2019

Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis

Biting Yu
University of Wollongong, by354@uowmail.edu.au

Luping Zhou
University of Sydney, lupingz@uow.edu.au

Lei Wang
University of Wollongong, leiw@uow.edu.au

Yinghuan Shi
Nanjing University

Jurgen Fripp
CSIRO Health and Biosecurity

See next page for additional authors

Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis

Abstract

Magnetic resonance (MR) imaging is a widely used medical imaging protocol that can be configured to provide different contrasts between the tissues in human body. By setting different scanning parameters, each MR imaging modality reflects the unique visual characteristic of scanned body part, benefiting the subsequent analysis from multiple perspectives. To utilize the complementary information from multiple imaging modalities, cross-modality MR image synthesis has aroused increasing research interest recently. However, most existing methods only focus on minimizing pixel/voxel-wise intensity difference but ignore the textural details of image content structure, which affects the quality of synthesized images. In this paper, we propose edge-aware generative adversarial networks (Ea-GANs) for cross-modality MR image synthesis. Specifically, we integrate edge information, which reflects the textural structure of image content and depicts the boundaries of different objects in images, to reduce this gap. Corresponding to different learning strategies, two frameworks are proposed, i.e., a generator-induced Ea-GAN (gEa-GAN) and a discriminator-induced Ea-GAN (dEa-GAN). The gEa-GAN incorporates the edge information via its generator, while the dEa-GAN further does this from both the generator and the discriminator so that the edge similarity is also adversarially learned. In addition, the proposed Ea-GANs are 3D-based and utilize hierarchical features to capture contextual information. The experimental results demonstrate that the proposed Ea-GANs, especially the dEa-GAN, outperform multiple state-of-the-art methods for cross-modality MR image synthesis in both qualitative and quantitative measures. Moreover, the dEa-GAN also shows excellent generality to generic image synthesis tasks on benchmark datasets about facades, maps, and cityscapes.

Disciplines
Engineering | Science and Technology Studies

Authors
Biting Yu, Luping Zhou, Lei Wang, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat

Ea-GANs: Edge-aware Generative Adversarial Networks for Cross-modality MR Image Synthesis

Biting Yu, Luping Zhou*, Lei Wang*, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat

Abstract—**Magnetic resonance imaging (MRI) is a widely used medical imaging protocol that can be configured to provide different contrast between the tissues in human body. By setting different scanning parameters, each MR imaging modality reflects the unique visual characteristic of scanned body part, benefiting the subsequent analysis from multiple perspectives. To utilise the complementary information from multiple imaging modalities, cross-modality MR image synthesis has aroused increasing research interest recently. However, most existing methods only focus on minimising pixel/voxel-wise intensity difference, but ignore the textural details of image content structure, which affects the quality of synthesised images. In this paper, we propose edge-aware generative adversarial networks (Ea-GANs) for cross-modality MR image synthesis. Specifically, we integrate edge information, which reflects the textural structure of image content and depicts the boundaries of different objects in images, to reduce this gap. Corresponding to different learning strategies, two frameworks are proposed, i.e., a generator-induced Ea-GAN (gEa-GAN) and a discriminator-induced Ea-GAN (dEa-GAN). The gEa-GAN incorporates the edge information via its generator, while the dEa-GAN further does this from both generator and discriminator so that the edge similarity is also adversarially learned. In addition, the proposed Ea-GANs are 3D based, and utilise hierarchical features to capture contextual information. Experimental results demonstrate that the proposed Ea-GANs, especially the dEa-GAN, outperform multiple state-of-the-art methods for cross-modality MR image synthesis in both qualitative and quantitative measures. Moreover, the dEa-GAN also shows excellent generality to generic image synthesis tasks on benchmark datasets about facades, maps, and cityscapes.**

Index Terms—**Neural networks, machine learning, magnetic resonance imaging (MRI), brain.**

I. INTRODUCTION

MEDICAL imaging, a technology creating the visual representations of anatomy or the function of human body, has diverse imaging protocols varying in characteristics and applications. Among them, magnetic resonance imaging (MRI), which produces different modalities of images by setting task-specific scanning parameters, has been broadly exploited in medical image analysis [1], [2]. During analysis, the images from multiple MR imaging modalities (e.g., T1-weighted, T2-weighted, FLAIR) are processed together, since each modality shows unique soft tissue contrast. For example, multi-modality MR images have been collectively utilised to study the neuroanatomy of human brains for disease diagnosis [3] or therapy planning [4]. The complementary information from multi-modalities demonstrates better predictive power than that from a single imaging modality. Also, the benefits of using multi-modality MR images for brain lesion segmentation have been widely recognised [5], [6]. At the same time, due to modality missing and modality inconsistency between different clinical centers [7], the high demand of employing multiple MR imaging modalities for analysis is not always met in clinic and research, which adversely affects the quality of diagnosis and treatment. Therefore, cross-modality MR image synthesis has recently aroused increasing research interest, and more and more investigations have been conducted to cope with the limitation of insufficient modalities in clinic and research [8], [9].

Medical image synthesis, which is defined as a mapping between the unknown target-modality images and the given source-modality ones, has shown wide applications, e.g., virtual dataset creation [10], missing image imputation [11], and image super-resolution [12]. Current approaches can be roughly grouped into two categories. The first category refers to atlas-based methods. These methods utilise the paired image atlases from source- and target-modalities, calculate the atlas-to-image transformation in source-modality, and then explore this transformation to synthesise target-modality-like images from their corresponding target-modality atlases [13]–[17]. Since most atlases are built upon healthy subjects, these methods perform less satisfactorily on the images with pronounced abnormalities. The second category, learning-based methods, can mitigate this issue. Specifically, these methods directly learn a mapping from source-modality to target-modality. Once a training set appropriately contains pathology, such information could be captured by the learned model, so that abnormalities, such as brain tumors, can also be synthesised in target-modality images.

A large category of the learning-based synthesis methods train a nonlinear model that maps each small source-modality patch to the pixel/voxel at the center of the corresponding patch having the same location in target-modality. For example, Huynh et al. train a structured random forest model to estimate CT images from MRI [18]; Wang et al. develop a semi-supervised tripled dictionary learning method to predict standard-dose PET images from low dose ones [19]; Ye et al. target cross-modality MR image synthesis via patch-based searching [8]. Meanwhile, all these mentioned patch-based methods have a limitation that the important spatial relation-

*Co-Corresponding Authors.

Biting Yu is with School of Computing and Information Technology, University of Wollongong.

Luping Zhou is with School of Electrical and Information Engineering, University of Sydney (e-mail: luping.zhou@sydney.edu.au).

Lei Wang is with School of Computing and Information Technology, University of Wollongong (e-mail: leiw@uow.edu.au).

Yinghuan Shi is with National Key Laboratory for Novel Software Technology, Nanjing University.

Jurgen Fripp and Pierrick Bourgeat are with CSIRO Health and Biosecurity, Brisbane, Australia.

ship among the small patches in the same image are ignored, leading to contrast inconsistency in synthesised image. To alleviate this issue, global spatial information is additionally captured by multi-resolution patch regression in [9] for cross-modality image synthesis. However, the handcrafted features used in the above methods [8], [9], [18], [19] have limited descriptive power to represent the complicated contextual details in images, which in turn affects synthesis quality. Moreover, in these methods, patch-based estimation is usually applied to each individual voxel, and the final estimation of a whole image is determined by a large number of highly overlapped patches. Therefore, such methods usually lead to over-smoothed synthesised images, and incur heavy computational cost.

To deal with the above problems, deep learning based models, such as convolutional neural networks (CNNs) and fully convolutional network (FCN), have been used to automatically learn features with better descriptive power [20], [21]. Moreover, researchers recently have found that using whole-image or large-patch based analysis can better learn the implicit dependencies among pixels/voxels at less expense of computational cost compared with small-patch-based approaches [22], [23]. Along this line, end-to-end deep learning models have been proposed to synthesise whole MR images in [7] and whole computed tomography (CT) images in [24], [25].

Moreover, conditional generative adversarial network (cGAN) [26] has recently achieved promising results in generic image synthesis [27]–[29]. Basically, a cGAN model consists of two modules, a generator (e.g., a traditional CNN-based image model) to synthesise images and a discriminator to differentiate the synthesised images from the ground-truth. These two modules compete with each other to achieve a Nash equilibrium. Very recently, cGAN-based image synthesis models have also been successfully applied to medical images, such as retinal images [30]–[32], CT images [33]–[35], PET images [36], [37], MR images [38]–[46], ultrasound images [47], and endoscopy images [48]. Most of these methods work on 2D image synthesis. For 3D medical imaging data, these methods estimate each axial slice independently, and then concatenate them to form a target 3D image. In this way, the coronal and sagittal slices of a synthesised 3D image are formed by the separately estimated lines from the axial planes, and therefore may show strong discontinuities, which impair the synthesis quality of the entire object. To mitigate the slice discontinuity problem, 3D-based cGAN models are proposed in [37], [38], [46]. However, similar to most 2D cGAN models, the 3D-based models in [37], [38] follow the work in [27] to simply minimise the pixel/voxel-wise difference between the synthesised and real images. This neglects the structural content in an image, such as the textures or shapes of objects, leading to less sharp synthesised images. The 3D-based model in [46] imposes additional constraints on the gradient similarity between real and synthesised images so that the sharpness of the synthesised images could be enhanced. Although this work is close to our method proposed in this paper, their fundamental differences will become clearer with the unfolding of our work.

