



A 3D multi-scale CycleGAN framework for generating synthetic PETs from MRIs for Alzheimer's disease diagnosis

M. Khojaste-Sarakhsi^{a,b}, Seyedhamidreza Shahabi Haghighi^{a,*}, S.M.T. Fatemi Ghomi^a, Elena Marchiori^b

^a Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

^b Institute for Computing and Information Sciences, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:

Cycle GAN
Multi-scale GAN
3D image-to-image translation
Image synthesis
Alzheimer's Disease diagnosis

ABSTRACT

This paper proposes a novel framework for generating synthesized PET images from MRIs to fill in missing PETs and help with Alzheimer's disease (AD) diagnosis. This framework employs a 3D multi-scale image-to-image CycleGAN architecture for the end-to-end translation of MRI and PET domains together. A hybrid loss function is also proposed to enforce structural similarity while preserving voxel-wise similarity and avoiding blurry images. As shown by the quantitative and visual assessment of the synthesized PETs, this framework is superior to the state-of-the-art. Moreover, using these synthesized PETs helps improve the ternary classification of AD subjects (AD vs. MCI vs. NC). Specifically, assuming an extreme case where none of the subjects has a PET, feeding the classifier with MRIs and their corresponding synthetic PETs results in a more accurate diagnosis than feeding it with just available MRIs. Accordingly, the proposed framework can help improve AD diagnosis, which is the final goal of the current study. Ablation investigation of the proposed multi-scale framework as well as the proposed loss function, is also conducted to study their contribution to the quality of synthesized PETs. Furthermore, other factors, such as stopping criteria, the type of normalization layer, the activation function, and dropouts, are examined, concluding that the appropriate use of these factors can significantly improve the quality of synthesized PETs.

1. Introduction

Alzheimer's Disease (AD), an irreversible neurodegenerative disease, is the leading cause of dementia, accounting for 70% of cases worldwide [1]. The prevalence of dementia is predicted to become triple by 2050 [2,3]. So, diagnosing AD by investigating its biomarkers is of great interest. Among these biomarkers, the physiological changes traceable through different neuroimage modalities are frequently examined. Magnetic Resonance Imaging (MRI) and its variants, including Structural MRI (sMRI) and functional MRI (fMRI), positron emission tomography (PET), and computerized tomography (CT) are the most common modalities utilized in this regard. PET is a functional imaging technique that provides a three-dimensional (3D) image at the molecular and cellular level, obtained by injecting a radio-tracer and computing a digital image with a scanner. A PET scan can often detect the abnormal metabolism of the tracer before the disease shows up on other imaging modalities [4].

The detection of early AD poses significant challenges, particularly in patients with mild cognitive impairment (MCI), where brain tissue appears nearly identical to that of normal individuals. This similarity complicates diagnosis for clinicians, emphasizing the need for more accurate detection methods. Computer-aided diagnostic techniques can address this need, aiming to enhance early AD detection while reducing the burden on physicians. Traditionally, research in AD has relied on single-modal data, providing limited localized information on brain abnormalities. Compared to single-modal data, multi-modal data can provide more useful information for differentiation and greatly improve the accuracy of AD classification [5].

It has been indicated in the literature that utilizing the brain functional information provided by PET along with the structural information embedded in MRIs can improve AD diagnosis compared to solely using the MRIs [6–8]. However, PET is more expensive than MRI and often not available in many hospitals due to the high cost associated with multiple examinations, poorly equipped hospitals, and difficulties

* Corresponding author.

E-mail address: shahabi@aut.ac.ir (S.S. Haghighi).

<https://doi.org/10.1016/j.imavis.2024.105017>

Received 9 August 2023; Received in revised form 2 March 2024; Accepted 4 April 2024

Available online 9 April 2024

0262-8856/© 2024 Elsevier B.V. All rights reserved.

in data collection [9]. Moreover, there is a growing concern regarding the potential radiation exposure associated with PET scans [10]. So, leveraging the more sensitive but less available PET to enhance AD diagnostic performance becomes challenging.

The broad success of deep generative models has prompted the development of cross-modal medical image synthesis that can directly synthesize the missing modality from available one for downstream tasks [9]. Since their introduction in 2014 [11], generative adversarial networks (GANs) have gained significant attention, especially in medical image-to-image translation. Specifically, by generating missing PET images from available patients' MRIs using a GAN, the crucial gaps in multi-modal datasets can be filled, resulting in advancing AD diagnosis [5]. Consequently, we can take the advantages offered by PETs while avoiding the extra time, cost, and exposure of acquiring them.

Synthetic PET is potentially useful for diagnosing degenerative disorders, such as AD, where grey matter atrophy, ventricular enlargement observed in MRI, and cerebral distribution of FDG in PET serve as crucial differentiating factors [12]. In a study conducted in [13], two experienced clinicians with 6 and 20 years of clinical neurology experience were asked to check a set of 100 images from 100 subjects in the test data, with an equal split between real and synthetic PETs generated based on the corresponding subjects' MRI and determine whether they believed the image was real or synthetic. According to the provided results, the clinicians had similar performance and could not tell the difference between the real and synthetic PETs.

Translating an MRI as a structural modality to PET which is a functional one is a challenging task and compared to translation tasks involving other modalities, such as CT, receives less attention. Based on recent surveys on medical image translation studies, the number of reviewed papers that investigated MRI to PET translation is significantly less than the number that investigated other modalities together. For instance, in a survey of 103 studies conducted by McNaughton et al. [14], only three studies were dedicated to PET generation from MRIs. Similarly, in the most recent survey on GAN-based generation of realistic 3D volumetric data by Ferreira et al. [15], only four out of 73 included studies focused on MRI to PET translation.

To address the challenges of MRI-to-PET translation within an end-to-end architecture, particularly for 3D images with their added computational complexity, various approaches have been explored in the literature. Employing low-dose PETs to provide some functional information helping better full-dose PET generation [16], simultaneous use of several MRI modalities as input to the framework [17], utilizing multi-modal or multi-scale inputs [18,19], focusing on some regions known to contribute more in AD diagnosis [20] and modifying the loss function to better guide the deep framework toward generating high-quality PETs [6] are some attempts to deal with this challenge. The current study approaches this task by proposing a 3D image-to-image CycleGAN network for an end-to-end translation of MRI to PET. The main contributions of the current study are as follows:

- **Novel 3D Multi-Scale Image-to-Image Translation Network** We introduce a novel approach for translating MRI to PET scans using a 3D multi-scale framework. Unlike conventional methods, our framework leverages multiple scales of the input image by employing some scale blocks. These blocks prepare the input at different resolutions, facilitating more effective information extraction during both the encoding and decoding stages. Additionally, we integrate a multi-scale discriminator, mirroring the encoder part of the generator, to force a balanced competition within the CycleGAN framework, resulting in higher-quality PETs. This framework significantly enhances the mapping between MRI and PET domains, yielding superior translation results compared to the state-of-the-art.
- **Hybrid Loss Function for Improved PET Generation** We propose a hybrid loss function designed to enforce the generation of PET scans, ensuring closer resemblance to ground truth images. This loss function supplements the conventional CycleGAN generator's loss terms

– adversarial loss, cycle-consistency loss, and identity loss. It comprises a weighted combination of L1, L2, and Multi-Scale Structural Similarity Measure (MS-SSIM) terms, preserving both voxel-wise and structural similarities. By integrating these components, we mitigate the risk of producing blurry images while ensuring the generation of realistic synthetic PETs resembling the real ones.

Furthermore, our study includes an extensive evaluation of various methodological aspects, including the impact of multi-scale inputs versus single-scale ones and the utilization of CycleGAN versus basic GAN architectures. Through an ablation study, we elucidate the contributions of each component to the quality of synthesized PETs. Besides, we investigate the effectiveness of our proposed loss function in enhancing PET quality compared to traditional loss terms for CycleGANs. We extend our analysis to practical applications in AD diagnosis, exploring the potential of generated PETs to fill missing data for subjects lacking PET scans in their medical records. Additionally, we conduct several comparisons and statistical tests to assess the influence of factors such as stopping criteria, normalization methods, activation functions, and dropout application on the quality of generated PET scans.

This paper is organized as follows. Section 2 provides a short review of the related literature and explains how the current study will fill the gap in the literature. Section 3 presents the proposed framework and loss function. Section 4 describes the utilized data and experimental details we will employ. Assessing the synthesized PETs, quantitatively and visually, and their contribution to AD diagnosis is also investigated in this section. An ablation study of the proposed framework and loss function as well as some numerical comparison regarding other components of the proposed framework, is provided in section 5. Finally, section 6 concludes the current study.

2. Literature review

GANs have found various applications in the field of medical image literature. [21] presents a systematic review of GANs for medical image classification and segmentation, while [22] provides a comprehensive review of different GAN architectures for medical image synthesis and segmentation. Other applications of GANs in medical imaging include data augmentation [23], medical image denoising [24], synthesizing high-resolution image modalities from low-resolution ones [25,26], and image-to-image translation of different types of PET obtained by different radio-tracers, as well as converting low-dose PETs to high-dose ones [27,28]. However, cross-modality image generation is the most notable application of GANs in medical imaging. Image-to-image translation of MRI to CT [29], CT to MRI [30], MRI and CT cyclic translation [31], MRI to PET [32], CT to PET [33,34] are some examples in this regard.

Since the introduction of GANs in [11], different variations of GAN architectures have been developed in the literature. Conditional GAN [35], image-to-image translation GAN [36], and Cycle GAN [37] are some of the most frequently used architectures. A review of the GAN variants adopted for medical imaging applications can be found in [22].

In addition to focusing on different GAN architectures employed in the literature, the supported dimensionality of these GANs should also be considered, i.e., adopting two-dimensional (2D) GANs for working on some slices of the image or synthesizing a whole modality by a 3D GAN. In the literature, several studies have utilized 2D GANs for image-to-image translation. Focusing on AD, here some studies are mentioned. In [33], a 2D cross-modality image generation for synthesizing PETs from CTs has been examined in a two-step procedure where initially, a PET-like image is generated by a Convolutional Neural Network (CNN). At the same time, a conditional GAN will refine it to generate a PET scan of the liver lesion. Most recently, a 2D GAN-based framework was proposed in [38] to translate one MRI type to another one, including T1-weighted, T2-weighted, T1-weighted post-contrast enhancement (T1c), and flow-sensitive alternating inversion recovery (FAIR). This

framework consists of a three-part generator and two discriminators. The first part of the generator initially generates a pseudo-target image, while the second and third contribute to making it better and generating the final target image. Some other 2D generators can be found in [39,40].

2D-based methods often have spatial discontinuities and feature inconsistencies in medical image generation. On the other hand, while 3D generative models are not prone to be affected by these issues, they require considerable training time due to numerous parameters, feature size, and model complexity [41]. In [42], to overcome the difficulties of employing 3D GANs, three 2D generators are responsible for generating feature maps of coronal, axial, and sagittal slices from a 3D PET which are then fused to construct a 3D MRI. Another paper trying to avoid the discontinuity of 2D slices and benefit from some advantages of 3D whole images can be found in [43]. In this study, four adjacent MRI slices are fed into a cycle GAN to generate the corresponding slices of the CT, which are subsequently recombined to synthesize the whole CT. To manage the computational cost of 3D GANs, utilizing cascade GANs was proposed in [44] to generate high-resolution images from the low-resolution ones where the proposed framework gradually increases the resolution. In [41], a combined 2D and 3D GAN framework is proposed to synthesize high-quality MRIs from low-quality ones. In this study, a 2D generator is in charge of generating 2D slices. In addition to passing a 2D discriminator, these slices will then compose a 3D image to be fed into a 3D discriminator. Another study combining 2D and 3D GAN is [45], where 2D high-quality slices are employed to generate high-quality 3D ultrasound images using a CycleGAN.

Translating a structural image modality to a functional one, such as synthesizing PETs from the corresponding MRIs, can be challenging. An instance of using a 3D GAN to generate high-dose PET scans by merging feature maps from low-dose PET and three MRI modalities is demonstrated in [16]. By incorporating functional information from low-dose PET and structural information from MRI, this approach can potentially address the challenge of converting structural image modalities to functional ones. Utilizing ultra-low-dose PET and two MRI modalities as inputs to a standard-dose PET generator is another example in this regard investigated in [46]. Similarly, [47] employed a fusion network to create a fused image from low-dose PET and MRI, which is then used as the input to a GAN for synthesizing 3D high-dose PET. In a recent study, a Transformer-GAN was proposed for standard-dose PET generation, with MRI and low-dose PET serving as inputs to this framework [48].

Other studies in the literature also explore the generation of PET scans from MRIs. For instance, in [17] a classifier is jointly trained by utilizing some shared features from the encoder and decoder parts of the generator. However, the information on the similarity assessment of the synthesized PETs regarding the original PETs is limited. Another study on generating synthetic PET from MRI is [32], which proposes a GAN with two generator modules. The first module takes a whole MRI as input to generate a whole PET, while the second module consists of independent generators working on separate patches of the input MRI to generate corresponding patches of PET. The output of these modules is concatenated and then passed through some additional layers for fusion. Although training such parallel generators can be computationally expensive, they can capture both local and global information for better image translation. Another relevant study is [20], which introduced a reversible GAN capable of converting MRIs to PETs and vice versa. Unlike a Cycle GAN, which uses two generators as in [49], the proposed reversible GAN uses a single generator with a reversible block in the middle to enable bidirectional conversions. As another example, a two-step approach using two 3D GANs was employed in [50] to generate PETs from MRIs, which involved sketching and refining the synthesized images. Another paper, [51], utilized a 3D basic GAN to generate PET scans from MRIs.

Adding an extra encoder network in addition to the generator and discriminator was considered in [52] where this encoder contributes to the introduced bidirectional mapping mechanism responsible for

embedding the semantic information of PET images for better preserving the diverse details of brain structures in synthetic PET images. In a most recent study, a 3D UNet-like network incorporating a self-attentive module was proposed for MRI to PET translation. The aim is to generate synthetic PET images, which, when combined with patients' MRIs, could enhance AD diagnosis [5]. Another instance of PET generation from MRI to improve AD diagnosis can be found in [9] where a joint learning framework of cross-modal synthesis and diagnosis is proposed.

The current study aims to contribute to the literature on MRI to PET image-to-image translation using the GAN framework, where a gap was identified. Specifically, in [16,47,48], low-dose PETs were in charge of providing some brain functional information for generating full-dose PETs, while in the current study, a single structural MRI will be used as the framework's input. Considering a single MRI modality is also opposed to [18,50,53] where four, two, and eight different contrast MRI modalities are used simultaneously for PET generation, respectively. On the other hand, the focus of PET generation in [5,17,49] is on the classification accuracy of AD subjects. The generated PETs are not assessed based on the common quality metrics for image synthesis. The current study, however, will do a comprehensive investigation on the quality of synthesized PETs as well as utilizing them for AD diagnosis. The proposed GAN in [20] mainly focuses on hippocampus region reversible MRI and PET generation, although full image generation is also considered for comparison. Joint MRI-PET translation and AD diagnosis where subjects' labels and a corresponding loss are incorporated in the framework proposed in [9] is opposed to our architecture that does not take the advantages of subjects' labels in PET generation.

