathbf{X} \in \mathbb{R}^{8:96 \times 96 \times 64}$ that includes data from T1w, T2-FLAIR, PD, SD-ASL, MD-ASL, SD-CBF, MD-CBF, and ATT. ^{15}O-water PET CBF is the target image. The number of channels is shown above each of the encoder and decoder blocks. Conv3D = 3D convolutional layer, GN = group normalization, PReLU = parametric rectified linear unit, Conv3DTranspose = transposed 3D convolution layer, and MaxPooling3D = Max pooling operation for 3D data.

4. Methodology

4.1. Proposed network architecture

Fig. 2 shows the architecture of the proposed 3D convolutional encoder–decoder network. The input to the network is an 8-channel tensor $\mathbf{X} \in \mathbb{R}^{h \times w \times d \times 8}$ that includes data from structural MRI (T1w, T2-FLAIR) and perfusion-related MRI scans (ASL difference images from SD-ASL and MD-ASL acquisitions, PD images obtained as part of quantification for the SD-ASL CBF calculations), as well as quantified metrics such as SD-CBF, MD-CBF, and ATT derived from MD-ASL. The output of the network $\mathbf{Y} \in \mathbb{R}^{h \times w \times d \times 1}$ denotes the ^{15}O-water PET CBF map. To enable the transformation of multi-contrast MRI into PET CBF, an attention-based encoder–decoder network is developed to serve as a non-linear mapping function f_θ, such that $\mathbf{Y} = f_\theta(\mathbf{X})$, where θ contains the network parameters to be learned.

In a previous study, we demonstrated the ability to predict the gold-standard ^{15}O-water PET CBF from a set of 16 input MRI contrasts using a 2D convolutional neural network (Guo et al., 2020). In this study, we improved the quality of synthetic PET by utilizing an attention-based 3D structure that capitalized on the spatial information across eight volumetric MRI scans and captured the long-range feature interactions necessary for accurate predictions. However, the application of 3D models to brain image-to-image translation problems is limited by the scarcity of annotated brain imaging data and the associated high computational cost. Therefore, we employed a larger cohort of healthy controls and cerebrovascular disease patients for the current study and applied several data augmentation strategies to further expand the overall number of PET/MRI data samples needed for improved translation performance. This included flipping (horizontally and vertically), shifting (horizontally and vertically), and rotating (clockwise and anti-clockwise) the input and output images, resulting in an eight-fold increase in the dataset size. Finally, a custom loss function was carefully designed to maintain contextual and structural information in input multi-contrast MRI scans and thus optimize the performance of the MRI-to-PET translation network.

The proposed attention-guided encoder–decoder network, illustrated in Fig. 2, utilizes 3D convolutional neural networks to integrate multiple MRI scans to generate high-quality synthetic PET scans. This network is composed of three modules: the encoder, decoder, and attention mechanisms. The encoder performs a series of consecutive 3D convolutions to compress the input multimodal MRI volumes into a low-resolution representation map, while the decoder applies 3D deconvolutions and upsampling operations to the representation map to generate the ^{15}O-water PET CBF maps. Attention mechanisms are also employed to identify relevant aspects of the input MRI scans at both the channel and spatial levels, resulting in a fine-grained quality prediction. To assess the influence of the input MRI sequences on the PET CBF prediction, the model was trained and tested using structural MRI only, perfusion MRI only (*i.e.*, the ASL sequences), and a combination of structural and perfusion MRI images.

4.2. Attention mechanisms

Convolutional encoder–decoder networks are widely utilized in image segmentation and image-to-image translation algorithms. However, the use of predefined convolutional filters restricts the encoder–decoder networks from learning global information while only leveraging local information (Zhang et al., 2019). This results in a non-trivial bias by discarding some of the essential features needed for accurate performance. Increasing the size of convolutional filters and deepening the encoder–decoder architectures are some of the naive solutions introduced to improve the image segmentation or synthesis performance. Nevertheless, these solutions may drastically increase the computational complexity without any discernible improvement in the results.

Attention mechanisms have been proposed as an advanced solution to capture long-range feature interactions and improve the performance of convolutional encoder–decoder networks (Oktay et al., 2018). In the context of MRI-to-PET synthesis, 3D MRI and PET images contain both brain and non-brain voxels (voxels outside the brain margins). Also, for patients with cerebrovascular disorders, abnormal lesions may only be present in specific regions of interest (ROIs) in certain MRI scans. To
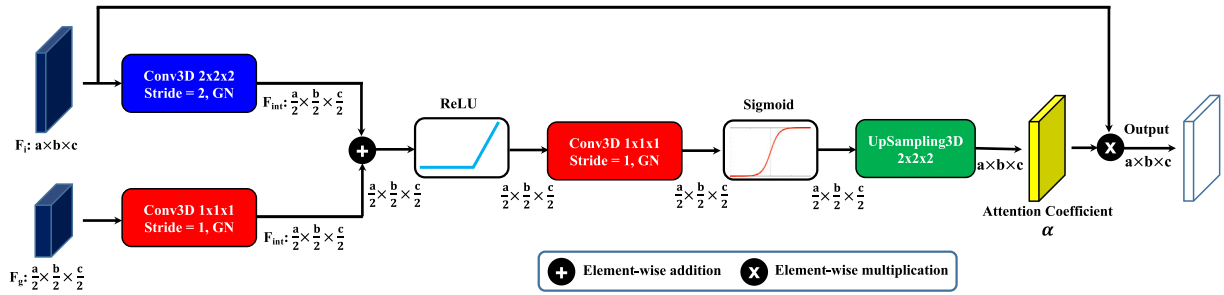
Fig. 3. The schematic of an attention mechanism used in the 3D convolutional encoder–decoder network. Input features (F_i) are multiplied element-wise by attention coefficients (α) computed in the attention module. The gating features (F_g) collected from a lower layer of the network are used to identify the spatial regions of interest with relevant activations and contextual information, GN denotes group normalization.

address this, we propose to use an attention mechanism, depicted in Fig. 3, to enable the encoder–decoder network to focus on both brain voxels and ROIs of varying sizes and appearances.

We opted for an additive soft attention mechanism, which gives different weights to the various regions of the feature map. The regions with higher relevance are given larger weights, while those with less relevance are assigned smaller weights. During the training process, the weights of the soft attention are optimized to enable the model to determine which regions should be given more attention. This attention mechanism takes two inputs, the input features (F_i) from the encoder and the gating features (F_g) from a coarser scale of the decoder. Since maximum pooling operations are applied recursively in the downsampling path of the network, F_g generally has smaller dimensions and better feature representation than F_i, as F_g originates from deeper layers in the network.

A strided 3D convolution is used to reduce the dimensions of F_i, producing an intermediate feature map F_{int}. Additionally, a regular $1 \times 1 \times 1$ 3D convolution is applied to the gating features, producing a separate intermediate feature map. These two feature maps are then element-wise summed, with aligned weights becoming larger and unaligned weights becoming relatively smaller. The resultant tensor is then passed through a rectified linear unit (ReLU) activation function, followed by a $1 \times 1 \times 1$ 3D convolution and a sigmoid function, which produces the attention coefficients (α). Lastly, the attention coefficients are upsampled to the original dimensions of F_i and then element-wise multiplied to F_i, thereby scaling the different regions of the input feature map according to their relevance.

Since different MRI sequences and spatial patterns impose different effects on the quality of synthesized PET CBF maps, the proposed encoder–decoder network was equipped with an additive soft attention mechanism, allowing it to investigate the pertinent aspects of the input MRI data at both the channel and spatial levels for a fine-grained quality prediction.

4.3. Image preprocessing and rescaling techniques

In order to preprocess the images, various rescaling methods, including data normalization and standardization, were explored. Through a rigorous analysis, it was determined that the most effective approach involved dividing all images by their mean value and multiplying them by a scalar factor. This rescaling technique yielded notable improvements in our experimental outcomes. The rationale behind selecting the scalar factor was to reduce pixel values for enhanced training of neural networks. Smaller pixel values offer several advantages, such as narrowing the overall value range, promoting convergence during training, and mitigating the impact of outliers. To determine the optimal scalar factor, an extensive grid search was conducted, involving systematic experimentation with different scalar values. The iterative process revealed that scalar values of either 0.05 or 0.1 consistently produced the most favorable outcomes in terms of performance metrics and model convergence.

4.4. Image quality assessment

The similarity between the synthetic and reference PET CBF images were evaluated using normalized root-mean-square error (NRMSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR), defined as:

$$NRMSE = \frac{1}{x_{max} - x_{min}} \sqrt{\frac{1}{mnp} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(x(i,j,k) - y(i,j,k) \right)^2} \quad (1)$$

where x and y refer to the reference and synthetic PET scans, x_{max} and x_{min} are the maximum and minimum intensity values of the reference PET, $x(i,j,k)$ and $y(i,j,k)$ are the reference and predicted voxel intensity values, and m, n, and p are the dimensions of reference or predicted PET images.

$$SSIM = \frac{\left(2\mu_x \mu_y + c_1 \right) \left(2\sigma_{xy} + c_2 \right)}{\left(\mu_x^2 + \mu_y^2 + c_1 \right) \left(\sigma_x^2 + \sigma_y^2 + c_2 \right)} \quad (2)$$

where μ_x and μ_y denote the mean values of reference and synthetic PET images, σ_x^2 and σ_y^2 denote the variance of reference and synthetic PET images, σ_{xy} is the covariance of both PET images, c_1 and c_2 are two constants used to stabilize the division with weak denominator, $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, $k_1 = 0.01$, $k_2 = 0.03$, and L is the dynamic range for the voxel intensity values of PET images.

$$PSNR = 10 \cdot \log_{10} \left(\frac{y_{max} \cdot mnp}{\sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(x(i,j,k) - y(i,j,k) \right)^2} \right) \quad (3)$$

where y_{max} is the maximum voxel intensity value of the image.

4.5. Custom loss function

In medical image translation problems, the loss function is used to quantify the accuracy of the predictive model in reproducing the target image. In the case of MRI-to-PET translation, the loss function is dependent on two inputs: the true PET image and the synthetic PET image produced by the encoder–decoder network. A lower value of the loss function indicates a more accurate prediction, while a higher value suggests a poor performance. The mean squared error (MSE) is a commonly used loss function for image synthesis, but it is not suitable for medical images that are prone to artifacts and ghosting. Additionally, MSE treats all regions of the medical image equally, which may result in fairly good overall prediction but poorer prediction in the regions of most interest (*e.g.*, abnormal lesions that are crucial for the disease assessment). To address this issue, more appropriate loss functions have been developed to better characterize the structure of the synthesized PET in comparison to the reference PET scan.

This study introduces a specifically designed loss function to improve the accuracy of the proposed encoder–decoder network. This loss function is a combination of several loss components that work cooperatively to drive the network toward the most accurate representation of PET images.

(1) Voxel-wise reconstruction loss: The mean absolute error (MAE) is used as a reconstruction loss that measures the voxel-wise similarity between actual and synthetic images, and it is defined as:

$$\mathcal{L}_r = \frac{1}{mnp} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left| x(i,j,k) - y(i,j,k) \right| \qquad (4)$$

where $x(i,j,k)$ and $y(i,j,k)$ are the voxel intensity values of reference and synthetic PET images.

Using the MAE as a reconstruction loss is beneficial in two ways: it minimizes the difference between the target and predicted images at a voxel level, and it regularizes the network to ensure a robust prediction performance when dealing with data containing outliers and artifacts.

(2) Perceptual loss: In addition to the reconstruction loss, the SSIM loss (Zhao et al., 2016) is used as a perceptual loss to maximize the structural similarity between both real and synthetic images.

