Medical Image Synthesis via Deep Learning

Biting Yu, Yan Wang, Lei Wang, Dinggang Shen, and Luping Zhou

Abstract

Medical images have been widely used in clinics, providing visual representations of under-skin tissues in human body. By applying different imaging protocols, diverse modalities of medical images with unique characteristics of visualization can be produced. Considering the cost of scanning high-quality single modality images or homogeneous multiple modalities of images, medical image synthesis methods have been extensively explored for clinical applications. Among them, deep learning approaches, especially convolutional neural networks (CNNs) and generative adversarial networks (GANs), have rapidly become dominating for medical image synthesis in recent years. In this chapter, based on a general review of the medical image synthesis methods, we will focus on introducing typical CNNs and GANs models for medical image synthesis. Especially, we will elaborate our recent work about low-dose to high-dose PET image synthesis, and cross-modality MR image synthesis, using these models.

B. Yu · L. Wang
School of Computing and Information Technology, University of Wollongong, Wollongong, NSW, Australia

Y. Wang
School of Computer Science, Sichuan University, Chengdu, China

D. Shen
IDEA Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

L. Zhou (✉)
School of Electrical and Information Engineering, University of Sydney, Camperdown, NSW, Australia
e-mail: luping.zhou@sydney.edu.au

Introduction

As a technology to produce the visual representations of anatomical and pathological structures and their functions in human body, medical imaging is widely applied in clinics for disease diagnosis and treatment planning. It consists of various imaging protocols which have their specific insights to produce different modality images. For example, computed tomography (CT)

creates the images of internal organs, bones, soft tissue, and blood vessels to show electron density and physical density [22]; magnetic resonance imaging (MRI) provides diverse contrasts of soft tissues through setting different scanning parameters [33]; positron emission tomography (PET) enables the visualization of metabolic processes of scanned body part [12]. Due to their different and sometimes complementary characteristics, multiple modalities are usually utilized in the analysis of clinical applications. However, the acquisition of some modality images, such as PET and CT, increases the risks of radiation exposure, especially when patients have to take these imaging scanning multiple times during the entire treatment [37]. Thus, the high-quality and human-safe medical images are not easy to acquire in the practical applications. Besides, due to the different imaging protocols and the cost of multi-modality image acquisition, sufficient and consistent modalities are not always accessible for every patient [57]. To handle these issues, medical image synthesis, which is defined as an approach to modeling a mapping from the given source images to the unknown target images, has been widely explored by researchers [18].

Medical image synthesis has been used in various applications, e.g., estimation of missing images [43], knowledge transformation across modalities [53], image super-resolution [24], and annotated dataset creation [14]. Here, according to its applications, we roughly classify medical image synthesis into two main categories, i.e., within-modality synthesis and cross-modality synthesis. Specially, the within-modality synthesis usually aims at generating the higher-quality images from the input within-modality images of relatively lower quality. In contrast, the cross-modality synthesis targets to capture the useful structuring information in the source-modality to generate the target-modality image. Although these two categories are applied in different practical tasks, the underlying synthesis principles are similar. The conventional synthesis approaches exploit diverse nonlinear models, e.g., dictionary learning [45] and random forest [26], to process the handcrafted medical image features which

are manually selected by professional experts during the synthesis. However, these handcrafted features have limited power to represent the complex visual information in medical images and therefore adversely affect the synthesis performance. Recently, deep learning based methods have mitigated this issue through automatically learning the task-specific features having sufficient descriptive power with the training of the mapping models [41, 55]. Through designing advanced deep learning models, the performance of medical image synthesis has been greatly improved.

In Table 1, a list of works that utilized deep learning models for medical image synthesis are presented. Here, we mainly focus on the synthesis applications for three major imaging modalities, i.e., CT, MR, and PET. The timeline for the development of these methods is summarized in Fig. 1. As shown in Table 1 and Fig. 1, deep learning approaches started to be popular for medical image synthesis in 2015 [42]. After two years of exploration, a large category of models, especially deep convolutional neural networks (CNNs) based architectures, became dominating for both within-modality and cross-modality synthesis in 2017 [10, 11, 20, 28, 31, 32, 51]. Before the end of 2017, a novel family of CNN based models, i.e., generative adversarial networks (GANs), attracted the attention of researchers and achieved promising results [3, 4, 7]. In 2018, more complicated CNN models were further explored in the conventional way [8, 9, 58]. At the same time, numerous GAN models with different frameworks were proposed in 2018 and 2019, and this research trend becomes more and more popular now.

In the rest of this chapter, we first discuss two typical types of deep learning models for medical image synthesis in section "Deep Learning Models for Medical Image Synthesis". Following that, we introduce four of our recent works for within-modality and cross-modality synthesis, respectively, in sections "Within-Modality Synthesis" and "Cross-Modality Synthesis." Finally, a brief conclusion about this chapter is given in section "Conclusion."