We propose edge-aware generative adversarial networks (Ea-GANs) to further overcome the slice discontinuity and the less sharp synthesis problems in most of existing cGAN models for medical image synthesis. Our methods are 3D-based, and they extract both voxel-wise intensity and image structure information to facilitate synthesis. To capture image structure, we extract the edges that contain critical textural information for visual recognition [5], [49], [50], and integrate the edge maps with the cGAN model to boost synthesis quality. Specifically, two frameworks, a generator-induced Ea-GAN (gEa-GAN) and a more advanced discriminator-induced Ea-GAN (dEa-GAN), are proposed to learn Ea-GANs via different learning strategies. The novelties and contributions of our paper are summarized as follows.

1) To mitigate the issue of incoherent estimation over small patches and across slices, our models provide a 3D based estimation and use the proposed cGAN to conduct cross-modality synthesis at the global level. It alleviates the discontinuous synthesis across slices encountered by the existing 2D cGAN models and captures the image context in a larger scope for the efficient synthesis of high-quality images.

2) More importantly, we point out that purely enforcing the voxel-wise intensity similarity is not sufficient for image synthesis, and propose Ea-GANs to preserve the edge information in order to improve synthesis quality.

3) To integrate the edge information, two variations of Ea-GANs, i.e., gEa-GAN and dEa-GAN, are proposed according to different learning strategies. In the gEa-GAN, edge information is incorporated into the objective function of the generator, enforcing the synthesised image to have a similar edge map as the real image. In the dEa-GAN, the edge information is innovatively incorporated into both generator and discriminator. In this way, the edge information will also be adversarially learned, which could further improve the synthesis performance.

4) Note that, our method is significantly different from the gradient loss based method in [46]. The edge information provided by the Sobel operator is less sensitive to noise and favours nearer neighbours, compared with the direct use of gradient information. More importantly, the gradient information in [46] was only used in the objective function of the generator, but not involved in the adversarial learning like our dEa-GAN. The latter, however, is proven to be a very effective strategy to improve the synthesis quality in this work.

5) The Ea-GANs are validated on two datasets of MR images containing brain lesions and skulls, respectively. The effectiveness of the proposed methods is validated by comparing them with a set of state-of-the-art image synthesis methods [7], [9], [27]. Moreover, to show the generality of our edge-aware approach, we also test the 2D variant of the dEa-GAN on multiple generic 2D image synthesis tasks, which demonstrates consistent improvements over the methods in comparison.

The rest of this paper is organised as follows. We first

introduce the background, especially the cGAN model, in Section II. Then, the two Ea-GAN frameworks, the gEa-GAN and the dEa-GAN, are proposed in Section III. After that, we present their experimental setting, results, and comparison with other methods in Section IV. In Section V, discussions about the reported experimental results are given. Finally, we conclude this paper in Section VI.

II. BACKGROUND

The original generative adversarial network (GAN) was first proposed in 2014 [51]. As a deep generative network, it succeeds in synthesising more realistic images than the traditional deep learning models do. Different from them, a GAN model is composed of a generator G and a discriminator D with an adversarial competition between them. During the training, the GAN model learns a mapping from random noise to an image that conforms to the distribution p_{data} of real images. Since the input is a random noise, there is no specific control on synthesised images with the model. To guide the image synthesis process, conditional generative network (cGAN) was proposed to condition the learning model on input data [26]. When given a source image $x \sim p_{data}(x)$, the cGAN model can be trained to generate x's corresponding target image $y \sim p_{data}(y)$, as a process of paired-image synthesis.

Among the existing cGAN models, Pix2pix [34] is a representative work, which has achieved very promising results for image synthesis, and is a comparison baseline of our work. Therefore, it is briefly introduced as follows. In Pix2pix [27], its generator G synthesises an image $G(x)$ that resembles a real target image y from the input source image x, while its discriminator D tries to distinguish the synthesised image pair $(x, G(x))$ from its corresponding real image pair (x, y). The objective of the generator G is formulated as follows:

$$\mathcal{L}^G_{cGAN} = \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x))) + \lambda_{l1}\mathbb{E}_{x,y \sim p_{data}(x,y)}[\|y - G(x)\|_1], \tag{1}$$

where $\mathbb{E}$ refers to the maximum likelihood estimation, and $G(\cdot)$ and $D(\cdot)$ indicate the outputs of the generator and the discriminator, respectively. In the first term of Equation 1, G tries to generate a realistic image which could be misclassified by D. In the second term of Equation 1, a L1-norm penalty on the pixel-wise intensity difference between the synthesised and real images is utilised to ensure intensity similarity. A hyper-parameter λ_{l1} is used to balance these two terms.

The objective of D in Pix2pix [27] is defined as:

$$\mathcal{L}^D_{cGAN} = -\mathbb{E}_{x,y \sim p_{data}(x,y)}[\log D(x, y)] - \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D(x, G(x)))]. \tag{2}$$

Different from G, D is to assign correct labels to the real and synthesised image pairs. Thus, D and G are following a two-player min-max game in Pix2pix [27].

To integrate the sub-tasks of image generation and discrimination, the final objective of Pix2pix [27] is defined as:

$$\mathcal{L}_{cGAN} = \mathcal{L}^G_{cGAN} + \mathcal{L}^D_{cGAN}. \tag{3}$$

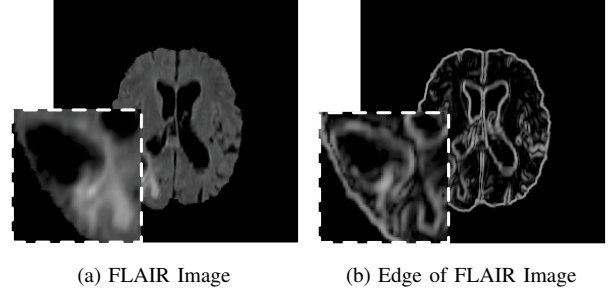(a) FLAIR Image   (b) Edge of FLAIR Image

Fig. 1: A brain FLAIR image (left), and the corresponding edge map (right) after the 3D Sobel edge detection. The contour of abnormal tissues can be depicted clearer by the edge map, which is shown as the zoomed regions.

As shown in the above objective functions, Pix2pix [27] only guarantees the pixel-wise intensity similarity between the synthesised and real images.

III. PROPOSED METHODS

In this section, we present the proposed edge-aware generative adversarial networks (Ea-GANs) for MR image synthesis. Two learning frameworks, the generator-induced Ea-GAN (gEa-GAN) and the discriminator-induced Ea-GAN (dEa-GAN), are presented in details, reflecting the different ways to preserve edge information in the cGAN models in Section III-A. Then, their architectures are presented in Section III-B, and some implementation details to enhance the robustness of Ea-GANs are discussed in Section III-C.

A. Ea-GANs

Most existing cGAN models, like Pix2pix [27], focus on the pixel-to-pixel/voxel-to-voxel image synthesis. They usually enforce the pixel/voxel-wise intensity similarity between the synthesised and real images. However, they ignore the structure of image content, such as the textural details in a MR image [52]. Since edges reflect the local intensity changes and show the boundaries between different tissues in a MR image, maintaining edges can capture the textural structure of image content and help sharpen the synthesised MR images. Especially, when lesions are contained in MR images, the edge information helps differentiate the lesion and the normal tissues, and contributes to better depicting the contour of abnormal regions, e.g. Gliomas tumors in brain MR images [53] (shown in the zoomed parts of Fig. 1). To enforce edge preservation during MR image synthesis, we add an extra constraint based on the similarity of the edge maps from synthesised and real images. The edge maps are computed using the commonly used Sobel operator due to its simplicity, and its derivative can easily be computed for the back-propagation.

As shown in Fig. 2, three Sobel filters, F_i, F_j, and F_k, are used to convolve an image A to generate three edge maps corresponding to the intensity gradients along i, j, and k directions, respectively. Then, these three edge maps are merged into a final edge map $S(A)$ by the following equation:
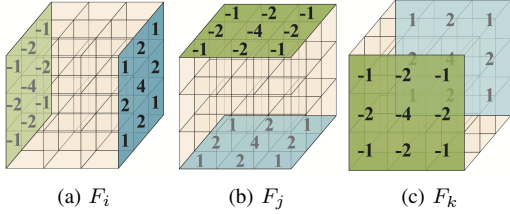
(a) F_i　　　(b) F_j　　　(c) F_k

Fig. 2: The three-dimensional Sobel operator includes three kernels as F_i, F_j, and F_k, respectively. The size of each kernel is $3 \times 3 \times 3$. Each empty cube without any number on its surface means the value of zero in the corresponding position of kernel. Similarly, the numbers in the blue or green cubes are the positive and negative values of three kernels.

$$S(A) = \sqrt{(F_i * A)^2 + (F_j * A)^2 + (F_k * A)^2}, \quad (4)$$

where $*$ means the convolution operation.

Based on different strategies to utilise the edge maps, two frameworks, i.e., gEa-GAN and dEa-GAN, are proposed (as shown in Fig. 3). Each of them consists of three modules, a generator G, a discriminator D, and a Sobel edge detector S. The details of these two frameworks are presented as follows.

1) Generator-induced Ea-GAN: For cross-modality MR image synthesis task, a source-modality image x and a target-modality image y are scanned on the same subject with different contrasts. The generator G of the proposed gEa-GAN aims to synthesise target-modality-like images $G(x)$ that can fool its discriminator D by training with the adversarial loss. Also, the L1-norm penalties are applied through G to discourage the dissimilarity between the real and synthesised images, and between their edge maps extracted by the Sobel edge detector S. In this way, both of the voxel-wise intensity similarity and the edge similarity are enforced during the synthesis. Accordingly, the objective of its generator G is defined as follows:

$$\begin{aligned}
\mathcal{L}_{gEa-GAN}^G = \mathbb{E}_{x \sim p_{data}(x)} &[\log(1 - D(x, G(x))) + \\
&\lambda_{l1} \mathbb{E}_{x,y \sim p_{data}(x,y)}[\|y - G(x)\|_1] + \\
&\lambda_{edge} \mathbb{E}_{x,y \sim p_{data}(x,y)}[\|S(y) - S(G(x))\|_1],
\end{aligned} \quad (5)$$

where the hyper-parameters, λ_{l1} and λ_{edge}, are used to balance the three terms in Equation 5.