The SSIM can efficiently measure the perceptual difference between images using three characteristics: luminance, contrast, and structure (Wang et al., 2004). Using it in training image translation models leads to better quality and more realistic synthetic images. The perceptual loss $\mathcal{L}_p$ is defined as:

$$\mathcal{L}_p = 1 - SSIM \qquad (5)$$

The overall MRI-to-PET translation loss, $\mathcal{L}$, is defined as the weighted sum of both reconstruction and perceptual loss components:

$$\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_p \mathcal{L}_p \qquad (6)$$

where λ_r and λ_p are the weights for reconstruction and perceptual loss terms, respectively.

Hyperparameter optimization through a grid search has revealed that a weighting of 0.2 for λ_r and 0.8 for λ_p yields prediction results that closely resemble subjective ratings.

4.6. Hyperparameter tuning and model selection

In our study, we employed a grid search combined with cross-validation to determine the optimal hyperparameters for our convolutional encoder–decoder network. The hyperparameters considered in the grid search included the depth of the network (with options of 3 and 4), the number of convolutional filters in both the encoder and decoder modules, the kernel size ($3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$), the weights assigned to the reconstruction and perceptual loss terms, the batch size (ranging from 4 to 32), the learning rate, various data normalization methods as described in Section 3, and different combinations of 16 input MRI contrasts.

During the grid search, we divided the "model development and primary analysis" dataset into training, validation, and test sets (see Fig. 1). The training set was used to train the encoder–decoder network, while the validation set enabled us to assess the performance of different hyperparameter configurations. Cross-validation was performed using a 5-fold setup, where each fold represented a distinct partition of the training set.

For each combination of hyperparameters, the neural network was trained on the training portion of the training set and evaluated on the validation set. This process was repeated for each fold, and the performance metrics (including NRMSE, SSIM, and PSNR) were averaged across the folds to obtain a representative measure for each hyperparameter configuration. By comparing the performance of different configurations based on the average evaluation metrics, we identified the optimal set of hyperparameters. The selected model, trained using the best hyperparameters, was then evaluated on the independent test set to assess its generalization ability.

To ensure balanced representation between healthy controls and patients, the dataset was stratified based on the number of individuals in each group during the partitioning process. The stratified splits were implemented to achieve proportional allocation of healthy controls

and patients in the training, validation, and test sets. However, it is important to note that demographic variables beyond the binary grouping of healthy controls and patients were not considered in the stratification process. Future studies could explore the influence of additional demographic factors, such as age or gender, on the model's performance.

This grid search combined with stratified cross-validation methodology allowed us to systematically explore a wide range of hyperparameter values and select the optimal configuration for our encoder–decoder network, enhancing the effectiveness and reliability of our proposed approach.

4.7. Network implementation

The proposed convolutional encoder–decoder network was implemented in Python with the TensorFlow framework. The network encoder had four convolutional layers, with the number of filters being 64, 128, 256, and 512, respectively. For each convolutional layer, the kernel size was $5 \times 5 \times 5$. The training and testing of the network were carried out on two NVIDIA Tesla V100-PCIE Volta GPUs. The custom loss function described in Section 4.5 was employed to optimize the network's weights and improve its predictive power.

In our experiments, the Nesterov Adam optimizer (Dozat, 2016), an improved variant of the Adam optimization algorithm (Kingma and Ba, 2014), was used with a learning rate of 0.0002 and a batch size of 4. The encoder–decoder network was trained using the proposed custom-designed loss function for 150 epochs, with an early stopping of 20 epochs. Early stopping is a form of regularization used to terminate training before the model starts to overfit (Raskutti et al., 2014).

4.8. Identifying regional CBF abnormalities

To assess the feasibility and clinical value of the proposed MRI-to-PET translation method, the utility of synthetic PET images was tested for identifying regional CBF abnormalities in cerebrovascular disease patients. CBF was measured in 10 brain regions, which were divided into anterior, three middle, and posterior parts in both the right and left hemisphere, based roughly on the Alberta Stroke Program Early CT Score (ASPECTS; see Fig. B.1) (Barber et al., 2000). The CBF measured in these 10 vascular territories was used to identify any regional CBF abnormalities and affected brain areas in the patients with cerebrovascular diseases. Since the global and regional values of CBF increase substantially after the administration of acetazolamide (Diamox), the threshold CBF used to detect abnormal vascular territories in the CBF maps obtained before the acetazolamide administration was relatively lower than that used after the acetazolamide administration.

For the pre-diamox and post-diamox measurements, CBF values were obtained from the ten ASPECTS brain regions from all healthy control participants in the "model development and primary analysis" dataset. The mean and standard deviation of CBF values were computed for each brain region across all healthy control participants. Subsequently, a set of thresholds, including the example threshold of three standard deviations below the mean CBF values, were established. These thresholds were used to generate ground truth labels for the corresponding ten brain regions of the patients in the "generalization set" by comparing the mean CBF values of these regions to their respective thresholds. A value below the threshold indicated an abnormal brain region with significantly reduced CBF (label '1'), while a value above or equal to the threshold indicated a normal brain region (label '0'). In the evaluation of synthetic PET CBF for the same patients, an iterative process involved varying threshold values (ranging from 20 to 100 ml/100 g/min, with an increment of 1) to create predicted labels by comparing the CBF values to each threshold. True Positive Rate (TPR) and False Positive Rate (FPR) were calculated for each threshold. By computing the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve using the TPR and FPR values at different

thresholds, the performance of the model in distinguishing abnormal brain regions with reduced CBF from normal regions was assessed separately for pre-diamox and post-diamox measurements.

The diagnostic performance of both synthetic PET CBF and ASL-derived CBF (SD-CBF and MD-CBF) maps was evaluated in both pre-diamox and post-diamox conditions using different threshold CBF values, which were defined as two, three, and four standard deviations below the mean CBF values in the healthy control participants. The ROC curves were used to demonstrate the classification performance of synthetic PET CBF, SD-CBF, and MD-CBF at different discrimination thresholds before and after acetazolamide administration. The AUC scores were also calculated to assess the diagnostic ability of synthetic PET CBF, SD-CBF, and MD-CBF to identify the vascular territories with abnormally low CBF. Additionally, the classification accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated at the model threshold that achieved the highest Youden's index (Youden, 1950).

4.9. Statistical analyses

The 10 ASPECTS brain territories were surveyed to compare the average CBF of healthy controls and cerebrovascular disease patients. Bland-Altman analyses were conducted to examine the agreement of regional CBF between the true ^{15}O-water PET CBF maps and the synthesized PET CBF maps generated by our model, as well as the SD-CBF maps and MD-CBF maps measured with SD-ASL and MD-ASL, respectively.

Joint intensity scatter plots were also used to study the relationship between the regional CBF measurements before and after acetazolamide administration. Pearson's correlation coefficient (r) was adopted to measure the strength of the linear association between the true ^{15}O-water PET CBF and each of the synthetic PET CBF, SD-CBF, and MD-CBF. It is worth noting that SD-CBF and MD-CBF are directly available from MRI, and any improvement from their prediction represents the added value of the trained network.

Lastly, cluster-adjusted heteroscedastic linear regression (Harvey, 1976) was employed to analyze the CBF measurement error and assess the influence of group type (control/patient) and scan time (PreDiamox/PostDiamox) on both bias and precision. Statistical analyses were performed using Stata 16.1 (StataCorp LP, College Station, TX) and R 4.1.1 (r-project.org), with a significance level of 0.05.

5. Experiments and results

We evaluate the quantitative and qualitative performance of the proposed MRI-to-PET translation approach, demonstrating how synthetic images can support clinical decisions by improving the diagnosis and assessment of neurological conditions. In addition, ablation experiments are conducted to assess the effectiveness and contributions of different loss functions and the role of attention mechanisms. Lastly, we study the relative impact of different input MRI scans on the overall quality of the synthetic PET images.

5.1. Experimental setup

Fig. 1 illustrates the analysis workflow for model training, validation, and independent testing. The database of both cohorts was divided into two separate sets. The first set, which included simultaneous PET/MRI data from 60 HCs and 45 PTs, was used for model development and primary analysis. Fivefold cross-validation was used to evaluate the model's performance on different portions of the data, with three folds used for training, one for validation, and one for testing. To avoid data leakage, all data from the same subject (both baseline and post-acetazolamide) were included in any of the training, validation, and testing sets. The MRI scans selected to be used as inputs

Table 2
PET synthesis results for healthy controls and cerebrovascular disease patients. The model is evaluated using five-fold cross-validation, and the quantitative metrics are computed for the whole brain region. ↑/↓ denotes that higher/lower values correspond to better quality of synthetic PET. Results are presented as mean ± standard deviation (std).

Participants	NRMSE ↓ mean ± std	PSNR (dB) ↑ mean ± std	SSIM ↑ mean ± std
Controls	0.024 ± 0.010	40.16 ± 1.09	0.940 ± 0.008
Patients	0.062 ± 0.015	37.32 ± 1.17	0.912 ± 0.012
Average	0.044 ± 0.015	38.80 ± 1.18	0.924 ± 0.014

for the model are T1-w, T2-FLAIR, SD-ASL, MD-ASL, PD, ATT, SD-CBF, and MD-CBF.

To assess the model's ability to reproduce results in practice, the second set of data, referred to as the "Generalization Set", was used to evaluate the model's performance on unseen data from 21 participants (10 HCs and 11 PTs). A set of agreement quantification methods (*i.e.*, Bland-Altman, joint intensity scatter plots) was used to study the relationship between the regional CBF values of actual and synthetic PET scans, and to determine if synthetic CBF maps can be used to detect regional CBF abnormalities.

5.2. Quantitative results

The quantitative performance of our model was evaluated using NRMSE, PSNR, and SSIM. Theoretically, lower NRMSE values and higher PSNR and SSIM values indicate better quality of the synthetic images. The average quantitative results for both healthy controls and cerebrovascular disease patients among the test sets were computed based on the whole three-dimensional brain region and reported in Table 2.

Results from the PET synthesis experiments indicated a considerable difference between healthy controls and patients in terms of all performance metrics, with the former performing better. This could be attributed to the structural abnormalities associated with vascular diseases, which have a different appearance in different MRI contrasts. Another possibility could be differences in baseline image quality between the two groups, though there were no obvious differences noted visually. The average results, shown in Table 2, demonstrated that our optimized encoder–decoder network can efficiently integrate multiple MRI exams and produce high-quality synthetic PET images.

In order to assess the efficacy of the proposed encoder–decoder network ("ours"), we compare its performance with other state-of-the-art image synthesis networks on the same dataset, including 2D U-Net, 2D c-GAN, 2D SC-GAN, 3D U-Net, 3D c-GAN, and 3D SC-GAN. The 2D U-Net was implemented with the same network parameters as in Guo et al. (2020). The 3D U-Net was implemented similarly with 3D convolutional filters. The 2D c-GAN and 3D c-GAN were implemented according to Wang et al. (2018). The 2D SC-GAN and 3D SC-GAN were developed based on the method described in Lan et al. (2021). To ensure a fair comparison, we used the same training and validation sets for both the baseline methods and our proposed method. The quantitative evaluation of these networks as well as our encoder–decoder network is presented in Table 3. The results reveal that 2D c-GAN and 2D SC-GAN achieve on-par or slightly better performance than the 2D U-Net with minimal differences in the quality assessment metrics. Moreover, 3D U-Net, 3D c-GAN, and 3D SC-GAN produce comparable PET synthesis results, suggesting that simple encoder–decoder networks may be more practical than unstable GANs for this particular medical image generation problem.