Table 1 Medical image synthesis publications with deep learning models

Publication	Method	Dataset	Organ
Within-modality synthesis			
CT (low-dose to full-dose)			
Chen et al. [11]	Custom three layer-CNN	NBIA[a]	Multiple body parts
Chen et al. [10]	CNN based residual encoder-decoder	NBIA[a]	Multiple body parts
Kang et al. [28]	CNN for wavelet domain denoising	low-dose CT[b]	Head, chest and abdomen
MRI (super-resolution or 3T to 7T)			
Zend et al. [58]	Residual CNN	Brainweb[c] NAMIC[d]	Brain
Chaudhari et al. [9]	Residual CNN	OAI [39]	Knee
Nie et al. [37]	Cascade GANs	–	Brain
PET (low-dose to full-dose)			
Xiang et al. [51]	Cascade CNNs	–	Brain
Wang et al. [46]	3D cGAN	–	Brain
Wang et al. [47]	3D cGAN with locality-adaptive module	–	Brain
Cross-modality synthesis			
MR to CT or CT to MR			
Nie et al. [36]	3D CNN-FCN	–	Pelvic
Han et al. [20]	U-net	–	Brain
Leynes et al. [31]	U-net	–	Pelvic
Liu et al. [32]	CNN based autoencoder	–	Brain
Chartsias et al. [7]	cycleGAN	MM-WHS[f]	Cardiac
Nie et al. [37]	Cascade GANs	ADNI[e]	Brain and pelvic
Emami et al. [17]	cGAN	–	Brain
Hiasa et al. [21]	cycleGAN with gradient loss	–	Musculoskeletal
Zhang et al. [59]	cycleGAN with segmentors	–	Cardiac
CT to PET or PET to CT			
Ben et al. [3]	FCN-cGAN	–	Liver
Bi et al. [4]	cGAN with tumor label input	–	Thorax
Armanious et al. [2]	cGAN with CasNet generator	–	Brain
MR to PET or PET to MR			
Choi et al. [13]	cGAN (pix2pix)	ADNI[e]	Brain
Wei et al. [49]	Cascade GANs	–	Brain
Cross-modality MR (T1, T2, FLAIR, and MRA)			
Van et al. [42]	Location-sensitive CNN	NAMIC[d]	Brain
Chartsias et al. [8]	CNN based encoder and decoder	ISLES2015[g] BRATS2015[h] IXI[i]	Brain
Dar et al. [16]	cGAN (pix2pix)	MIDAS [5] BRATS2015[h] IXI[i]	Brain
Olut et al. [38]	cGAN	IXI[i]	Brain
Mok et al. [35]	cGAN (two generators and four multi-scale discriminators)	BRATS2015[h]	Brain

Table 1 (continued)

Publication	Method	Dataset	Organ
Yang et al. [52]	cGAN	BRATS2015[h]	Brain
Welander et al. [50]	cycleGAN and UNIT	Human Connectome[j]	Brain
Yu et al. [56]	3D cGAN	BRATS2015[h]	Brain
Yu et al. [57]	3D cGAN with edge map adversarial learning	BRATS2015[h] IXI[i]	Brain

[a] https://dcm.bmia.nl/ncia/login.jsf
[b] https://www.aapm.org/GrandChallenge/LowDoseCT/
[c] http://www.bic.mni.mcgill.ca/brainweb/
[d] http://hdl.handle.net/1926/1687
[e] www.adniinfo.org
[f] http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/
[g] http://www.isles-challenge.org/ISLES2015/
[h] https://sites.google.com/site/braintumorsegmentation/home/brats2015
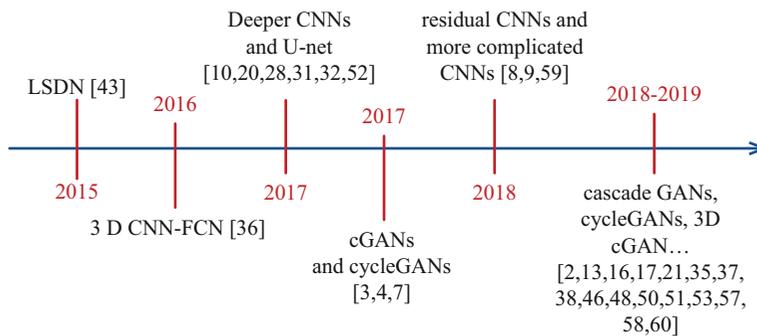[i] http://brain-development.org/ixi-dataset/
[j] https://ida.loni.usc.edu/login.jsp.

Fig. 1 The development of deep learning models for medical image synthesis

Deep Learning Models for Medical Image Synthesis

The state-of-the-art medical image synthesis methods usually use convolutional neural networks (CNN) [29]. With the delicate design, they can be exploited for whole-image or large-patch based synthesis to capture the implicit dependency among pixels/voxels in the same input during the end-to-end training. Among them, the most typical CNN architecture is U-net [40]. More recently, a number of CNN based generative adversarial networks (GANs) further improve the medical image synthesis results [55]. Therefore, this section will present the details of the typical conventional CNN model, i.e., U-net, and the basic GAN model in the research area of medical image synthesis.