Following that in Pix2pix [27], the objective function of the discriminator D is defined as follows:

$$\begin{aligned}
\mathcal{L}_{gEa-GAN}^D = -\mathbb{E}_{x,y \sim p_{data}(x,y)}&[\log D(x, y)] - \\
\mathbb{E}_{x \sim p_{data}(x)}&[\log(1 - D(x, G(x)))].
\end{aligned} \quad (6)$$

Finally, the gEa-GAN integrates its generator G and discriminator D together by training these two modules simultaneously with the following objective:

$$\mathcal{L}_{gEa-GAN} = \mathcal{L}_{gEa-GAN}^G + \mathcal{L}_{gEa-GAN}^D. \quad (7)$$

2) Discriminator-induced Ea-GAN: The gEa-GAN enforces the voxel-wise intensity similarity and the edge similarity by its generator for image synthesis. However, as the edge term only appears on the generator side, edge information is not perceived by the discriminator. Inspired by the mechanism of adversarial learning between generator and discriminator, we further propose a dEa-GAN framework to incorporate the edge maps into this battle. Both generator and discriminator could benefit from the synthesised image and its edge map. Thus, the discriminator is also able to utilise the edge details to differentiate real and synthesised images, and this in turn enforces the generator to produce the better edge details for synthesis.

Similarly to the gEa-GAN, the generator G in the dEa-GAN model is trained using the adversarial loss, the voxel-wise intensity difference loss, and the edge difference loss for synthesis, according to the following objective:

$$\begin{aligned}
\mathcal{L}_{dEa-GAN}^G = \mathbb{E}_{x \sim p_{data}(x)}&[\log(1 - D(x, G(x), S(G(x)))) + \\
&\lambda_{l1} \mathbb{E}_{x,y \sim p_{data}(x,y)}[\|y - G(x)\|_1] + \\
&\lambda_{edge} \mathbb{E}_{x,y \sim p_{data}(x,y)}[\|S(y) - S(G(x))\|_1].
\end{aligned} \quad (8)$$

Compared with the gEa-GAN model, the edge map $S(G(x))$ also implicitly appears in the first term of Equation 8 through the output of the discriminator D.

The objective of the discriminator D now becomes:

$$\begin{aligned}
\mathcal{L}_{dEa-GAN}^D = -\mathbb{E}_{x,y \sim p_{data}(x,y)}&[\log D(x, y, S(y))] - \\
\mathbb{E}_{x \sim p_{data}(x)}&[\log(1 - D(x, G(x), S(G(x))))].
\end{aligned} \quad (9)$$

As can be seen, the discriminator takes a triplet as its input by adding the edge map $S(G(x))$ or $S(y)$. For a synthesised triplet composed of x, $G(x)$, and $S(G(x))$, the label is zero; for a real triplet composed of x, y, $S(y)$, the label is one. The discriminator tries to differentiate these two types of triplets.

Again, the final objective of this dEa-GAN model is:

$$\mathcal{L}_{dEa-GAN} = \mathcal{L}_{dEa-GAN}^G + \mathcal{L}_{dEa-GAN}^D. \quad (10)$$

B. Detailed architectures

The proposed two Ea-GANs are three-dimensional to capture the local and global contextual information along all three directions. Both of them consist of three modules, the generator G, the discriminator D, and the edge detector S. Among these three modules, the edge detector S is the Sobel operator that is presented in Section III-A, and the generator and discriminator are built upon CNN architectures to extract deep features from images. Due to limited GPU memory and the required training batch-size, we utilise large overlapped patches ($128 \times 128 \times 128$) rather than a whole image to train our Ea-GAN models, which could provide a sufficient number of samples to train a good model. The detailed architectures of our generator and discriminator are designed as follows.
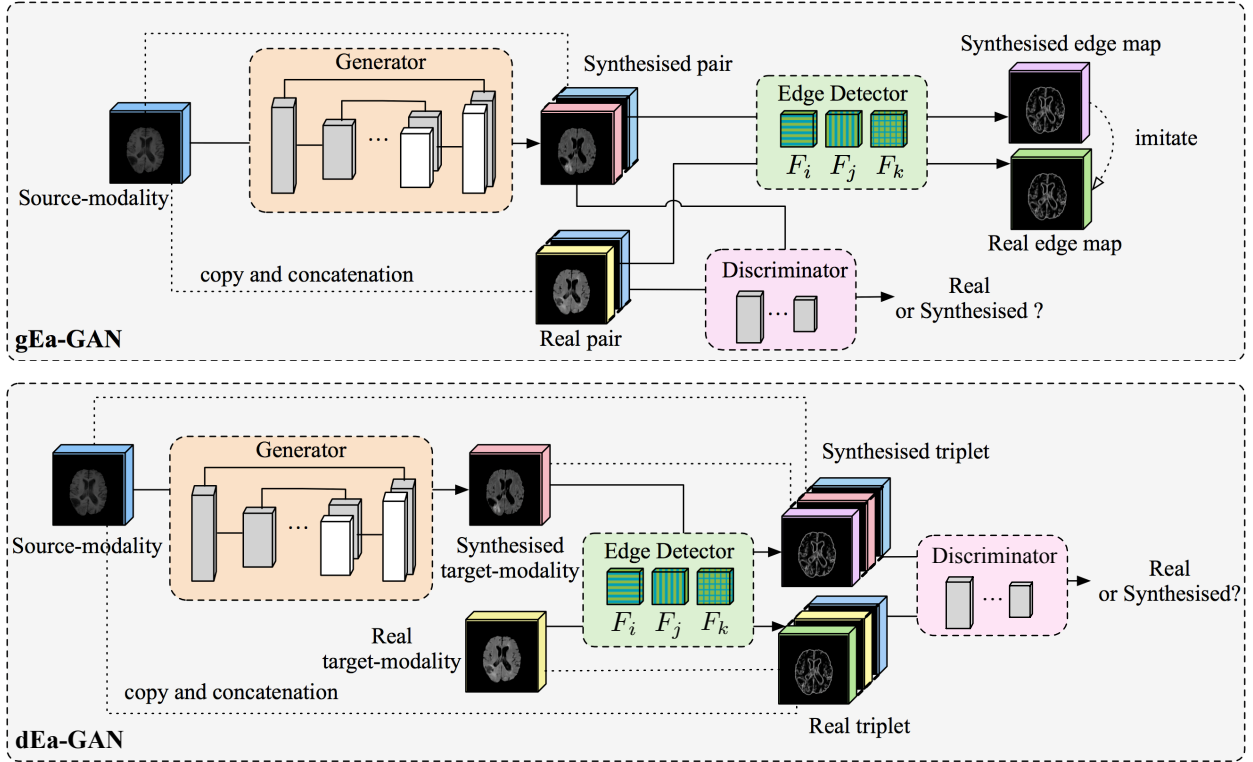
Fig. 3: Frameworks of Ea-GANs. Both gEa-GAN and dEa-GAN include a generator G, a discriminator D, and a Sobel edge detector S. The generator G is trained to synthesise a realistic target-modality image with its edge map detected by the Sobel edge detector S, while the discriminator D is learned to distinguish between the synthesised and real pair/triplet for sharp synthesis. In the back-propagation step of training, the generator G of gEa-GAN is affected by the gradients from the dissimilarity between the synthetic and real edge maps, while both generator G and discriminator D of dEa-GAN are affected by the detected edge maps.

1) Generator architecture: U-net, as a CNN-based model, has been propsoed to analyse whole images or large image patches in the literature [22]. It acquires global contextual information from the input, and ensures the spatial contiguity of the output. The typical characteristic of U-net architecture is the contracting and expanding paths with multiple skip connections between them. Using this structure, U-net can capture the hierarchical features of an input image, and mitigate the gradient vanishing caused by the long back-propagation when training deep networks [22]. It has been extended into 3D variants to better deal with 3D medical images [23], [54]. To benefit from the structure of U-net, we design the generator of Ea-GANs as a 3D U-net-like network. It is symmetric with seven convolutional (conv) blocks in its contracting path and seven up-convolutional (up-cov) blocks in its expanding path. Between each conv block and the corresponding up-conv block, skip connection is applied to capture multi-depth information of source-modality images effectively. The specific construction of this generator is shown in Fig. 4.

2) Discriminator architecture: For the proposed gEa-GAN model, the input of its discriminator is a pair of images, so the discriminator takes in two channels of 3D large patches. Meanwhile, the dEa-GAN model processes triplets with three channels of 3D large patches. Thus, the designed discriminators vary in the two Ea-GAN models, the gEa-GAN and

the dEa-GAN, by their first layers to involve the different numbers of input channels. Our discriminator consists of four conv blocks, another convolutional layer with one kernel of size $1 \times 1 \times 1$, and a sigmoid layer to target the class index one or zero that means real or synthesised. The details of each layer in the discriminator are illustrated in Fig. 5.

