



# A new VAE-GAN model to synthesize arterial spin labeling images from structural MRI

Feihong Li <sup>a,b</sup>, Wei Huang <sup>a,b,\*</sup>, Mingyuan Luo <sup>c</sup>, Peng Zhang <sup>d</sup>, Yufei Zha <sup>d</sup>

<sup>a</sup> Department of Computer Science, School of Information Engineering, Nanchang University, Nanchang 330022, China

<sup>b</sup> China Mobile-NCU AI&IOT Jointed Lab, Informatization Office, Nanchang University, Nanchang 330022, China

<sup>c</sup> Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518061, China

<sup>d</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

## ARTICLE INFO

### MSC:

41A05  
41A10  
65D05  
65D17

### Keywords:

Image synthesis  
Generative adversarial network  
Variational auto-encoder

## ABSTRACT

Arterial spin labeling (ASL) is a relatively new MRI technique that can measure cerebral blood flow, which is of great importance for the diagnosis of dementia diseases. Besides, this valuable imaging modality does not need exogenous tracers and has no radiation, which makes it favorable for elder patients. However, ASL data does lack in many contemporary image-based dementia diseases datasets, which include popular ADNI-1/GO/2/3 datasets. In order to supplement the valuable ASL data, a new Generative adversarial network (GAN)-based model is proposed to synthesize ASL images in this study. This new model is unique, as the popular variational auto-encoder (VAE) has been utilized as the generator of the GAN-based model. Hence, a new VAE-GAN architecture is introduced in this study. In order to demonstrate its superiority, dozens of experiments have been conducted. Experimental results demonstrate that, this new VAE-GAN model is superior to other state-of-the-art ASL image synthesis methods, and the accuracy improvement after incorporating synthesized ASL images from the new model can be as high as 42.41% in dementia diagnosis tasks.

## 1. Introduction

Dementia denotes a more and more commonly seen type of disease in the elderly, and its symptoms include but not limit to forgetfulness, glassy-eyed, dullness of speech, etc. Moreover, dementia will badly influence the perceptual nerve of patients at the late stage, making patients lose their normal self-caring ability eventually. According to the World Health Organization, one of each 20 people over the age of 65 suffers from the dementia disease. It is also widely acknowledged that, the dementia disease has become the fourth leading cause of death in the elderly after the cardiovascular disease, the cerebrovascular disease and the malignant tumor, for the time being. Therefore, the accurate diagnosis of dementia is important.

In order to fulfill the clinical diagnosis of dementia, clinicians often take the cortical atrophy, the ventricular enlargement, the thickness, the volume of medial temporal lobe decrease, the hippocampal cortical reduction, etc., as bio-markers. Also, various imaging tools including MRI (Magnetic Resonance Imaging), CT (Computed Tomography), SPECT (Single-Photon Emission Computed Tomography), PET (Positron Emission Tomography), have been widely incorporated in the task of dementia diagnosis. It is also valuable to point out that, ASL (Arterial

Spin Labeling) is a relatively new MRI technique, which begins to attract more and more research attention in dementia studies in recent years.

Generally speaking, ASL is one and the only one perfusion technology that can be compared with the gold standard <sup>15</sup>O-PET in principle, with hydrogen of free diffusion as endogenous contrast agent [1]. Since ASL does not need exogenous tracers and has no radiation, it has been widely incorporated for measuring the perfusion signal. Recent studies have demonstrated that, the reduction of regional blood perfusion in brain tissues can be utilized as an important factor in the pathogenesis of the dementia disease. The cerebral blood flow, the cerebral blood volume and the oxygen metabolism rate can all be obtained from ASL images, and these metrics are significant in the dementia diagnosis [2]. Although ASL has many advantages as mentioned above, it does lack in many well-known image-based dementia datasets. For instance, the ADNI-1/GO/2/3 datasets include sMRI (structural MRI), DTI-MRI (Diffusion Tensor Imaging-MRI), PET and resting-state fMRI (functional MRI) are incorporated as their main imaging modalities [3]. However, ASL images are not among them. In order to mitigate the dilemma, a series of recent studies dedicate to synthesizing ASL images to supplement the “missing” modality in these well-known datasets.

\* Corresponding author at: China Mobile-NCU AI&IOT Jointed Lab, Informatization Office, Nanchang University, Nanchang 330022, China.  
E-mail address: [n060101@e.ntu.edu.sg](mailto:n060101@e.ntu.edu.sg) (W. Huang).

It is also necessary to point out that, the synthesis from one commonly seen modality (e.g., sMRI) to ASL is often considered as a sophisticated and non-linear mapping. In order to obtain such a mapping, deep learning techniques are often incorporated. Recent studies have demonstrated that, various modalities including MRI images [4–12], PET images [13–15], CT images [16], ultrasound images [17], breast X-ray images [18,19], eye images [20–23], endoscopic images [24], etc., can be successfully synthesized. These medical image synthesis studies are important, as they can provide an alternative way to obtain medical images that are not acquired beforehand. Furthermore, medical image synthesis is also important for realizing the training of computer-aided tools that should often be driven by ample training data.

In this study, the synthesis task from sMRI to ASL images is emphasized. Technically, a new GAN (Generative Adversarial Network)-based model is proposed, and it incorporates the famous VAE (Variational Auto-Encoder) as the generator of the GAN model for synthesizing medical images. Advantages of this new GAN-based model include that, it can effectively alleviate the dilemma of unstable trainings that are commonly seen in contemporary GAN models, and the quality of synthesized ASL images is high. Experimental results also demonstrate that, the new GAN model is superior to other recently proposed models for ASL image synthesis. Also, synthesized ASL images obtained by the new GAN model can significantly improve the dementia diagnosis accuracy by 42.41% based on datasets in this study.

The organization of the paper is described as follows. In Section 2, representative studies that are published in recent years in the field of medical image synthesis are reviewed. Section 3 elaborates technical details of the new model in this study. Section 4 conducts a series of experiments and their corresponding comprehensive analyses. The superiority of the new model is suggested, therein. The conclusion of this study is drawn in Section 5.

## 2. Related works

As mentioned above, the task of medical image synthesis can be considered as to find a good mapping, which is capable to convert the source image into the target image. Since this mapping is sophisticated and often non-linear, deep learning models that are capable to implicitly represent such a sophisticated and non-linear mapping are often relied on. Generally speaking, the GAN model and the VAE model are two of the most commonly utilized models in image synthesis task. Technically, for the original GAN model [25], it is often composed of a generator and a discriminator. The role of the generator is to synthesize images, while the role of the discriminator is to discern synthesized images from real images. When the above competition continues, the performance of both the generator and the discriminator can be improved. In this way, synthesized images obtained from the GAN model can become more and more “real”. For the original VAE model [26], its architecture is illustrated in Fig. 1, where  $X$  denotes the input image;  $Z \sim N(0, 1)$  represents the latent representation of  $X$  following a Gaussian distribution (with the mean of 0 and the variance of 1);  $X'$  indicates the generated/synthesized image. It is also necessary to point out that, the VAE model is often composed of an encoder and a decoder, in which the encoder aims to encode the input image  $X$  to generate the latent representation  $Z$ . The decoder tries to reconstruct the latent representation  $Z$  into new image  $X'$ . The advantage of the original VAE model is that, its training is not prone to collapse. However, the disadvantage of the original VAE model is also obvious, as its synthesized images can become blurry. It is also important to highlight that, the original GAN model is capable to synthesize images with quality details, but its training is widely acknowledged to be quite unstable. Hence, characteristics of these two models can be complementary.