Furthermore, the quantitative measures presented in Table 3 illustrate that the 3D implementations of U-Net and GANs outperform their 2D counterparts, attributed to the heightened level of spatial information present in 3D images. Our attention-based encoder–decoder
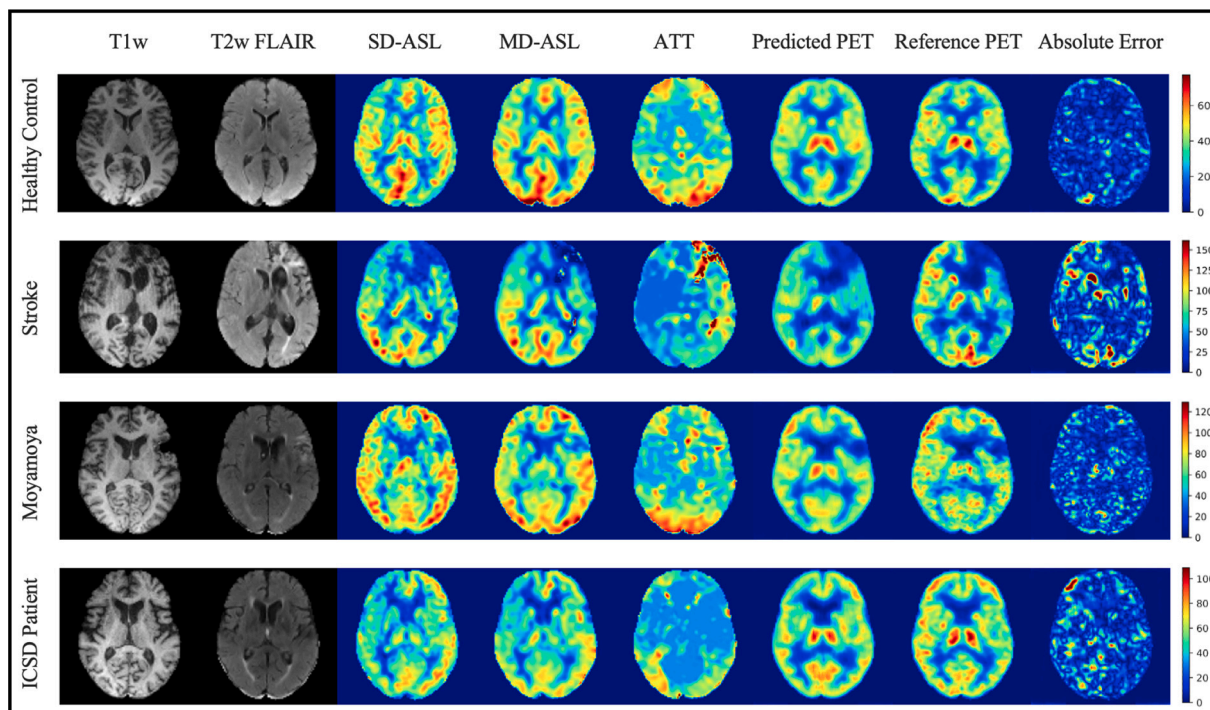
Fig. 4. MRI-to-PET prediction results for healthy control and cerebrovascular disease patients in the axial plane: Examples of input multi-contrast MRI scans, output synthetic PET, reference PET, and corresponding magnified (×3) absolute error maps. PET CBF is quantified in milliliters per 100 g of brain tissue per minute (ml/100 g/min).

Table 3
Quantitative comparison between our model and baseline models.

Method	NRMSE ↓ mean ± std	PSNR (dB) ↑ mean ± std	SSIM ↑ mean ± std
U-Net (2D)	0.204 ± 0.034	30.45 ± 1.92	0.862 ± 0.030
c-GAN (2D)	0.205 ± 0.032	31.70 ± 2.04	0.865 ± 0.028
SC-GAN (2D)	0.198 ± 0.032	32.05 ± 2.02	0.865 ± 0.033
U-Net (3D)	0.168 ± 0.024	33.86 ± 1.35	0.880 ± 0.015
c-GAN (3D)	0.168 ± 0.026	34.00 ± 1.42	0.878 ± 0.016
SC-GAN (3D)	0.150 ± 0.21	35.04 ± 1.39	0.888 ± 0.022
Ours	**0.044 ± 0.015**	**38.80 ± 1.18**	**0.924 ± 0.014**

network, incorporating attention mechanisms and a tailored loss function, exhibits superior performance compared to other methods across all performance metrics, as evidenced by our visual assessment (Fig. 7). The performance of our PET CBF synthesis approach highlights its capacity to harness the structural and contextual information inherent in multi-contrast MRI data, thereby enhancing the quality of the inferenced PET images.

5.3. Qualitative results

We demonstrate the significance of combining structural and perceptual information in multi-contrast MRI data through optimizing a custom loss function and utilizing attention mechanisms. To illustrate our findings, we present a qualitative comparison between the synthetic and actual PET images for healthy controls and cerebrovascular disease patients in Fig. 4. The top panel demonstrates the model's successful performance on a normal subject without any abnormal brain lesions, with the generated PET image closely resembling the ground-truth PET image. The remaining three panels illustrate the model's performance on patients with ischemic stroke, Moyamoya disease, and intracranial atherosclerotic steno-occlusive disease (ICSD). The synthetic PET images and corresponding magnified absolute error maps show that the proposed network can accurately synthesize abnormal brain lesions in synthetic PET images.

Fig. 5 illustrates the qualitative visualizations of synthetic PET scans for the same representative subjects along coronal plane. The PET CBF results for the healthy control group demonstrate accurate predictions in the coronal view of the brain. However, due to the limited abnormal brain regions relative to the whole brain volume, the model tends to overfit to normal regions for cerebrovascular disease patients with impaired regional CBF. Consequently, the prediction performance for areas with altered CBF (*e.g.*, ischemic penumbra and infarcted tissue) is somewhat inferior to that in the normal brain areas. Overall, our PET synthesis approach can effectively improve the quality and clinical utility of the structural and perfusion MRI exams, producing high-quality synthetic ^{15}O-water PET CBF maps without the use of radioactive tracers.

5.4. Ablation study

Ablation studies are a valuable method for investigating knowledge representations in encoder–decoder networks and are especially helpful in examining network performance and reliability against structural artifacts and ghosting. We conducted experiments to evaluate the effectiveness and contributions of different network settings and training strategies. We experimented with several loss functions, including MSE, MAE, and SSIM, as well as a custom loss function with a weighted summation of different metrics, to optimize the quality of synthesized PET CBF maps. We further studied the importance of the attention mechanisms used by the network's decoder for PET CBF prediction solely based on a subset of the encoder's feature maps. To be precise, in this section, we not only removed parts of the network but also substituted them with more appropriate alternative constructs.

Fig. 6 illustrates the PET prediction performance of different loss functions and network settings, as well as the incremental performance gain of each component. The reference PET, synthetic PET, and corresponding magnified absolute error map produced at different network settings are displayed for both healthy subjects and patients with cerebrovascular diseases (Fig. 6(a)). It was observed that the use of more appropriate loss functions led to a steady improvement in
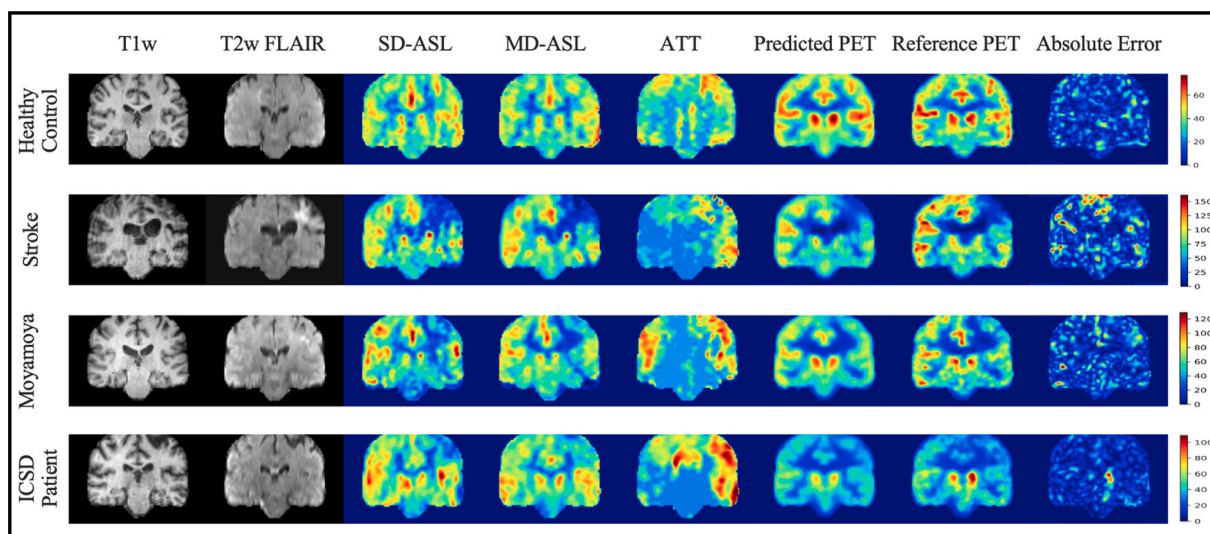
Fig. 5. MRI-to-PET prediction results for healthy control and cerebrovascular disease patients in the coronal plane: Examples of input multi-contrast MRI scans, output synthetic PET, reference PET, and corresponding magnified (×3) absolute error maps. PET CBF is quantified in ml/100 g/min.
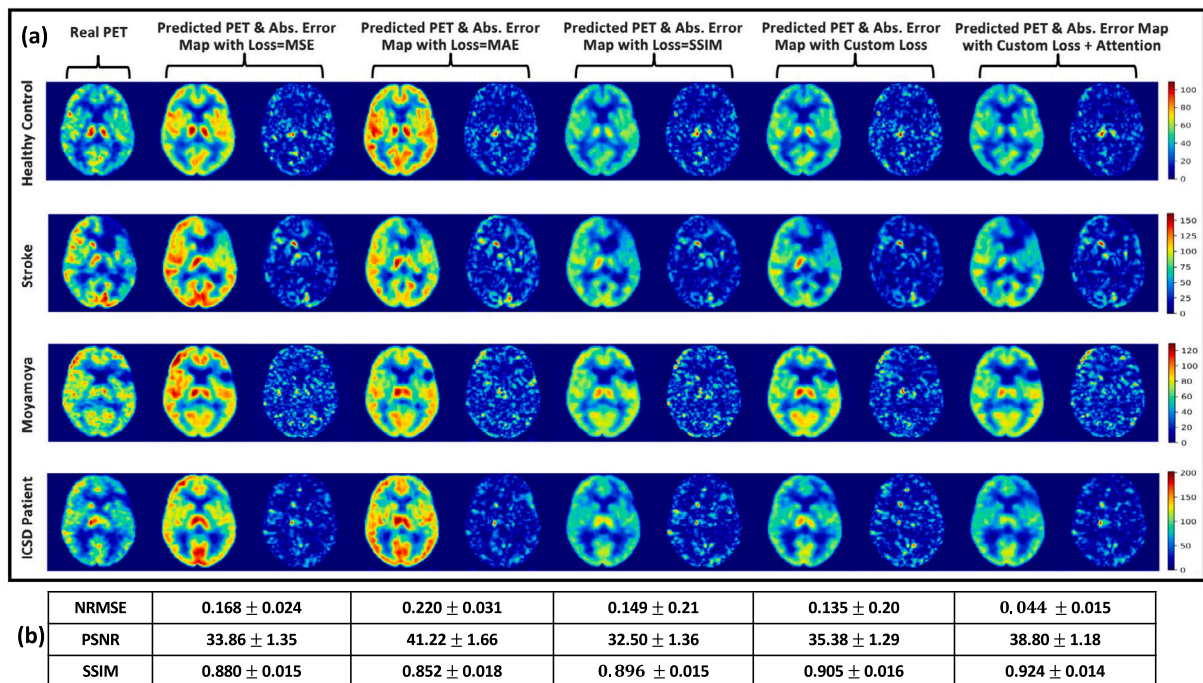