C. Implementation

When training a GAN model, a common issue is that it can become unstable or even experience a mode collapse easily [55]. For example, the discriminator tends to be more powerful than the generator, which is reflected by different decreasing speeds of their loss functions. In this case, the whole model is unstable, and cannot synthesise high-quality images. Many techniques have been discussed in [55] to improve the stability of training the GAN models. In our work, we consider two strategies. First, the labels used by the discriminator are smoothed to raise the difficulty of differentiation, and further reduce the vulnerability of adversarial learning. For a synthesised pair/triplet, the target label of the discriminator is set as a random number between 0 and 0.3, while for real pair/triplet, the target label is set as a random value between 0.7 and 1.2, inspired by [56]. In this way, the task of discriminator becomes more challenging to match the difficulty of the task of generator, so that the adversarial
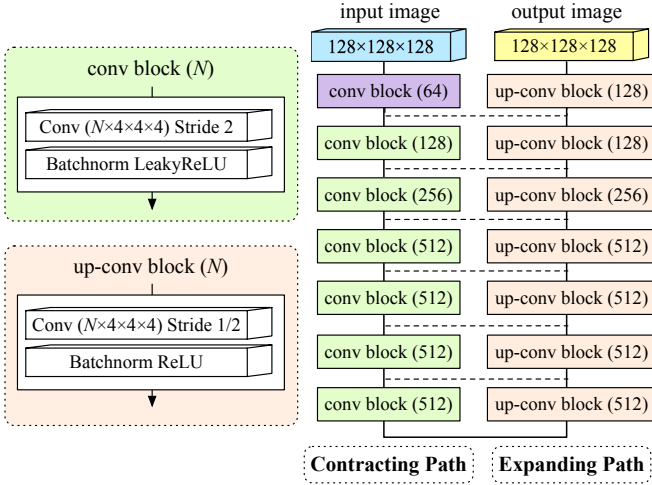
Fig. 4: Generator architecture. All the conv and up-conv blocks contain convolutional, batch normalisation, and ReLU layers. In addition to these three layers, drop-out is applied to the first three blocks on the expanding path. Batch normalisation is not used in the first block of the contracting path. Dashed lines mean the skip connections between contracting and expanding paths by copy and concatenation. The slope of LeakyReLUs is 0.2.
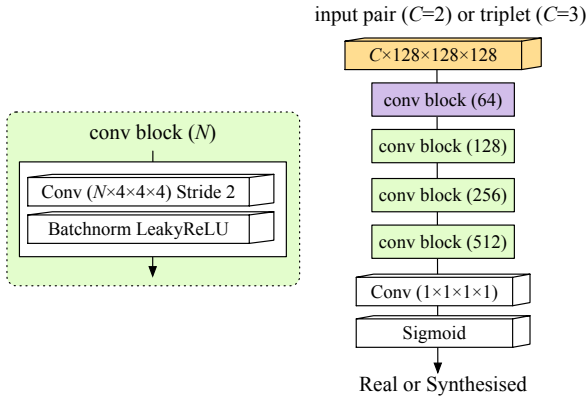

Fig. 5: Discriminator architecture. Convolutional, and LeakyReLU layers with the sllope of 0.2 are applied to all conv blocks. Batch normalisation is not used in the first blocks of D.

training becomes balanced.

The second strategy is used to better utilise the edge information in MR images. At the initial stage of training, the quality of the extracted edge maps, which highly relies on the generated images, is not good enough to effectively guide the synthesis. To mitigate this issue, the value of the hyper-parameter λ_{edge} is initially set to be small and then gradually increased to adjust the importance of edge information. Specifically, we linearly increase λ_{edge} from zero to 100 in the first 20 epochs, and then fix it at 100 in the following epochs. In this way, the Ea-GANs can effectively utilise the edge information to synthesise sharp and realistic target-modality-like images.

IV. EXPERIMENTAL RESULTS

A. Dataset and experimental setting

We use two datasets, i.e. the brain tumor contained BRATS2015 [57] and the non-skull stripped IXI [58], to evaluate two proposed Ea-GANs.

The BRATS2015 dataset consists of 274 subjects with four modalities of co-registered MR images: T1-weighted (T1), T1-weighted and contrast-enhanced (T1c), T2-weighted (T2) and FLAIR, with the image size $240 \times 240 \times 155$ (voxels). In this paper, we use T1 as the source-modality since it is the most commonly used modality for structural imaging, and test two synthesis tasks with FLAIR and T2 as the target-modality, respectively. Five-fold cross-validation is conducted to effectively evaluate different methods on the entire dataset. For each cross-validation split, we divide the dataset into a training set (consisting of 4/5 samples) and a test set (consisting of 1/5 samples). The original intensity values of all used images are linearly scaled to $[-1, 1]$ without any additional contrast change before processed by Ea-GANs. For each image, eight large patches (size: $128 \times 128 \times 128$) are extracted, and the overlapped regions are averaged to form the final estimation. Note that, in order to increase the number of training samples, we use large patches rather than whole images for training, which is essentially different from voxel-wise regression used in the traditional small patch-based synthesis methods.

The IXI dataset includes 578 subjects of the non-skull stripped brain MR images from five modalities, i.e., T1, T2, PD, MRA, and DTI, with the image size $256 \times 256 \times N$ (N is different for each subject). We synthesise T2 images from PD images following [7]. The dataset is utilised by a five-fold cross-validation, so the training set and test set consists of samples from 4/5 and 1/5 subjects, respectively, for each cross-validation split. We also linearly scale the original intensity values into $[-1, 1]$ without any additional contrast change in the pre-processing. For each 3D image, non-overlapped large patches (size: $128 \times 128 \times 128$) are extracted along the trans-coronal and trans-sagittal directions. Along the trans-axial direction, patches are padded with -1 if $N < 128$.

For all synthesis tasks, we conduct 150 epochs to train the models. In the first 100 epochs, the leaning rate is fixed as 0.0002, and then it linearly decays to zero in the next 50 epochs. Adam solver with a batch-size of six is applied to minimise the objectives. During training, λ_{l1} is fixed as 300, while λ_{edge} linearly increases from zero to 100 in the first 20 epochs, and then stays at 100 in the next 130 epochs.

B. Methods in comparison

The proposed Ea-GANs are compared with two state-of-the-art cross-modality MRI synthesis methods, Replica [9] and Multimodal [7], and a generic-image-synthesis cGAN-based model, Pix2pix [27].

1) Replica [9] uses the handcrafted multi-resolution 3D patch features to train random forests for synthesis.
2) Multimodal [7] is a 2D CNN-based model to synthesise the MR image slice by slice with the constraint of pixel-wise intensity difference.

3) Pix2pix [27] is a 2D cGAN model, which synthesises whole 2D images by focusing on maintaining the pixel-wise intensity similarity.

We directly run these three models using their publicly released codes, and follow the original papers for both image pre-processing and model setting. The two 2D models, i.e., Multimodal [7] and Pix2pix [27], are trained using axial slices. Then the synthesised axial slices of each subject are concatenated to form a 3D volume.

Moreover, to facilitate the comparison, we propose a 3D cGAN model, which has the same objective function as Pix2pix [27]. This 3D cGAN model follows the same architecture and parameter setting as the two proposed models, i.e., gEa-GAN and dEa-GAN. All of the 3D models work on large 3D patches to increase the number of training samples.

In addition, to verify the advantages of using edge maps over directly using image gradients, we also build a cGAN model that follows the network architecture of gEa-GAN and dEa-GAN, but use the image gradient difference loss in [46] instead of the edge similarity loss. This model is denoted as a gradient cGAN. To balance each term in the objective function, the added image gradient difference loss is normalized by the number of output voxels and multiplied by 3000.

C. Evaluation measures

Following literatures, we employ three measures in the experiments to evaluate the synthesis performance of the proposed Ea-GAN models and other methods in comparison: peak signal-to-noise ratio (PSNR), normalised mean squared error (NMSE), and structural similarity index (SSIM) [59]. These three evaluation metrics are widely applied to the whole synthesised image [7], [19]. Given the ground-truth target-modality 3D image y and the synthesised 3D image $G(x)$, PSNR is defined as:

$$\text{PSNR}(y, G(x)) = 10 \log_{10} \frac{MAX_{range}^2(y, G(x))}{N_{voxel}^{-1} \|y - G(x)\|_2^2}, \quad (11)$$

where $MAX_{range}(y, G(x))$ means the maximum intensity range of y and $G(x)$, and N_{voxel} denotes the total number of voxels of y or $G(x)$. PSNR is applied to measure the estimation accuracy in terms of the logarithmic decibel scale. Thus, higher PSNR indicates better synthesis performance.

NMSE is calculated as:

$$\text{NMSE}(y, G(x)) = \frac{\|y - G(x)\|_2^2}{\|y\|_2^2}. \quad (12)$$

It is exploited to evaluate the voxel-wise intensity distance between y and $G(x)$, so lower NMSE shows higher synthesis quality.

Different from the above two metrics which evaluate absolute errors between y and $G(x)$, SSIM is used to measure the similarity of the two images considering image degradation as perceived change in structural information. It is computed as:

$$\text{SSIM}(y, G(x)) = \frac{(2\mu_y \mu_{G(x)} + c_1)(2\sigma_{yG(x)} + c_2)}{(\mu_y^2 + \mu_{G(x)}^2 + c_1)(\sigma_y^2 + \sigma_{G(x)}^2 + c_2)}, \quad (13)$$

where μ_y, $\mu_{G(x)}$, σ_y, and $\sigma_{G(x)}$ are the means and variances of image y and $G(x)$, and $\sigma_{yG(x)}$ is the covariance of y and $G(x)$. Higher SSIM demonstrates better estimation.

Before the evaluation, the intensity values of all the synthesised and real images are added by one, and then divided by two. Thus, the intensity values of all images are between zero and one. It can be easily seen that, the value of SSIM is between -1 and 1, while PSNR and NMSE values are unbounded.

D. Results on BRATS2015

The synthesis results on the BRATS2015 dataset are reported, including those from Replica [9], Multimodal [7], Pix2pix [27], 3D cGAN, gradient cGAN, and the two proposed methods of gEa-GAN and dEa-GAN. We show their synthesis performances evaluated on whole images (including a brain and a background) in Table I and on tumor regions in Table II, respectively. To test if the proposed dEa-GAN is statistically significantly better than a compared method, paired t-test is conducted, following [60]–[62]. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined in the tables.