Representative studies in medical image synthesis based on GAN techniques that are proposed in recent years are introduced and discussed as follows. Guibas et al. proposed a Dual GAN in [27], and its

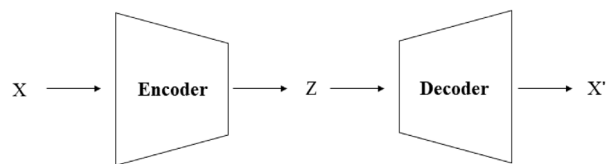


Fig. 1. The architecture of the original variational auto-encoder.

image synthesis task was divided into two stages, i.e., the geometry stage and the photorealism stage. The first stage was used to synthesize the geometric structure of the gray-scale retinal fundi images, and the second stage was to synthesize the real and color retinal fundi images. In [28], Emami et al. proposed a new cGAN network to synthesize CT images from MR images. MR images were added to both the generator and the discriminator. Experiments demonstrated that, the original cGAN [29] model is more suitable for cross-modal medical image translation tasks. Harms et al. incorporated the cycle-GAN model to synthesize high-quality CBCT images and integrated the concept of the popular residual block into the new cycle-GAN model, which was named as res-cycle-GAN [30]. In order to minimize the potential health risks caused by inherent tracer radiation within PET scans, Wang et al. proposed a locality adaptive multi-modality generative adversarial network (LA-GANs) in [31], whose main task is to synthesize full-dose high quality PET images from low-dose ones.

Besides the above-mentioned studies that fulfill image synthesis based on a variety of medical image modalities, the ASL image is also emphasized for the synthesis purpose in recent years. In [32], a novel unbalanced deep discriminant learning-based network equipped with residual network-based sub-structures was proposed to realize the synthesis of ASL images from sMRI images. It was also the first attempt to synthesize ASL images. This work synthesizes ASL images from the perspective of the global image plane, while other successive studies try to fulfill the synthesis task from the perspective of both the global image plane and the local image regions. In [33], a novel locally-constrained WGAN-GP ensemble was introduced, and diverse local constraints were incorporated. Furthermore, the GMM-based noise was generated from the Glow [34] model, and a WGAN-GP-based network was proposed in [35] which is believed to better reflect the characteristics of heterogeneity commonly seen in medical images. Also, an improved capsule-based network was proposed to synthesize ASL images in [36]. The advantage of incorporating capsule networks in medical image synthesis tasks is that, the capsule network does not adopt pooling operations, and spatial details of images can be better preserved. To sum up, the original GAN model and other related techniques (i.e., multiple channels, the exclusion of pooling operations, the ensemble model, the multi-Gaussian-distributed noise generation, etc.) have been adopted in recent ASL image synthesis studies.

In this study, the synthesis from sMRI images to ASL images is realized via a new GAN-based model, which incorporates the famous VAE model into its GAN architecture. In this way, advantages of the stable training in the VAE model and the high quality of synthesized outcomes in the GAN model can both be enjoyed. Technical details of the new model are elaborated in Section 3.

## 3. Methodology

In this section, technical details of the newly introduced model in this study are elaborated. The main architecture and the optimization of the new model are discussed in Sections 3.1, and 3.2, respectively.

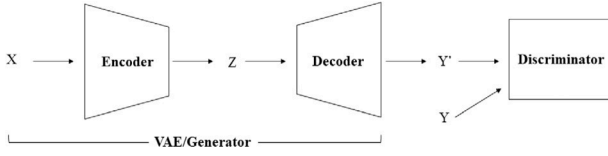


Fig. 2. The main architecture of the new VAE-GAN model in this study.

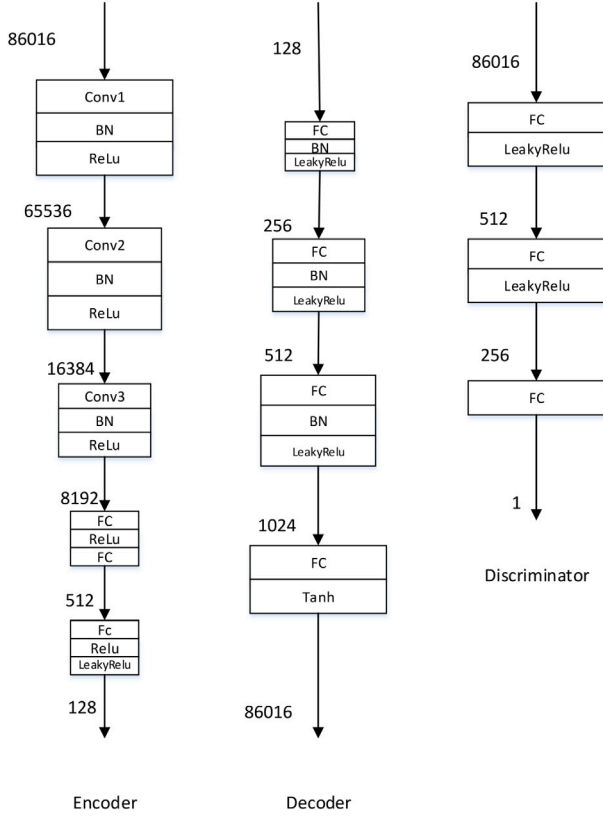


Fig. 3. Detailed architectures of the encoder, the decoder, and the discriminator of the new VAE-GAN model.

### 3.1. The main architecture of the new VAE-GAN model

The main architecture of the new VAE-GAN model is illustrated in Fig. 2, in which the generator of the new model is realized by the VAE model. To be specific, the generator is composed of an encoder and a decoder. Technically,  $X$  in Fig. 2 denotes the input of the encoder, and  $Z$  represents the latent feature after encoding. After that,  $Y'$  is decoded from  $Z$ , and then fed into the discriminator together with the real data  $Y$ . The above process is inspired by [26], and its mathematical form is explicitly represented in Eq. (1).

$$Z \sim \text{Encoder}(X) = p(Z|X), Y' \sim \text{Decoder}(Z) = q(Y|Z) \quad (1)$$

Moreover, detailed architectures of the encoder, the decoder, and the discriminator of the new VAE-GAN model are illustrated in Fig. 3, in which the number of parameters in each individual layer is annotated. Detailed explanations are as follows.

For the encoder, it adopts the convolution with the size of  $3 \times 3$ . The stride and the padding are set as 2 and 1, respectively. After that, the conventional batch normalization is added to avoid the notorious problems of gradient vanishment and gradient explosion. Then, the popular ReLU function is adopted to prevent the overfitting problem. The above-mentioned flow has been repeated 3 times in the encoder, so that the number of parameters in the encoder can be effectively

reduced. Finally, the full connection layer, the batch normalization layer and the LeakyRelu layer are further added to the encoder. For the decoder, it incorporates the “FC+BN+leakyReLU” architecture as the backbone. Technically, its full connection layer can enhance the fitting capability of the whole model. The leakyRelu function is incorporated here as the activation function, to avoid the “dead neuron” phenomenon commonly seen in deep learning models when the input is negative. Also, the Tanh function is utilized, in order to speed up the convergence of the decoder during the training. For the discriminator, it is also valuable to mention that, the “FC + LeakyRelu” architecture is utilized as the backbone. This structure has been repeated 2 times within the discriminator.