Convolutional Neural Networks

The U-net model can extract the global contextual information from the source image and also reserve the spatially continuous details in the target image. As illustrated in Fig. 2, the original U-net model consists of the contracting and expanding paths. The number of convolutional layers in the contracting path is same to that in the expanding path. Between these two paths, multiple skip-connections are built to bridge them. With this structure, the U-net model can acquire the multi-depth information of the input source image. In addition, the gradient vanishing problem which is commonly shown during the training of deep learning models is mitigated, since the gradient of the deeper layers can be directly back-propagated to the shallower layers via the skip-connections.
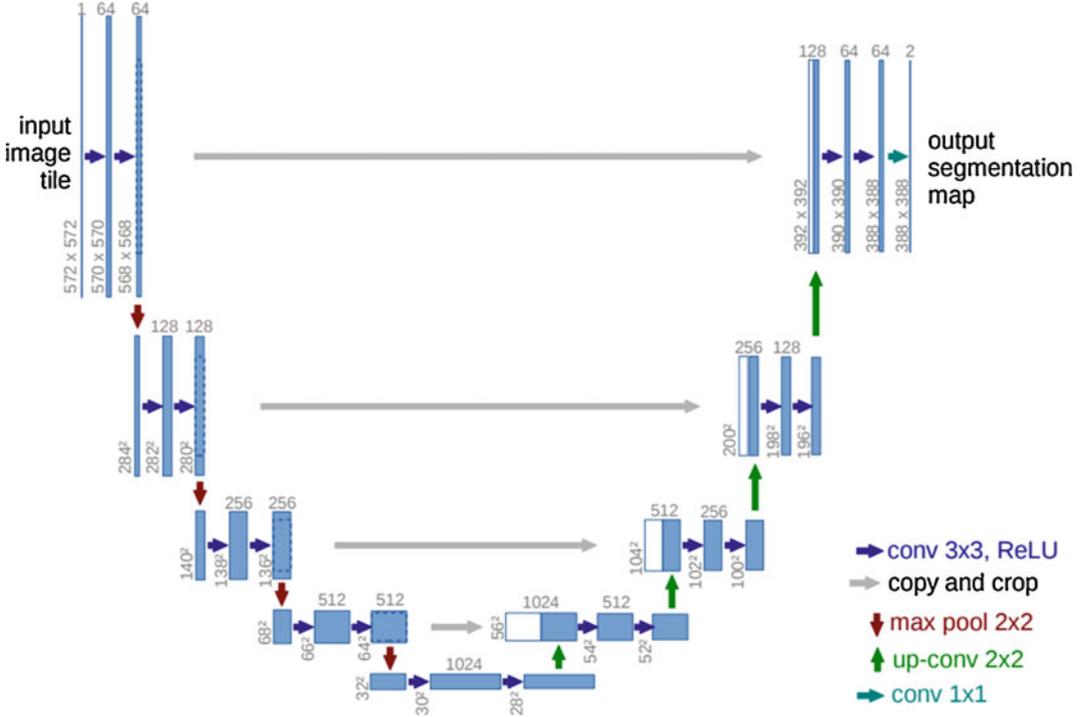
Fig. 2 Original 2D U-net model taken from [40]

This specific CNN architecture extracts the hierarchical visual clues from the input and is particularly suitable for the medical image synthesis tasks.

Generative Adversarial Networks

In 2014, the original GANs were first proposed for the generic image synthesis tasks [19]. Different from the common CNN based deep learning models, a GAN model consists of two agents, i.e., a generator G and a discriminator D, and is trained by the adversarial learning, as shown in Fig. 3. The original GAN model aims to learn a mapping from an input random noise to a target image that follows the distribution p_{data} of real images. In order to condition the GAN model on an input data of auxiliary information which could guide the mapping processing, conditional generative adversarial networks (cGANs) were then proposed [34]. When the input data is a source image $\mathbf{x} \sim p_{data}(\mathbf{x})$, the cGANs can

be trained to synthesize its corresponding target image $\mathbf{y} \sim p_{data}(\mathbf{y})$ with the specific control from $\mathbf{x}$. This is a process of paired image-to-image synthesis.