		Predicted PET & Abs. Error Map with Loss=MSE	Predicted PET & Abs. Error Map with Loss=MAE	Predicted PET & Abs. Error Map with Loss=SSIM	Predicted PET & Abs. Error Map with Custom Loss	Predicted PET & Abs. Error Map with Custom Loss + Attention
(b)	NRMSE	0.168 ± 0.024	0.220 ± 0.031	0.149 ± 0.21	0.135 ± 0.20	0.044 ± 0.015
	PSNR	33.86 ± 1.35	41.22 ± 1.66	32.50 ± 1.36	35.38 ± 1.29	38.80 ± 1.18
	SSIM	0.880 ± 0.015	0.852 ± 0.018	0.896 ± 0.015	0.905 ± 0.016	0.924 ± 0.014

Fig. 6. Examples of PET CBF prediction (in ml/100 g/min) for healthy controls and cerebrovascular disease patients at different loss functions and network settings: (a) example results of the reference PET CBF against synthetic PET CBF and magnified (×3) absolute error maps at different settings in the axial plane, (b) quantitative comparison between different loss functions and network elements.

both quantitative and qualitative CBF prediction results (Fig. 6(b)). Furthermore, attention mechanisms were found to be essential in allowing the network to focus more on the relevant aspects of the input data at the channel and spatial levels, thus resulting in improved synthetic PET CBF quality. In conclusion, a 3D convolutional encoder–decoder network utilizing attention mechanisms and a custom-designed loss function was able to effectively exploit both channel-level and spatial-level information from multi-contrast MRI inputs to generate high-quality synthetic PET CBF maps.

5.5. CBF quantification assessment

The statistical significance of experimental results was evaluated using a set of paired comparison analyses. The generalization set was used to examine the levels of agreement and correlation between regional CBF values in true ^{15}O-water PET CBF measurements and those of synthetic PET CBF, SD-CBF, and MD-CBF maps. The generalization set included 60 simultaneous PET/MRI scans acquired from 10 healthy controls and 11 cerebrovascular disease patients. The 10 control cases had a total of 32 PET/MRI scans (18 pre-acetazolamide and 14 post-acetazolamide) and the 11 patients had a total of 28 PET/MRI scans (17 pre-acetazolamide and 11 post-acetazolamide). Overall, 600 vascular territories from the 60 PET/MRI observations were used to compare synthesized PET CBF obtained by an encoder–decoder network and MRI-derived CBF measurements.

Fig. 8 displays the Bland-Altman plots of regional CBF values for healthy control and cerebrovascular disease patients at PreDiamox and
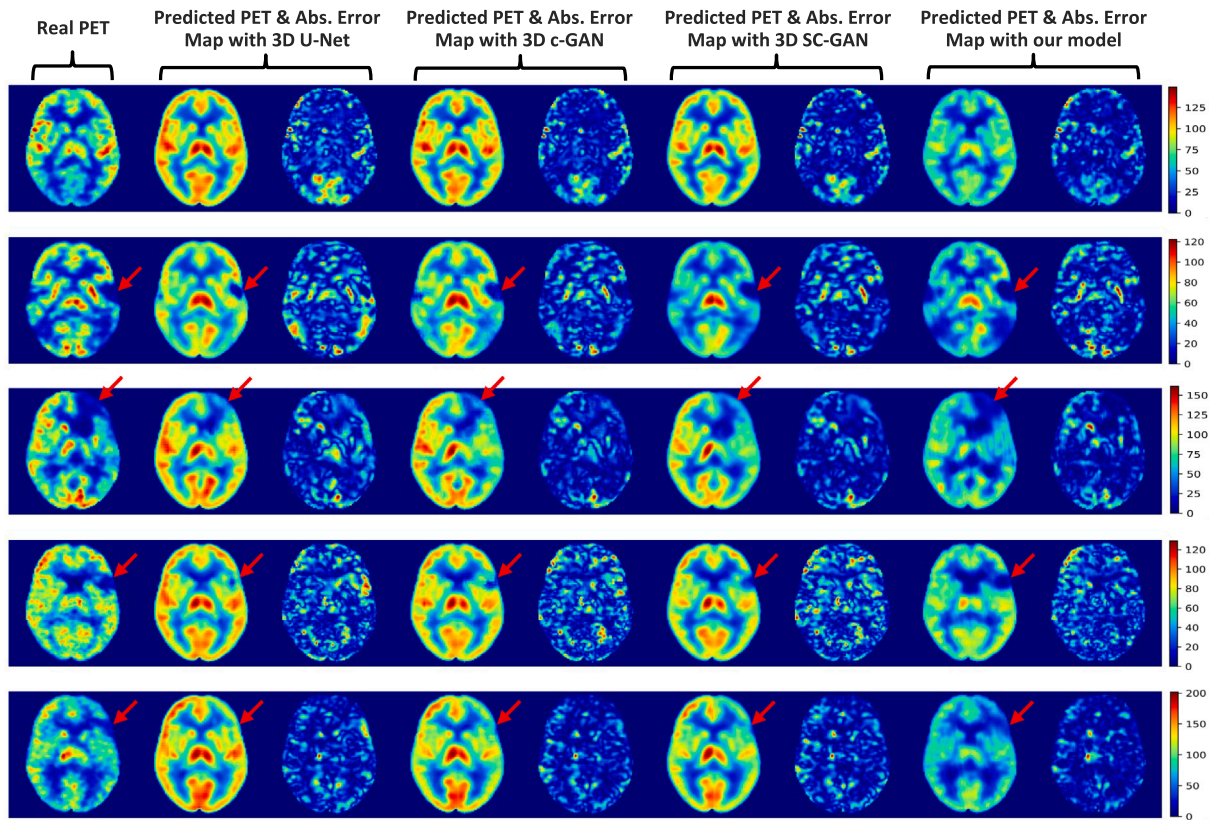
Fig. 7. Qualitative comparisons of different deep learning models in synthesizing PET CBF maps from multi-contrast MR images. CBFs are quantified in ml/100 g/min. The proposed method produces more accurate predictions than the standard 3D U-Net, 3D c-GAN, and 3D SC-GAN, particularly for those with abnormal lesions.
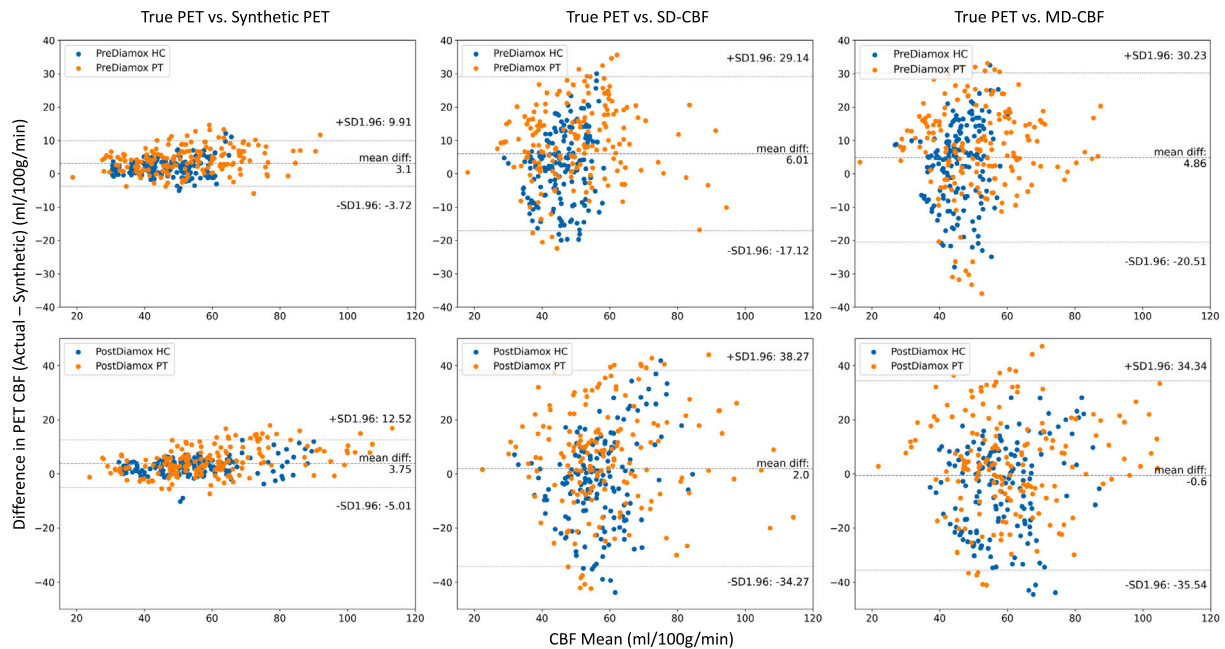

Fig. 8. Bland-Altman plots of the mean CBF in the ASPECTS vascular territories for PreDiamox (top panel) and PostDiamox (bottom panel) measurements. Each panel includes three plots showing the agreement between the reference PET CBF (True PET) and (i) the PET CBF produced by our model (Synthetic PET, left), (ii) CBF derived from single-delay ASL (SD-CBF, middle), and (iii) CBF derived from multi-delay ASL (MD-CBF, right).

PostDiamox measurements. The plots indicate the level of agreement between the reference PET CBF maps and each of the synthetic PET CBF, SD-CBF, and MD-CBF maps. At PreDiamox measurements, the regional CBF values of the synthetic PET had lower bias and variance than the ASL-derived CBF maps, demonstrating the added value of the