### 3.2. The optimization of the new VAE-GAN model

The energy function to optimize the new VAE-GAN model is represented in Eq. (2).

$$\mathcal{L}_{total} = L_{GAN} + L_1 + L_{KL} \quad (2)$$

in which  $L_{GAN}$  represents the energy function of the GAN model;  $L_1$  and  $L_{KL}$  denote two loss functions for the VAE model. Furthermore,  $L_{GAN}$  is described in Eq. (3).

$$L_{GAN} = \mathbb{E}_{x_s \sim \mathbb{P}_s} [D(G(x_s))] - \mathbb{E}_{x_t \sim \mathbb{P}_t} [D(x_t)] + \gamma \mathbb{E}_{x_t \sim \mathbb{P}_t} [\|\nabla_{x_t} D(x_t)\|_2 - 1]^2 \quad (3)$$

where the idea of the recent WGAN-GP model [37] is inherited. To be specific, the discriminator loss, the generator loss and the gradient penalty are explicitly represented on the right hand side of Eq. (3). Moreover,  $D$  and  $G$  in Eq. (3) indicate the discriminator and the generator of the GAN model, respectively.  $x_s$  in Eq. (3) denotes the sMRI data in this study, which follows the data distribution  $\mathbb{P}_s$ .  $x_t$  in Eq. (3) represents the ASL data in this study, which follows the data distribution  $\mathbb{P}_t$ . Furthermore,  $x_t = \epsilon x_t + (1 - \epsilon)G(x_s)$ , in which  $\epsilon$  is a random number between [0, 1].  $\gamma$  in Eq. (3) represents the weight of the gradient penalty.

Additionally, the specific form of  $L_1$  in Eq. (2) is illustrated in Eq. (4).

$$L_1 = \frac{1}{n} \sum_{i=1}^n H(a_i, a'_i) = \frac{1}{n} \sum_{i=1}^n H(a_i, S(s_i)) \quad (4)$$

Detailed explanations of Eq. (4) are as follows. Provided  $\tau_1$  represents the set of real sMRI images and  $\tau_2$  represents the set of real ASL images,  $s_i \in \tau_1$  and  $a_i \in \tau_2$  denote real sMRI image  $s_i$  and real ASL image  $a_i$ , respectively. Also,  $a'_i$  of Eq. (4) indicates the synthesized ASL image. The mapping from the sMRI image to the ASL image is represented by  $S(\cdot)$  in Eq. (4), and  $a'_i = S(s_i)$ . Moreover,  $H$  in Eq. (4) is represented by Eq. (5).

$$H(x, y) = \frac{1}{k} \sum_{i=1}^k \begin{cases} \frac{1}{2}(x_i - y_i)^2, & \text{if } |x_i - y_i| \leq 1. \\ |x_i - y_i| - \frac{1}{2}, & \text{otherwise.} \end{cases} \quad (5)$$

in which  $k$  denotes the number of sMRI (or ASL) images. Eqs. (4) and (5) suggest that, voxel-wise deviations between real ASL images and synthesized ASL images are incorporated to reflect the ASL synthesis performance in this study. Also, the smoothed L1 loss is added into the optimization of the new VAE-GAN model. Since real ASL images usually contain outliers, the L1 loss can be utilized as a regression loss, rather than adopting the conventional L2 loss that is more sensitive to outliers. In addition, since the original L1 loss is not smooth at zero points, the smoothed L1 loss can be selected as an alternative in this study.

Finally, the explicit form of the last term  $L_{KL}$  in the right-hand side of Eq. (2) can be described in Eq. (6).

$$L_{KL} = D(q(Z|Y) \| p(Z|X)) \quad (6)$$

where  $Z$  indicates the latent representation;  $X$  denotes the sMRI data;  $Y$  describes the real ASL data;  $p(Z|X)$  and  $q(Z|Y)$  reflect the distribution of  $Z$  under the condition of the sMRI data  $X$  and the ASL data  $Y$ , respectively. It can be noticed that, Eq. (6) actually aims to measure the similarity between two distributions  $p(Z|X)$  and  $q(Z|Y)$ .

#### 4. Experimental analyses

In this study, a series of experiments have been conducted and comprehensive analyses have been fulfilled from the statistical point of view. Experiments and their corresponding analyses demonstrate the superiority of the newly introduced VAE-GAN model. Details of them are as follows.

##### 4.1. Datasets and data preprocessing

Two datasets are taken into consideration for experimental evaluations in this study. Their descriptions are as follows.

The first dataset is popular in recent ASL image synthesis studies [32,33,35,36]. To be specific, the first dataset contains sMRI images and their corresponding ASL images. The sMRI images belong to the high-resolution MPRAGE T1-weighted MRI modality, and the ASL images are obtained by pseudo-continuous ASL scanning without background suppressions. These raw MRI data in this dataset were collected using a SIEMENS 3T TIM Trio magnetic resonance scanner, and the key acquisition parameters include: label time = 1500 ms, label delay = 1500 ms, post-label delay = 1500 ms, time delay = 1500 ms, and post-label delay = 1500 ms. In addition, spatial resolutions of sMRI images and ASL images in this dataset are set to  $64 \times 64 \times 21$ . Moreover, there are 355 patients in this dataset, which includes 38 patients with Alzheimer's disease (AD), 185 patients with mild cognitive impairment (MCI), and 132 patients with non-cognitive impairment (NCI) as normal controls. The average age of all patients in this database is  $70.56 \pm 7.20$  years old, and writing informed consents have been obtained for fulfilling this study.

The second dataset is the ADNI-1 dataset, which is from the well-known ADNI-1/GO/2/3 datasets family and has been widely utilized in contemporary image-based dementia diseases studies. It is important to highlight that, there is no ASL image in the ADNI-1 dataset. Therefore, the synthesis model learned from the first dataset is incorporated to synthesize ASL images from available sMRI images of the ADNI-1 dataset. In this way, ASL images can be supplemented and their effectiveness in dementia diseases diagnosis can be further investigated. In order to facilitate the ASL image synthesis, the spatial resolution of sMRI images in the ADNI-1 dataset is also set as  $64 \times 64 \times 21$ , which is consistent with that of sMRI images in the first dataset.

In addition, a series of preprocessing steps are carried out on all raw data, which include the motion correction, the brain extraction (skull removal), intra-modality registrations within sMRI images and ASL images (using the first slice as the reference plane), and the inter-modality registration between sMRI images and ASL images. In order to implement these preprocessing steps, the popular SPM [38] and FSL [39] toolboxes are utilized. Also, the IBA-SPM toolbox [40] are further incorporated to realize the segmentation of eight brain regions (i.e. left/right hippocampus, left/right parahippocampal gyrus, left/right putamen and left/right thalamus), whose local features are extracted and utilized in diverse shallow learning-based diagnosis tools. Details of them are introduced in Section 4.4.