Most existing GAN models for medical image synthesis [14, 16, 46, 57] follow a representative work pix2pix [25] and achieve very promising results. As a cGAN model, it synthesizes an image $G(\mathbf{x})$ from the given source image $\mathbf{x}$ to resemble the real target image $\mathbf{y}$ by its generator G. At the same time, its discriminator D is trained to differentiate between the synthesized image pair $(\mathbf{x}, G(\mathbf{x}))$ and the corresponding real image pair $(\mathbf{x}, \mathbf{y})$. Therefore, the synthesis performance can be improved through the adversarial competition between these two agents. The training loss of the generator G is formulated as follows:

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

where the symbol $\mathbb{E}$ denotes mathematical expectation, and $G(\cdot)$ and $D(\cdot)$ accordingly refer to the
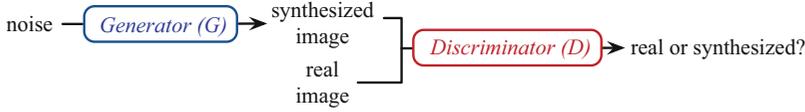
Fig. 3 Original GAN model

outputs of the generator and the discriminator. In Eq. (1), the generator G is trained to synthesize a realistic image which could fool the discriminator D via the first term. In its second term, the generator G tries to enforce the intensity similarity between the synthesized and real images through an L1-norm penalty on the pixel-wise intensity difference. The symbol λ_{l1} is a hyper-parameter to balance these two terms.

The loss function of the discriminator D is defined as follows:

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

In Eq. (2), different from the generator G, the discriminator D is trained to estimate the correct labels (0 or 1) for the synthesized or real image pairs. Thus, the adversarial competition between G and D conforms to a two-player min-max game.

In this cGAN model, the two sub-jobs of image generation and image discrimination are achieved together. Therefore, the final loss function integrates the above two objectives as follows:

$$\mathcal{L}_{cGAN} = \mathcal{L}_{cGAN}^G + \mathcal{L}_{cGAN}^D. \qquad (3)$$

Both the generator and the discriminator in this cGAN model are CNN based to capture the powerful deep learning features. Specially, the generator has a U-net-like architecture to obtain the hierarchical contextual information from the input source images and then generate the better target images.

Within-Modality Synthesis

In this section, we present our two recent works for within-modality synthesis. These two works

aim to synthesize high-quality positron emission tomography (PET) images to reduce the dose of radioactive tracer during the PET scanning. Since PET is widely exploited to visualize metabolism processes of human in clinics and research, it is important to get the clear PET images for patients. Before the PET image scanning, a full-dose radioactive tracer on a biologically active molecule is injected into the patient's body. During the scanning, the gamma rays which are emitted from the radioactive tracer in the body can be detected by the PET scanner. After that, the PET scanner analyzes the detected gamma rays of the full-dose tracer and constructs a high-quality three-dimensional (3D) PET image. However, the injected full-dose radioactive tracer brings up the risk of radioactive exposure and also raises the concerns about potential health hazards. As reported in "Biological Effects of Ionizing Radiation (BEIR VII),"[1] one full-dose radioactive tracer for every brain PET scan will improve the potential of lifetime cancer by 0.04%. When patients should take multiple times of PET scanning during their treatment, these risks will even accumulate, especially for the pediatric patients. To handle the radiation exposure issue, some researchers have lowered the injected dose of the tracer to the half of the full-dose, which inevitably decreases the quality of scanned PET images. The comparison of the full-dose PET image (F-PET) and the low-dose PET image (L-PET) is given in Fig. 4. Therefore, this high-quality PET image synthesis task aims to estimate the F-PET images from the given L-PET images.

---

[1] http://www.nap.edu/catalog/11340/health-risks-from-exposure-tolowlevels-of-ionizing-radiation.
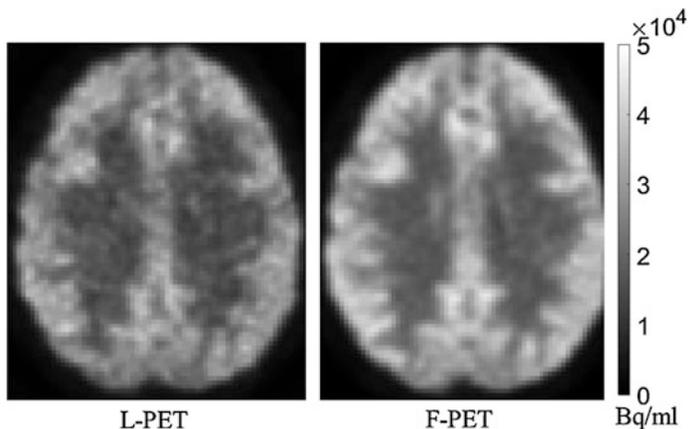
Fig. 4 Comparison between the low-dose PET (L-PET) image and the corresponding full-dose PET (F-PET). Taken from [47]

3D cGAN

Many cGAN models for medical image synthesis are 2D based [4,14,16,54]. When they are applied to the 3D medical imaging data, like PET and MRI, these methods first separate the 3D source image into axial slices and then separately map these source slices to the 2D target slices. After the concatenation of these synthesized 2D slices, a 3D target image can be reconstructed. Thus, the coronal and the sagittal slices of the reconstructed 3D target image are formed by the independently synthesized lines from the estimated axial planes. This will inevitably cause the loss of contextual information along the sagittal and coronal directions and the strong discontinuities in the final image. To mitigate this issue, our work in [46] proposes a 3D cGAN model to estimate the high-quality F-PET image from the L-PET image.