trained network. The synthetic PET CBF maps were only 3.1 ml/100 g/min lower than the true PET CBF maps with 95% confidence intervals of −3.7 & 9.9 ml/100 g/min. For PostDiamox measurements, the synthetic PET CBF maps had an average bias of 3.7 ml/100 g/min
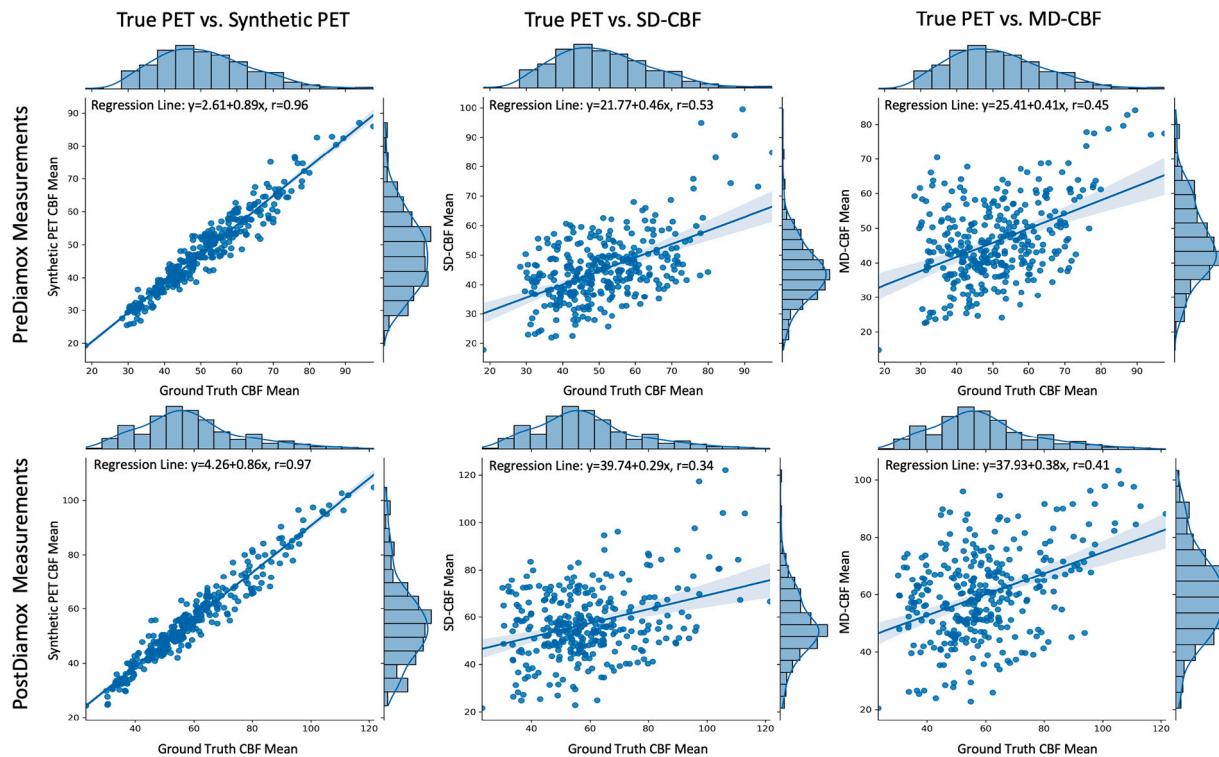
Fig. 9. Joint plots of mean CBF in ASPECTS vascular territories for PreDiamox (top panel) and PostDiamox (bottom panel) measurements: Each panel displays three plots for the relationship and distribution histogram between True PET and Synthetic PET (left), SD-CBF (middle), and MD-CBF (right). The regression line and Pearson correlation coefficient (r) are added to each of the joint plots.

Table 4

PET synthesis error (true PET CBF − synthetic PET CBF): Bias and precision in synthetic PET measurements for healthy controls and patients at different scan times. Mean and SD represent the bias and variability in the measurements, respectively. No. refers to the number of PET/MRI observations.

Group	Controls			Patients			Average		
Scan time	PreDiamox	PostDiamox	Total	PreDiamox	PostDiamox	Total	PreDiamox	PostDiamox	Total
Mean	1.82	2.62	2.22	4.25	4.88	4.56	3.10	3.75	3.42
SD	2.56	3.20	2.90	3.68	5.25	4.56	3.47	4.48	4.04
No.	180	140	320	170	110	280	350	250	600

with 95% confidence intervals of −5.0 & 12.5 ml/100 g/min, indicating better synthesis results than ASL-derived CBF maps.

Results of the PET CBF quantification performance for both healthy control and patient groups are outlined in Appendix C. Fig. C.1 indicates that the PET synthesis performance in patients was inferior to that in healthy control participants, with an average bias of 4.6 ml/100 g/min and 95% confidence intervals of −4.4 to 13.5 ml/100 g/min, in comparison to an average bias of 2.2 ml/100 g/min and 95% confidence intervals of −3.5 to 7.9 ml/100 g/min. The Bland-Altman plots in Fig. C.1 also demonstrate that the mean CBF was not significantly different among the CBF measurement types; however, the SD-CBF and MD-CBF measurements had a greater variability than the synthetic PET CBF measurements.

Table 4 presents the PET synthesis error for healthy controls and patients before and after acetazolamide administration. The results indicate that both bias and precision in synthetic PET measurements were noticeably different between groups and timepoints (all four marginal comparisons with p<0.001). Healthy controls exhibited significantly lower bias and precision than patients for both baseline and post-acetazolamide timepoints. Additionally, the PET synthesis performance was moderately better at baseline conditions.

The correlation between regional CBF measurements in acquired ^{15}O-water PET and those of synthetic PET and ASL-derived CBF maps

was also investigated in this study. Fig. 9 describes the density and joint scatter plots of regional CBF values in pre-acetazolamide (top panel) and post-acetazolamide (bottom panel) measurements. Each panel in Fig. 9 displays three plots for the relationship and distribution histogram between true ^{15}O-water PET CBF and synthetic PET CBF (left), SD-CBF (middle), and MD-CBF (right). The regression line and Pearson's correlation coefficient (r) are included in each of the joint plots. The results indicate a high positive correlation between true and synthetic PET's regional CBF values, with Pearson's correlation coefficient of 0.96 and 0.97 for the pre-acetazolamide and post-acetazolamide scans, respectively. The ASL-derived CBF maps, on the other hand, showed a moderate positive correlation (r = 0.45–0.53) at baseline and a weak positive correlation (0.34–41) for post-acetazolamide measurements.

5.6. Clinical significance – abnormal region identification

The diagnostic performance of synthetic PET CBF and ASL-derived CBF maps was evaluated by testing three threshold CBF values, which were 2, 3, and 4 standard deviations below the mean CBF values in healthy control participants. Fig. 10 demonstrates the classification performance of synthetic PET CBF, SD-CBF, and MD-CBF to identify reduced CBF regions in pre-acetazolamide and post-acetazolamide CBF
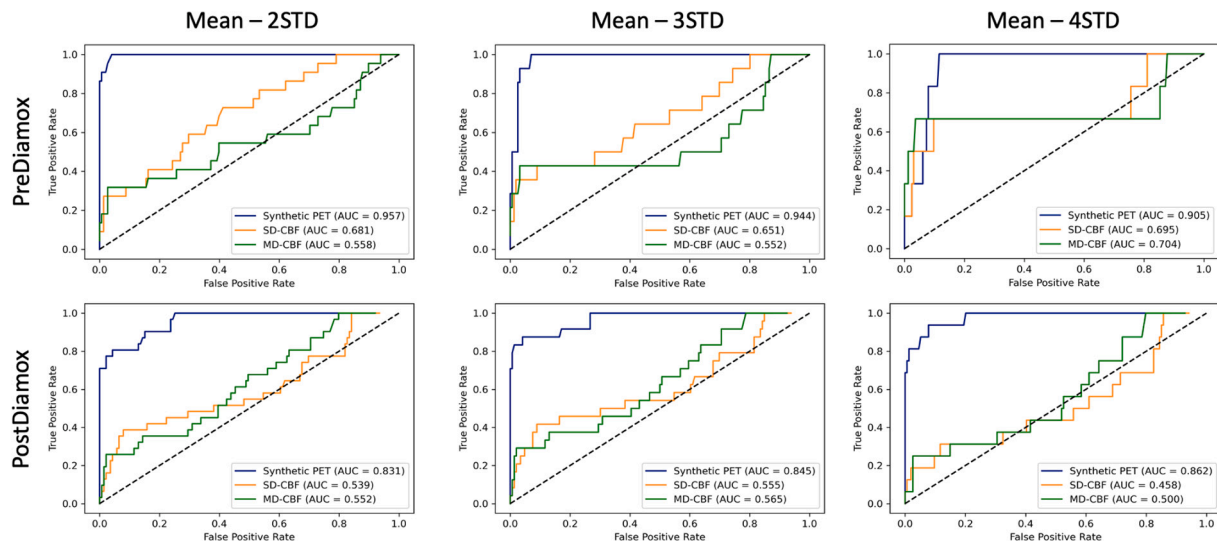
Fig. 10. ROC curves and AUC scores for identifying vascular territories with reduced CBF in Prediamox (top panel) and PostDiamox (bottom panel) measurements for cerebrovascular patients: Each panel includes three plots showing the classification performance at three different threshold values, (i) Threshold at 2 standard deviation (STD) below the mean CBF of healthy control participants (left), (ii) Threshold at 3 STD below mean CBF (middle), and (iii) Threshold at 4 STD below mean CBF (right). Each plot includes three ROC curves showing the classification performance of Synthetic PET (blue curve), SD-CBF (red curve), and MD-CBF (green curve).

measurements. The plots show the ROC curves and AUC scores to differentiate between the vascular territories with abnormally low CBF and those with normal CBF. The synthetic PET CBF maps generated by our model demonstrate superior performance to ASL-derived CBF maps at all tested thresholds before and after acetazolamide administration.

The threshold values for impaired PET CBF were defined as three standard deviations below the mean PET CBF values in the healthy control participants. The AUC scores for synthetic PET CBF, SD-CBF, and MD-CBF in pre-acetazolamide measurements were 0.94, 0.65, and 0.55 respectively, with a threshold CBF of three standard deviations below the mean in healthy control participants. Similarly, in post-acetazolamide measurements, the AUC scores for synthetic PET CBF, SD-CBF, and MD-CBF were 0.84, 0.55, and 0.56 for the same threshold CBF. Irrespective of the threshold of impaired PET CBF, the AUC score of synthetic PET CBF was higher than that of SD-CBF and MD-CBF. This pattern was observed for milder or more severe CBF thresholds as well, with synthetic PET CBF consistently outperforming ASL-derived CBF maps.

Fig. 11 also shows radar charts of the classification performance measures for SD-CBF, MD-CBF, and synthetic PET CBF at the threshold of three standard deviations below the mean CBF. The metrics of classification accuracy, sensitivity, specificity, PPV, and NPV were used to evaluate the detection performance of abnormal brain regions in pre-acetazolamide and post-acetazolamide CBF measurements. These results demonstrate the diagnostic value of synthetic PET and how the proposed model improved the clinical utility of MRI-derived CBF measurements at both baseline and after vasodilator administration. Our PET synthesis model offers a great promise for medical diagnostics, showing accurate identification of regional CBF abnormalities in patients with cerebrovascular diseases.

5.7. PET synthesis from perfusion MRI without structural information

In this section, we investigate the PET synthesis performance of the proposed encoder–decoder network when provided with ASL MRI images only as inputs. Yousefi et al. (2021) have previously studied the ASL-to-PET translation problem, where residual CNN was used to generate PET data from 2D ASL and T1w images. Yousefi et al. conducted a seven-fold cross-validation to evaluate the performance of their PET synthesis network on a dataset consisting of healthy control

participants, obtaining an average SSIM of 0.85 ± 0.08 and PSNR of 21.8 ± 4.5 dB. We implemented and tested the same CNN network on our dataset, which included both healthy controls and patients, and found that it produced lower prediction results of 0.82 ± 0.04 SSIM and 23.6 ± 3.1 dB PSNR. Our method utilized single-delay and multi-delay ASL images as the sole input to the encoder–decoder network, eliminating the need for anatomical information from any of the structural MRI scans. Our experiments yielded improved quantitative results, with an average SSIM and PSNR of 0.86 ± 0.03 and 30.4 ± 2.3 dB, respectively.

Fig. 12 shows examples of ASL-to-PET prediction for both healthy controls and cerebrovascular disease patients in axial and coronal planes. It can be observed that our model produced adequate PET CBF maps for healthy controls, in which the magnified absolute error maps show an insignificant difference between the true and synthetic PET CBF maps. However, an inferior PET synthesis performance was seen in patients, showing overestimation for the brain regions with reduced CBF. In normal brain territories, the CBF was either underestimated or overestimated, showing a non-trivial discrepancy between true and synthetic PET images. The absence of anatomical structure (from T1w or T2-FLAIR) is probably the reason behind the performance deterioration. Further, the Bland-Altman plots of Fig. 13 illustrate the agreement between the mean CBF in true and synthetic PET CBF maps for both healthy controls and cerebrovascular disease patients in PreDiamox and PostDiamox measurements when only ASL data is used as an input to our network. It is evident that the lack of anatomical data results in higher variability in the predicted CBF values at both PreDiamox and PostDiamox measurements.