The proposed GAN framework in [32] is in line with the aim of our study. It has global and some local or region of interest (ROI)-based generators, which makes it computationally expensive. However, in the current study, employing different scales of the input MRI does not drastically impact the computational cost of the framework. Moreover, the proposed GAN framework in [32] needs to determine the optimal number of ROIs. This number may differ depending on the dataset and the intensity distribution of the images. On the contrary, assessing how many downscalings should be performed in the current proposed framework is unnecessary.

Besides, in addition to the general complexities of the medical images, including class imbalance and data scarcity [54], translating structural features to functional ones can result in more complexity, which needs a more cautious training of the framework. Therefore, modifying the loss function to guide the deep framework toward better image-to-image translation for medical applications is a common practice in the literature. In a recent study [19], a manifold projection operator and a corresponding generator loss are proposed to project the actual data distribution and the generator distribution into the low-dimensional subspace to prevent mode collapse that might happen in GAN architectures. As in some other instances, a hybrid loss combining 3D gradient profile loss, SSIM, adversarial loss, Kullback-Leibler divergence constraint, and L1 loss was considered in [6], or in [42] a loss function including Jensen Shannon distance (JSD) and L2 loss was employed. Adding an MSE loss to other standard terms of GAN loss [55], a hybrid loss of MS-SSIM, L1, and an ROI-based loss [32], and a weighted average of the employed losses in the proposed subnetworks in [17] are some other examples in this regard. In order to preserve the structural similarity while preserving the voxel-wise similarity and avoiding generating blurry images, the current study will propose a hybrid loss function having a similarity loss term in addition to what is common in the literature for CycleGANs.

3. Methodology

3.1. Proposed multi-scale CycleGAN framework

In this study, a GAN framework consisting of two similar generators

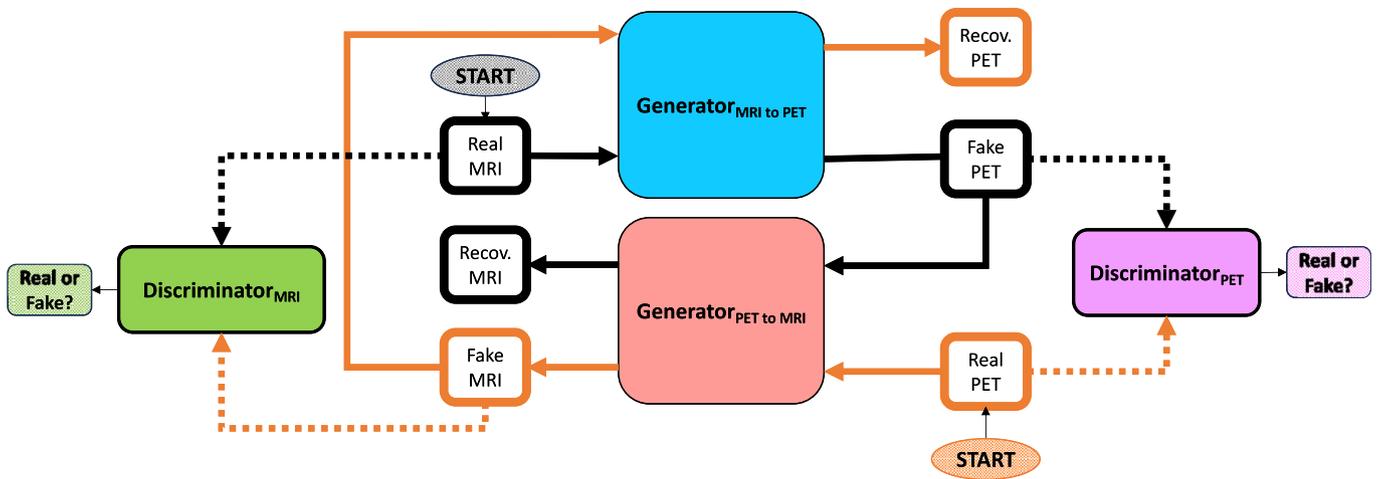


Fig. 1. Overview of the CycleGAN framework for MRI to PET translation.

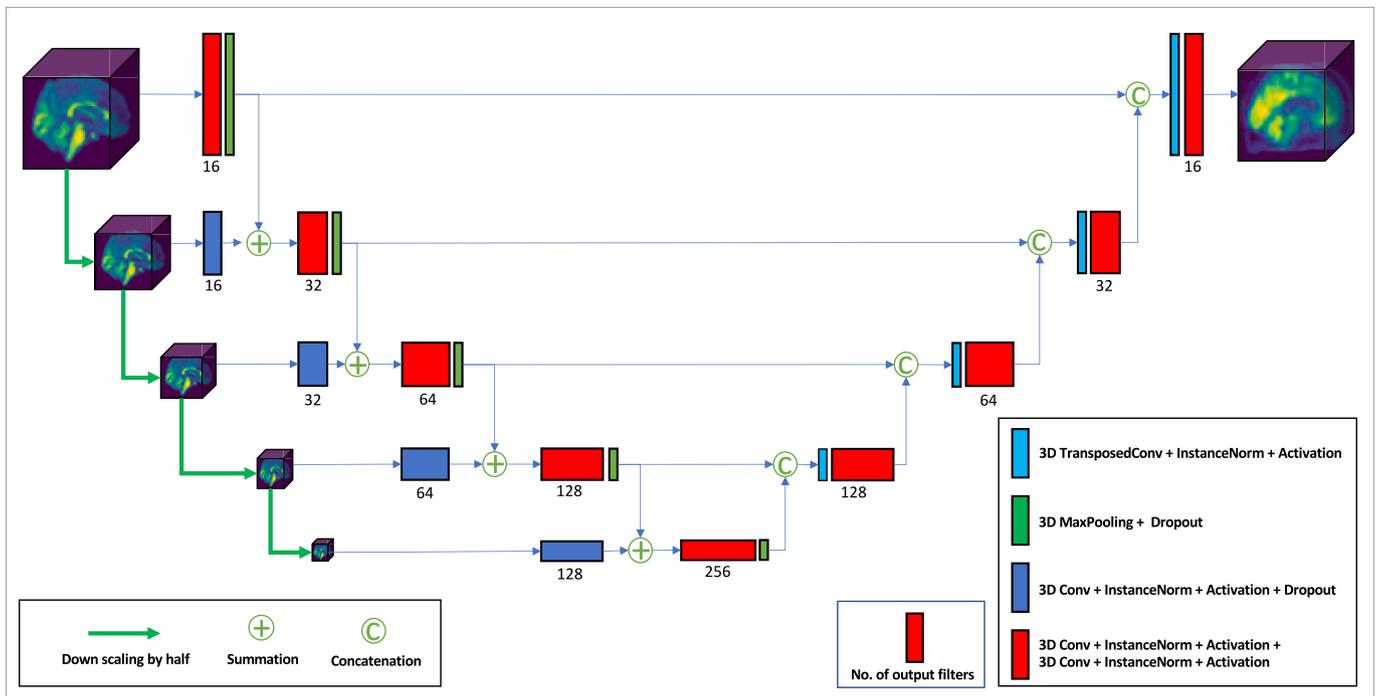


Fig. 2. Proposed 3D Multi-scale generator network (MultiG).

and two similar discriminators is employed as depicted in Fig. 1. The generator network comprises convolutional encoder and decoder parts in a UNet-like architecture. The UNet model is a successful architecture commonly used in medical image segmentation and generation due to its encoder-decoder structure and skip connections [56,57]. It extracts image features using downsampling convolutions in the encoder part and recovers the image to its original size using the decoder. The skip connection combines low-level and high-level features to provide better target output. However, the UNet model has some limitations, such as inflexibility in training with different size input images and incomplete feature exploitation [58]. In the literature, especially for image segmentation, redesigning skip connections and adding more connections between the encoder and decoder parts [59,60] or cascading UNets [61] or modifying the convolutional blocks [62] are investigated to help better exploiting image features.

The proposed approach in this study involves a multi-scale UNet architecture, where the downsampling steps of the encoder will receive

input from different scales of the input modality. The benefit of using this multi-scale input is the integration of information from input images at various scales to help better recover the low-level information, which is prone to be lost during the downsampling convolutions in the encoder part [58]. The proposed generator is indicated in Fig. 2. In this network, the size of the first input MRI is $64 \times 64 \times 64$. For down-scaling this MRI by half, we employed trilinear interpolation with a scale factor equal to 0.5. Then, to prepare it for the summation with the output of the upper convolution layer, a 3D scale block is designed.

Despite the potential advantages of stacking several convolution layers with an incrementally increasing number of filters for better mapping the features, considerable computational and memory requirements will be challenging [63,64]. In the current study, where two 3D deep networks compete against each other, managing the computational cost is crucial. Accordingly, we employed convolution blocks consisting of two consecutive convolution layers. The first layer is a convolution operation with a kernel size of $3 \times 3 \times 3$ where the number

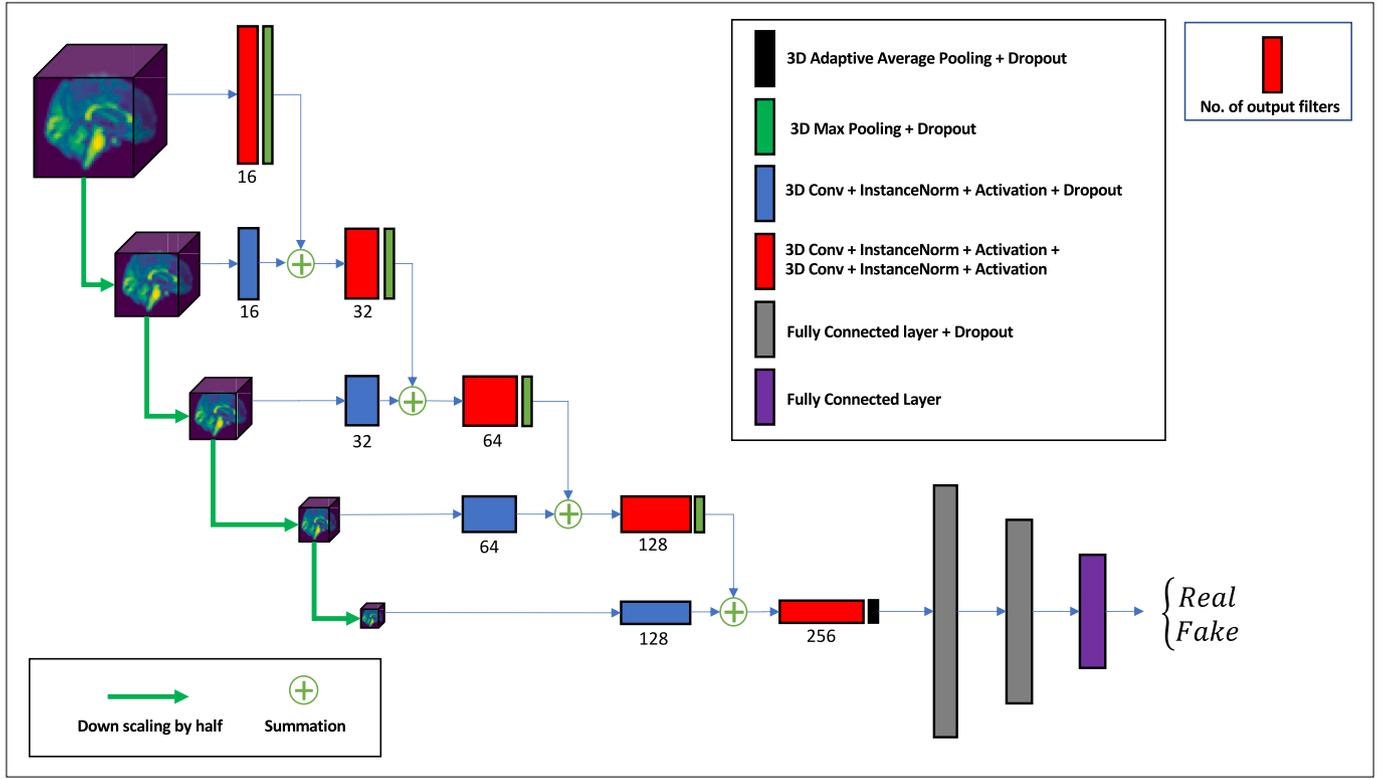


Fig. 3. Proposed 3D Multi-scale discriminator network.

of its output filters is twice the input ones. We use the same kernel for the second layer but do not double the number of filters. So, without a drastic change in the number of parameters, we can benefit from an extra convolution operation.

Similar to some very recent studies [6,17], we used instance normalization instead of batch normalization in the proposed convolution blocks. After passing the normalized features to the activation function, we applied dropout to avoid over-fitting issues that frequently happen in deep architectures. Fig. 2 illustrates the scale block's components. For the discriminator, we employed a similar network to the down-sampling part of the generator. This similarity is supposed to help reach a more balanced game between the generator and discriminator in the proposed GAN framework. The implemented discriminator network is illustrated in Fig. 3. We embed the proposed generator and discriminator networks in a CycleGAN framework, which will be mentioned as CycleGAN-MultiG in the current study. So, CycleGAN-MultiG contains two generators to map MRI to PET and PET to MRI. Accordingly, this architecture has two discriminators: one for distinguishing between ground truth and synthesized PETs and the other for discriminating the real and synthesized MRIs.

3.2. Proposed loss function

For a CycleGAN, a combined loss function was introduced in [37] to enforce the generation of more realistic and similar images. This loss function comprises three individual losses: adversarial, cycle-consistency, and identity loss. Following a short description and their corresponding terms are provided.

Adversarial Loss: The adversarial loss function is designed to minimize the difference between the generated and the real images, making the generated images indistinguishable from the ground truth. The mathematical term for the adversarial loss is the sum of the negative log-likelihoods of the real and synthesized images as given in eq. 1 for generating and discriminating PETs.

$$L_{Adv}(G_{PET}, D_{PET}) = -\frac{1}{N} \sum_{i=1}^N \log(D_{PET}(PET)) + \log(1 - D_{PET}(G_{PET}(MRI))) \quad (1)$$

where G_{PET} is the PET generator from MRI and D_{PET} is the PET discriminator. A similar term should also be considered for minimizing adversarial loss for the pair of MRI's generator and discriminator.