Framework

The framework of 3D cGAN model is illustrated in Fig. 5. Similar to the aforementioned cGAN model, this 3D cGAN consists of two agents: a 3D U-net-like generator G and a 3D CNN based discriminator D. The generator G processes a given L-PET image which is the source image and generates a synthesized F-PET image to approximate a real F-PET image. Simultaneously, the discriminator D is trained to distinguish between the synthesized F-PET-like image pair

and the real F-PET image pair. The adversarial learning in the 3D cGAN follows the competition between two sub-tasks, i.e., the image generation of G and the image discrimination of D.

Experimental Results

As reported in [46], the 3D cGAN is evaluated on a real human brain dataset, which contains two categories: eight normal subjects and eight subjects diagnosed as mild cognitive impairment (MCI). Experiments are conducted in the widely used "Leave-One-Subject-Out cross-validation" strategy, i.e., in each experiment one subject is used as test data and the other 15 subjects are applied for training. To acquire sufficient 3D training data, 125 large image patches of size $64 \times 64 \times 64$ are extracted from every original PET image of size $128 \times 128 \times 128$ with the stride of 16. In the final synthesized 3D PET image, the overlapped regions are averaged from the estimated large patches. To evaluate the PET synthesis performance, peak signal-to-noise (PSNR) and normalized mean squared error (NMSE) are used.

To validate the effectiveness of the 3D model, 2D cGANs are compared with the 3D cGAN model. These 2D cGANs are separately trained with the 2D slices from the corresponding axial, coronal, and sagittal views. One visual example of synthesized results by the compared methods
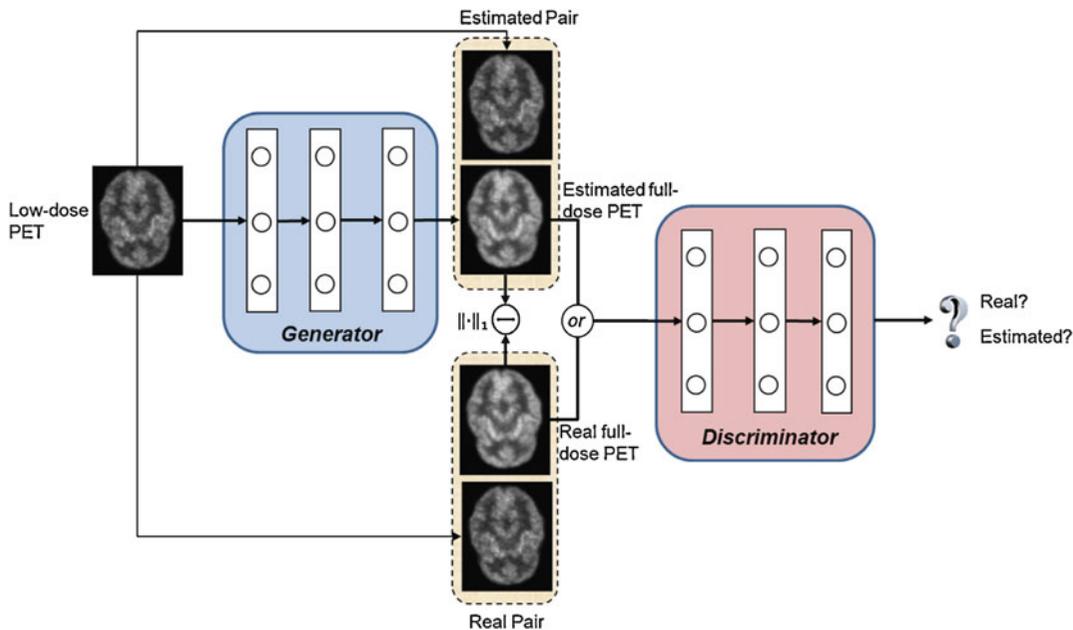
Fig. 5 Framework of training a 3D cGAN to estimate the full-dose PET image from low-dose counterpart. Taken from [46]
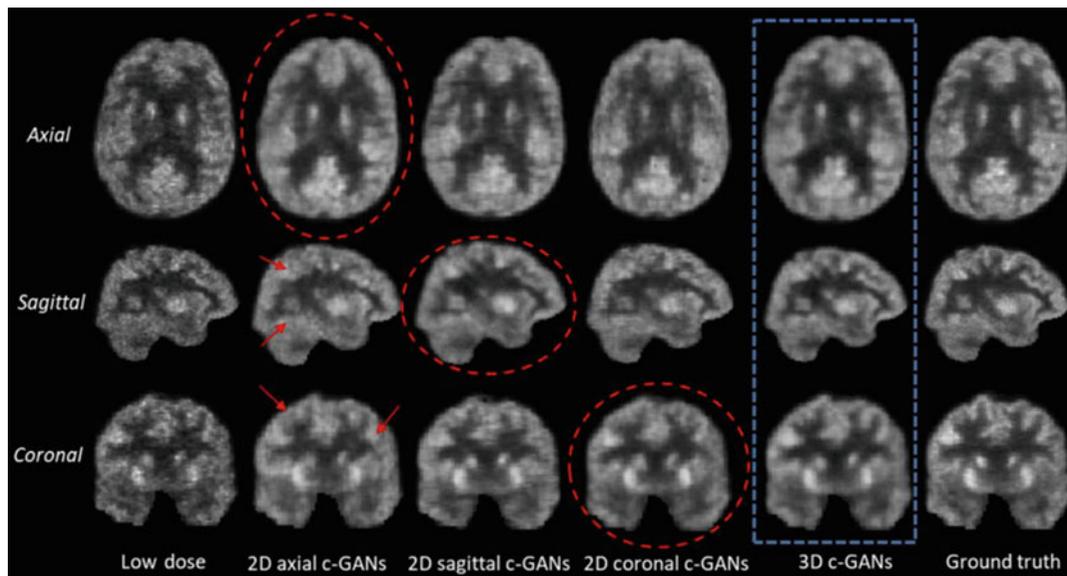