5.8. PET synthesis from structural MRI without ASL imaging

Lastly, we evaluate the feasibility of synthesizing PET CBF maps from structural MRI including T1w and T2-FLAIR exams, but excluding data from the perfusion imaging sequences. Fig. 12 shows examples of structural MRI-to-PET translation results for healthy controls and cerebrovascular disease patients. Our model yields a reasonable similarity between true and synthetic PET CBF maps, with the whole-brain region being slightly overestimated. This is expected, as gray and white matter perfusion in normal subjects is known to differ in a reproducible way.

On the other hand, a serious performance degradation was observed in PET image generation for cerebrovascular disease patients. Both
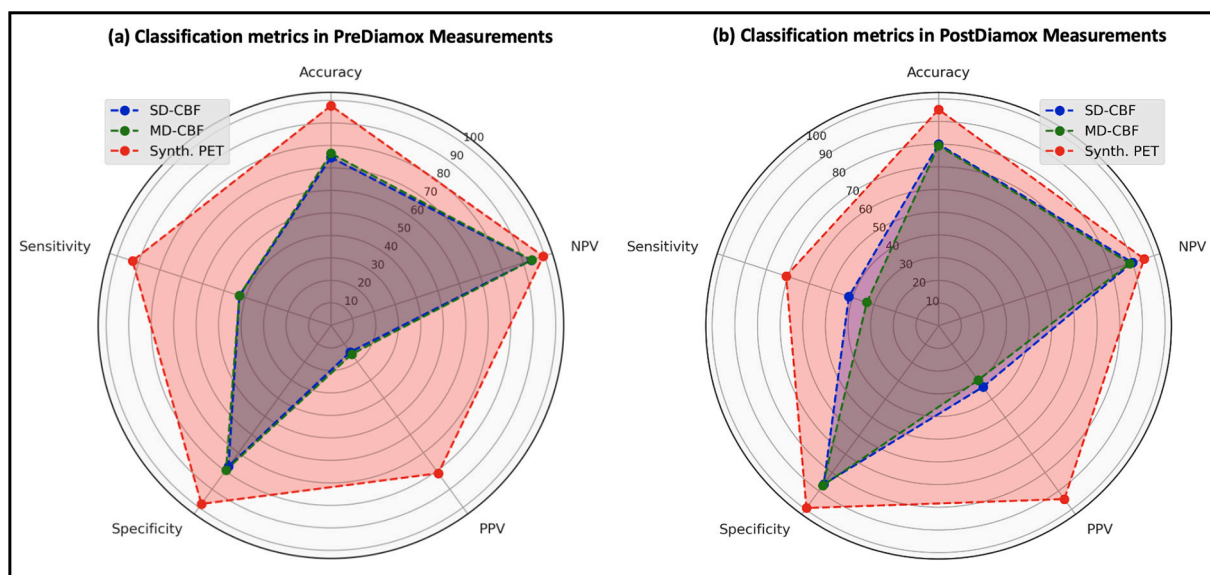
Fig. 11. Radar charts of classification performance measures for SD-CBF, MD-CBF, and synthetic PET CBF at Threshold = Mean − 3 STD: (a) and (b) Evaluation metrics for detecting abnormal regions (*i.e.*, regions with reduced CBF) in PreDiamox and PostDiamox measurements, respectively.
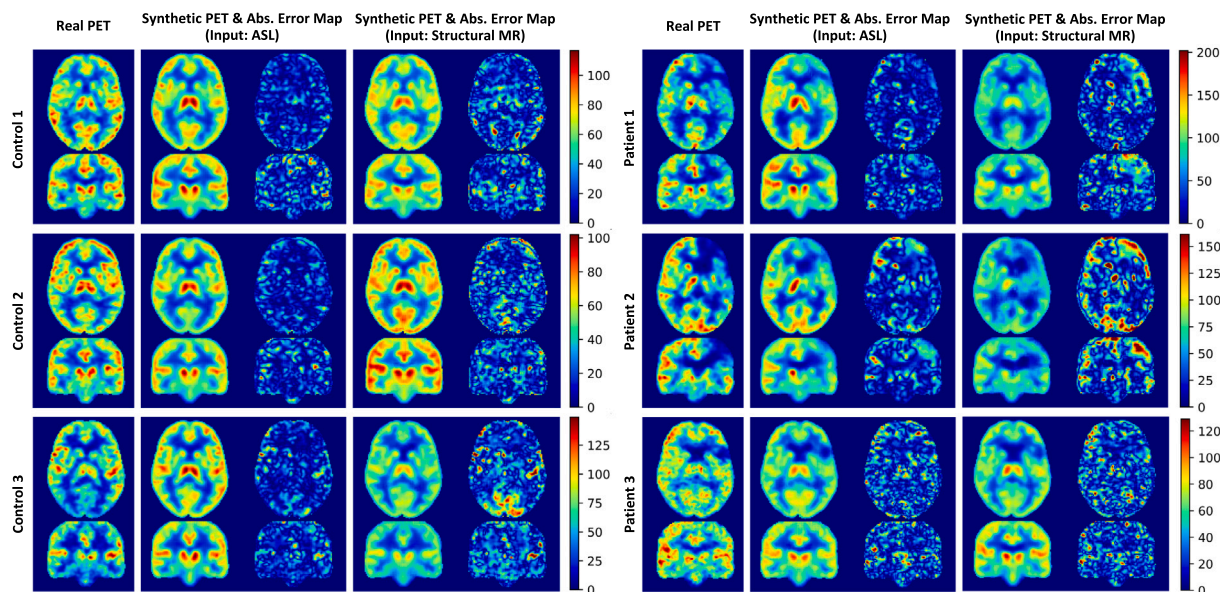

Fig. 12. MRI-to-PET CBF prediction using either ASL data or structural MRI. Each panel illustrates the data of a separate subject, with real axial and coronal images as well as synthetic images and magnified absolute error maps.

axial and coronal visualizations reveal the limited performance and the inability of this approach to produce acceptable CBF maps when provided with structural MRI as inputs. None of the brain regions with reduced CBF was properly predicted, showing clearly that ASL scans are crucial for PET image synthesis in patients with cerebrovascular disease. Additionally, quantitative analysis demonstrates limited PET synthesis performance, with an average SSIM and PSNR of 0.78 ± 0.08 and 20.1 ± 3.6, respectively. Fig. 14 also demonstrates the large degree of variability in synthesized PET CBF values for both the control and patient groups due to the lack of perfusion data.

Given the established importance of both structural MRI and ASL perfusion MRI contrasts, including SD-ASL and MD-ASL, for accurate PET CBF synthesis, we aim to investigate whether omitting either SD-ASL or MD-ASL from the acquisition process has a discernible impact on PET CBF synthesis performance. The rationale behind this investigation is to potentially reduce acquisition time and associated costs while maintaining the integrity of PET CBF synthesis. As shown in Appendix D, Fig. D.1, omitting either MD-ASL or SD-ASL from the PET CBF synthesis process significantly degrades the quality of the CBF maps compared to using both contrasts in conjunction with structural MRI data. These findings demonstrate that both ASL contrasts provide complementary information and are essential for generating high-quality CBF maps.
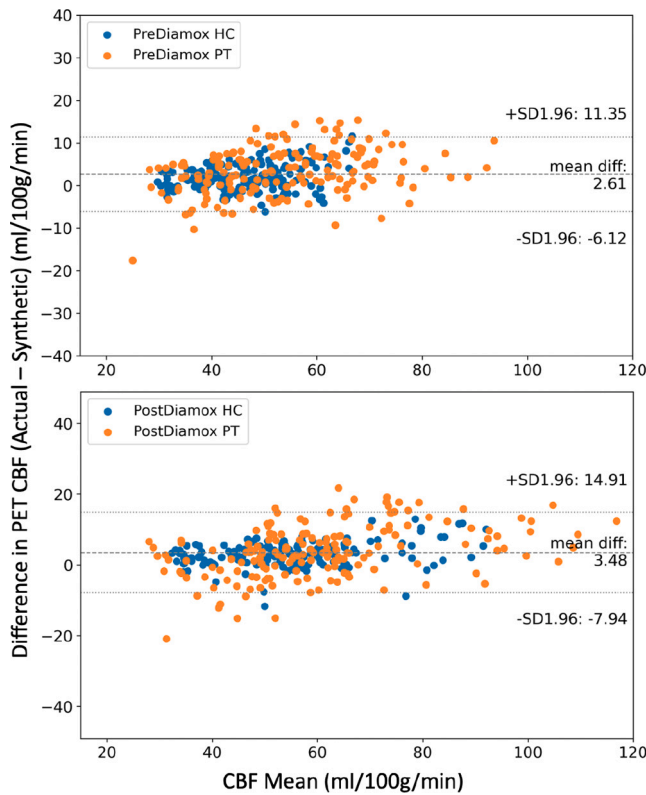
Fig. 13. Bland-Altman plots of the mean CBF in True PET and Synthetic PET for healthy controls and cerebrovascular disease patients at PreDiamox (top panel) and PostDiamox (bottom panel) measurements when ASL data is used as an input to our network.

Fig. 14. Bland-Altman plots of the mean CBF in True PET and Synthetic PET for healthy controls and cerebrovascular disease patients at PreDiamox (top panel) and PostDiamox (bottom panel) measurements when structural MR data is used as an input to our network.

6. Discussion

This study proposed and evaluated a 3D attention-based encoder–decoder network for brain MRI-to-PET translation. The network architecture effectively integrates structural MRI and ASL scans to capture both anatomical and perfusion features, thus improving the quality of synthesized PET scans. A custom loss function was developed to optimize the PET synthesis performance in both normal and abnormal brain regions. This loss function is a combination of multiple components that work cooperatively on driving the network toward the most representative distribution of actual PET images. A reconstruction loss based on the mean absolute error was used to ensure high voxel-wise similarity between real and synthetic PET images. A perceptual loss based on SSIM was also used to supplement the global loss and maximize the contextual and visual similarity between real and synthetic PET images. Attention mechanisms were also incorporated to capture long-range feature interactions and help the encoder–decoder network learn the underlying multimodal data distribution. Results demonstrate that 3D convolutional encoder–decoder networks with attention mechanisms and a well-designed loss function can accurately synthesize PET CBF maps from multi-contrast MRI images without the use of radioactive tracers.