1) Results on the whole images:
- **Comparison between 2D and 3D cGAN**
  As shown in Table I, 3D cGAN which is a 3D variant of Pix2pix [27], significantly outperforms Pix2Pix [27] with better synthesis results. Specifically, this 3D model improves the quality of T1-to-FLAIR synthesis and T1-to-T2 synthesis by (1) 1.8dB PSNR, 0.025 NMSE, and 0.018 SSIM (T1 to FLAIR), and (2) 1.22dB PSNR, 0.015 NMSE, and 0.011 SSIM (T1 to T2), respectively, compared with the 2D model of Pix2Pix [27]. This demonstrates the importance of considering 3D contextual information during the synthesis. When looking into the coronal and sagittal slices in Fig. 6 and Fig. 7, it is found that the discontinuity along these two directions is more salient in Pix2pix [27] than in 3D cGAN (especially the regions between two red arrows in the zoomed parts).
- **Comparison between 3D cGAN and the two proposed Ea-GANs**
  As a model only focusing on the voxel-wise intensity similarity, 3D cGAN produces lower-quality images than both of the proposed Ea-GANs which jointly consider the intensity similarity and edge similarity during training. Concretely, for the synthesis task of FLAIR, PSNR and SSIM increase from 29.26dB (3D cGAN) to 30.11dB (dEa-GAN) and from 0.958 (3D cGAN) to 0.963 (dEa-GAN) respectively, and NMSE decreases from 0.119 (3D cGAN) to 0.105 (dEa-GAN). For T2 synthesis results, dEa-GAN improves PSNR, NMSE, and SSIM by 0.61db, 0.007, and 0.003 from 3D cGAN, respectively. These improvements consistently demonstrate the necessity of preserving edge details in image synthesis. Meanwhile, from all the three views in Fig. 6 and Fig. 7, the proposed two Ea-GANs synthesise sharper edges than the 3D cGAN (indicated by two blue arrows in the zoomed parts).

TABLE I: Quantitative evaluation results of the synthesised FLAIR-like and T2-like images from T1 on the BRATS2015 dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	T1 to FLAIR			T1 to T2		
	PSNR	NMSE	SSIM	PSNR	NMSE	SSIM
Replica [9]	27.17±2.60	0.171±0.267	0.939±0.013	26.92±2.36	0.158±0.324	0.946±0.015
Multimodal [7]	27.26±2.82	0.184±0.284	0.950±0.014	27.31±2.39	0.140±0.229	0.951±0.016
Pix2pix [27]	27.46±2.55	0.144±0.189	0.940±0.015	28.12±2.45	0.110±0.220	0.953±0.014
3D cGAN (ablation study)	29.26±3.21	0.119±0.205	0.958±0.016	29.34±3.23	0.095±0.199	0.964±0.017
gradient cGAN (ablation study)	29.38±3.25	0.116±0.204	0.960±0.017	29.43±3.28	0.097±0.210	0.966±0.017
Proposed gEa-GAN	29.55±3.24	0.115±0.199	0.960±0.017	29.58±3.29	0.093±0.218	0.966±0.018
Proposed dEa-GAN	**30.11±3.22**	**0.105±0.174**	**0.963±0.016**	**29.98±3.37**	**0.088±0.223**	**0.967±0.016**

TABLE II: Quantitative evaluation results of the synthesised FLAIR-like and T2-like tumor parts from T1 on the BRATS2015 dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	T1 to FLAIR			T1 to T2		
	PSNR	NMSE	SSIM	PSNR	NMSE	SSIM
Replica [9]	13.34±3.41	0.137±0.068	0.601±0.083	14.93±3.17	0.123±0.084	0.650±0.139
Multimodal [7]	13.82±3.66	0.131±0.076	0.638±0.096	15.50±3.75	0.109±0.117	0.689±0.138
Pix2pix [27]	14.48±3.12	0.127±0.093	0.618±0.084	16.03±3.10	0.099±0.084	0.703±0.095
3D cGAN (ablation study)	15.95±3.52	0.098±0.094	0.681±0.090	16.79±3.56	0.089±0.093	0.725±0.099
gradient cGAN (ablation study)	15.67±3.63	0.104±0.123	0.682±0.090	16.87±3.40	0.085±0.089	0.752±0.098
Proposed gEa-GAN	16.37±3.49	0.090±0.101	0.697±0.092	17.23±3.50	0.083±0.099	0.752±0.100
Proposed dEa-GAN	**16.90±3.59**	**0.084±0.099**	**0.705±0.093**	**18.02±3.55**	**0.079±0.098**	**0.766±0.098**

TABLE III: PSNR evaluation results of the synthesised edge maps on the BRATS2015 dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	T1 to FLAIR			T1 to T2		
	Sobel	Prewitt	Canny	Sobel	Prewitt	Canny
Replica [9]	31.08±1.85	30.91±1.81	15.40±1.03	30.39±1.59	30.21±1.55	15.20±0.78
Multimodal [7]	31.40±1.95	31.27±1.93	15.61±1.11	30.16±1.72	29.96±1.73	15.18±0.80
Pix2pix [27]	31.64±1.63	31.46±1.60	15.73±0.84	31.51±1.46	31.34±1.47	15.22±0.76
3D cGAN (ablation study)	32.87±1.95	32.73±1.95	16.48±0.95	32.53±1.84	32.37±1.87	15.70±0.90
gradient cGAN (ablation study)	33.06±2.07	32.91±2.05	16.64±0.98	32.63±1.97	32.67±2.00	15.77±0.92
Proposed gEa-GAN	33.10±2.04	32.96±2.03	16.64±0.97	32.87±1.98	32.70±2.01	15.78±0.92
Proposed dEa-GAN	**33.25±2.08**	**33.11±2.08**	**16.72±0.99**	**32.88±1.98**	**32.71±2.01**	**16.16±0.93**

- **Comparison between gradient cGAN and gEa-GAN**
  The gradient cGAN, which directly extracts the image gradient difference for training, performs worse than the proposed gEa-GAN, as shown in Table I. These two methods only differ in whether gradient similarity or edge similarity is used. This indicates the superiority of using the Sobel edge similarity loss over directly using the image gradient loss. The performance is further improved when the edge similarity is also adversarially learned in dEa-GAN.
- **Comparison between the two proposed Ea-GANs**
  When comparing between the two proposed Ea-GANs, the dEa-GAN significantly improves the averaged PSNR and SSIM values by approximately 0.6dB and 0.003 respectively, and lowers the NMSE value by about 0.01 from the gEa-GAN for the FLAIR synthesis task. Similarly, the dEa-GAN improves PSNR, NMSE, and SSIM by 0.4dB, 0.005, and 0.01, respectively, for the T2 synthesis task. Those validate that integrating the edge information into both generator and discriminator can

significantly enhance the learning of edge similarity, and further improves the whole image synthesis performance.
- **Comparison between the state-of-the-art models and the two proposed Ea-GANs**
  When comparing our results with the state-of-the-art methods in literature, Replica [9] obtains the worst PSNR and SSIM evaluation results, which indicates that the small-patch-based method with the handcrafted features may not be able to capture the image contextual information for MR synthesis. For the 2D models, Pix2pix [27] can get slightly better quantitative results than Multimodal [7]. However, when comparing them with the proposed Ea-GANs, our methods outperform these two methods in terms of all the three measures. As shown in the yellow circles of Fig. 6 and Fig. 7, our methods produce the FLAIR-like and T2-like images with the more local details along all the three directions: axial, sagittal, and coronal. It is worth noting that, the gradient cGAN can be regarded as the best performing one in the compared methods, in terms of either the closest mean

TABLE IV: NMSE evaluation results of the synthesised edge maps on the BRATS2015 dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	T1 to FLAIR			T1 to T2		
	Sobel	Prewitt	Canny	Sobel	Prewitt	Canny
Replica [9]	0.381±0.180	0.370±0.180	1.013±0.180	0.452±0.334	0.448±0.335	1.306±0.342
Multimodal [7]	0.372±0.230	0.367±0.232	1.019±0.117	0.445±0.428	0.444±0.436	1.111±0.181
Pix2pix [27]	0.369±0.225	0.366±0.228	1.021±0.250	0.290±0.185	0.284±0.183	1.099±0.177
3D cGAN (ablation study)	0.292±0.229	0.287±0.231	0.971±0.253	0.238±0.172	0.233±0.171	0.992±0.197
gradient cGAN (ablation study)	0.279±0.217	0.275±0.218	0.937±0.245	0.226±0.170	0.222±0.169	0.976±0.211
Proposed gEa-GAN	0.277±0.237	0.274±0.238	0.939±0.252	0.225±0.177	0.221±0.176	0.969±0.206
Proposed dEa-GAN	**0.269±0.225**	**0.266±0.228**	**0.937±0.256**	**0.224±0.173**	**0.220±0.172**	0.954±0.200

TABLE V: SSIM evaluation results of the synthesised edge maps on the BRATS2015 dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	T1 to FLAIR			T1 to T2		
	Sobel	Prewitt	Canny	Sobel	Prewitt	Canny
Replica [9]	0.938±0.008	0.939±0.008	0.897±0.016	0.954±0.008	0.955±0.008	0.900±0.015
Multimodal [7]	0.947±0.012	0.949±0.014	0.905±0.016	0.955±0.014	0.956±0.013	0.900±0.016
Pix2pix [27]	0.959±0.009	0.960±0.008	0.909±0.015	0.963±0.009	0.964±0.009	0.905±0.015
3D cGAN (ablation study)	0.965±0.010	0.966±0.010	0.923±0.015	0.948±0.010	0.969±0.010	0.913±0.017
gradient cGAN (ablation study)	0.965±0.010	0.966±0.010	0.925±0.015	0.970±0.011	0.971±0.010	0.915±0.017
Proposed gEa-GAN	0.968±0.010	0.969±0.010	0.924±0.015	0.970±0.010	0.971±0.010	0.915±0.017
Proposed dEa-GAN	**0.968±0.010**	**0.970±0.010**	**0.926±0.016**	**0.971±0.010**	**0.972±0.010**	**0.915±0.017**

TABLE VI: Quantitative evaluation results of the synthesised T2-like images and their edge maps from PD on the IXI dataset (mean±standard deviation). The paired t-test is conducted between dEa-GAN and a compared method at the significance level of 0.05. When the improvement of dEa-GAN over the method is statistically significant, the result of that compared method will be underlined.