##### 4.2. Experimental settings

As mentioned above, the new model incorporates both VAE and GAN to synthesize medical images. Therefore, advantages of rich synthesized images' details (obtained by GAN) and stable trainings (obtained by VAE) can be expected from this new VAE-GAN model. In

order to verify its effectiveness, the new VAE-GAN model is compared with state-of-the-art studies in ASL image synthesis. To be specific, the locally-constrained WGAN-GP ensemble introduced in [33] and the WGAN-GP+Glow model proposed in [35] are implemented and compared. Besides, several GAN-based image synthesis models including CycleGAN [42], WGAN-GP [37], and LSGAN [41] are further compared for the comprehensive purpose. Also, the well-known ResNet [43] and CNN [44] models from the "Non-flow&GAN-based" category are also implemented and compared in this experiment. It is valuable to point out that, parameters of these compared methods are determined through trial-and-errors for optimal synthesis performance, and details of them are elaborated in Table 1.

Moreover, implementation details of the new VAE-GAN model introduced in this study are described as follows. In order to make the optimization of the new model converge more efficiently, an adaptive learning rate scheme is incorporated. The initial learning rate of the generator is set as  $Lo_G = 1e-06$ , and the initial learning rate of the discriminator is set as  $Lo_D = 3e-08$ . When  $epoch = i$ , the learning rate of the generator changes into  $Lr_G = Lo_G \times 0.95^{(i/1000)}$ , and the learning rate of the discriminator becomes  $Lr_D = Lo_D \times 0.95^{(i/1000)}$ . The total epoch equals to 10 000, and the batch size is set as 8. In order to realize the optimization, the RMSprop [45] optimizer is incorporated, and the average training time of the new model is around 12 h. It is also valuable to mention that, all above-mentioned parameters of the new VAE-GAN model is also determined through trial-and-errors for obtaining optimal synthesis performance. All experiments in this study are carried out based on a workstation mainly equipped with the Intel Xeon Silver 4110 CPU, 128G RAM, two NVIDIA Titan V GPU cards. The programming is realized via PyTorch 3.7.0 based on the CentOS 7.4 operating system.

##### 4.3. Qualitative evaluations

Fig. 4 illustrates an example of synthesized ASL images obtained by all compared methods based on one patient from the first dataset of this study. There are 9 rows in Fig. 4, in which the first row includes real ASL images obtained by actual scanning (i.e., utilized as the golden standard) and the rest 8 rows demonstrate synthesized ASL images obtained by different compared methods (i.e., Rows 2: the new VAE-GAN model, 3: locally-constrained WGAN-GP ensemble [33], 4: WGAN-GP+Glow [35], 5: CycleGAN [42], 6: WGAN-GP [37], 7: LSGAN [41], 8: ResNet [43], and 9: CNN [44]). Furthermore, Column I in Fig. 4 demonstrates example 2D slices of synthesized ASL images (Rows 2–9) / real ASL images (Row 1). Column II in Fig. 4 indicates difference maps that are calculated as the voxel-wise absolute difference between synthesized ASL images (Column I of Rows 2–9) and their corresponding golden standard (Column I of Row 1). Several qualitative conclusions can be drawn here. First, the ideal case of difference maps certainly belongs to Column II of Row 1, in which no difference exists when subtracting the golden standard from itself. Second, difference maps of VAE-GAN (Row 2) are significantly less obvious than those of other rows, which suggests that synthesized ASL images of the new VAE-GAN model are the most similar towards the golden standard with respect to all compared methods. Moreover, the visualization of example synthesized ASL images obtained by the new VAE-GAN method from real sMRI images of one patient in the first dataset is illustrated in Fig. 5. It is necessary to mention that, the visualization is realized by the 3D slicer software with the grayscale rendering.

##### 4.4. Quantitative evaluations

In this section, synthesized ASL images obtained from all compared methods are incorporated together with their corresponding real sMRI images, to fulfill the important multi-modal dementia diseases diagnosis task. Generally, the main purpose is to accurately differentiate

**Table 1**  
Implementation details of all compared synthesis methods in this study.

No.	Models	Refs.	Details in implementations	Epoch	Learning rate	Batch size	Patch number	Parameters ( $1M = 10^6$ )
1	WGAN-GP+Glow	[35]	8 flows; each flow contains 8 bijections; 5 layers in the generator; 3 layers in the discriminator; the weight of the gradient penalty is 10.	10 000	0.0001	32	$64 \times 64 \times 21$	152.37M
2	Locally-constrained WGAN-GP ensemble	[33]	8 local regions; the weight of the regional constraint is [3, 5, 5, 5, 5, 2, 2]; 5 layers in the generator; 3 layers in the discriminator; The weight of the gradient penalty is 10; Weights of all local constraints are 0.05.	50	0.0002	4	$64 \times 64 \times 21$	399.05M
3	WGAN-GP	[37]	5 layers in the generator; 3 layers in the discriminator; the weight of the gradient penalty is 10.	200	0.0002	4	$64 \times 64 \times 21$	144.04M
4	LSGAN	[41]	5 layers in the generator; 5 layers in the discriminator.	200	0.0002	4	$64 \times 64 \times 21$	0.51M
5	CycleGAN	[42]	9 layers in the generator; 5 layers in the discriminator.	200	0.0002	4	$64 \times 64 \times 21$	28.55M
6	ResNet	[43]	19 layers.	200	0.0002	4	$64 \times 64 \times 21$	0.24M
7	CNN	[44]	7 layers.	200	0.0002	4	$64 \times 64 \times 21$	0.20M

**Table 2**  
Statistics on accuracies of dementia diseases diagnosis based on real sMRI images and synthesized ASL images obtained from all compared synthesis methods based on the first dataset (i.e., accuracies of dementia diseases diagnosis based on NCI/MCI/AD patients are described via colors of blue, orange, and green in corresponding entries).