Fig. 6 Qualitative comparison between the results estimated by 2D cGANs and 3D cGANs. In the axial and coronal images, the left side of the image is the right side of the brain, and the right side of the image is the left side of the brain, taken from [46]

is given in Fig. 6. As shown, the results by 3D cGAN, which are presented in the blue block, have high visual quality in all three views. In contrast, these three 2D cGANs only produce good results in their corresponding trained views as indicated in the red circles, but get blurred synthesized views along the other two directions. The given example shows that the 2D cGANs
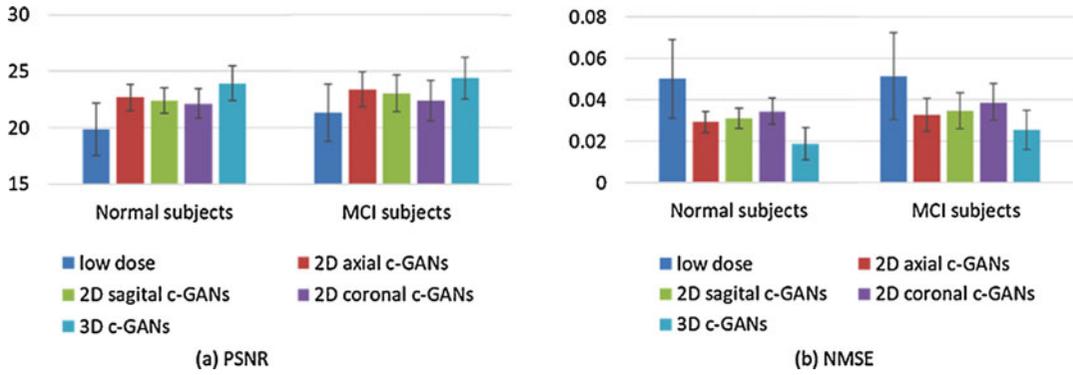
Fig. 7 Quantitative comparison between 2D cGANs and 3D cGANs, in terms of PSNR and NMSE, taken from [46]. Error bar indicates the standard deviation
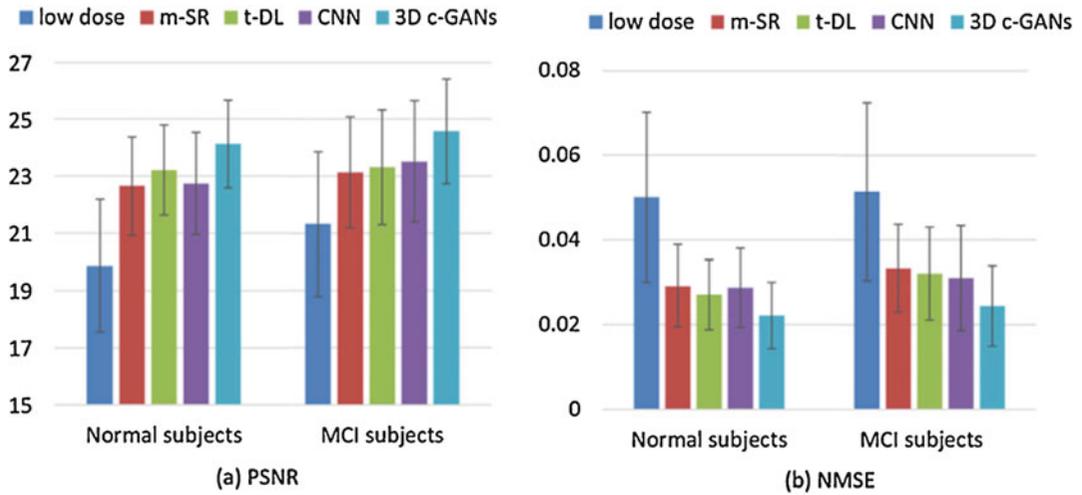

Fig. 8 Quantitative comparison between the existing PET estimation methods and the proposed method, in terms of PSNR and NMSE, taken from [46]. Error bar indicates the standard deviation

cause the discontinuous estimation across slices and lose the 3D structural information during the synthesis. The quantitative results of PSNR and NMSE are separately reported on the normal and the MCI data in Fig. 7. 3D cGAN achieves the best PSNR and NMSE results on both two categories of PET data, which consistently indicates that the 3D information captured in 3D cGAN can boost the synthesis.