Methods	Whole image			PSNR of edge maps		
	PSNR	NMSE	SSIM	Sobel	Prewitt	Canny
Replica [9]	27.99±1.65	0.087±0.034	0.947±0.013	29.43±1.77	29.35±1.76	13.33±0.57
Multimodal [7]	29.77±2.46	0.078±0.080	0.953±0.019	30.87±1.96	30.76±1.97	14.25±0.72
Pix2pix [27]	30.80±1.90	0.054±0.031	0.964±0.012	31.62±1.49	31.59±1.45	14.48±0.72
3D cGAN (ablation study)	32.10±2.02	0.038±0.039	0.973±0.011	32.61±1.50	32.56±1.48	14.58±0.77
gradient cGAN (ablation study)	32.51±2.18	0.036±0.036	0.974±0.011	32.91±1.69	32.86±1.67	14.59±0.77
Proposed gEa-GAN	32.81±2.11	0.035±0.036	0.975±0.010	33.10±1.60	33.06±1.59	14.79±0.78
Proposed dEa-GAN	**33.25±2.20**	**0.031±0.032**	**0.977±0.011**	**33.59±1.69**	**33.55±1.67**	**15.20±0.78**

Methods	NMSE of edge maps			SSIM of edge maps		
	Sobel	Prewitt	Canny	Sobel	Prewitt	Canny
Replica [9]	0.154±0.039	0.153±0.039	0.682±0.080	0.947±0.013	0.949±0.012	0.860±0.018
Multimodal [7]	0.124±0.072	0.124±0.074	0.554±0.086	0.956±0.014	0.958±0.014	0.884±0.020
Pix2pix [27]	0.097±0.041	0.096±0.041	0.538±0.087	0.963±0.009	0.964±0.009	0.886±0.020
3D cGAN (ablation study)	0.079±0.044	0.078±0.045	0.492±0.082	0.968±0.009	0.969±0.009	0.897±0.018
gradient cGAN (ablation study)	0.076±0.037	0.075±0.037	0.463±0.074	0.970±0.010	0.971±0.009	0.903±0.017
Proposed gEa-GAN	0.067±0.039	0.066±0.039	0.454±0.074	0.973±0.009	0.974±0.008	0.904±0.017
Proposed dEa-GAN	**0.064±0.034**	**0.063±0.034**	**0.445±0.073**	**0.975±0.009**	**0.976±0.008**	**0.906±0.017**

or the smallest average difference to our proposed dEa-GAN. Still, our dEa-GAN performs statistically significantly better than it as analysed above.

2) Results on tumor regions: The above seven methods are also compared on the lesion-contained regions in Table II. The Ea-GANs obtain the best values of PSNR, NMSE, and SSIM over all the methods in comparison. This is consistent with our observations on whole images. Also, the dEa-GAN shows statistically significant improvements over the best-performing one, i.e., the gradient cGAN, among the compared methods. These show the capacity of the proposed Ea-GANs on preserving the critical pathological information in the synthesised images, since such information could be correlated to edges.

E. Results on IXI dataset

As can be seen in Table VI, the proposed Ea-GANs outperform the other five methods in comparison according to all the three measures. The proposed dEA-GAN model demonstrates considerable improvements, with NMSE dropping from 0.087 (Replica) to 0.031 (dEa-GAN), SSIM rising from 0.947 (Replica) to 0.977 (dEa-GAN), and PSNR rising from 27.99dB (Replica) to 33.25dB (dEa-GAN), respectively. The second best is the proposed gEa-GAN model. These results validate that the two proposed Ea-GANs can also synthesise the non-skull stripped MR images with higher quality. Example images are shown in Fig. 8. Although all the methods produce the high-quality synthesised T2 images, the visual results generated by the two proposed Ea-GANs
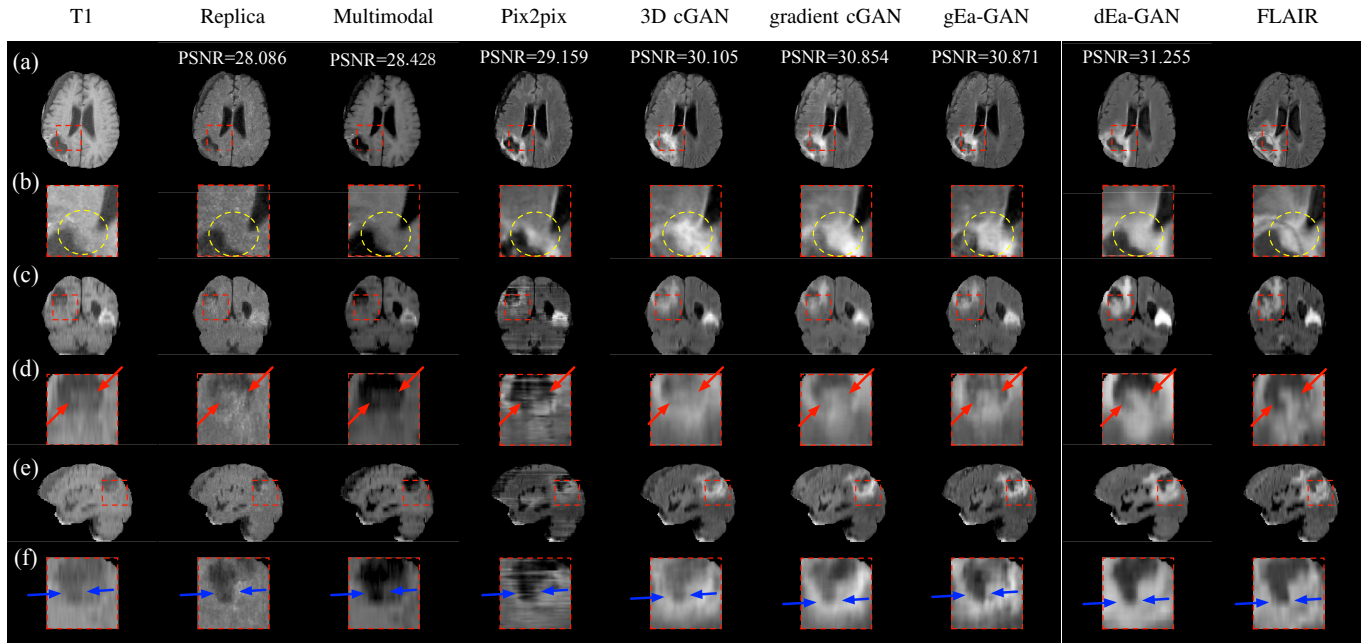
Fig. 6: Comparison between the two proposed Ea-GANs and other state-of-the-art methods (T1 to FLAIR on the BRATS2015 dataset): (a) axial slices, (b) zoomed parts of axial slices, (c) coronal slices, (d) zoomed parts of coronal slices, and (e) sagittal slices, (f) zoomed parts of sagittal slices.

show the sharper edges (indicated by the two red arrows in the zoomed parts), which is consistent with the observation from quantitative evaluation.

F. Results on synthesised image edge maps

To directly show the edge preserving performance of the proposed Ea-GANs, three kinds of edge maps, i.e. Sobel, Prewitt, and Canny binary edge maps, extracted from the synthesised and real images are compared via PSNR, NMSE, and SSIM in Tables III, IV, V, and VI. As shown, the proposed Ea-GANs produce the edge maps that are closest to the ground-truth. These results directly verify the effectiveness of maintaining edge similarity by the two proposed Ea-GANs.

G. Generic image synthesis results

To evaluate the generality and the effectiveness of our edge-aware approach, we extend the dEa-GAN into its 2D variant, 2D dEa-GAN, and compare it with Pix2pix [27]. Following the literature, three generic image-to-image translation benchmark datasets are used. For the facades dataset [63], the label-to-photo translation is conducted with 400 training samples and 206 test samples. For the maps dataset, 1096 training images and 1098 test images that are scraped by [27] are used, and the map-to-aerial translation is processed. For the cityscapes dataset [64], photos are synthesised from the cityscapes labels with 2975 training images and 500 test images. All the image pre-processing steps and experimental setting follow the work in [27].

In Table. VII, we report the quantitative comparisons of the synthesised images by Pix2pix [27] and our 2D dEa-GAN

model on these three datasets. As consistently seen, our 2D dEa-GAN significantly outperforms Pix2pix [27] with higher PSNR and SSIM, and lower NMSE. It validates that preserving edge information is essential for the different generic image synthesis tasks.

TABLE VII: PSNR, NMSE, and SSIM on the generic image synthesis datasets (mean±standard deviation). The paired t-test is conducted between dEa-GAN and Pix2pix to the significance level of 0.05. When the improvement of dEa-GAN is statistically significant, the result of Pix2pix will be underlined.