Cycle Consistency Loss: Cycle consistency loss ensures that if an image from one domain is translated to another and then returned to the original domain, it remains similar to the original image. So, the aim is to minimize the difference between the original and generated images after these two translations. The regarding loss term for translating MRI and PET domains to each other is given in eq. 2.

$$L_{Cyc}(G_{PET}, G_{MRI}) = \|G_{PET}(G_{MRI}(PET)) - PET\|_1 + \|G_{MRI}(G_{PET}(MRI)) - MRI\|_1 \quad (2)$$

where G_{PET} is the PET generator from MRI and G_{MRI} is the generator for translating PET to MRI. In the equation, instead of the employed L1 distance, L2 can also be employed.

Identity Loss: Identity loss is used to preserve the identity of the input image in the output image during the training of the generators to ensure that the generator produces images similar to the input rather than producing quite new and different images. The corresponding loss term for the PET generator is presented in the following equation.

$$L_{Id}(G_{PET}) = \|G_{PET}(PET) - PET\|_1 \quad (3)$$

Similarly, the MRI generator's identity loss term would be:

$$L_{Id}(G_{MRI}) = \|G_{MRI}(MRI) - MRI\|_1 \quad (4)$$

However, due to the challenges of cross-modality image synthesis task, from a structural scan to a functional one [32], we hypothesize that incorporating an additional term to the standard losses used in a CycleGAN could enhance the quality of synthesized PET images and

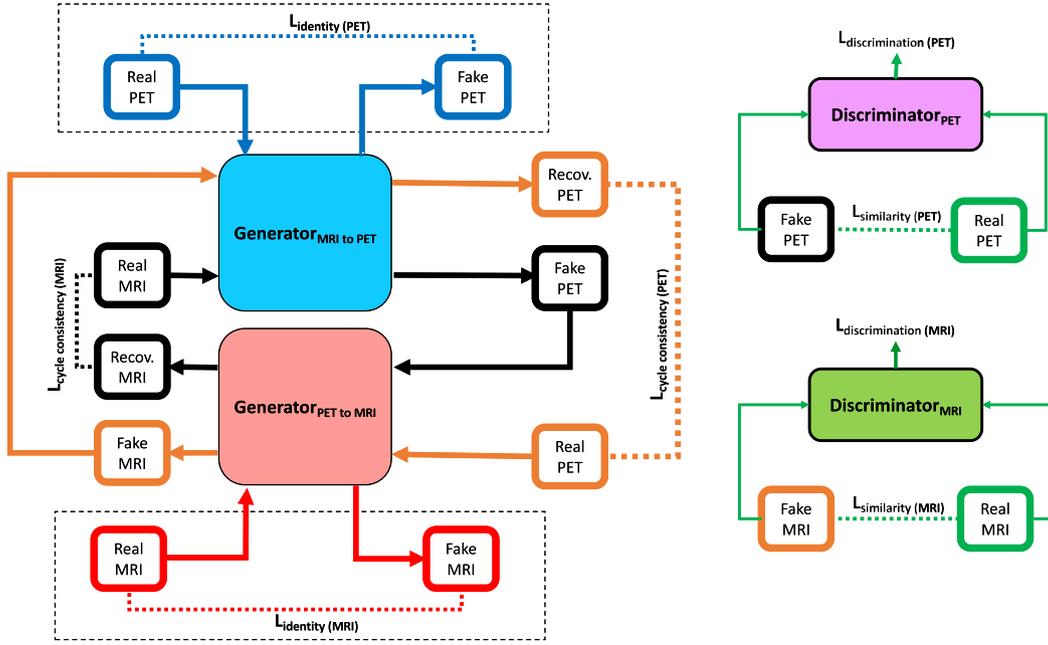


Fig. 4. Overview of the proposed losses in the proposed CycleGAN framework for MRI to PET translation.

improve the values of similarity metrics. As a result, we propose to add an extra loss term that will assess the similarity between the synthesized PET images and their respective targets.

The use of various loss functions to measure the dissimilarity between the ground truth image and a synthesized image has been explored in the literature. Most of these loss functions are either pixel or voxel-wise, used to quantify the pixel or voxel level discrepancy. Mean absolute error (MAE) and mean square error (MSE), also called L1 and L2 distances, are commonly employed in image reconstruction applications. However, it has been established that these standard measures do not always align with human perceptions of quality [65] and may result in a synthesized image that appears blurry [24]. So, in some studies in the literature, utilizing multi-scale structural similarity measure (MS-SSIM) was proposed to overcome this challenge [24,32,65]. Moreover, due to the multi-scale nature of MS-SSIM, it can better capture the strong feature correlations in medical images than mean-based methods [24]. Since MS-SSIM, given in eq. 5, is differentiable, it can be easily employed for gradient-descent-based learning.

$$S(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (8)$$

where μ_x and σ_x are the mean and the variance of x , while σ_{xy} refers to the covariance of x and y . In these equations, C_1 , C_2 , and C_3 are small constants.

This study proposes to use a hybrid similarity loss as given in eq. 9. By adding two voxel-wise losses in addition to $L_{MS-SSIM}$ and tuning their relative weights (λ_{MAE} , λ_{MSE} , and $\lambda_{MS-SSIM}$), we hypothesized that the network would be forced to generate more realistic and indistinguishable images. Then, this similarity loss will incorporate along with cycle consistency, identity, and adversarial losses in the total loss term of the PET generator as mentioned in eq. 10. The proposed framework also uses a similar loss term for the MRI generator.

$$L_{Sim}(G_{PET}) = \lambda_{MAE} \times L_{MAE}(G_{PET}) + \lambda_{MSE} \times L_{MSE}(G_{PET}) + \lambda_{MS-SSIM} \times L_{MS-SSIM}(G_{PET}) \quad (9)$$

In general, Big-O notation is frequently employed. Big-O character-

$$L_{MS-SSIM}(G_{PET}) = 1 - I_M(G_{PET}(MRI), PET)^{\alpha_M} \prod_{j=1}^M C_j(G_{PET}(MRI), PET)^{\beta_j} S_j(G_{PET}(MRI), PET)^{\gamma_j} \quad (5)$$

where I , C , and S are luminance, contrast, and structure terms illustrated in eq. 6, 7, and 8, respectively; M is the employed scale for iteratively down-sampling x and y by a factor of 2. α , β , and γ adjust the relative importance of these terms which are usually set to 1 [66].

$$I(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (6)$$

$$C(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (7)$$

izes the worst-case scenario and provides insights into the execution time or space requirements, such as memory usage or disk space, for a given function or algorithm [67]. To examine the computational complexity of the proposed similarity loss, MSE, MAE, and MS-SSIM should be individually investigated. The corresponding computations of MSE and MAE are very similar, each having a computational complexity typically considered $O(n)$, where n represents the input data size (length \times height \times width). So, their computational cost will grow linearly based on the input data size. For MS-SSIM, the computational complexity is $O(kn)$, where k is the number of utilized downsampling scales in the MS-SSIM calculation. In the current study, the value of k is set to 4, following the literature [68]. Therefore, MS-SSIM also exhibits a linear computational complexity. So, for the proposed similarity loss, a

Table 1
Demographic and MMSE score of the included subjects in this study.

Diagnosis	No. of Subjects	Age (mean \pm std)	MMSE (mean \pm std)
CN	80	75.56 \pm 4.72	28.91 \pm 1.13
MCI	129	75.17 \pm 7.37	27.13 \pm 1.70
AD	73	75.24 \pm 7.32	23.52 \pm 2.15

weighted summation of MSE, MAE, and MS-SSIM, the overall computational complexity remains linear, ensuring a modest impact on the computational cost of the proposed framework.

Accordingly, the new term for PET generator loss will be as eq. 10, where W_{Sim} , W_{Cycle} , W_{Id} , and W_{Adv} are the corresponding weights of similarity, cycle, identity, and adversarial losses, respectively. Fig. 4 illustrates all the considered losses in the proposed CycleGAN.

$$L(G_{PET}) = W_{Sim}L_{Sim}(G_{PET}) + W_{Cycle}L_{Cycle}(G_{PET}, G_{MRI}) + W_{Id}L_{Id}(G_{PET}) + W_{Adv}L_{Adv}(G_{PET}, D_{PET}) \quad (10)$$

4. Experiments and results

4.1. Data acquisition and preprocessing

Our study is based on a popular AD dataset which is one of the few publicly available AD datasets and has been previously used in various other works. The Alzheimer’s Disease Neuroimaging Initiative (ADNI)¹ is a longitudinal multi-center study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. There are three groups of imaging modalities in the ADNI dataset. Structural MRIs, including T1-weighted, T2-weighted, and Fluid-attenuated inversion recovery (FLAIR), functional MRIs taken in the resting state, and two types of PETs, including Amyloid and Fluorodeoxyglucose (FDG) ones.

Among the available neuroimages in the ADNI dataset, a pair of baseline MRI and PET are selected for those subjects having them at the baseline. These subjects are labeled based on their conversion in 36 months, similar to [69]. Patients who stay normal control (NC) and AD during the follow-up are labeled NC and AD, respectively. For mild cognitive impairment (MCI), different types of cognitive impairment, such as late MCI and early MCI, are labeled as MCI. Moreover, if a subject’s label is NC (normal control) and in 36-month screening turns into MCI or AD, the considered label would be MCI or AD, respectively. Subjects initially diagnosed as AD or MCI and then converted to NC in the 36th month were excluded.

To shorten the preprocessing steps, we download those scans that have undergone the maximum level of correction from the ADNI website. These images have been passed through grad-warping and intensity correction steps and have been scaled for gradient drift using the phantom data. Moreover, for skull-stripping of sMRIs, we applied the brain extraction (BET) package of FSL library.² Then, for linear MRI registration to the Montreal Neurological Institute (MNI) template, we use the FreeSurfer package.³ Afterward, each PET was registered to the same patient’s MRI using FreeSurfer linear registration. Then, we resize all MRIs and PETs to $64 \times 64 \times 64$ using tensor interpolation of Pytorch to reduce the computational cost.

After excluding all the mislabeled patients and those having mis-registered images, in total, 282 subjects remained to be studied in this paper. Table 1 provides a summary of these subjects, their diagnosis label, as well as their age and Mini-Mental State Examination (MMSE) score. MMSE is a known cognitive test for investigating the orientation to time and place, the immediate and delayed recall of three words, the

attention and calculations, language, and visuoconstructional functions [70]. Despite the extensive application of MMSE in clinical application due to its correlation with cognitive decline due to AD, it may add confounding information to the diagnosis procedure [71]. However, combined with other AD biomarkers, MMSE can help more accurate diagnosis.

To have a completely unseen test set, 10% of the subjects were randomly selected, and the remaining 254 subjects were utilized for training and validation in a five-fold cross-validation setting. For data augmentation, blurring data by Gaussian filters was used, i.e., after putting the test set aside, the training set was tripled using Gaussian filters by a Sigma randomly picked between 0 and 1.5. Scipy,⁴ an open-source Python library, was used for applying these filters.

4.2. Implementation details

Pytorch was utilized on an NVIDIA GeForce RTX3090 GPU to train all the proposed networks. Adam optimizer with an initial learning rate of $5e-4$ was adopted for GAN frameworks’ training, while a step scheduler decays the learning rate every 50 epochs by half. We considered a batch size of 4 for the GANs. Parametric ReLU was applied (PReLU) as the activation function of all layers except the generator’s last layer, for which we employed Sigmoid to get an output PET whose voxel intensities are between zero and one. Based on several experiments we conducted, the best combination of weights for L_{Sim} in the current study is 1, 1, and 100 for L_{MAE} , $L_{MS-SSIM}$, and L_{MSE} , respectively. Moreover, the corresponding weights of L_{Sim} , L_{Cycle} , L_{Id} , and L_{Adv} were set to 100, 10, 10, and 1, respectively. For BasicGAN, the same weights were used, except W_{Cycle} and W_{Id} , which are excluded from the generator’s loss.

Assessing the quality of synthesized images has always been controversial. For non-medical applications, a visual inspection can provide good information at a glance. However, it is more complicated for medical applications. The human eye cannot be guaranteed to distinguish distortions and unusual patterns, even if the synthesized images are inspected by specialists [72]. On the other hand, there are different similarity measures in the literature to assess the quality of synthesized images. However, it is hard to pick one of them as the primary measure to choose the best generative network. This study considers three similarity measures to quantify the quality of the generated PETs. These measures are Peak-Signal-Noise-Ratio (PSNR), Structural Similarity Measure (SSIM), and Mean Absolute Error (MAE), which are frequently utilized in the literature [33,40]. The corresponding equations of these measures are as follows: Real refers to the ground truth image, and Fake refers to the generated image by feeding the real image into the framework. It is worth mentioning that MAE is a scale-dependent measure. So, the scale of the intensities of real and fake PETs should be the same.

$$MAE(real, fake) = \frac{1}{N} \sum_{i=1}^N |real(i) - fake(i)| \quad (11)$$

$$PSNR(real, fake) = 20 \log_{10} \frac{Max_{Intensity}}{\sqrt{MSE(real, fake)}} \quad (12)$$

where MSE is the mean squared error between the real and fake PETs and $Max_{Intensity}$ stands for the maximum possible intensity for a voxel in the real or generated image. In this study, all the images are scaled between zero and one before passing through the GAN network. Moreover, the activation function of the last layer of GAN is Sigmoid, and its output ranges between zero and one. So, we set the value of $Max_{Intensity}$ equal to one. The value of the SSIM measure is a number between -1 and 1 , where 1 shows ideal similarity, 0 means no similarity, and in the

¹ <https://adni.loni.usc.edu/>

² <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>

³ <https://surfer.nmr.mgh.harvard.edu/>

⁴ <https://scipy.org/>

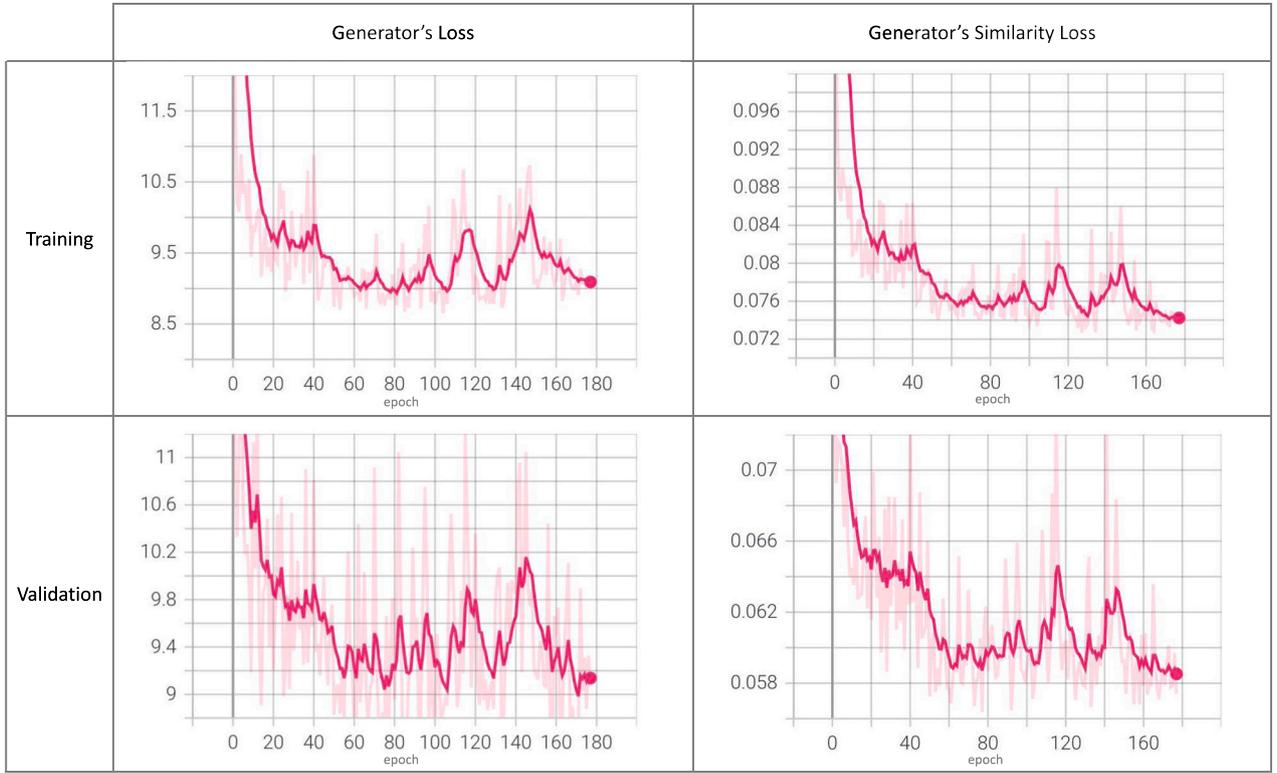


Fig. 5. The convergence dynamics of CycleGAN-MultiG's loss over training epochs.