Three state-of-the-art PET synthesis methods are compared with the 3D cGAN. They are (1) mapping based sparse representation (m-SR) [44], (2) semi-supervised tripled dictionary learning method (t-DL) [45], and (3) common CNN based method [51]. The

quantitative comparison results are reported in Fig. 8. 3D cGAN performs best among all four methods in terms of both PSNR and NMSE, which demonstrates the superiority of the 3D cGAN in full-dose PET image synthesis.

Locality Adaptive Multi-Modality GANs

Recent research reports that using multiple modalities, like PET and MRI, benefits the medical image quality enhancement [23]. In addition, different from PET, scanning MRI would not raise the risks of radioactive exposure
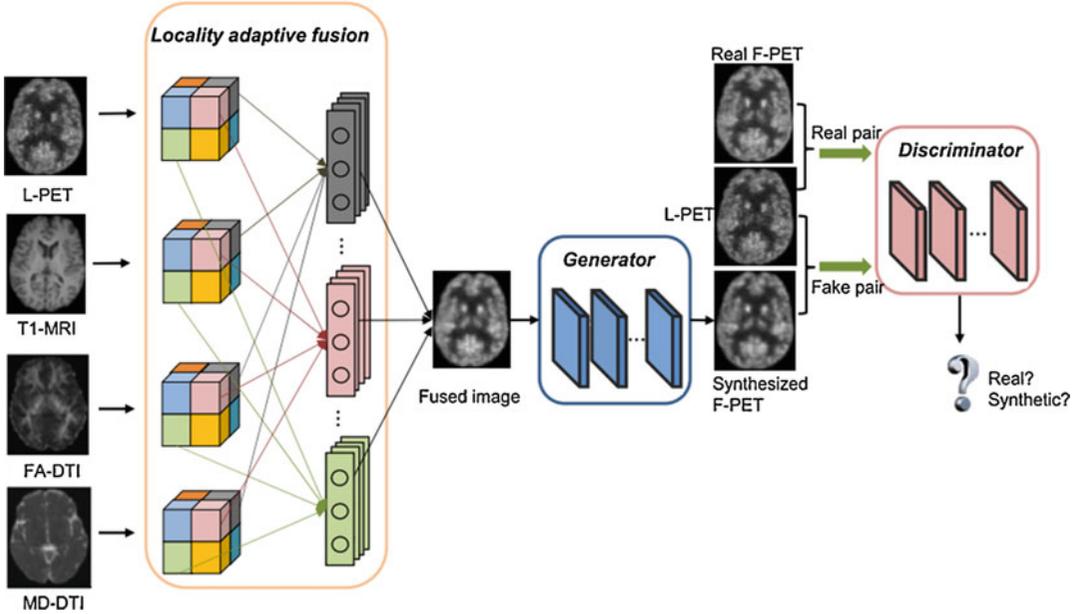
Fig. 9 Overview of locality adaptive multi-modality GANs, taken from [47]

for patients. Thus, medical images of T1-weighted MRI (T1-MRI), fractional anisotropy diffusion tensor image (FA-DTI), and mean diffusivity DTI (MD-DTI) are applied to assist the synthesis of high-quality F-PET images from L-PET images. Traditionally, the image convolution in CNN based GANs is performed on these multiple images (input channels) in a global manner. That is to say, the common multi-channel based GANs apply the same convolution filter to all image locations of each input modality for producing the feature maps which will be combined in deeper layers. As a result, these multi-channel models would not consider the location-varying contributions from the various input modalities. To handle this issue, locality adaptive multi-modality GANs (LA-GANs) are proposed in [47] for PET image synthesis.