Methods	facades		
	PSNR	NMSE	SSIM
Pix2pix [27]	13.21±1.71	0.993±0.056	0.246±0.079
Proposed 2D dEa-GAN	**13.36±1.67**	**0.984±0.061**	**0.260±0.083**

Methods	maps		
	PSNR	NMSE	SSIM
Pix2pix [27]	15.06±2.08	0.878±0.065	0.203±0.083
Proposed 2D dEa-GAN	**15.60±2.09**	**0.870±0.068**	**0.237±0.083**

Methods	cityscapes		
	PSNR	NMSE	SSIM
Pix2pix [27]	15.98±2.41	0.851±0.097	0.421±0.083
Proposed 2D dEa-GAN	**16.46±2.55**	**0.844±0.100**	**0.435±0.086**

V. DISCUSSION

Our work aims to synthesise high-quality MR images by cGAN-based models. The proposed 3D-based Ea-GANs enforce the voxel-wise intensity similarity during training, and additionally integrate the edge maps as the image contextual information to improve synthesis performance. Two strategies
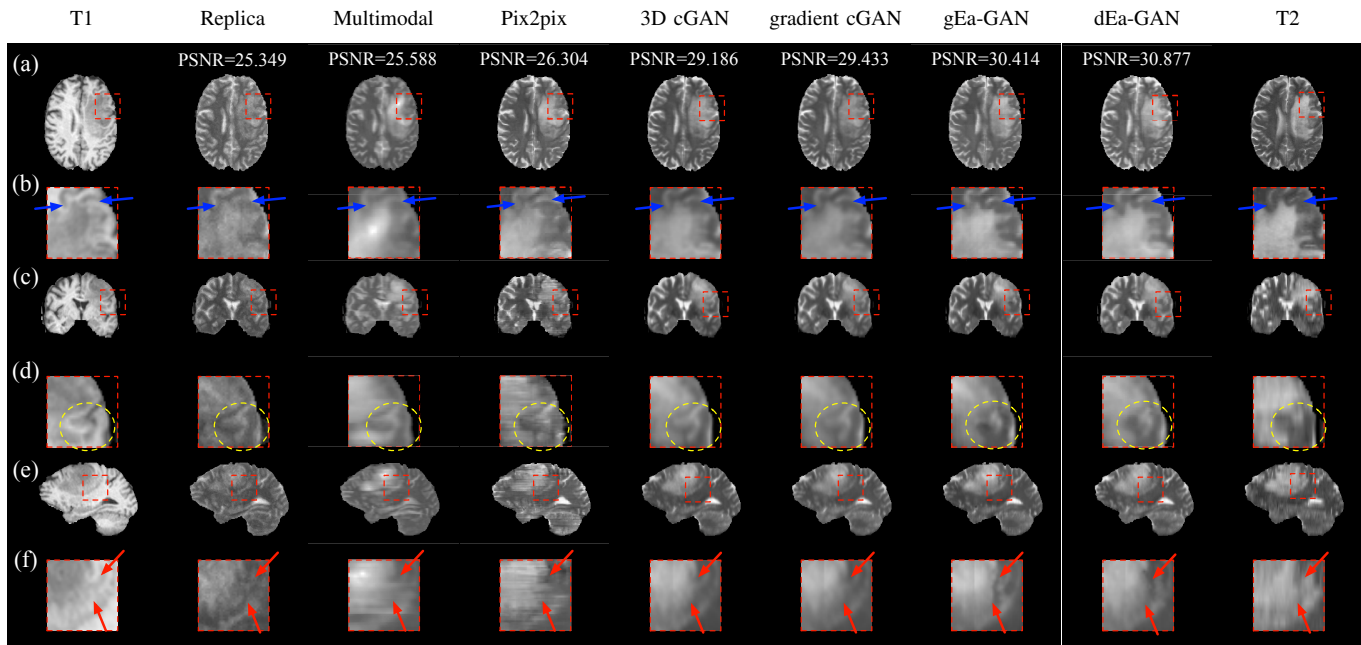
Fig. 7: Comparison between the two proposed Ea-GANs and other state-of-the-art methods (T1 to T2 on the BRATS2015 dataset): (a) axial slices, (b) zoomed parts of axial slices, (c) coronal slices, (d) zoomed parts of coronal slices, and (e) sagittal slices, (f) zoomed parts of sagittal slices.

have been proposed for this purpose. The first one is the proposed gEa-GAN model which extracts the Sobel edge maps of both synthesised and real target-modality images, and minimises their distance during the training of generator. The second one is the proposed dEa-GAN. It further enforces this edge similarity via the adversarial learning between generator and discriminator. Our experimental results fully demonstrate the importance of perceiving the edge details during synthesis with the consistent improvements in terms of different evaluation measures, and across all the datasets that have been tested. Moreover, by jointly acquiring the edge information via both of the generator and discriminator, the dEa-GAN is found to also outperform the proposed gEa-GAN that only incorporates edge details on the generator side.

In addition, our Ea-GANs prove to also work well in lesion regions, which beats all the other methods in comparison according to three measurements on brain tumor synthesis. Last but not the least, edge-aware GANs well generalise to other generic image synthesis tasks, as shown on a variety of benchmark datasets about facades, maps, and cityscapes, demonstrating the power of our Ea-GAN model as a general image synthesiser.

Most previous works for cGAN-based medical image synthesis [30]–[48] are 2D models that separately estimate each slice along the image's trans-axial direction. They ignore the 3D image contextual information and result in the discontinuous estimation. To overcome this limitation, the idea of using 3D conditional GANs has been exploited in some existing works [37], [38], [46]. However, there are distinct differences between those methods and the proposed models. First, [37], [38] only maintain the voxel-wise similarity during training.

In contrast, our work innovatively explores the edge-aware idea in cGAN models to synthesise the higher-quality images. Second, whereas a gradient difference loss is used in [46], we apply a Sobel operator to extract the edge details. Applying the Sobel filters has more advantages than directly using the image gradients. By the averaging operation of Sobel operator, the filter is less sensitive to noise than directly using the image gradients. Also, the Sobel filter assigns higher weights to its nearer neighbours and lower weights to its farther neighbours, which cannot be achieved by directly utilising image gradients. The superiority of our gEa-GAN over the gradient cGAN justifies that the Sobel filters are more effective than the simple image gradients for MR image synthesis. More importantly, we innovatively integrate the edge information into adversarial learning (rather than simply putting it in the cost function of generator) to significantly improve the synthesis quality, which was not touched at all in [46].

Our current work has the following limitations. First, due to the limited number of training images, the validation set is not used to optimally choose training parameters. To increase the practical number of training data, large 3D patches rather than whole images are used for training. In our future work, we will try to augment more samples into the study to choose parameters, and evaluate our methods on 3D whole image synthesis. Second, enhancing the edge similarity is only one of the potential approaches to maintaining image content. In our current study, due to the difficulty of optimisation, some other image structure related measurements, such as local normalised cross-correlation, are not studied. This would be thoroughly explored in our future study. Third, current MR image synthesis frameworks usually require the training
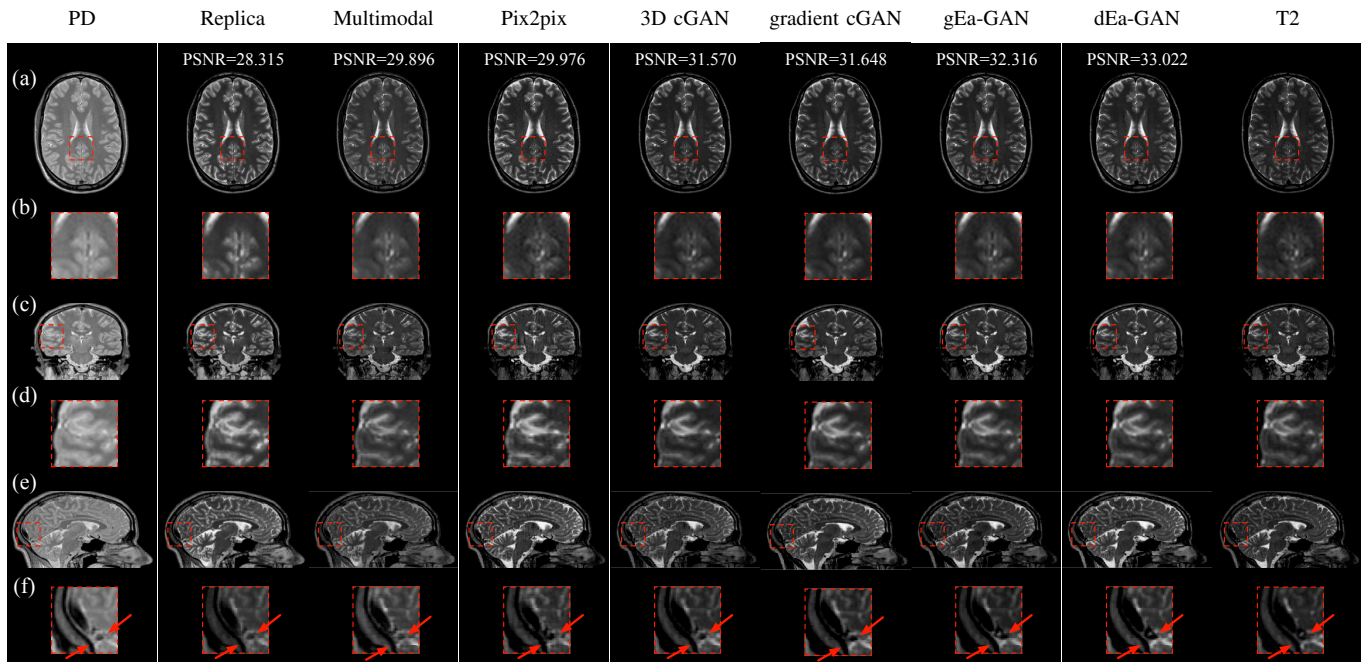
Fig. 8: Comparison between the two proposed Ea-GANs and other state-of-the-art methods (PD to T2 on the IXI dataset): (a) axial slices, (b) zoomed parts of axial slices, (c) coronal slices, (d) zoomed parts of coronal slices, and (e) sagittal slices, (f) zoomed parts of sagittal slices.

samples to have both source- and target-modalities. However, such a requirement is often difficult to meet in many medical applications, which further reduces the number of usable training samples, and limits the performance of models for synthesis. In our future work, we are going to relax this requirement by developing the semi-supervised Ea-GANs to make our models more adaptive to various scenarios in medical image synthesis. Fourth, our work demonstrates the effectiveness of the proposed methods in conducting learning in the same dataset. In the future, we will try to exploit domain adaptation techniques to further extend them to tackle cross-dataset learning problems.