Table 2

Quality comparison of the synthesized PETs using different architectures (best in bold).

Method	PSNR	SSIM	MAE
FCN [73]	22.89	0.5838	0.0516
UNet [56]	22.92	0.6343	0.0517
Pix2Pix [36]	22.93	0.6061	0.0514
BicycleGAN [74]	24.84	0.6503	0.0409
BPGAN [6]	25.08	0.6646	0.0396
JSRL [17]	26.25	0.7709	0.0304
PA-Net [5]	26.41	0.7689	0.0293
CycleGAN-MultiG	27.25	0.7997	0.02791

case of a perfect anti-correlation, this value will be -1 .

$$SSIM(real, fake) = I(real, fake)^{\alpha} C(real, fake)^{\beta} S(real, fake)^{\gamma} \quad (13)$$

4.3. Image quality assessment

This section is dedicated to assessing the quality of synthesized PETs using CycleGAN-MultiG. Beforehand, to assess the learning trajectory of the proposed framework and its generalization ability to new instances, the convergence dynamics of the generator's loss and the generator's similarity loss over training epochs should be carefully monitored and evaluated to ensure robust performance. Fig. 5 illustrates the training and validation curves for one of the cross-validation folds, revealing a stable convergence without significant divergence or over-fitting. The observed fluctuations and minor peaks in these curves can be attributed to dynamic adjustments made by the learning rate scheduler during training. It's important to note that, due to computational constraints, early stopping was implemented before achieving a perfectly smooth curve, maintaining a balance between computational efficiency and satisfactory model convergence given our resource limitations.

This section assesses the quality of synthesized PETs by employing different similarity measures and providing some visual inspections. For

Table 3

P-values for testing the hypothesis of the superior performance of CycleGAN-MultiG over JSRL and PA-Net.

Similarity Measure	H0: CycleGAN-MultiG equals JSRL H1: CycleGAN-MultiG Better than JSRL	H0: CycleGAN-MultiG equals PA-Net H1: CycleGAN-MultiG Better than PA-Net
PSNR	9.765e-3	1.102e-2
SSIM	6.367e-4	2.247e-4
MAE	2.984e-4	9.653e-4

comparison purposes, since the only available source code among all the reviewed papers focusing on MRI to PET translation is the JSRL framework introduced in [17], we used this architecture as a basis for comparison. Moreover, we have implemented the Pyramid-Attentive GAN (PA-Net) architecture as described in [5], which represents the most recent published study in the literature. In this study, the number of channels for different layers of the generator is not mentioned. So, we assumed that the mentioned number of channels for the discriminator is also valid for the generator. To ensure a fair comparison, we replaced our proposed network with JSRL and PA-Net, feeding them with our data, and keeping all other hyperparameters consistent with our framework. We conducted the experiment with the JSRL and PA-Net frameworks twice and collected the numerical results.

Besides, we use the numerical results of implemented GANs in [6] due to a similar dataset structure and image preprocessing to our study. These architectures are FCN [73], UNet [56], Pix2Pix [36], BicycleGAN [74], and BPGAN [6]. These results, along with the obtained mean value of JSRL [17], PA-Net [5], and our proposed framework, CycleGAN-MultiG, are given in Table 2.

Regarding the numerical results given in Table 2, employing JSRL and PA-Net for our data resulted in similar outcomes. Notably, PA-Net exhibited better performance in terms of PSNR and MAE metrics, while JSRL demonstrated higher SSIM. Meanwhile, CycleGAN-MultiG outperforms JSRL and PA-Net, the most recently introduced image-to-

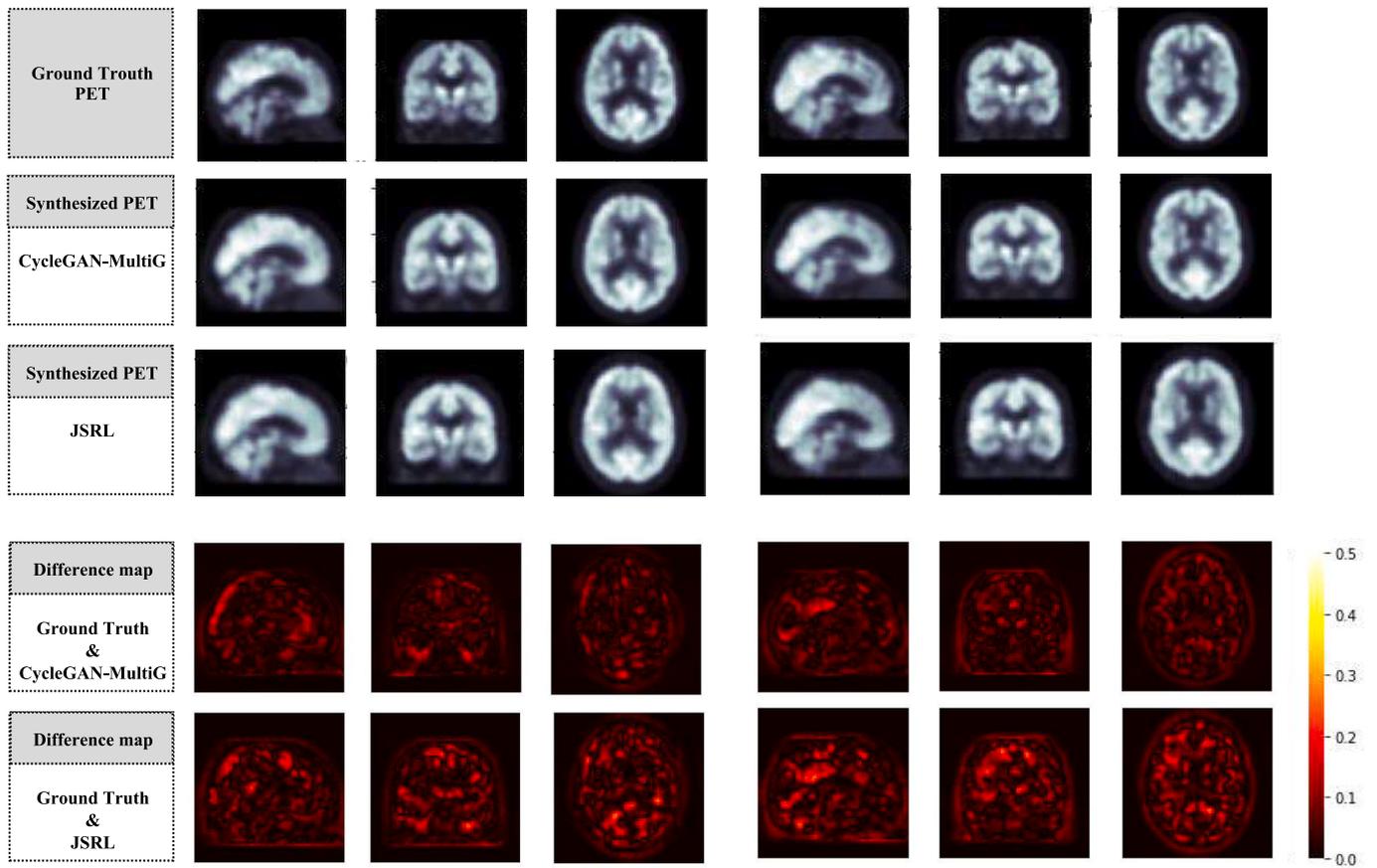


Fig. 6. Comparison of ground truth PETs and their corresponding Synthesized ones and their absolute difference maps.

image translation architectures for PET generation from MRIs, across all investigated similarity metrics. Specifically, CycleGAN-MultiG achieves a mean PSNR value that is 3.8% and 3.1% higher than JSRL and PA-Net, respectively. Additionally, for SSIM, there is an increase of 3.7% and 3.8% from 0.7709 and 0.7689 to 0.7997, respectively. Moreover, CycleGAN-MultiG results in a decrease in the mean MAE value from 0.0304 and 0.0293 to 0.02791, representing reductions of 8.19% and 4.78%, respectively, for this metric.

However, to check for a statistically significant difference between CycleGAN-MultiG and JSRL as well as CycleGAN-MultiG and PA-Net, we used Wilcoxon rank-signed test [75]. The corresponding p-values of these two-sided non-parametric tests for this comparison are given in Table 3. According to this table, for all similarity measures (PSNR, SSIM, and MAE), we can reject the null hypothesis at a confidence level of 5%, confirming the proposed framework's superior performance over state-of-the-art architectures for 3D image-to-image translation of MRI to PET.

In addition to numerically comparing the quality of synthesized PETs, a visual assessment of the generated images can help illustrate how similar these images are to the ground truth. Accordingly, two subjects from the test set are considered the ground truth. Then, we use the trained CycleGAN-MultiG framework to generate their corresponding synthesized PETs. For comparison purposes, the corresponding generated PETs by the trained JSRL [17] for these subjects are also considered. Moreover, to better understand the absolute difference between the ground truth PETs and their synthesized ones, the corresponding difference maps have also been illustrated in Fig. 6.

Since the difference between the synthesized PETs using PA-Net [5] were not noticeable compared to JSRL [17] due to their similar performance, JSRL, which demonstrated slightly better SSIM, is depicted as the representative image-to-image translation architecture from the literature. This visual assessment confirms the performance of the

proposed CycleGAN-MultiG framework to generate diverse PET scans that can resemble the ground truth and outperform the state-of-the-art competing framework. Therefore, CycleGAN-MultiG has the potential to be used for filling missing PETs when MRI is the only available neuroimage to make a diagnostic decision for a subject.

4.4. Task-based quality assessment

Medical images are typically acquired for specific purposes. While traditional similarity measures such as SSIM, MAE, and PSNR can provide a decent estimation of the image quality, they may not always be aligned with the medical goal of the image. Hence, task-based image quality assessment is necessary for medical imaging [76,77]. Some studies have demonstrated that some state-of-the-art GANs trained on medical image datasets can generate images that appear realistic but contain potentially impactful errors [78,79]. Therefore, to safely use GANs in medical imaging applications, synthesized images must be evaluated based on the specific task they are intended for. In line with this, in addition to quantitative and visual assessment of synthesized PETs, it is essential to investigate their contribution to improving AD diagnosis, which is the primary objective of generating missing PETs in the present study.

To study to which extent these synthesized PETs can help improve the performance of AD diagnosis, we investigate an extreme setting. In this setting, all the subjects in the test set have synthesized PETs and compare it to the AD diagnosis when subjects have real PETs and when there is no PET at all. For comparison purposes, the synthesized PETs using the implemented JSRL and PA-Net frameworks are also investigated.

Since this study focuses on image-to-image translation, we did not design or fine-tune a classifier. Instead, a classifier similar to the

Table 4
Comparison of AD diagnosis using ground truth and synthesized PET (best in bold).

Method	Accuracy	Precision	Recall	F1	AUC
MRI + s*	0.528	0.698	0.528	0.421	0.751
MRI + 100% real PET + s	0.635	0.681	0.635	0.612	0.823
MRI + 50% real PET + 50% synthesized PET (JSRL) [17] + s	0.569	0.678	0.542	0.512	0.765
MRI + 50% real PET + 50% synthesized PET (PA-Net) [5] + s	0.572	0.632	0.549	0.517	0.772
MRI + 50% real PET + 50% synthesized PET (Ours) + s	0.577	0.641	0.558	0.521	0.781
MRI + 100% synthesized PET (JSRL) [17] + s	0.592	0.683	0.592	0.534	0.797
MRI + 100% synthesized PET (PA-Net) [5] + s	0.601	0.675	0.603	0.542	0.802
MRI + 100% synthesized PET (Ours) + s	0.614	0.681	0.614	0.572	0.818

* Supplementary information (age and MMSE).

introduced discriminator network was utilized. However, for the sake of simplicity, we removed the multi-scale blocks and modified the last FC layer to gain a 3-class classifier (AD vs. MCI vs. NC). The learning rate for this network was set to 5e-5 without any scheduler and a batch size of 8 was considered. To train this network in a five-fold cross-validation, we used subjects' MRIs and real PETs. We also utilized subjects' age and MMSE score which are the most commonly used supplementary information in AD diagnosis literature, as mentioned in a recent survey [80]. The concatenated tensor of MRI, PET, age, and MMSE of each subject was the input to this classifier. In the test phase, we utilized this trained network to classify unseen test subjects (pairs of MRI and PETs along with the subjects' age and MMSE, as well as MRIs and synthesized PETs along with their corresponding age and MMSE). The AD diagnosis performance was measured using some well-known metrics: Accuracy, Precision, Recall, F1-score, and Area under the receiver operating characteristic curve (AUC) [81]. We averaged the corresponding classification metrics obtained for the unseen test regarding these folds as presented in Table 4.