Model	sMRI (as the baseline)			sMRI + synthesized ASL by Ours			sMRI + synthesized ASL by Locally-constrained WGAN-GP ensemble			sMRI + synthesized ASL by WGAN-GP + Glow			sMRI + synthesized ASL by CNN		
LR	0.8771 ± 0.0310			0.9343 ± 0.0286			0.9171 ± 0.0278			0.8871 ± 0.0369			0.9057 ± 0.0213		
	0.8307	1.0000	0.4057	0.9579	1.0000	0.5556	0.9234	1.0000	0.5238	0.8538	1.0000	0.4290	0.8752	1.0000	0.5238
SVM	0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517		
	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
SVR	0.9314 ± 0.0277			0.9343 ± 0.0292			0.9314 ± 0.0277			0.9371 ± 0.0302			0.9314 ± 0.0226		
	0.9334	1.0000	0.5238	0.9579	1.0000	0.5556	0.9607	1.0000	0.5317	0.9456	1.0000	0.5000	0.9334	1.0000	0.5238
Ranking	0.8971 ± 0.0517			0.9086 ± 0.0224			0.9000 ± 0.0508			0.8971 ± 0.0517			0.8971 ± 0.0517		
	1.0000	1.0000	0.0000	0.8142	1.0000	0.6984	1.0000	1.0000	0.0159	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
CNN-2	0.4740 ± 0.0783			0.4704 ± 0.0650			0.4602 ± 0.0513			0.4930 ± 0.0841			0.4676 ± 0.0427		
	0.3333	0.6486	0.1316	0.2955	0.6811	0.0263	0.3863	0.5784	0.1316	0.4545	0.5946	0.1316	0.5076	0.5081	0.1316
CNN-20	0.4820 ± 0.0607			0.5296 ± 0.0077			0.5158 ± 0.0679			0.4880 ± 0.0503			0.5014 ± 0.0603		
	0.3561	0.6378	0.1842	0.4697	0.6324	0.1316	0.5227	0.5838	0.1579	0.5984	0.4810	0.1316	0.3030	0.7189	0.1316
ResNet-56	0.4900 ± 0.0602			0.4873 ± 0.0551			0.4664 ± 0.0417			0.4830 ± 0.0429			0.4648 ± 0.0467		
	0.4697	0.6054	0.0000	0.4848	0.5730	0.1579	0.4848	0.5243	0.1316	0.5757	0.4864	0.1579	0.5000	0.5135	0.1053
Average	0.7212			0.7374			0.7269			0.7261			0.7236		
	0.7033	0.8417	0.1779	0.7114	0.8409	0.3036	0.7540	0.8124	0.2132	0.7754	0.7946	0.1929	0.7313	0.8201	0.2023
Boost	0.00%			+42.41%			+14.92%			+12.83%			+6.28%		
Models	sMRI + real ASL ASL by ResNet			sMRI + synthesized ASL by WGAN-GP			sMRI + synthesized ASL by LSGAN			sMRI + synthesized ASL by CycleGAN			sMRI + real ASL (as the golden standard)		
LR	0.8971 ± 0.0330			0.8829 ± 0.0263			0.8714 ± 0.0309			0.8886 ± 0.0240			0.8914 ± 0.0269		
	0.9034	1.0000	0.3810	0.8333	1.0000	0.4057	0.8327	1.0000	0.3437	0.8693	1.0000	0.3095	0.8671	1.0000	0.4444
SVM	0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517			0.8971 ± 0.0517		
	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000
SVR	0.8971 ± 0.0517			0.9314 ± 0.0270			0.9257 ± 0.0238			0.9229 ± 0.0289			0.9371 ± 0.0216		
	1.0000	1.0000	0.0000	0.9334	1.0000	0.5238	0.9482	1.0000	0.4286	0.9497	1.0000	0.4444	0.9456	1.0000	0.5000
Ranking	0.9029 ± 0.0482			0.9000 ± 0.0508			0.8971 ± 0.0517			0.8971 ± 0.0517			0.9000 ± 0.0492		
	1.0000	1.0000	0.0238	0.9034	1.0000	0.2857	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0079
CNN-2	0.4930 ± 0.0683			0.4060 ± 0.0280			0.4410 ± 0.0388			0.4540 ± 0.0486			0.5634 ± 0.0589		
	0.3712	0.6811	0.0000	0.4015	0.4864	0.0263	0.4318	0.5351	0.0263	0.4697	0.5351	0.0000	0.5530	0.6811	0.0263
CNN-20	0.4648 ± 0.0386			0.4060 ± 0.0174			0.4410 ± 0.0459			0.5200 ± 0.0491			0.5718 ± 0.0214		
	0.2955	0.6649	0.0789	0.3712	0.4864	0.1316	0.3712	0.5567	0.1316	0.4545	0.6486	0.0789	0.3939	0.8054	0.0526
ResNet-56	0.5014 ± 0.0708			0.4340 ± 0.0377			0.4220 ± 0.0132			0.4190 ± 0.0492			0.5549 ± 0.0275		
	0.4242	0.6324	0.1316	0.4848	0.4810	0.0263	0.4545	0.4594	0.1316	0.3485	0.5459	0.0526	0.4470	0.7189	0.1316
Average	0.7219			0.6939			0.6993			0.7141			0.7594		
	0.7135	0.8541	0.0879	0.7039	0.7791	0.1999	0.7198	0.7930	0.1517	0.7274	0.8185	0.1265	0.7438	0.8865	0.1661
Boost	+1.83%			-71.47%			-57.33%			-18.59%			100%		

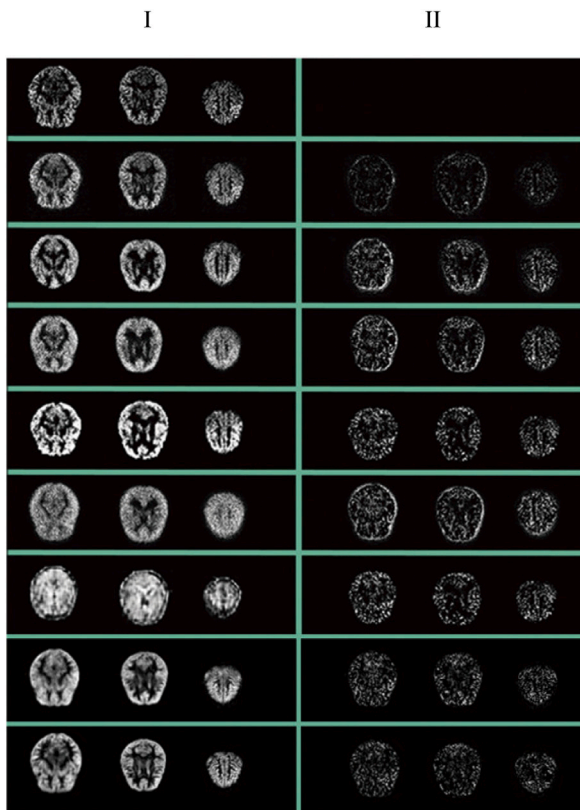


Fig. 4. An illustration of synthesized ASL images obtained from all compared methods based on one patient of the first dataset in this study.

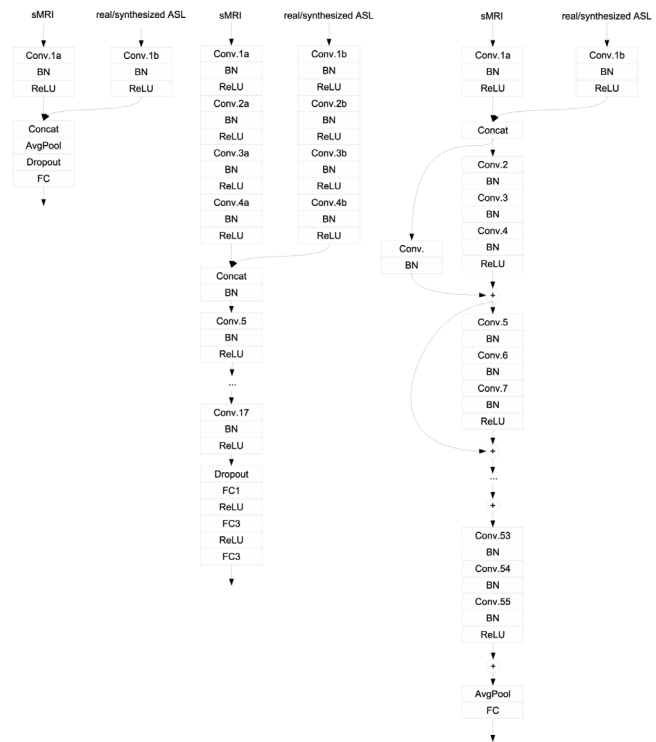


Fig. 6. Architectures of three deep learning-based diagnosis tools utilized in this study (i.e., from left to right: CNN-2, CNN-20, ResNet-56).