The potential of a network that can take a widely available modality, such as MRI, and predict results that are only available at specialized centers, such as ^{15}O-water PET, is immense. This could enable gold-standard CBF measurements in sites without access to short half-life PET agents, thus opening up whole new avenues of research. Furthermore, it would democratize PET by allowing imaging in economically challenged areas that lack the expensive infrastructure required to support a PET scanner. This would enable more accurate studies of CBF in a wider range of patients and disease classes, rather than simply limiting them to those from urban areas with chronic conditions.
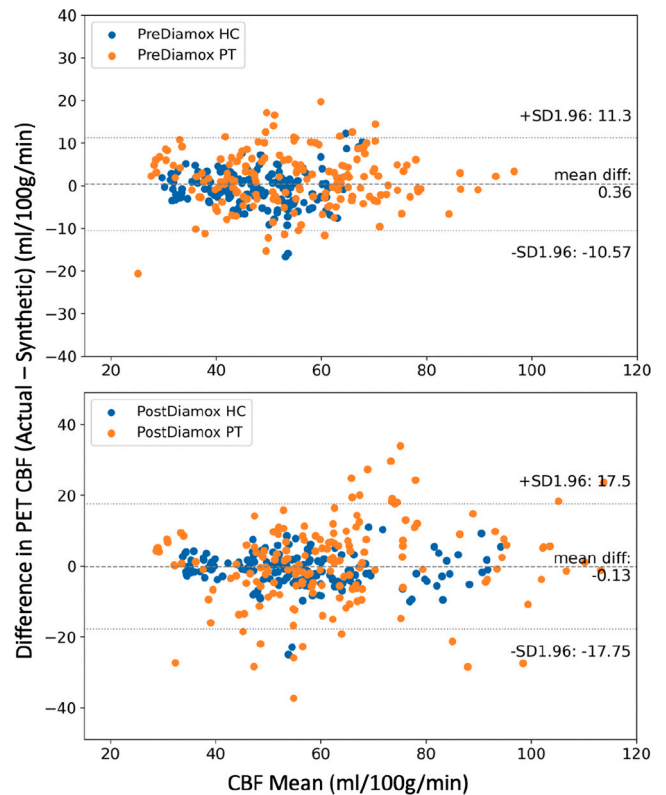
Several ablation studies were conducted to illustrate the impact of different loss functions, network elements, and input MRI contrasts on the quality of generated PET images. Experimental results revealed that both anatomical and functional information from structural and perfusion MRI exams are essential for synthesizing realistic and high-quality PET scans, particularly for patients with cerebrovascular diseases. Single-delay and multi-delay ASL scans had the most significant effect on the accuracy of PET synthesis. Further, pairwise comparison methods such as Bland-Altman analyses and density scatter plots demonstrated a high level of agreement and correlation between regional CBF values in actual and synthetic PET images. In comparison to ASL-derived CBF measurements, the synthetic PET CBF maps exhibited comparable bias, significantly better precision, and a notably higher positive correlation with true ^{15}O-water PET CBF measurements.

The proposed work has several potential clinical applications. We report results on a task of clinical importance, namely discriminating between vascular territories with and without CBF abnormalities. To do this, mean CBF values were computed for 10 brain regions in healthy control participants before and after administration of acetazolamide, a short-acting vasodilator. Subsequently, different threshold CBF values based on mean CBF and its variability in healthy controls were used to identify abnormal regions with low CBF in cerebrovascular disease patients. The improved performance of the network over MRI-only imaging indicates that the network is effective not only at characterizing the overall pattern of CBF, but also accurately the severity and location of abnormal regions, which may only occupy a small fraction of the overall image volume. This information is not always captured in summary statistics often used for quantitative assessment, such as PSNR, NRMSE, and SSIM.

Data curation was one of the major limitations in this study. The co-registration of input multi-contrast brain MRI images and quantification of ^{15}O-water PET CBF maps are laborious and time-consuming procedures. Subjective evaluations are also needed to ensure acceptable image and associated information quality. To address this, future work will investigate the deployment of automated deep learning algorithms that can process neuroimages in the native space. Also, the generalizability of the model will be examined using multi-center data acquired from different populations at different sites and scanners, as well as with different underlying diseases.

7. Conclusion

PET imaging of CBF is a critical component in the diagnosis and assessment of cerebrovascular diseases. However, its use is limited because of its prohibitive cost and the use of ionizing radiation. This study introduces an attention-based convolutional encoder–decoder network for synthesizing ^{15}O-water PET CBF maps from multi-contrast MRI scans without using radioactive tracers. The performance of the proposed image-to-image translation network is examined for different network settings and input MRI sequence combinations. Quantitative evaluations show improved PET synthesis results compared to previous MRI-to-PET CBF prediction models. Additionally, qualitative results also reveal that regional CBF values in synthetic PET are in strong agreement with those of the ground-truth PET, with no statistically significant difference between them. In patients with cerebrovascular diseases, brain regions with abnormally low CBF were accurately identified in synthetic PET CBF maps. This technique has the potential to increase the accessibility of cerebrovascular disease assessment for underserved populations, underprivileged communities, and developing nations, without the need for expensive and radiation-emitting PET imaging.

CRediT authorship contribution statement

Ramy Hussein: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **David Shin:** Conceptualization, Formal analysis, Investigation, Resources, Software, Validation, Visualization. **Moss Y. Zhao:** Data curation, Formal analysis, Software, Validation. **Jia Guo:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation. **Guido Davidzon:** Investigation, Project administration, Supervision, Validation. **Gary Steinberg:** Methodology, Resources, Supervision. **Michael Moseley:** Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Visualization. **Greg Zaharchuk:** Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Validation, Visualization, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ramy Hussein reports financial support was provided by National Institutes of Health. Greg Zaharchuk reports financial support was provided by National Institutes of Health. Moss Zhao reports financial support was provided by American Heart Association.

Data availability
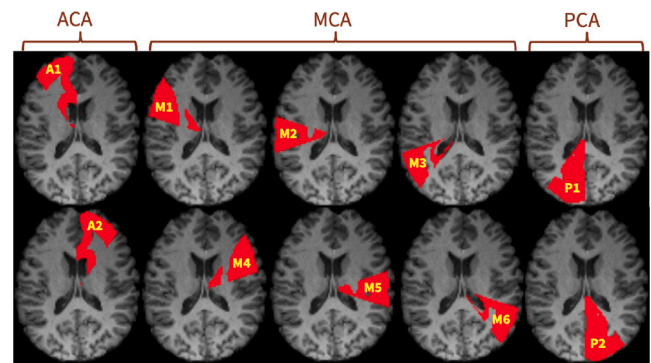
The authors do not have permission to share data.

Fig. B.1. Brain arterial vascular territories of ASPECTS.

Appendix A. MRI parameters

The list of parameters used for the structural MRI and ASL imaging are reported in Table A.1. MRA stands for the magnetic resonance angiography, AIF is the arterial input function, and GRE is the gradient echo sequence. SD-ASL and MD-ASL correspond to the single-delay and multi-delay ASL, respectively.

Appendix B. ASPECTS vascular territories

Fig. B.1 shows the 10 regions of interest (ROIs) of the ASPECTS mask. Each hemisphere has 5 ROIs: one anterior cerebral artery (ACA), three middle cerebral artery (MCA) and one posterior cerebral artery (PCA).

Appendix C. Bland-Altman analysis of the mean CBF in the ASPECTS regions for healthy control and cerebrovascular disease patients

In this appendix, the Bland-Altman plots of the mean cerebral blood flow in the ASPECTS vascular territories are presented for both healthy control and cerebrovascular disease patients. The plots in Fig. C.1 compare the agreement between the true PET CBF and each of (1) the synthetic PET CBF produced by our encoder–decoder network, (2) the CBF derived from single-delay ASL (named SD-CBF), and (3) the CBF derived from multi-delay ASL (named MD-CBF).

Appendix D. PET CBF synthesis performance in the absence of MD-ASL and SD-ASL techniques

Fig. D.1 demonstrates the PET synthesis performance under three conditions: without SD-ASL, without MD-ASL, and with both SD-ASL and MD-ASL. The findings reveal a complementary relationship between the two ASL contrasts, emphasizing their collective necessity alongside structural MRI data for the generation of high-quality CBF maps. This investigation provides valuable insights for refining imaging protocols and resource utilization in neuroimaging studies.

Table A.1
List of parameters used for MRI acquisition.

| Parameter	MRA - AIF	GRE	Parameter	SD-ASL	MD-ASL
TR/TE	22/2.4 ms	667/15 ms	Labeling pulse shape	Hanning	Hanning
No. of slices	120	30	Labeling pulse duration	0.5 ms	0.5 ms
Flip angle	15°	20°	Labeling pulse spacing	1.22 ms	1.22 ms
Slice thickness	1.2 mm	5 mm	RF pulse strength	0.014 Gauss	0.018 Gauss
Matrix	512×512	256×256	Mean gradient strength	0.7 mT/m	0.7 mT/m
FOV	220×220 mm^2	24×24 mm^2	Maximal gradient strength	7 mT/m	4.5 mT/m
Voxel size	0.43×0.43 mm^2	–	Bolus duration	1450 ms	2000 ms
Scan duration	4:03 min	1:56 min	TR/TE	4854/10.7 ms	6691/10.7 ms
Parameter	**T1w**	**T2-FLAIR**	PLD	2025 ms	700, 1325, 1950, 2575, 3200 ms
TR/TE	9600/3800 ms	9500/140 ms	No. of slices	36	36
No. of slices	–	30	FOV	24 cm^3	24 cm^3
Matrix	256×256	512×512	Acquisition voxelsize	$3.73 \times 3.73 \times 4$ mm^3	$5.77 \times 5.77 \times 4$ mm^3
FOV	180×180 mm^2	–	Reconstruction voxelsize	$1.875 \times 1.875 \times 4$ mm^3	$1.875 \times 1.875 \times 4$ mm^3
Resolution	$1 \times 1 \times 1$ mm^3	$0.47 \times 0.47 \times 5$ mm^3	Readout planing	Axial	Axial
Scan duration	3:22 min	3:33 min	Scan duration	4:13 min	4:39 min
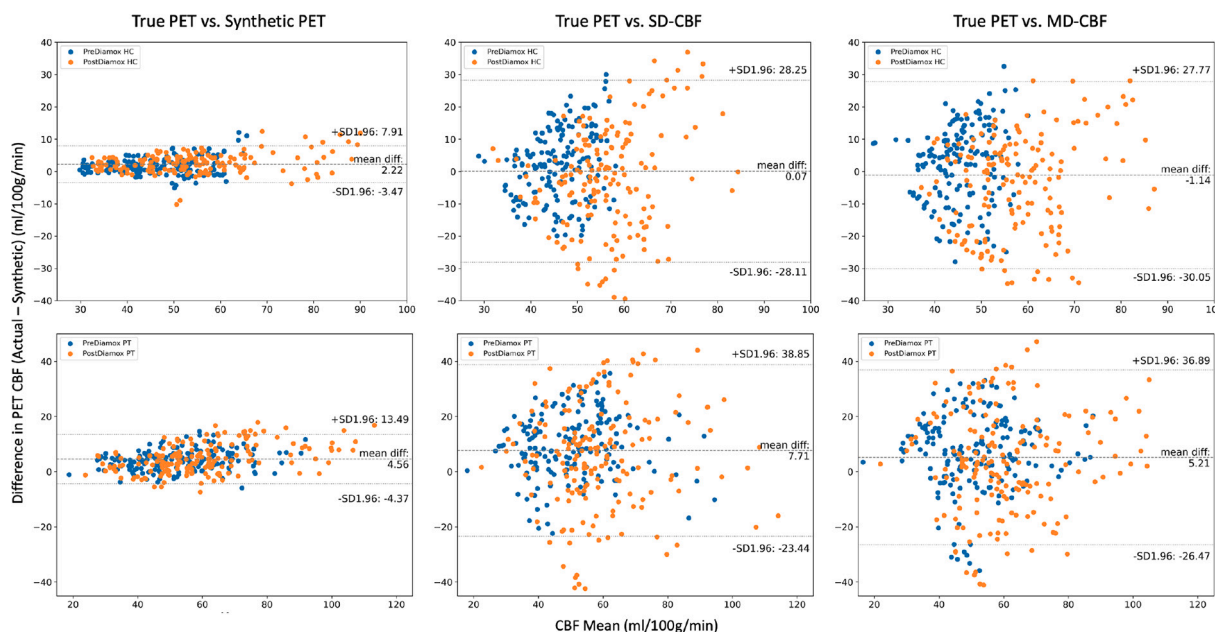
Fig. C.1. Bland-Altman plots of the mean CBF in the ASPECTS vascular territories for healthy controls (HC, top panel) and cerebrovascular disease patients (PT, bottom panel). Each panel includes three plots showing the agreement between the True PET CBF and (i) the Synthetic PET (left), (ii) SD-CBF (middle), and (iii) MD-CBF (right).
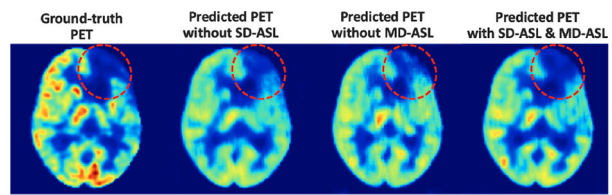

Fig. D.1. Comparison of PET CBF synthesis performance: Impact of MD-ASL and SD-ASL contrasts.