According to Table 4, PET scans can considerably contribute to AD

diagnosis performance. Considering AUC as the most comprehensive metric that is less sensitive to class imbalance [82], using ground truth PETs along with MRIs can increase the AUC metric from 0.751 to 0.823. Even with synthesized PETs that are generated based on the corresponding MRIs, the performance of the AD classifier will increase. Using synthesized PETs of the CycleGAN-MultiG framework can enhance the value of AUC to 0.818, which is 8.9% higher than the corresponding AUC when MRI is the only input to the classifier. However, this table indicates that the model performs poorly when trained on a combination of synthesized and real data compared to models trained solely on real or synthesized PETs. Slight distribution differences between synthetic and real data can result in challenges during model training that hinder the model's overall performance when integrating both data types. A similar finding is reported in [20]. Furthermore, a comparison between the classification performance of our proposed framework, JSRL [17] and PA-Net [5] confirms that our framework outperforms these state-of-the-art frameworks for image-to-image translation of MRI to PET, which results in PET images more closely resembling ground truth PETs.

5. Discussion

This section is dedicated to some ablation studies and numerical comparisons to figure out the contribution of the proposed GAN's elements to its superior performance compared to the literature. We conducted ablation studies corresponding to the GAN type (BasicGAN or CycleGAN), multi-scale inputs, and the proposed loss function. Moreover, the impact of different network components and implementation details, such as the normalization layer, activation function, dropout, and considered stopping criterion are numerically compared.

5.1. GAN framework ablation study

To ablate the impact of multi-scale inputs and the CycleGAN, three alternative frameworks are implemented in this study: BasicGAN-PlainG, where the generator and discriminator lack multi-scale inputs and blocks shown in Fig. 7 and 8, resembling a standard GAN; BasicGAN-MultiG, featuring a multi-scale generator and discriminator in a basic GAN setup; and CycleGAN-PlainG, a CycleGAN variant without any consideration for multi-scale inputs.

To assess computational costs in these GAN frameworks, the number

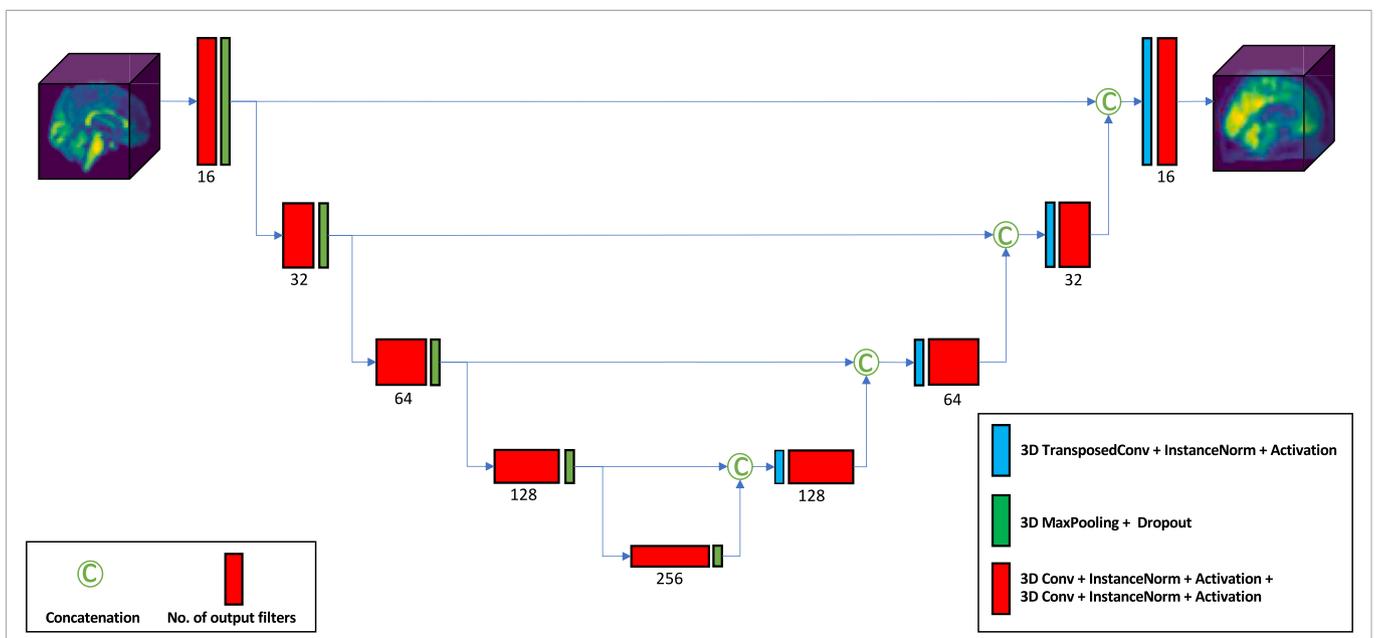


Fig. 7. 3D generator network (PlainG).

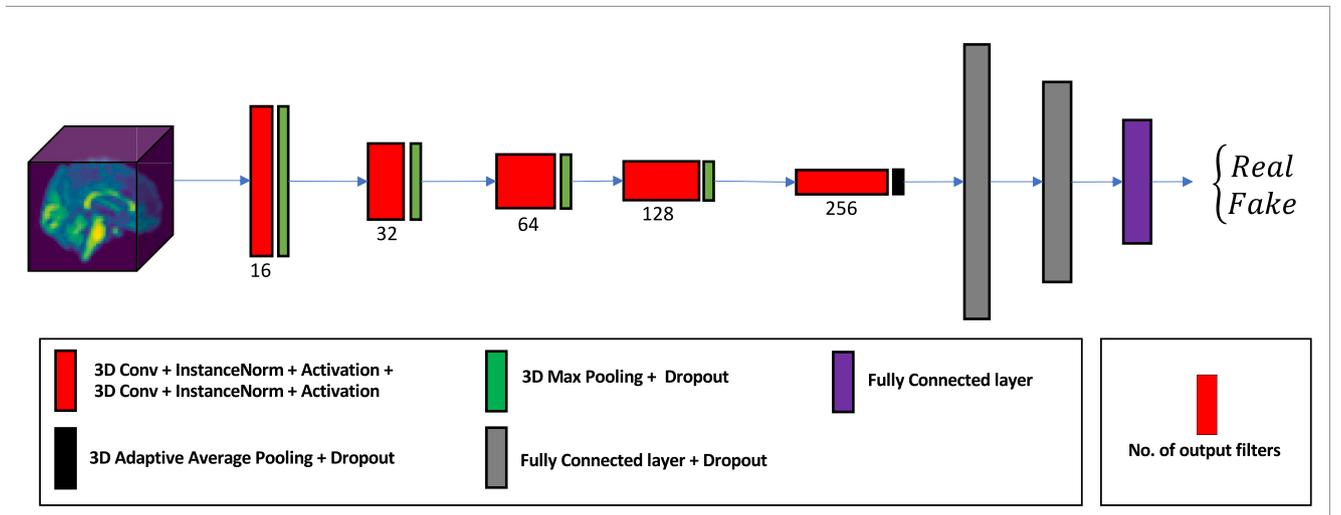


Fig. 8. 3D discriminator network (PlainD).

Table 5
Computational efficiency of different GAN frameworks.

Architecture	No. of Trainable Parameters ($\times 10^6$)			No. of MACs ($\times 10^9$)		
	Generator(s)	Discriminator(s)	Total	Generator(s)	Discriminator(s)	Total
BasicGAN-UNet [83]	16.313	7.319	23.632	236.701	33.747	270.448
BasicGAN-PlainG	8.234	0.904	9.138	48.025	1.727	49.752
BasicGAN-MultiG	8.245	0.907	9.152	54.925	2.334	57.259
CycleGAN-UNet [83]	32.626	14.638	47.264	473.402	67.494	540.896
CycleGAN-PlainG	16.468	1.808	18.276	96.050	3.454	99.504
CycleGAN-MultiG	16.490	1.814	18.304	109.850	4.668	115.518

Table 6
Quality comparison of the synthesized PETs using different architectures (best in bold).

Method	PSNR (mean \pm std)	SSIM (mean \pm std)	MAE (mean \pm std)
BasicGAN-PlainG	27.314 \pm 0.104	0.7981 \pm 3.7e-3	0.02803 \pm 2.4e-4
BasicGAN-MultiG	27.434 \pm 0.184	0.7994 \pm 2.4e-3	0.02792 \pm 2.6e-4
CycleGAN-PlainG	27.287 \pm 0.074	0.7994 \pm 2.4e-3	0.02796 \pm 2.0e-4
CycleGAN-MultiG	27.251 \pm 0.093	0.7997 \pm 2.9e-3	0.02791 \pm 2.7e-4

of trainable parameters and Multiply–Accumulate (MAC) operations are provided in Table 5. The presence of two generators and two discriminators in CycleGAN doubles the trainable parameters. Besides, we also report the corresponding metrics for the introduced 3D UNet in [83]. This UNet resembles our PlainG with convolution blocks having two consecutive convolutions. Like the current study, the entire UNet is the generator, and its encoder part acts as the discriminator. Metrics are based on a $64 \times 64 \times 64$ input size, providing a fair comparison. According to this table, the proposed PlainG and MultiG networks exhibit superior computational efficiency compared to the UNet and have reasonable trainable parameters and MACs.

These GAN frameworks are investigated based on PSNR, SSIM, and MAE to study the impact of employing different architectures on the quality of synthesized images. To ensure a fair comparison, these frameworks were fed by the same set of images with the same training and validation folds and identical test data. Moreover, to help mitigate the impact of randomness in the training process, the five-fold cross-validation is repeated twice, and the numerical results for the unseen test set are recorded. According to the mean and standard deviation (std)

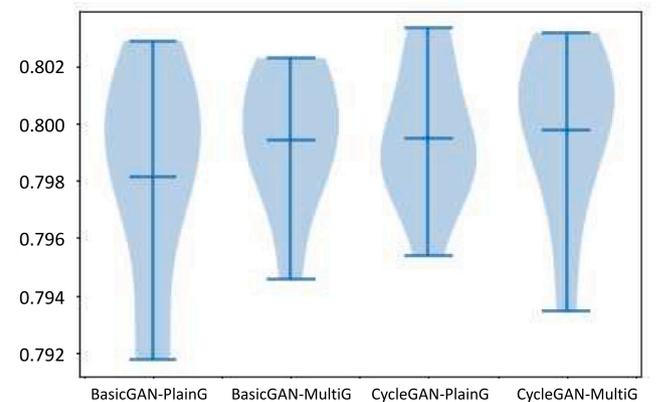


Fig. 9. Comparison of obtained SSIM distributions using different GANs.

values given in Table 6, CycleGAN-MultiG has generated the most similar PETs, on average, regarding SSIM and MAE measures, but not PSNR. Different studies focused on investigating similarity measures for medical image applications have argued that MSE and PSNR, which is also based on MSE, are not able to provide a good measure in terms of the image features [84,85]. In medical applications, it has been studied that the image of higher PSNR value may look worse than that of lower value [86], and utilizing SSIM, which more closely resembles human visual perception, was proposed instead.

Focusing on SSIM as the primary metric, Fig. 9 depicts the distributions of the SSIM values obtained from the various GANs, presented as violin plots. The results indicate that CycleGAN-MultiG achieved a slightly better median compared to the other frameworks. Notably, CycleGAN-MultiG exhibits a more favorable distribution, although

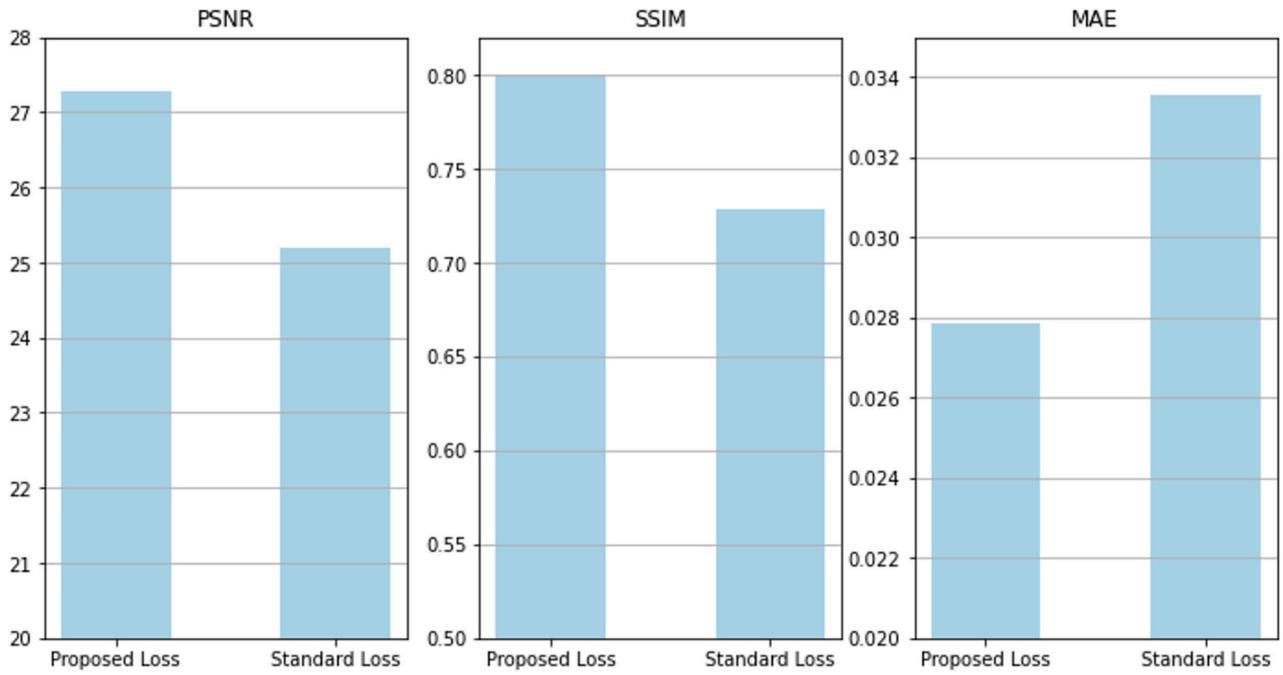


Fig. 10. Comparing the impact of the proposed loss function on the quality of synthesized PETs.