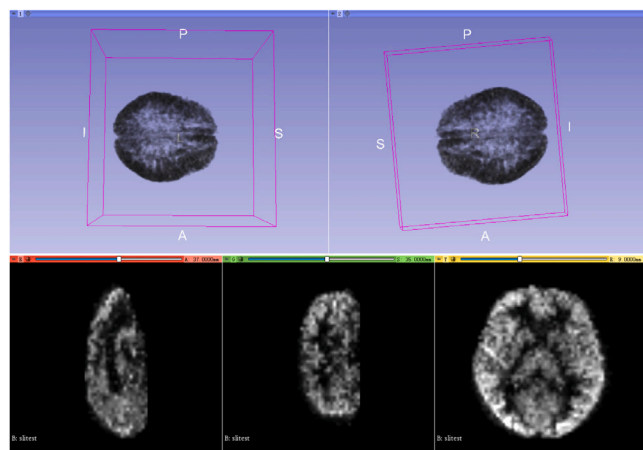


Fig. 5. The visualization of synthesized ASL images obtained by the new VAE-GAN method from real sMRI images of one patient in the first dataset (i.e., the visualization is realized by the 3D slicer software with the grayscale rendering).

progressions of dementia diseases (i.e., AD, MCI, NCI). In order to realize the diagnosis, there are 7 diagnosis tools implemented, including 4 shallow learning-based diagnosis tools and 3 deep learning-based diagnosis tools. Details of them are described as follows.

First, 4 shallow learning-based diagnosis tools include the linear regression (LR), the triple-class SVM (SVM), the support vector regression (SVR), and the support vector ranking (Ranking). Since hand-crafted features are usually necessary to be extracted as inputs in shallow learning-based methods, the 8 segmented regions mentioned in Section 4.1 are emphasized and their means as well as standard deviations are extracted to construct low-level visual features. Also, SVM, SVR,

and Ranking are implemented using the SVM-light toolbox [46], and the Gaussian radial basis function (Gaussian RBF) is adopted as the non-linear kernel to fulfill the well-known “kernel trick” in these tools. The Gaussian width is automatically determined via the classic radius margin bound method [47]. Second, 3 deep learning-based diagnosis tools are CNN-2, CNN-20, and ResNet-56, whose architectures are illustrated in Fig. 6. It can be observed that, all these deep learning-based diagnosis tools have two-channel inputs, in which one channel is for inputting sMRI images and the other is for inputting real/synthesized ASL images. It is also necessary to mention that, when only sMRI is to be input (i.e., as the baseline in Table 2), only the channel for incorporating sMRI images is active.

Table 2 demonstrates statistics of accuracies of dementia diseases diagnosis when incorporating real sMRI images and their corresponding synthesized ASL images obtained from all compared synthesis methods, based on the first dataset. Since diagnosis experiments are conducted following the classic 5-fold cross-validation strategy from the statistical perspective, the entry (i.e.,  $\mu \pm \sigma$ ) in Table 2 represents the average accuracy ( $\mu$ ) of the 5-fold cross-validation and its corresponding standard deviation ( $\sigma$ ). For the two bottom lines “Average” and “Boost” in Table 2, “Average” calculates the mean accuracy from all diagnosis outcomes in each individual column of Table 2, and “Boost” indicates the percentage of diagnosis performance improvement. Since the diagnosis of only using real sMRI images is considered as the baseline (i.e., the second column of Table 2) and the diagnosis of using both real sMRI images and real ASL images is regarded as the golden standard (i.e., the last column of Table 2), entries in “Boost” are quantitatively calculated as:  $\text{Boost} = \frac{\text{mean of each individual method} - \text{mean of the baseline}}{\text{mean of the golden standard} - \text{mean of the baseline}} \times 100\%$ .

It can be summarized from Table 2 that, synthesized ASL images obtained by the new VAE-GAN model are capable to provide the highest diagnosis performance improvement, with the largest increasing percentage of +42.41% among all compared synthesis methods. Besides the new VAE-GAN model, the locally-constrained WGAN-GP ensemble and WGAN-GP+Glow are two other methods that are capable to improve

**Table 3**

Statistics on accuracies of dementia diseases diagnosis based on real sMRI images and synthesized ASL images obtained by the new VAE-GAN model based on the ADNI-1 dataset.

Model	sMRI	sMRI + synthesized ASL by VAE-GAN
LR	$0.7555 \pm 0.0491$	$0.7545 \pm 0.0525$
SVM	$0.8090 \pm 0.0369$	$0.8090 \pm 0.0369$
SVR	$0.8545 \pm 0.0432$	$0.8636 \pm 0.0389$
Ranking	$0.8090 \pm 0.0369$	$0.8090 \pm 0.0369$
CNN-2	$0.7876 \pm 0.0365$	$0.8160 \pm 0.0526$
CNN-20	$0.8984 \pm 0.0430$	$0.9000 \pm 0.0447$
ResNet-56	$0.8640 \pm 0.0873$	$0.8560 \pm 0.0404$
Average	0.8254	<b>0.8297</b>
Boost		<b>11.26%</b>

the diagnosis performance after incorporating their synthesized ASL images. For other compared methods, their synthesized ASL images unfortunately deteriorate the diagnosis performance, whose results are lower than the baseline.

In Fig. 7, examples of synthesized ASL images obtained by CycleGAN, LSGAN, and WGAN-GP are illustrated. It can be noticed that these synthesized outcomes have shortcomings of partial volume effects (Fig. 7-a), blurring (Fig. 7-b), noisy data (Fig. 7-c), etc. The quality of these synthesized ASL images is not satisfactory. Hence, diagnosis performance based on them is not satisfactory, either. Since the non-linear mapping from structural MRI to ASL images should be implicitly represented by the deep learning model whose generalization capability can be suggested by its hyper-parameters, models with significantly less number of hyper-parameters (e.g., 0.51M for LSGAN) may not be capable to implicitly represent the above non-linear mapping well. Thus, the synthesis performance as well as its corresponding diagnosis performance based on synthesized data will be deteriorated therein.

Moreover, after incorporating synthesized ASL images together with their corresponding original sMRI images in the ADNI-1 dataset, the accuracy of dementia diseases diagnosis can be significantly improved as well. Table 3 lists statistics on accuracies of dementia diseases diagnosis using real sMRI images and synthesized ASL images obtained by the new VAE-GAN model, based on the ADNI-1 dataset. Since no real ASL images exist in the ADNI-1 dataset, no golden standards can be referred to. However, an absolute increase of 0.043 (i.e.,  $0.8297 - 0.8254 = 0.0043$ ) in the average accuracy can be obtained, which brings about approximately +11.26% performance improvement in Table 3 (i.e., it is calculated proportional to  $(0.7374 - 0.7212) = 0.0162$  of the new VAE-GAN model that brings about +42.41% performance improvement based on the first dataset in Table 2). Hence, incorporating synthesized ASL images obtained from the new VAE-GAN model is also beneficial to improve the dementia diseases diagnosis, compared with only adopting sMRI images, in the ADNI-1 dataset.