References

Alsop, D.C., Detre, J.A., Golay, X., Günther, M., Hendrikse, J., Hernandez-Garcia, L., Lu, H., MacIntosh, B.J., Parkes, L.M., Smits, M., et al., 2015. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. Magn. Reson. Med. 73 (1), 102–116.

Armanious, K., Jiang, C., Abdulatif, S., Küstner, T., Gatidis, S., Yang, B., 2019. Unsupervised medical image translation using cycle-medGAN. In: 2019 27th European Signal Processing Conference (EUSIPCO). IEEE, pp. 1–5.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684.

Barber, P.A., Demchuk, A.M., Zhang, J., Buchan, A.M., Group, A.S., et al., 2000. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. Lancet 355 (9216), 1670–1674.

Bazangani, F., Richard, F.J., Ghattas, B., Guedj, E., 2022. FDG-PET to T1 weighted MRI translation with 3D elicit generative adversarial network (E-GAN). Sensors 22 (12), 4640.

Ben-Cohen, A., Klang, E., Raskin, S.P., Soffer, S., Ben-Haim, S., Konen, E., Amitai, M.M., Greenspan, H., 2019. Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. Eng. Appl. Artif. Intell. 78, 186–194.

Buxton, R.B., Frank, L.R., Wong, E.C., Siewert, B., Warach, S., Edelman, R.R., 1998. A general kinetic model for quantitative perfusion imaging with arterial spin labeling. Magn. Reson. Med. 40 (3), 383–396.

Chen, K.T., Gong, E., de Carvalho Macruz, F.B., Xu, J., Boumis, A., Khalighi, M., Poston, K.L., Sha, S.J., Greicius, M.D., Mormino, E., et al., 2019. Ultra–low-dose 18f-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology 290 (3), 649.

Chen, D.Y., Ishii, Y., Fan, A.P., Guo, J., Zhao, M.Y., Steinberg, G.K., Zaharchuk, G., 2020. Predicting PET cerebrovascular reserve with deep learning by using baseline MRI: a pilot investigation of a drug-free brain stress test. Radiology 296 (3), 627.

Chen, C., Raymond, C., Speier, B., Jin, X., Cloughesy, T.F., Enzmann, D., Ellingson, B.M., Arnold, C.W., 2021. Synthesizing MR image contrast enhancement using 3D high-resolution ConvNets. arXiv preprint arXiv:2104.01592.

Dai, W., Robson, P.M., Shankaranarayanan, A., Alsop, D.C., 2012. Reduced resolution transit delay prescan for quantitative continuous arterial spin labeling perfusion imaging. Magn. Reson. Med. 67 (5), 1252–1265.

Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Çukur, T., 2019. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 38 (10), 2375–2388.

Detre, J.A., Leigh, J.S., Williams, D.S., Koretsky, A.P., 1992. Perfusion imaging. Magn. Reson. Med. 23 (1), 37–45.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dozat, T., 2016. Incorporating nesterov momentum into adam.

Gao, X., Shi, F., Shen, D., Liu, M., 2021. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in alzheimers disease. IEEE J. Biomed. Health Inf..

Gong, E., Pauly, J.M., Wintermark, M., Zaharchuk, G., 2018. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J. Magn. Reson. Imaging 48 (2), 330–340.

Grade, M., Hernandez Tamames, J., Pizzini, F., Achten, E., Golay, X., Smits, M., 2015. A neuroradiologist's guide to arterial spin labeling MRI in clinical practice. Neuroradiology 57 (12), 1181–1202.

Guo, J., Gong, E., Fan, A.P., Goubran, M., Khalighi, M.M., Zaharchuk, G., 2020. Predicting 15o-water PET cerebral blood flow maps from multi-contrast MRI using a deep convolutional neural network with evaluation of training cohort bias. J. Cereb. Blood Flow Metab. 40 (11), 2240–2253.

Harvey, A.C., 1976. Estimating regression models with multiplicative heteroscedasticity. Econometrica 461–465.

Hu, S., Shen, Y., Wang, S., Lei, B., 2020. Brain MR to PET synthesis via bidirectional generative adversarial network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 698–707.

Ito, H., Kanno, I., Kato, C., Sasaki, T., Ishii, K., Ouchi, Y., Iida, A., Okazawa, H., Hayashida, K., Tsuyuguchi, N., et al., 2004. Database of normal human cerebral blood flow, cerebral blood volume, cerebral oxygen extraction fraction and cerebral metabolic rate of oxygen measured by positron emission tomography with 15o-labelled carbon dioxide or water, carbon monoxide and oxygen: a multicentre study in Japan. Eur. J. Nucl. Med. Mol. Imaging 31 (5), 635–643.

Iturria-Medina, Y., Sotero, R.C., Toussaint, P.J., Mateos-Pérez, J.M., Evans, A.C., 2016. Early role of vascular dysregulation on late-onset alzheimer's disease based on multifactorial data-driven analysis. Nat. Commun. 7 (1), 1–14.

Jin, C.-B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X., 2019. Deep CT to MR synthesis using paired and unpaired data. Sensors 19 (10), 2361.

Kearney, V., Ziemer, B.P., Perry, A., Wang, T., Chan, J.W., Ma, L., Morin, O., Yom, S.S., Solberg, T.D., 2020. Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks. Radiol. Artif. Intell. 2 (2).

Khalighi, M.M., Deller, T.W., Fan, A.P., Gulaka, P.K., Shen, B., Singh, P., Park, J.-H., Chin, F.T., Zaharchuk, G., 2018. Image-derived input function estimation on a TOF-enabled PET/MR for cerebral blood flow mapping. J. Cereb. Blood Flow Metab. 38 (1), 126–135.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lan, H., Initiative, A.D.N., Toga, A.W., Sepehrband, F., 2021. Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis. Magn. Reson. Med. 86 (3), 1718–1733.

Leijenaar, J.F., van Maurik, I.S., Kuijer, J.P., van der Flier, W.M., Scheltens, P., Barkhof, F., Prins, N.D., 2017. Lower cerebral blood flow in subjects with alzheimer's dementia, mild cognitive impairment, and subjective cognitive decline using two-dimensional phase-contrast magnetic resonance imaging. Alzheimer's Dementia: Diagn. Assess. Dis. Monit. 9, 76–83.

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S., 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 305–312.

Liu, J., Pasumarthi, S., Duffy, B., Gong, E., Zaharchuk, G., Datta, K., 2022. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. arXiv preprint arXiv:2204.13738.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al., 2001. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). Philos. Trans. R. Soc. Lond. Ser. B 356 (1412), 1293–1322.

Mukherjee, D., Patil, C.G., 2011. Epidemiology and the global burden of stroke. World Neurosurg. 76 (6), S85–S90.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention U-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D., 2018. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 455–463.

Preetha, C.J., Meredig, H., Brugnara, G., Mahmutoglu, M.A., Foltyn, M., Isensee, F., Kessler, T., Pflüger, I., Schell, M., Neuberger, U., et al., 2021. Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study. Lancet Digit. Health 3 (12), e784–e794.

Raskutti, G., Wainwright, M.J., Yu, B., 2014. Early stopping and non-parametric regression: an optimal data-dependent stopping rule. J. Mach. Learn. Res. 15 (1), 335–366.

Shin, H.-C., Ihsani, A., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., Initiative, A.D.N., 2020. GANbert: Generative adversarial networks with bidirectional encoder representations from transformers for MRI to PET synthesis. arXiv preprint arXiv:2008.04393.

Sikka, A., Peri, S.V., Bathula, D.R., 2018. MRI to FDG-PET: cross-modal synthesis using 3D U-net for multi-modal alzheimer's classification. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 80–89.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219.

Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., Kandel, B.M., van Strien, N., Stone, J.R., Gee, J.C., et al., 2014. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 99, 166–179.

Villringer, A., Rosen, B.R., Belliveau, J.W., Ackerman, J.L., Lauffer, R.B., Buxton, R.B., Chao, Y.-S., Wedeenand, V.J., Brady, T.J., 1988. Dynamic imaging with lanthanide chelates in normal brain: contrast due to magnetic susceptibility effects. Magn. Reson. Med. 6 (2), 164–174.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L., 2018. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage 174, 550–562.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., Colliot, O., 2019. Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis. Med. Image Anal. 58, 101546.

Wolterink, J.M., Dinkla, A.M., Savenije, M.H., Seevinck, P.R., van den Berg, C.A., Išgum, I., 2017. Deep MR to CT synthesis using unpaired data. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 14–23.

Xie, H., Lei, Y., Wang, T., Roper, J., Axente, M., Bradley, J.D., Liu, T., Yang, X., 2022. Magnetic resonance imaging contrast enhancement synthesis using cascade networks with local supervision. Med. Phys. 49 (5), 3278–3287.

Yaakub, S.N., McGinnity, C.J., Clough, J.R., Kerfoot, E., Girard, N., Guedj, E., Hammers, A., 2019. Pseudo-normal PET synthesis with generative adversarial networks for localising hypometabolism in epilepsies. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 42–51.

Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.I., Xu, Y., et al., 2020b. MRI cross-modality image-to-image translation. Sci. Rep. 10 (1), 1–18.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J.L., Xu, Z., 2020a. Unsupervised MR-to-CT synthesis using structure-constrained cycleGAN. IEEE Trans. Med. Imaging 39 (12), 4249–4261.

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35.

Yousefi, S., Sokooti, H., Teeuwisse, W.M., Heijtel, D.F., Nederveen, A.J., Staring, M., van Osch, M.J., 2021. ASL to PET translation by a semi-supervised residual-based attention-guided convolutional neural network. arXiv preprint arXiv:2103.05116.

Yusuf, S., Reddy, S., Ôunpuu, S., Anand, S., 2001. Global burden of cardiovascular diseases: Part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. Circulation 104 (23), 2855–2864.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 7354–7363.

Zhang, J., He, X., Qing, L., Gao, F., Wang, B., 2022. BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal alzheimer's disease diagnosis. Comput. Methods Programs Biomed. 217, 106676.

Zhao, M.Y., Fan, A.P., Chen, D.Y.-T., Sokolska, M.J., Guo, J., Ishii, Y., Shin, D.D., Khalighi, M.M., Holley, D., Halbert, K., et al., 2021. Cerebrovascular reactivity measurements using simultaneous 15o-water PET and ASL MRI: Impacts of arterial transit time, labeling efficiency, and hematocrit. NeuroImage 233, 117955.

Zhao, H., Gallo, O., Frosio, I., Kautz, J., 2016. Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging 3 (1), 47–57.

Zhou, Y., Huang, S.-C., Bergsneider, M., 2001. Linear ridge regression with spatial constraint for generation of parametric images in dynamic positron emission tomography studies. IEEE Trans. Nucl. Sci. 48 (1), 125–130.