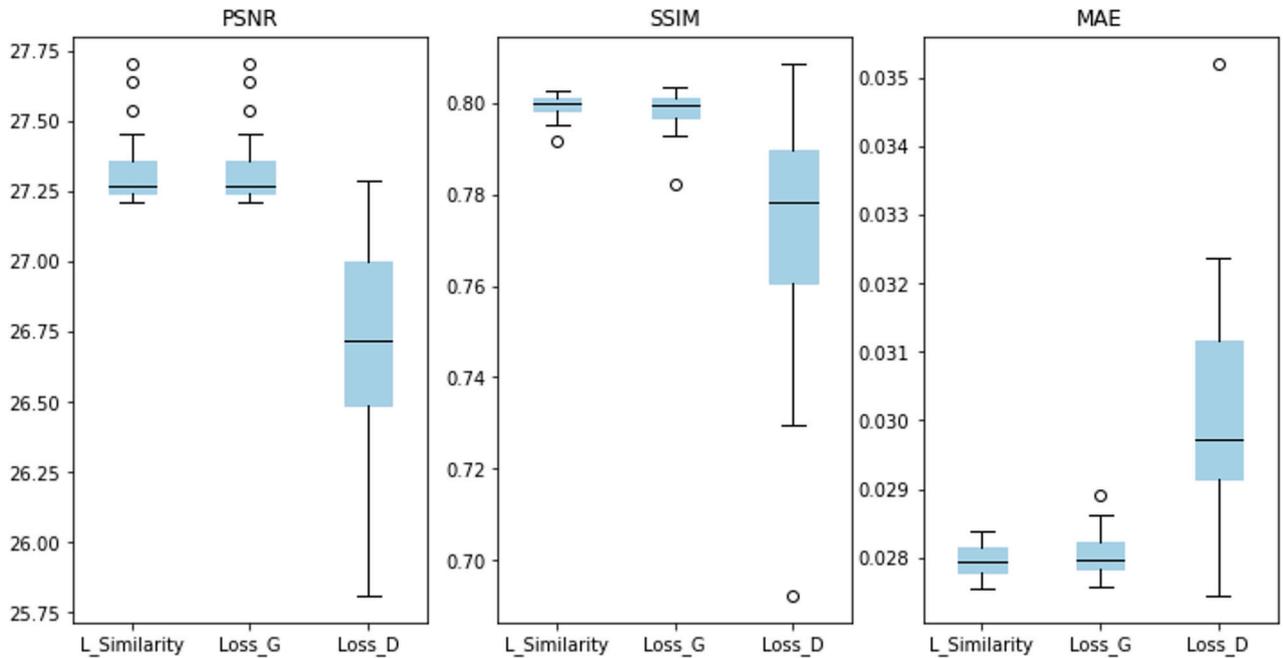


Fig. 11. Comparing the impact of different stop criterion on the quality of synthesized PETs.

CycleGAN-PlainG also performs well. Regarding SSIM distribution, CycleGAN-MultiG reveals a longer but very thin tail for smaller SSIM values. Assuming this extended tail is not attributed to an outlier, there are relatively fewer instances where the SSIM is lower than CycleGAN-PlainG. Conversely, it displays a substantial concentration of data points with higher SSIM values, which is desirable. Overall, these findings suggest that CycleGAN-MultiG generally outperforms CycleGAN-PlainG in terms of SSIM. However, it is worth mentioning that due to a larger number of MACs and slightly more trainable parameters, selecting the optimal choice among them needs an investigation into the trade-off between computational cost and the quality of generated images.

5.2. Loss function ablation study

In the second ablation study, we investigate the impact of the proposed similarity loss term by setting its weight in Eq. 10 to zero while maintaining other weights, parameters, and implementation details. This transforms the loss function into a standard CycleGAN loss comprising adversarial, identity, and cycle-consistency components. Fig. 10 depicts the average similarity metric values for the unseen test set. The inclusion of the similarity term noticeably influences the quality of synthesized PETs across all metrics. Specifically, the average PSNR for the test set increases by 8.2% from 25.19 to 27.27, SSIM improves by 9.8% from 0.728 to 0.800, and MAE decreases by 17.01% from 0.0335

Table 7

P-values of testing the hypothesis if there is a difference in the quality of synthesized PETs using different stop criterion (statistically significant at 5% confidence level are **bold**).

	L_{Sim} vs. L_G	L_{Sim} vs. L_D	L_G vs. L_D
PSNR	0.36827	5.722e-06	1.907e-06
SSIM	0.31179	2.670e-05	1.335e-4
MAE	0.08969	1.335e-4	4.768e-05

to 0.0278. These results underscore the significant contribution of the proposed loss function to enhancing the quality of synthesized PETs, indicating a substantial impact on the overall framework performance.

5.3. Stopping criterion

Training GANs is inherently challenging due to the dynamic competition between generators and discriminators. The delicate balance between the two is crucial for realistic data generation, making it difficult to establish a clear stopping point for training. While visual inspection suffices for non-medical applications, it becomes impractical for medical imaging. Most studies commonly employ a maximum epoch number as a stopping criterion, as observed in [6]. Introducing novel metrics to be used as a stop criterion is also considered in the literature, especially for segmentation [87–89]. To understand how different stop criteria impact the quality of synthesized images in the current study, three stop criteria, i.e., discriminator loss, generator loss, and generator similarity loss, are investigated. Fig. 11 presents different similarity measures for the unseen test set. It is worth noting that for CycleGAN, the corresponding losses of G_{PET} and D_{PET} are considered in this investigation.

Based on these results, L_D appears to be less favorable as a stop criterion. However, to check for a statistically significant difference Wilcoxon rank-signed test is used. The corresponding p-values of this test are given in Table 7. According to this table, for all similarity measures (PSNR, SSIM, and MAE), we can reject the null hypothesis at a confidence level of 5% for comparison of L_{Sim} vs. L_D and L_G vs. L_D . So, using L_{Sim} and L_G over L_{Sim} as a stop criterion would make a difference. On the other hand, based on the corresponding p-values, the null hypothesis of L_{Sim} vs. L_G cannot be rejected for any of the similarity

measures, suggesting no statistically significant difference in the synthesized image's quality. In the current study, L_{Sim} is used for our experiments in this study.

5.4. Normalization layer

Normalization layers are the most widely used components in deep networks, and utilizing them can accelerate the training and boost performance. This normalization can be applied at batch, channel, or instance levels [90]. While batch normalization is a common choice in many studies [25,42,91], it has been shown that batch normalization can induce vulnerability to models and decrease the model robustness, especially in medical images [92]. To investigate the impact of normalization type, we replaced all Instance Normalization layers with Batch Normalization ones in the proposed framework, resulting in the box plots illustrated in Fig. 12. Although this plot represents the superior performance of Instance Normalization, a statistical test can confirm it. To do so, two Wilcoxon rank-signed tests are conducted, and their p-values are gathered in Table 8. According to this table, the null hypothesis, Instance Normalization equals Batch Normalization, would be rejected at a confidence level of 5%. Based on the p-values provided in the second column, the null hypothesis can be rejected at a confidence level of 5% in favor of the alternative hypothesis. So, Instance Normalization works better than Batch Normalization regarding all the similarity measures, i.e., PSNR, SSIM, and MAE.

5.5. Activation function

The rectified linear activation function (ReLU) is a function that outputs the input if it is positive and zero in case it is negative. A

Table 8

P-values of testing the hypothesis corresponds to employed normalization type.

Similarity Measure	H0: Instance equals Batch H1: Instance Not equals Batch	H0: Instance equals Batch H1: Instance Better than Batch
PSNR	1.301e-4	2.613e-4
SSIM	6.675e-6	1.335e-5
MAE	9.536e-7	1.907e-6

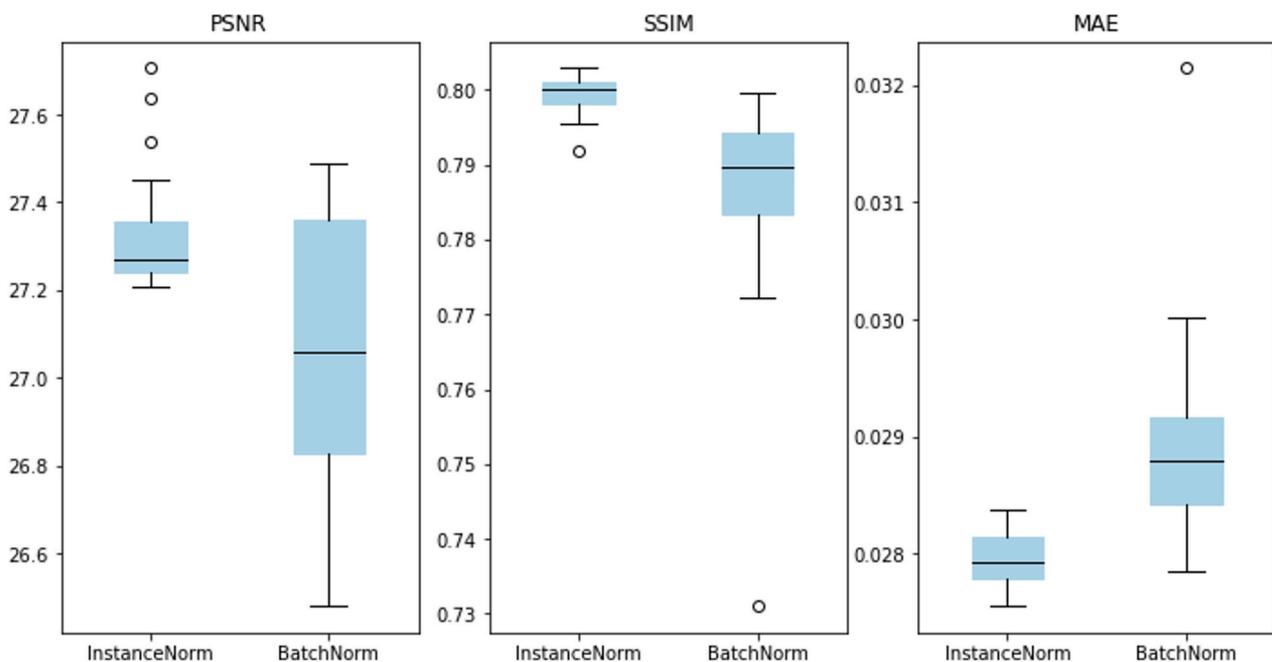


Fig. 12. Comparing the impact of employed normalization type on the quality of synthesized PETs.

Table 9

Mean and standard deviation of the similarity measures corresponds to different activation functions (best in **bold**).

Activation Function	PSNR (mean \pm std)	SSIM (mean \pm std)	MAE (mean \pm std)
PReLU	27.3337 \pm 0.1426	0.79936 \pm 0.0027	0.02791 \pm 0.0002
LeakyReLU(0.2)	27.2719 \pm 0.0857	0.79820 \pm 0.0022	0.02800 \pm 0.0001
ReLU	27.3167 \pm 0.1363	0.79910 \pm 0.0025	0.02795 \pm 0.0002

modified ReLU called LeakyReLU allows small negative values by introducing a slight slope that should be tuned. To skip the tuning step, Parametric ReLU (PReLU), given in eq. 14, was introduced [93] in which the slope is learned through training the deep network. Despite the introduction of PReLU in 2015, almost all of the employed deep architectures, especially for medical applications, use the LeakyReLU activation function, frequently with a slope equal to 0.2. In the few studies utilizing PReLU, performance improvement compared to ReLU or LeakyReLU has been obtained [94]. To assess the impact of activation functions on synthesized PET quality, we replaced all PReLU activations in our proposed frameworks first with LeakyReLU(0.2) and then with ReLU. Mean and standard deviation values for PSNR, SSIM, and MAE are

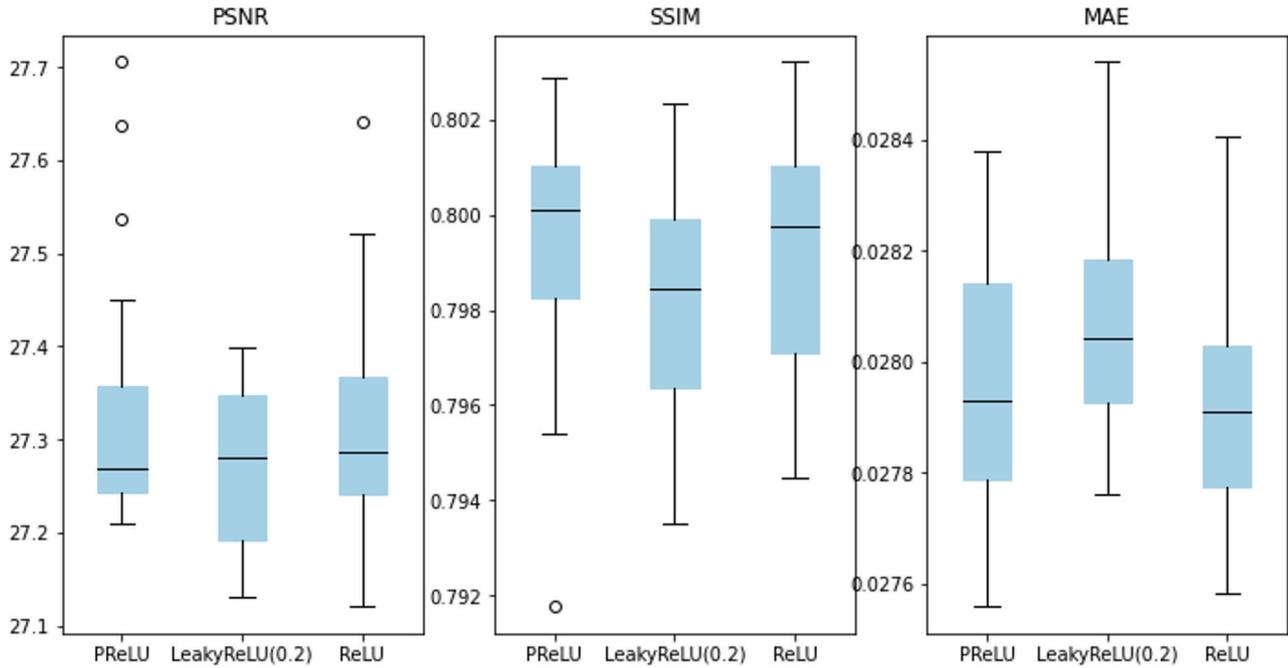


Fig. 13. Comparing the impact of activation function on the quality of synthesized PETs.