## 5. Conclusions

In this study, a new GAN-based model is introduced to fulfill the valuable ASL image synthesis task. Technically, VAE is incorporated as the generator of the GAN model. Experiments based on two image-based dementia datasets suggest that, the new VAE-GAN model outperforms other recently proposed state-of-the-arts in ASL image synthesis as well as other popular image synthesis methods. It is necessary to point out that, this study is different from other previous related VAE-GAN studies. For instance, for Kwon et al.'s work in [48], another encoder and code discriminator of the VAE model are further incorporated besides the discriminator and the generator of the GAN model. However, ours replaces the generator of the GAN model with the VAE model. For Sengupta et al.'s work in [49], their input to the VAE model is random noise, and the VAE model does not replace any component of the GAN model. Nevertheless, ours utilizes the real structural MRI images as the input of the VAE model and considers the

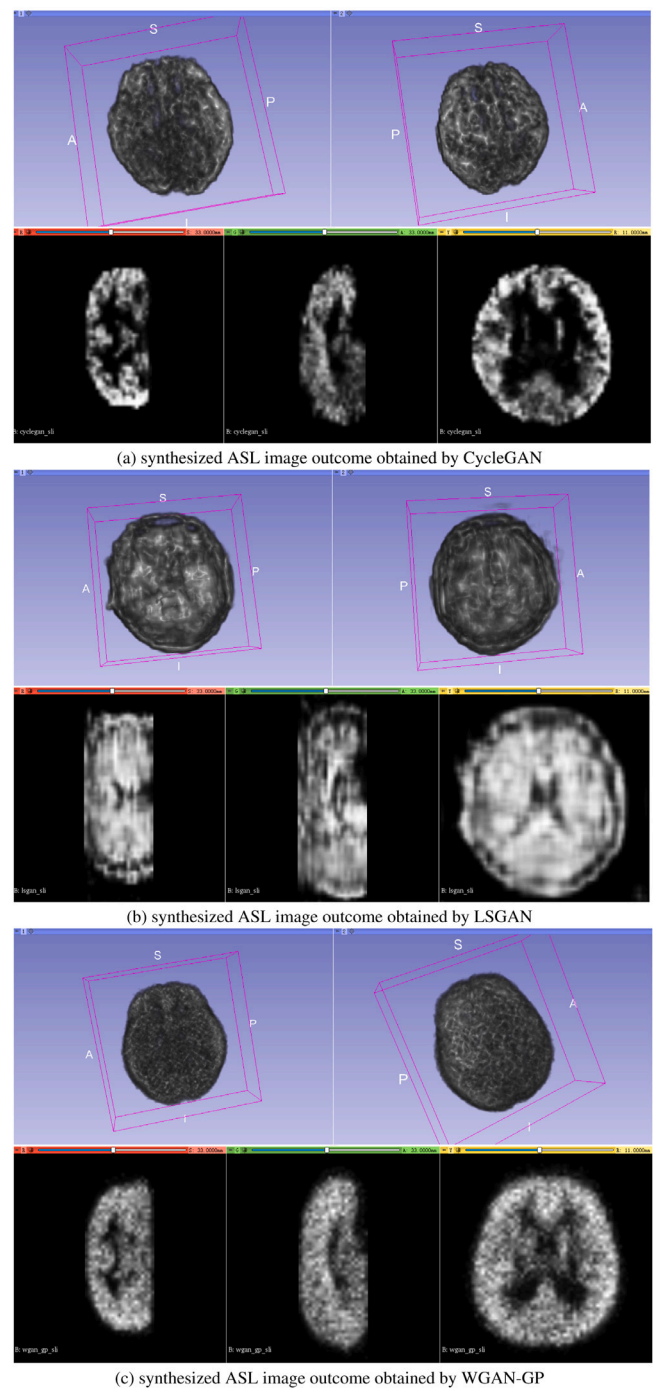


Fig. 7. The visualization of examples of synthesized ASL images obtained by CycleGAN, LSGAN, and WGAN-GP based on patients of the first dataset (i.e., the visualization is realized by the 3D slicer software with the grayscale rendering).

VAE model as the generator of the GAN model. Hence, the difference between these related studies and the new method proposed in this study is significant. Future efforts will be incorporated to investigate VAE-GAN ensemble models for ASL image synthesis. Also, new model architectures including the scale-estimated deep networks [50] will be investigated for the ASL image synthesis task.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by grants 61862043 & 61971352 approved by National Natural Science Foundation of China, the key grant 20204BCJ22011 approved by Natural Science Foundation of Jiangxi Province, and the grant 2018JM6015 approved by Natural Science Foundation of Shaanxi Province.