VI. CONCLUSION

In this paper, we proposed two novel end-to-end frameworks of Ea-GANs, i.e., gEa-GAN and dEa-GAN, to synthesise the target-modality MR images from the given source-modality. Jointly preserving the voxel-wise intensity similarity and the edge similarity during synthesis, the proposed Ea-GANs, especially the dEa-GAN, achieve significantly better results than multiple state-of-the-art methods for MR image synthesis. In addition, the excellent generality of dEa-GAN has been demonstrated on different generic image synthesis tasks.

REFERENCES

[1] L. Clarke, R. Velthuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, and M. Silbiger, "Mri segmentation: methods and applications," *Magnetic resonance imaging*, vol. 13, no. 3, pp. 343–368, 1995.

[2] C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. OBrien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen, "Prediction of ad with mri-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1397, 1999.

[3] M. Dadar, T. A. Pascoal, S. Manitsirikul, K. Misquitta, V. S. Fonov, M. C. Tartaglia, J. Breitner, P. Rosa-Neto, O. T. Carmichael, C. Decarli *et al.*, "Validation of a regression technique for segmentation of white matter hyperintensities in alzheimer's disease," *IEEE transactions on medical imaging*, vol. 99, pp. 1–1, 2017.

[4] M. Lê, H. Delingette, J. Kalpathy-Cramer, E. R. Gerstner, T. Batchelor, J. Unkelbach, and N. Ayache, "Personalized radiotherapy planning based on a computational tumor growth model," *IEEE transactions on medical imaging*, vol. 36, no. 3, pp. 815–825, 2017.

[5] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.

[6] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[7] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, "Multi-modal mr synthesis via modality-invariant latent representation," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 803–814, 2018.

[8] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, "Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 606–613.

[9] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Medical image analysis*, vol. 35, pp. 475–488, 2017.

[10] N. Cordier, H. Delingette, M. Lê, and N. Ayache, "Extended modality propagation: image synthesis of pathological cases," *IEEE transactions on medical imaging*, vol. 35, no. 12, pp. 2598–2608, 2016.

[11] G. van Tulder and M. de Bruijne, "Why does synthesized data improve multi-sequence classification?" in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 531–538.

[12] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding," *arXiv preprint arXiv:1705.02596*, 2017.

[13] M. Hofmann, F. Steinke, V. Scheel, G. Charpiat, J. Farquhar, P. Aschoff, M. Brady, B. Scholkopf, and B. J. Pichler, "Mri-based attenuation correction for pet/mri: a novel approach combining pattern recognition

and atlas registration," *Journal of nuclear medicine*, vol. 49, no. 11, p. 1875, 2008.

[14] M. Hofmann, I. Bezrukov, F. Mantlik, P. Aschoff, F. Steinke, T. Beyer, B. J. Pichler, and B. Schölkopf, "Mri-based attenuation correction for whole-body pet/mri: quantitative evaluation of segmentation-and atlas-based methods," *Journal of Nuclear Medicine*, vol. 52, no. 9, pp. 1392–1399, 2011.

[15] S. Roy, A. Carass, and J. L. Prince, "Magnetic resonance image example-based contrast synthesis," *IEEE transactions on medical imaging*, vol. 32, no. 12, pp. 2348–2363, 2013.

[16] N. Burgos, M. J. Cardoso, K. Thielemans, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, R. Ahmed, C. J. Mahoney, J. M. Schott *et al.*, "Attenuation correction synthesis for hybrid pet-mr scanners: application to brain studies," *IEEE transactions on medical imaging*, vol. 33, no. 12, pp. 2332–2341, 2014.

[17] M. Chen, A. Jog, A. Carass, and J. L. Prince, "Using image synthesis for multi-channel registration of different image modalities," in *Medical Imaging 2015: Image Processing*, vol. 9413. International Society for Optics and Photonics, 2015, p. 94131Q.

[18] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian, and D. Shen, "Estimating ct image from mri data using structured random forest and auto-context model," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 174–183, 2016.

[19] Y. Wang, G. Ma, L. An, F. Shi, P. Zhang, D. S. Lalush, X. Wu, Y. Pu, J. Zhou, and D. Shen, "Semisupervised tripled dictionary learning for standard-dose pet image prediction using low-dose pet and multimodal mri," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 3, pp. 569–579, 2017.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[23] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[24] J. Fu, Y. Yang, K. Singhrao, D. Ruan, D. A. Low, and J. H. Lewis, "Male pelvic synthetic ct generation from t1-weighted mri using 2d and 3d convolutional neural networks," *arXiv preprint arXiv:1803.00131*, 2018.

[25] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2017, pp. 14–23.

[26] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv preprint arXiv:1611.07004*, 2016.

[28] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint arXiv:1703.10593*, 2017.

[29] H. Huang, P. S. Yu, and C. Wang, "An introduction to image synthesis with generative adversarial nets," *arXiv preprint arXiv:1803.04469*, 2018.

[30] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 781–791, 2018.

[31] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Medical image analysis*, vol. 49, pp. 14–26, 2018.

[32] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," *arXiv preprint arXiv:1709.01872*, 2017.

[33] X. Yi and P. Babyn, "Sharpness-aware low dose ct denoising using conditional generative adversarial network," *arXiv preprint arXiv:1708.06453*, 2017.

[34] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using cyclegan: Effects of gradient consistency loss and training data size," *arXiv preprint arXiv:1803.06629*, 2018.

[35] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Mr-to-ct synthesis using cycle-consistent generative adversarial networks," 2017.

[36] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, "Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans)," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. Springer, 2017, pp. 43–51.

[37] Y. Wang, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3d conditional generative adversarial networks for high-quality pet image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, 2018.

[38] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, "3d cgan based cross-modality mr image synthesis for brain tumor segmentation," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 626–630.

[39] S. U. H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *arXiv preprint arXiv:1802.01221*, 2018.

[40] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *International Conference on Artificial Neural Networks*. Springer, 2017, pp. 626–634.

[41] H. Choi and D. S. Lee, "Generation of structural mr images from amyloid pet: Application to mr-less quantification," *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1111–1117, 2018.

[42] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I. Chang, Y. Xu *et al.*, "Mri image-to-image translation for cross-modality image registration and segmentation," *arXiv preprint arXiv:1801.06940*, 2018.

[43] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "Gan-based synthetic brain mr image generation," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 734–738.

[44] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, "Adversarial synthesis learning enables segmentation without target modality ground truth," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 1217–1220.

[45] M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur, S. Han, W. Dally, J. M. Pauly *et al.*, "Deep generative adversarial networks for compressed sensing automates mri," *arXiv preprint arXiv:1706.00051*, 2017.

[46] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Transactions on Biomedical Engineering*, 2018.

[47] Y. Hu, E. Gibson, L.-L. Lee, W. Xie, D. C. Barratt, T. Vercauteren, and J. A. Noble, "Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*. Springer, 2017, pp. 105–115.

[48] F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised reverse domain adaption for synthetic medical images via adversarial training," *arXiv preprint arXiv:1711.06606*, 2017.

[49] K. Suzuki, I. Horiba, and N. Sugie, "Neural edge enhancer for supervised edge enhancement from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1582–1596, 2003.

[50] H. Khotanlou, O. Colliot, J. Atif, and I. Bloch, "3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy sets and systems*, vol. 160, no. 10, pp. 1457–1473, 2009.

[51] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[52] A. Bustin, D. Voilliot, A. Menini, J. Felblinger, C. de Chillou, D. Burschka, L. Bonnemains, and F. Odille, "Isotropic reconstruction of mr images using 3d patch-based self-similarity learning," *IEEE Transactions on Medical Imaging*, 2018.

[53] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, T. Okuda, L. Liang, Y. Ge, Y. Komohara *et al.*, "Usefulness of diffusion-weighted mri with echo-planar technique in the evaluation of cellularity in gliomas," *Journal of magnetic resonance imaging*, vol. 9, no. 1, pp. 53–60, 1999.

[54] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D*

Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016, pp. 565–571.

[55] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[56] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.

[57] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[58] I. C. London. The ixi dataset. [Online]. Available: http://brain-development.org/ixi-dataset/

[59] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[60] Y. Huang, L. Shao, and A. F. Frangi, "Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 815–827, 2018.

[61] K.-H. Thung, P.-T. Yap, E. Adeli, S.-W. Lee, and D. Shen, "Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion," *Medical image analysis*, vol. 45, pp. 68–82, 2018.

[62] L. Fang, L. Zhang, D. Nie, X. Cao, I. Rekik, S.-W. Lee, H. He, and D. Shen, "Automatic brain labeling via multi-atlas guided fully convolutional networks," *Medical image analysis*, 2018.

[63] R. Tyleček and R. Šára, "Spatial pattern templates for recognition of objects with regular structure," in *German Conference on Pattern Recognition.* Springer, 2013, pp. 364–374.

[64] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.