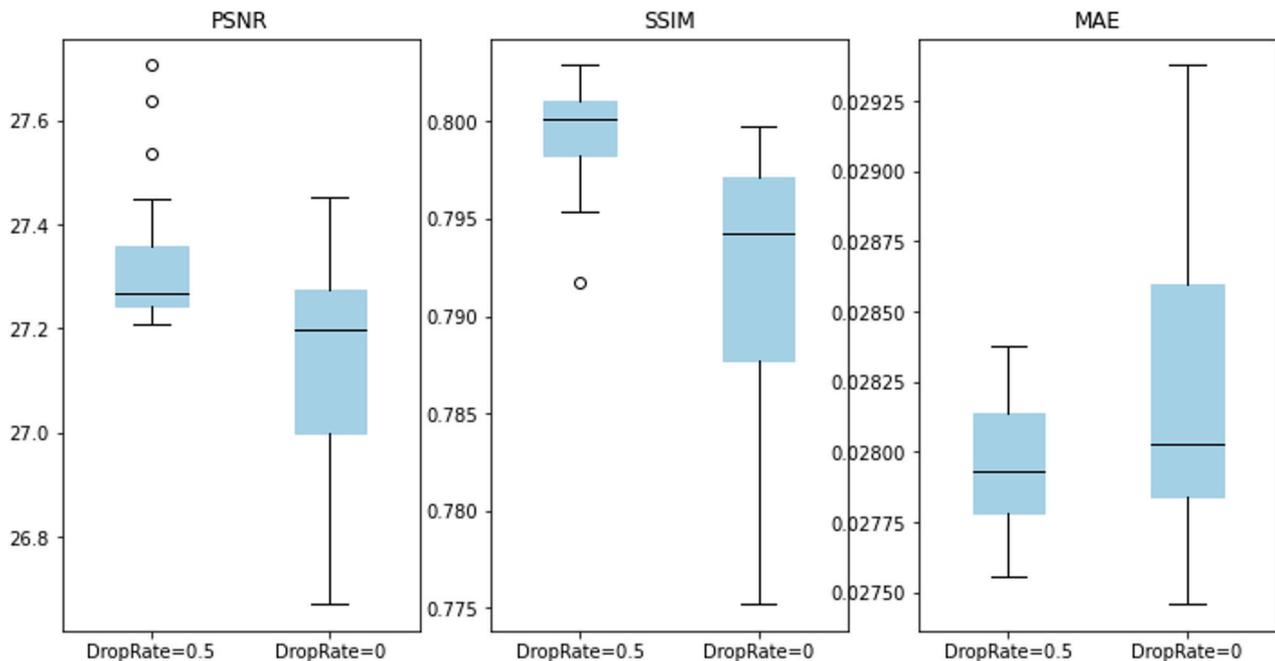


Fig. 14. Comparing the impact of dropout on the quality of synthesized PETs.

presented in Table 9. According to this table, PReLU outperformed LeakyReLU(0.2) and ReLU, although ReLU resulted in very close values to PReLU. Fig. 13 illustrates these results, indicating the optimum slope's value obtained during training is smaller than 0.2 but very close to zero, which makes it almost equivalent to ReLU. Consequently, utilizing PReLU over LeakyReLU or ReLU is crucial for the framework to optimize this negative slope effectively based on the data.

$$PReLU(x) = \max(0, x) + a \times \min(0, x) \quad (14)$$

5.6. Dropout

Over-fitting is a common pitfall in which AI models capture noise or superficial information rather than genuinely distinguishing features [95]. Data scarcity can increase the risk of over-fitting due to the huge number of network parameters compared to the number of training data [54], especially in medical image applications with 3D inputs and limited labeled data. Dropout, introduced in [96], is one of the known methods to prevent over-fitting. The key idea behind dropout is randomly dropping some neurons and their corresponding connections during training to avoid excessive co-adaptation [96].

In the current study with a limited number of paired MRI-PET samples, we employed several dropouts in the proposed framework. For the generator, the multi-scale blocks have an embedded dropout layer, and we also apply dropout after pooling operations. For the discriminator, FC layers are followed by dropouts in addition to the mentioned dropouts for the generator. A common dropout rate of 0.5 [28] is used in our study. To investigate the impact of dropout on synthesized PET quality, we repeat our experiments by setting this rate equal to zero, indicating no dropout usage in the entire framework. Fig. 14 illustrates the corresponding numerical results for PSNR, SSIM, and MAE metrics. According to this figure, applying a drop rate of 0.5 can increase the quality of synthesized PETs. To check for a statistically significant difference in employing dropout, we conducted a Wilcoxon rank-signed test, hypothesizing that dropout improves results (higher PSNR and SSIM, lower MAE) compared to no dropout. The corresponding p-values testing this hypothesis are $2.24e-4$, $8.39e-5$, and 0.041 for PSNR, SSIM, and MAE, respectively. So, the null hypothesis of having equal medians will be rejected, supporting the conclusion that applying dropouts for multi-scale blocks, FC layers, and after pooling operations in this study leads to improved synthesized PET quality.

6. Conclusion

This Study proposed a 3D multi-scale image-to-image CycleGAN framework to generate synthesized PETs from MRIs to fill missing PETs and help better diagnose AD using real MRIs and synthesized PETs. Feeding this framework by different scales of the input MRI, employing convolution blocks having two consecutive convolution operations, applying multiple dropouts, and using skip connections, enabled this framework to reach an effective mapping between MRI and PET domains. In addition, to push the framework to generate more similar and realistic functional PETs from structural MRIs, a hybrid loss function was proposed to enforce the structural similarity while preserving the voxel-wise similarity and avoiding generating blurry images. The quality assessments of these PET quantitatively and visually confirmed the superior performance of the proposed framework compared to the state-of-the-art methods in the literature. Moreover, according to the numerical results obtained for the ternary classification of AD subjects (AD vs. MCI vs. NC), these synthesized PETs can help improve AD diagnosis. Specifically, assuming an extreme case where none of the subjects has a PET, feeding the classifier with MRIs and their corresponding synthesized PETs reached a more accurate diagnosis than feeding the classifier with just available MRIs. Conducted ablation investigation regarding the proposed GAN framework as well as the proposed loss function unravels their contribution to the quality of generated PETs. Accordingly, it was

shown that using a stopping criterion based on the generator's loss over the discriminator's, applying instance normalization instead of batch normalization, utilizing parametric ReLU rather than LeakyReLU or ReLU, and appropriate use of dropout can considerably boost the quality of synthesized PETs.

While the primary application focuses on synthesizing PET images from MRIs for AD diagnosis, the design of the framework and the modified loss function in the current study have the potential to extend to some other medical imaging domain translation scenarios due to some shared patterns and characteristics across different modalities. Although the outstanding performance of this framework for a complex task of mapping structural to functional domain shows its potential for some less complicated tasks such as low-dose to high-dose mapping of PETs, low-resolution to high-resolution translation of CTs, or converting different structural MRIs to each other, extended experiments are needed to fully validate this potential. So, a future research direction is employing this framework for other types of image translation to assess its performance for other tasks. On the other hand, since employing the proposed loss term to measure the similarity of translated images to their ground truth caused a considerable quality improvement according to the provided discussion, it seems that utilizing it along with common CycleGAN losses, adversarial loss, cycle-consistency loss, and identity loss, would boost the performance of CycleGANs for different applications. As a future study, it is of great interest to investigate the impact of this proposed loss function on the performance of CycleGAN-based frameworks previously studied in the literature. In the current study, we faced computational limitations for embedding a classifier in the proposed GAN framework. However, developing the current framework into a more comprehensive one, integrating missing modality generation and diagnosis can be another direction for future studies.

CRedit authorship contribution statement

M. Khojaste-Sarakhsi: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Seyedhamidreza Shahabi Haghighi:** Supervision, Project administration, Investigation. **S.M.T. Fatemi Ghomi:** Supervision, Investigation. **Elena Marchiori:** Supervision, Investigation.

Declaration of competing interest

None

Data availability

The authors do not have permission to share data.

References

- [1] F.R. Farina, D.D. Emek-Savaş, L. Rueda-Delgado, R. Boyle, H. Kiiski, G. Yener, R. Whelan, A comparison of resting state EEG and structural MRI for classifying Alzheimer's disease and mild cognitive impairment, *NeuroImage* 215 (April) (2020), 116795, 7.
- [2] Matej Mihelčić, Goran Šimić, Mirjana Babić Leko, Nada Lavrač, Sašo Žderoski, Tomislav Šmuc, Using redescription mining to relate clinical and biological characteristics of cognitively impaired and Alzheimer's disease patients, *PLoS One* 12 (10) (2017) 10, e0187364.
- [3] Alzheimer's Association, Alzheimer's disease facts and figures, *Alzheimers Dement.* 14 (3) (2018) 367–429, 3 2018.
- [4] Ane Alberdi, Asier Aztiria, Adrian Basarab, On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey, 7, 2016.
- [5] Mengyi Zhang, Lijing Sun, Zhaokai Kong, Wenjun Zhu, Yang Yi, Fei Yan, Pyramid-attentive GAN for multimodal brain image complementation in Alzheimer's disease classification, *Biomed. Sign. Process. Control* 89 (2024) 3.
- [6] Jin Zhang, Xiaohai He, Linbo Qing, Feng Gao, Bin Wang, BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal Alzheimer's disease diagnosis, *Comput. Methods Prog. Biomed.* 217 (2022) 4.
- [7] V.P. Subramanyam Rallabandi, Krishnamoorthy Seetharaman, Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging, *Biomed. Sign. Process. Control* 80 (2023), 104312, 2.