## References

- [1] J. Detre, J. Leigh, D. Williams, A. Koretsky, Perfusion imaging, *Magn. Reson. Med.* 23 (1) (1992) 37–45.
- [2] E. Musiek, Y. Chen, M. Kerczykowski, B. Saboury, P. Martinez, J. Reddin, A. Alavi, D. Kimberg, D. Wolk, P. Julin, A. Newberg, S. Arnold, J. Detre, Direct comparison of fluorodeoxyglucose positron emission tomography and arterial spin labeling magnetic resonance imaging in Alzheimer's disease, *Alzheimer's Dement.* 8 (1) (2012) 51–59.
- [3] S. Mueller, M. Weiner, L. Thal, R. Petersen, C. Jack, W. Jagust, J. Trojanowski, A. Toga, L. Beckett, The Alzheimer's disease neuroimaging initiative, *Neuroimaging Clin. N. Am.* 15 (4) (2005) 869–877.
- [4] N. Cordier, H. Delingette, M. Le, N. Ayache, Extended modality propagation: image synthesis of pathological cases, *IEEE Trans. Med. Imaging* 35 (12) (2016) 2598–2608.
- [5] H. Li, J. Paetzold, A. Sekuboyina, F. Kofler, J. Zhang, J. Kirschke, B. Wiestler, B. Menze, DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 795–803.
- [6] Y. Huang, L. Shao, A. Frangi, Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning, *IEEE Trans. Med. Imaging* 37 (3) (2018) 815–827.
- [7] A. Chartsias, T. Joyce, M. Giuffrida, S. Tsafaris, Multimodal MR synthesis via modality-invariant latent representation, *IEEE Trans. Med. Imaging* 37 (3) (2018) 803–814.
- [8] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, P. Bourgeat, Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis, *IEEE Trans. Med. Imaging* 38 (7) (2019) 1750–1762.
- [9] T. Zhang, H. Fu, Y. Zhao, J. Cheng, M. Guo, Z. Gu, B. Yang, Y. Xiao, S. Gao, J. Liu, SkrGAN: sketching-rendering unconditional generative adversarial networks for medical image synthesis, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019.
- [10] S. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, T. Cukur, Image synthesis in multi-contrast MRI with conditional generative adversarial networks, *IEEE Trans. Med. Imaging* 38 (10) (2019) 2375–2388.
- [11] N. Duchateau, M. Sermesant, H. Delingette, N. Ayache, Model-based generation of large databases of cardiac images: synthesis of pathological cine MR sequences from real healthy cases, *IEEE Trans. Med. Imaging* 37 (3) (2018) 755–766.
- [12] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with deep convolutional adversarial networks, *IEEE Trans. Biomed. Imag.* 65 (12) (2018) 2720–2730.
- [13] I. Polycarpou, G. Soutanidis, C. Tsoumpas, Synthesis of realistic simultaneous positron emission tomography and magnetic resonance imaging data, *IEEE Trans. Med. Imaging* 37 (3) (2018) 703–711.
- [14] Y. Wang, L. Zhou, B. Yu, L. Wang, C. Zu, D. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, 3D auto-context-based locality adaptive multi-modality gans for pet synthesis, *IEEE Trans. Med. Imaging* 38 (6) (2019) 1328–1339.
- [15] Y. Wang, B. Yu, L. Wang, C. Zu, D. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, L. Zhou, 3D conditional generative adversarial networks for high-quality PET image estimation at low dose, *NeuroImage* 174 (2018) 550–562.
- [16] G. Zeng, G. Zheng, Hybrid generative adversarial networks for deep MR to CT synthesis using unpaired data, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 759–767.
- [17] Y. Zhou, S. Giffard-Roisin, M. Craene, S. Camarasu-Pop, J. D'Hooge, M. Alessandrini, D. Friboulet, M. Sermesant, O. Bernard, A framework for the generation of realistic synthetic cardiac ultrasound and magnetic resonance imaging sequences from the same virtual patients, *IEEE Trans. Med. Imaging* 37 (3) (2018) 741–754.
- [18] Y. Ren, Z. Zhu, Y. Li, D. Kong, R. Hou, L. Grimm, J. Marks, J. Lo, Mask embedding for realistic high-resolution medical image synthesis, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 422–430.
- [19] G. Jiang, Y. Lu, J. Wei, Y. Xu, Synthesize mammogram from digital breast tomosynthesis with gradient guided cgans, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 801–809.
- [20] P. Costa, A. Galdran, M. Meyer, M. Niemeijer, M. Abramoff, A. Mendonca, A. Campilho, End-to-end adversarial retinal image synthesis, *IEEE Trans. Med. Imaging* 37 (3) (2018) 781–791.
- [21] Y. Zhou, X. He, S. Cui, F. Zhu, L. Liu, L. Shao, High-resolution diabetic retinopathy image synthesis manipulated by grading and lesions, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 505–513.
- [22] X. Wang, M. Xu, L. Li, Z. Wang, Z. Guan, Pathology-aware deep network visualization and its application in glaucoma image synthesis, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 423–431.
- [23] A. Diaz-Pinto, A. Colomer, V. Naranjo, S. Morales, Y. Xu, A. Frangi, Retinal image synthesis and semi-supervised learning for glaucoma assessment, *IEEE Trans. Med. Imaging* 38 (9) (2019) 2211–2218.
- [24] T. Kanayama, Y. Kurose, K. Tanaka, K. Aida, S. Satoh, M. Kitsuregawa, T. Harada, Gastric cancer detection from endoscopic images using synthesis by GAN, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 530–538.
- [25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks. [Online]. Available: <https://arxiv.org/abs/1406.2661>, 0000.
- [26] D. Kingma, M. Welling, Auto-encoding variational bayes. [Online]. Available: <https://arxiv.org/abs/1312.6114>, 0000.
- [27] J. Guibas, T. Virdi, P. Li, Synthetic medical images from dual generative adversarial networks. [Online]. Available: <https://arxiv.org/pdf/1709.01872.pdf>, 0000.
- [28] H. Emami, M. Dong, S. Nejad-Davaran, C. Glide-Hurst, Generating synthetic CTs from magnetic resonance images using generative adversarial networks, *Med. Phys.* 45 (8) (2018) 3627–3636.
- [29] M. Mirza, S. Osindero, Conditional generative adversarial nets. [Online]. Available: <https://arxiv.org/abs/1411.1784>, 0000.
- [30] J. Harms, Y. Lei, T. Wang, R. Zhang, J. Zhou, X. Tang, W. Curran, T. Liu, X. Yang, Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography, *Med. Phys.* 46 (9) (2019) 3998–4009.
- [31] Y. Wang, L. Zhou, L. Wang, B. Yu, C. Zu, D. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, Locality adaptive multi-modality GANs for high-quality PET image synthesis, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2018.
- [32] W. Huang, M. Luo, X. Liu, P. Zhang, H. Ding, W. Xue, D. Ni, Arterial spin labeling images synthesis from sMRI using unbalanced deep discriminant learning, *IEEE Trans. Med. Imaging* 38 (10) (2019) 2338–2351.
- [33] W. Huang, M. Luo, X. Liu, P. Zhang, H. Ding, D. Ni, Arterial spin labeling images synthesis via locally-constrained WGAN-GP ensemble, in: *Medical Image Computing and Computer-Assisted Intervention*, Springer, Shenzhen, China, 2019, pp. 768–776.
- [34] D. Kingma, P. Dhariwal, Glow: generative flow with invertible 1x1 convolutions, in: *Annual Conference on Neural Information Processing Systems*, Montreal, Canada, 2018, pp. 10236–10245.
- [35] W. Huang, M. Luo, X. Liu, P. Zhang, H. Ding, D. Ni, Novel bi-directional images synthesis based on WGAN-GP with GMM-based noise generation, in: *International Workshop on Machine Learning in Medical Imaging*, Springer, Cham, 2019.
- [36] W. Huang, M. Luo, X. Liu, P. Zhang, H. Ding, D. Ni, Arterial spin labeling image synthesis from structural MRI using improved capsule-based networks, *IEEE Access* 8 (2020) 181137–181153.
- [37] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of wasserstein GANs. [Online]. Available: <https://arxiv.org/abs/1704.00028>, 0000.
- [38] Statistical parametric mapping. [Online]. Available: <https://www.fil.ion.ucl.ac.uk/spm/>, 0000.
- [39] FMRIB software library v6.0. [Online]. Available: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>, 0000.
- [40] Individual brain atlases using statistical parametric mapping (IBA-SPM) software. [Online]. Available: <http://www.thomaskoenig.ch/Lester/ibaspm.htm>, 0000.
- [41] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, S. Smolley, Least squares generative adversarial networks, 2017, [Online]. Available: <https://arxiv.org/abs/1611.04076>.
- [42] J. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks. [Online]. Available: <https://arxiv.org/abs/1703.10593>, 0000.
- [43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, [Online]. Available: <https://arxiv.org/abs/1512.03385>.
- [44] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.



- [45] G. Hinton, N. Srivastava, K. Swersky, Neural networks for machine learning, [Online]. Available: [http://www.cs.toronto.edu/tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](http://www.cs.toronto.edu/tijmen/csc321/slides/lecture_slides_lec6.pdf), 0000.
- [46] SVM light - an implementation of support vector machine in C. [Online]. Available: <http://svmlight.joachims.org>, 0000.
- [47] S. Keerthi, Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms, *IEEE Trans. Neural Netw.* 13 (5) (2002) 1225–1229.
- [48] G. Kwon, C. Han, D. Kim, Generation of 3D brain MRI using auto-encoding generative adversarial networks, in: *International Workshop on Machine Learning in Medical Imaging*, Springer, Cham, 2019.
- [49] S. Sengupta, A. Athwale, T. Gulati, J. Zelek, V. Lakshminarayanan, FunSyn-NEt: enhanced residual variational auto-encoder and image-to-image translation network for fundus image synthesis, in: *Medical Imaging: Image Processing*, International Society for Optics and Photonics, 2020.
- [50] H. Huang, Y. Han, P. Zhang, W. Huang, Tracking based on scale-estimated deep networks with hierarchical correction ensembling for cross-media understanding, *Displays* 69 (2021) 102055.