- [8] Fangyu Liu, Shizhong Yuan, Weimin Li, Xu Qun, Bin Sheng, Patch-based deep multi-modal learning framework for Alzheimer's disease diagnosis using multi-view neuroimaging, *Biomed. Sign. Process. Control* 80 (2023), 104400, 2.
- [9] Chenhui Wang, Sirong Piao, Zhizhong Huang, Qi Gao, Junping Zhang, Yuxin Li, Hongming Shan, Joint learning framework of cross-modal synthesis and diagnosis for Alzheimer's disease by mining underlying shared modality information, *Med. Image Anal.* 91 (2024) 1.
- [10] Shijie Chen, Xin Tian, Yuling Wang, Yunfeng Song, Ying Zhang, Jie Zhao, Jyh Cheng Chen, DAEGAN: Generative adversarial network based on dual-domain attention-enhanced encoder-decoder for low-dose PET imaging, *Biomed. Sign. Process. Control* 86 (2023) 9.
- [11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Xu Bing, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, *Adv. Neural Inf. Proces. Syst.* 3 (January) (2014) 2672–2680, 6.
- [12] Kh. Sanuwani Dayarathna, Tohidul Islam, Sergio Uribe, Guang Yang, Munawar Hayat, Zhaolin Chen, Deep learning based synthesis of MRI, CT and PET: Review analysis 2 (2024).
- [13] Fernando Vega, Abdoljalil Addeh, Aravind Ganesh, Eric E. Smith, M. Ethan MacDonald, Image translation for estimating two-dimensional axial amyloid-beta PET from structural MRI, *J. Magn. Reson. Imaging* 3 (2023).
- [14] Jake McNaughton, Justin Fernandez, Samantha Holdsworth, Benjamin Chong, Vickie Shim, Alan Wang, Machine Learning for Medical Image Translation: A Systematic Review 9, 2023.
- [15] André Ferreira, Jianning Li, Kelsey L. Pomykala, Jens Kleesiek, Victor Alves, Jan Egger, GAN-based generation of realistic 3D volumetric data: a systematic review and taxonomy, *Med. Image Anal.* 93 (2024), 103100, 4.
- [16] Yan Wang, Luping Zhou, Yu Biting, Lei Wang, Zu Chen, David S. Lalush, Weili Lin, Wu Xi, Jiliu Zhou, Dinggang Shen, 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis, *IEEE Trans. Med. Imaging* 38 (6) (2019) 1328–1339, 6.
- [17] Yunbi Liu, Ling Yue, Shifu Xiao, Wei Yang, Dinggang Shen, Mingxia Liu, Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages, *Med. Image Anal.* 75 (2022) 1.
- [18] Yan Jin, Jonathan DuBois, Chongyue Zhao, Liang Zhan, Audrey Gabelle, Neda Jahanshad, Paul M. Thompson, Arie Gafson, Shibeshih Belachew, Brain MRI to PET Synthesis and Amyloid Estimation in Alzheimer's Disease via 3D Multimodal Contrastive GAN, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 14348 LNCS, Springer Science and Business Media Deutschland GmbH, 2024, pp. 94–103.
- [19] Tu Yue, Shukuan Lin, Jianzhong Qiao, Yilin Zhuang, Zhiqi Wang, Dai Wang, Multimodal fusion diagnosis of Alzheimer's disease based on FDG-PET generation, *Biomed. Sign. Process. Control* 89 (2024) 3.
- [20] Wanyun Liu, Weiming Lin, Gang Chen, Hejun Zhang, Qinquan Gao, Yechong Huang, Tong Tong, Du. Min, Bidirectional mapping of brain MRI and PET With 3D reversible GAN for the diagnosis of Alzheimer's disease, *Front. Neurosci.* 15 (2021) 4.
- [21] Jiwoong J. Jeong, Amara Tariq, Tobiloba Adejumo, Hari Trivedi, Judy W. Gichoya, Imon Banerjee, Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation, *J. Digit. Imaging* 1 (2022).
- [22] Salome Kazemina, Christoph Baur, Arjan Kuijper, Bram van Ginneken, Nassir Navab, Shadi Albarqouni, Anirban Mukhopadhyay, GANs for medical image analysis, *Artif. Intell. Med.* 109 (October 2019) (2020), 101938, 9.
- [23] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, Hayit Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, *Neurocomputing* 321 (2018) 321–331, 12.
- [24] Chenyu You, Wenxiang Cong, Ge Wang, Qingsong Yang, Hongming Shan, Lars Gjestebj, Guang Li, Ju Shenghong, Zhuizhang Zhang, Zhen Zhao, Yi Zhang, Structurally-Sensitive Multi-Scale Deep Neural Network for Low-Dose CT Denoising, *IEEE Access* 6 (2018) 41839–41855, 7.
- [25] Xiao Zhou, Shangran Qiu, Prajakta S. Joshi, Chonghua Xue, Ronald J. Killiany, Asim Z. Mian, Sang P. Chin, Au Rhoda, Vijaya B. Kolachalama, Enhancing magnetic resonance imaging-driven Alzheimer's disease classification performance using generative adversarial learning, *Alzheimer's Res. Therapy* 13 (1) (2021) 12.
- [26] Chenyu You, Wenxiang Cong, Michael W. Vannier, Punam K. Saha, Eric A. Hoffman, Ge Wang, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Ju Shenghong, Zhen Zhao, Zhuizhang Zhang, CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE), *IEEE Trans. Med. Imaging* 39 (1) (2020) 188–203, 1.
- [27] Seung Kwan Kang, Hongyoon Choi, Jae Sung Lee, Translating amyloid PET of different radiotracers by a deep generative model for interchangeability, *NeuroImage* 232 (2021) 5.
- [28] Yan Wang, Yu Biting, Lei Wang, Zu Chen, David S. Lalush, Weili Lin, Wu Xi, Jiliu Zhou, Dinggang Shen, Luping Zhou, 3D conditional generative adversarial networks for high-quality PET image estimation at low dose, *NeuroImage* 174 (March) (2018) 550–562.
- [29] Nie Dong, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, Dinggang Shen, Medical Image synthesis with context-aware generative adversarial networks, in: Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, Simon Duchesne (Eds.), *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2017, pp. 417–425.
- [30] Jue Jiang, Hu Yu-Chi, Neelam Tyagi, Pengpeng Zhang, Andreas Rimmer, Gig S. Mageras, Joseph O. Deasy, Harini Veeraraghavan, Tumor-Aware, Adversarial Domain Adaptation from CT to MRI for Lung Cancer Segmentation, in: volume 11071 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2018, pp. 777–785.
- [31] Ying Chen, Hongping Lin, Wei Zhang, Wang Chen, Zonglai Zhou, Ali Asghar Heidari, Huihui Chen, Xu. Guohui, ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation, *Biomed. Sign. Process. Control* 92 (2024), 106100, 6.
- [32] Apoorva Sikka, Skand, Jitender Singh Virk, Deepti R. Bathula, MRI to PET Cross-Modality Translation using Globally and Locally Aware GAN (GLA-GAN) for Multi-Modal Diagnosis of Alzheimer's Disease 8, 2021.
- [33] Avi Ben-Cohen, Eyal Klang, Stephen P. Raskin, Shelly Soffer, Simona Ben-Haim, Eli Konen, Michal Marianne Amitai, Hayit Greenspan, Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection, *Eng. Appl. Artif. Intell.* 78 (November 2018) (2019) 186–194.
- [34] Lei Bi, Jinman Kim, Ashnil Kumar, Dagan Feng, Michael Fulham, Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs), in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, pp. 43–51, volume 10555 LNCS.
- [35] Mehdi Mirza, Simon Osindero, Conditional Generative Adversarial Nets 11, 2014.
- [36] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks 11, 2016.
- [37] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired Image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 10 2017, pp. 2242–2251, volume 2017-October.
- [38] Bo Zhan, Luping Zhou, Zhiang Li, Wu Xi, Pu Yifei, Jiliu Zhou, Yan Wang, Dinggang Shen, D2FE-GAN: Decoupled dual feature extraction based GAN for MRI image synthesis, *Knowl.-Based Syst.* 252 (2022) 9.
- [39] Jyoti Islam, Yanqing Zhang, GAN-based synthetic brain PET image generation, *Brain Inform.* 7 (1) (2020) 12.
- [40] Hajar Emami, Ming Dong, Siamak P. Nejad-Davarani, Carri K. Glide-Hurst, Generating synthetic CTs from magnetic resonance images using generative adversarial networks, *Med. Phys.* 45 (8) (2018) 3627–3636, 8.
- [41] Euijin Jung, Miguel Luna, Sang Hyun Park, Conditional GAN with 3D discriminator for MRI generation of Alzheimer's disease progression, *Pattern Recogn.* 133 (2023) 1.
- [42] Farideh Bazangani, Frédéric J.P. Richard, Badih Ghattas, Eric Guedj, FDG-PET to T1 weighted MRI translation with 3D elicit generative adversarial network (E-GAN), *Sensors* 22 (12) (2022), 6.
- [43] Reda Oulbache, Samuel Kadoury, MRI to CT Synthesis of the Lumbar Spine from a Pseudo-3D Cycle GAN, in: *Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society*, 2020, pp. 1784–1787, volume 2020-April. 4.
- [44] Hristina Uzunova, Jan Ehrhardt, Fabian Jacob, Alex Frydrychowicz, Heinz Handels, Multi-scale GANs for Memory-efficient Generation of High Resolution Medical Images, in: volume 11769 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2019, pp. 112–120.
- [45] Jaeyoung Huh, Shujaat Khan, Sungjin Choi, Dongkuk Shin, Jeong Eun Lee, Eun Sun Lee, Jong Chul Ye, Tunable image quality control of 3-D ultrasound using switchable CycleGAN, *Med. Image Anal.* 83 (2023), 102651, 1.
- [46] K.T. Chen, R. Tesfay, M.E.L. Koran, J. Ouyang, S. Shams, C.B. Young, G. Davidzon, T. Liang, M. Khalighi, E. Mormino, G. Zaharchuk, Generative adversarial network-enhanced ultra-low-dose [18F]-PI-2620 s PET/MRI in aging and neurodegenerative populations, *Am. J. Neuroradiol.* 44 (9) (2023) 1012–1020, 9.
- [47] Yan Wang, Luping Zhou, Lei Wang, Biting Yu, Zu Chen, David S. Lalush, Weili Lin, Xi Wu, Jiliu Zhou, Dinggang Shen, Locality adaptive multi-modality GANs for high-quality PET image synthesis, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, pp. 329–337, volume 11070 LNCS.
- [48] Yan Wang, Yanmei Luo, Zu Chen, Bo Zhan, Zhengyang Jiao, Wu Xi, Jiliu Zhou, Dinggang Shen, Luping Zhou, 3D multi-modality Transformer-GAN for high-quality PET reconstruction, *Med. Image Anal.* 91 (2024) 1.
- [49] Yongsheng Pan, Mingxia Liu, Chunfeng Lian, Tao Zhou, Yong Xia, Dinggang Shen, Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis, in: *MICCAI (oral segmentation) volume 1*, Springer International Publishing, 2018, pp. 455–463.
- [50] Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Nicholas Ayache, Bruno Stankoff, Olivier Colliot, Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training, in: *MICCAI (oral segmentation) vol. 3*, Springer International Publishing, 2018, pp. 514–522.
- [51] Apoorva Sikka, Skand Vishwanath Peri, Deepti R. Bathula, MRI to FDG-PET: Cross-modal synthesis using 3d u-net for multi-modal alzheimer's classification, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11037 LNCS, Springer Verlag, 2018, pp. 80–89.
- [52] Hu Shengye, Baiying Lei, Shuqiang Wang, Yong Wang, Zhiguang Feng, Yanyan Shen, Bidirectional mapping generative adversarial networks for brain MRI to PET synthesis, *IEEE Trans. Med. Imaging* 41 (1) (2022) 145–157, 8.
- [53] Rami Hussein, David Shin, Moss Y. Zhao, Jia Guo, Guido Davidzon, Gary Steinberg, Michael Moseley, Greg Zaharchuk, Turning brain MRI into diagnostic PET: 150-water PET CBF synthesis from multi-contrast MRI via attention-based encoder-decoder networks, *Med. Image Anal.* 103072 (2023) 4.
- [54] Su Tongxue Zhou, Ruan, and Stéphane Canu., A review: deep learning for medical image segmentation using multi-modality fusion, *Array* 100004 (9) (2019) 3–4.
- [55] Daria Zotova, Julien Jung, Carole Lartizien, GAN-based synthetic FDG PET images from T1 brain MRI can serve to improve performance of deep unsupervised anomaly detection models, in: *Lecture Notes in Computer Science (including subseries*

- Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), volume 12965 LNCS, Springer Science and Business Media Deutschland GmbH, 2021, pp. 142–152.
- [56] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: convolutional networks for biomedical image segmentation, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* volume 9351, Springer Verlag, 2015, pp. 234–241.
- [57] Jiaming Luo, Yongzhe Tang, Jie Wang, Lu Hongtao, USMLP: U-shaped Sparse-MLP network for mass segmentation in mammograms, *Image Vis. Comput.* 137 (2023), 104761, 9.
- [58] Song-Toan Tran, Ching-Hwa Cheng, Thanh-Tuan Nguyen, Minh-Hai Le, Don-Gey Liu, TMD-Unet: Triple-Unet with multi-scale input features and dense skip connection for medical image segmentation, *Healthcare* 9 (1) (2021), 54, 1.
- [59] Bing Huang, Feng Yang, Mengxiao Yin, Xiaoying Mo, Cheng Zhong, A review of multimodal medical image fusion techniques, *Comput. Math. Methods Med.* 1–16 (4) (2020) 2020.
- [60] Md Zongwei Zhou, Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, UNet++: redesigning skip connections to exploit multiscale features in image segmentation, *IEEE Trans. Med. Imaging* 39 (6) (2020) 1856–1867, 6.
- [61] Ahmed Awad Albishri, Syed Jawad Hussain Shah, Yugyung Lee, CU-Net: cascaded U-net model for automated liver and lesion segmentation and summarization, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 11 2019, pp. 1416–1423.
- [62] Yilong Chen, Kai Wang, Xiangyun Liao, Yinling Qian, Qiong Wang, Zhiyong Yuan, Pheng-Ann Heng, Channel-Unet: a spatial channel-wise convolutional neural network for liver and tumors segmentation, *Front. Genet.* 10 (2019) 11.
- [63] Anda Tang, Pei Quan, Lingfeng Niu, Yong Shi, A Survey for Sparse Regularization Based Compression Methods, 8, 2022.
- [64] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, Kaori Togashi, Convolutional neural networks: an overview and application in radiology, *Insights into Imag.* 9 (4) (2018) 611–629, 8.
- [65] Jake Snell, Karl Ridgeway, Renjie Liao, Brett D. Roads, Michael C. Mozer, Richard S. Zemel, Learning to generate images with perceptual similarity metrics, in: 2017 IEEE International Conference on Image Processing (ICIP) 9, IEEE, 2017, pp. 4277–4281.
- [66] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment, in: *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, IEEE, 2003, pp. 1398–1402.
- [67] S. Gayathri Devi, K. Selvam, S.P. Rajagopalan, An abstract to calculate big o factors of time and space complexity of machine code, *IET Conference Publications* 2011 (583 CP) (2011) 844–847.
- [68] Abhinav K. Venkataraman, Wu Chengyang, Alan C. Bovik, Ioannis Katsavounidis, Zafar Shahid, A Hitchhiker's guide to structural similarity, *IEEE Access* 9 (2021) 28872–28896.
- [69] Simeon Spasov, Luca Passamonti, Andrea Duggento, Pietro Liò, Nicola Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, *NeuroImage* 189 (August 2018) (2019) 276–287.
- [70] Daoqiang Zhang, Dinggang Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, *NeuroImage* 59 (2) (2012) 895–907, 1.
- [71] Francisco J. Martinez-Murcia, Andres Ortiz, Juan Manuel Gorriz, Javier Ramirez, Diego Castillo-Barnes, Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders, *IEEE J. Biomed. Health Inform.* 24 (1) (2020) 17–26.
- [72] Matthias S. Treder, Ryan Codrai, Kamen A. Tsvetanov, Quality assessment of anatomical MRI images from generative adversarial networks: Human assessment and image quality metrics, *J. Neurosci. Methods* 374 (2022) 5.
- [73] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [74] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A. Efros, Oliver Wang, Eli Shechtman, Toward Multimodal Image-to-Image Translation 11, 2017.
- [75] Janez Demsar, Statistical Comparisons of Classifiers over Multiple Data Sets, Technical report, 2006.
- [76] Harrison H. Barrett, Kyle J. Myers, Christoph Hoeschen, Matthew A. Kupinski, Mark P. Little, Task-based measures of image quality and their relation to radiation dose and patient risk, *Phys. Med. Biol.* 60 (2) (2015) R1–R75, 1.
- [77] Xiaohui Zhang, Varun A. Kelkar, Jason Granstedt, Hua Li, Mark A. Anastasio, Impact of deep learning-based image super-resolution on binary signal detection, *J. Med. Imag.* 8 (06) (2021) 11.
- [78] Joseph Paul Cohen, Margaux Luck, Sina Honari, Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. 5, 2018.
- [79] Varun A. Kelkar, Dimitrios S. Gotsis, Frank J. Brooks, K.C. Prabhat, Kyle J. Myers, Rongping Zeng, Mark A. Anastasio, Assessing the ability of generative adversarial networks to learn canonical medical image statistics, *IEEE Trans. Med. Imaging* 42 (6) (2023) 1799–1808, 6.
- [80] M. Khojaste-Sarakhsi, Seyedhamidreza Shahabi Haghighi, S. M.T. Fatemi Ghomi, and Elena Marchiori. Deep learning for Alzheimer's disease diagnosis: A survey 8, 2022.
- [81] Shuangshuang Gao, Dimas Lima, A review of the application of deep learning in the detection of Alzheimer's disease, *Int. J. Cogn. Comput. Eng.* 3 (2022) 1–8, 6.
- [82] Alaa Tharwat, Classification assessment methods, *Appl. Comput. Inform.* 17 (1) (2021) 168–192, 1.
- [83] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 6, 2016.
- [84] T. Venkat, Narayana Rao, Assessment of Diverse Quality Metrics for Medical Images including Mammography. Technical Report 11, 2014.
- [85] Li Sze Chow and Raveendran Paramesran, Review of medical image quality assessment, *Biomed. Sign. Process. Control* 27 (2016) 145–154, 5.
- [86] Junqing Liu, Lei Ma, Jianfeng He, Qian Zhang, Honglei Chen, Lipeng Pan, A comparative study of assessment methods for medical image quality, in: 2012 5th International Conference on BioMedical Engineering and Informatics, IEEE, 10 2012, pp. 131–134.
- [87] Muhammad Nadeem Cheema, Anam Nazir, Po Yang, Bin Sheng, Ping Li, Huating Li, Xiaoe Wei, Jing Qin, Jinman Kim, David Dagan Feng, Modified GAN-CAED to minimize risk of unintentional liver major vessels cutting by controlled segmentation using CTA/SPET-CT, *IEEE Trans. Industrial Inform.* 17 (12) (2021) 7991–8002, 12.
- [88] Pedro P. Rebouças Filho, Antonio C. da Silva Barros, Jefferson S. Almeida, J.P. C. Rodrigues, Victor Hugo C. de Albuquerque, A new effective and powerful medical image segmentation algorithm based on optimum path snakes, *Appl. Soft Comput.* 76 (2019) 649–670, 3.
- [89] Daniel Franco-Barranco, Julio Pastor-Tronch, Aitor González-Marfil, Arrate Muñoz-Barrutia, Ignacio Arganda-Carreras, Deep learning based domain adaptation for mitochondria segmentation on EM volumes, *Comput. Methods Prog. Biomed.* 222 (2022) 7.
- [90] Xiao Yun Zhou, Peichao Li, Zhao Yang Wang, Guang Zhong Yang, U-net training with instance-layer normalization, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11977 LNCS, Springer, 2020, pp. 101–108.
- [91] Wei Huang, Mingyuan Luo, Jing Li, Peng Zhang, Yufei Zha, A novel locally-constrained GAN-based ensemble to synthesize arterial spin labeling images, *Inf. Sci.* 609 (2022) 691–710, 9.
- [92] Fei Kong, Fangqi Liu, Kaidi Xu, Xiaoshuang Shi, Why does batch normalization induce the model vulnerability on adversarial images? *World Wide Web* 26 (2023) 1073–1091, 3.
- [93] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, in: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [94] Yu-Dong Zhang, Chichun Pan, Junding Sun, Chaosheng Tang, Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU, *J. Comput. Sci.* 28 (2018) 1–10, 9.
- [95] Simukayi Mutasa, Shawn Sun, Richard Ha, Understanding artificial intelligence based radiology studies: What is overfitting?, 9, 2020.
- [96] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ruslan Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Technical report, 2014.