Framework

The LA-GANs model includes three modules: (1) the locality-adaptive fusion network, (2) the generator network, and (3) the discriminator network, as illustrated in Fig. 9. The newly added locality-adaptive fusion network processes L-PET, T1-MRI, FA-DTI, and MD-DTI images as input channels and estimates a fused image by learning different convolutional kernels at different image locations. Specifically, the module of locality-adaptive fusion network first separately partitions the entire input L-PET, T1-MRI, FA-DTI, and MD-DTI images into N non-overlapped small patches which are accordingly denoted by P_i^L, P_i^{T1}, P_i^{FA}, and P_i^{MD} ($i = 1, \ldots, N$). These small patches from different locations are indicated by different colors in Fig. 9. After that, the patches at the same location from the four input modalities are separately convolved by four different $1 \times 1 \times 1$ filters with parameters ω_i^L, ω_i^{T1}, ω_i^{FA}, and ω_i^{MD}, respectively. Through this locality-adaptive convolution, a fused patch P_i^C can be calculated as follows:

$$P_i^C = \omega_i^L P_i^L + \omega_i^{T1} P_i^{T1} + \omega_i^{FA} P_i^{FA} + \omega_i^{MD} P_i^{MD},$$
$$s.t. \ \omega_i^L + \omega_i^{T1} + \omega_i^{FA} + \omega_i^{MD} = 1,$$
$$\omega_i^L, \omega_i^{T1}, \omega_i^{FA}, \omega_i^{MD} > 0, i = 1, \ldots, N.$$
$$(4)$$

Therefore, N groups of different convolution filters for the $N * 4$ small patches at N locations from four modalities can be learned.
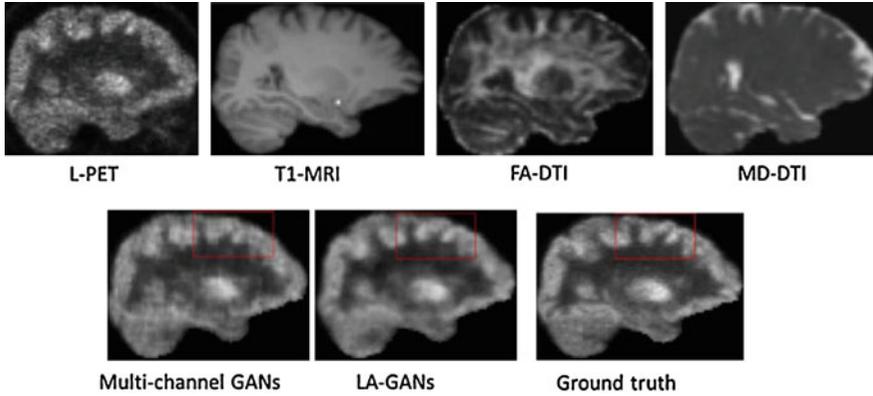
Fig. 10 Visual comparison with multi-channel GANs method, taken from [47]

Table 2 Quantitative comparison with the multi-channel GANs method on normal and MCI subjects

	Methods	PSNR		SSIM	
		Mean (std.)	Med.	Mean (std.)	Med.
Normal subjects	L-PET	19.88 (2.34)	20.68	0.9790 (0.0074)	0.980
	Multi-channel	24.36 (1.93)	24.78	0.9810 (0.0065)	0.983
	LA-GANs	**24.61 (1.79)**	**25.32**	**0.9860 (0.0053)**	**0.987**
MCI	L-PET	21.33 (2.53)	21.62	0.9760 (0.0102)	0.979
subjects	Multi-channel	24.99 (2.03)	25.36	0.9795 (0.0098)	0.982
	LA-GANs	**25.19 (1.98)**	**25.54**	**0.9843 (0.0097)**	**0.988**

Mean (standard deviation), Median. The paired t-test of PSNR shows that our improvement against the multi-channel one is statistically significant with $p < 0.05$ ($p = 0.048$ for NC subjects and $p = 0.016$ for MCI subjects). For SSIM, our method also presents the significant improvement, with p-value 0.051 for NC subjects and 0.037 for MCI subjects, respectively

After the above locality-adaptive fusion, the final fused image is applied as the input of the generator to generate F-PET-like images. The generator and the discriminator in our LA-GANs work similarly to those in the aforementioned 3D cGAN.

Experimental Results

To evaluate the effectiveness of the newly added locality-adaptive fusion network module, the common multi-channel GANs model is compared with the LA-GANs. Figure 10 gives an example of visual results obtained by these two methods. We can observe that the LA-GANs model synthesizes the F-PET-like image with less artifacts than the compared multi-channel model, which are clearly indicated by red rectangles. The quantitative results of these two methods are reported in Table 2 via the evaluation

measures of PSNR and structural similarity index (SSIM) [48]. The top part gives the results on normal subjects, and the bottom part reports the results on MCI subjects. These results show the superiority of the locality-adaptive fusion network module over the common multi-channel processing, in terms of both PSNR and SSIM. In addition, through conducting the paired t-test, all these improvements are statistically significant at the significance level of 0.05. Both the visual and quantitative results demonstrate the effectiveness of locality-adaptive fusion network in cGAN models for the full-dose PET synthesis task.

Moreover, the LA-GANs model is compared with four state-of-the-art methods, i.e., (1) mapping based sparse representation method (m-SR) [44], (2) tripled dictionary learning method (t-DL) [45], (3) multi-level CCA method (m-CCA) [1], and (4) auto-context CNN method
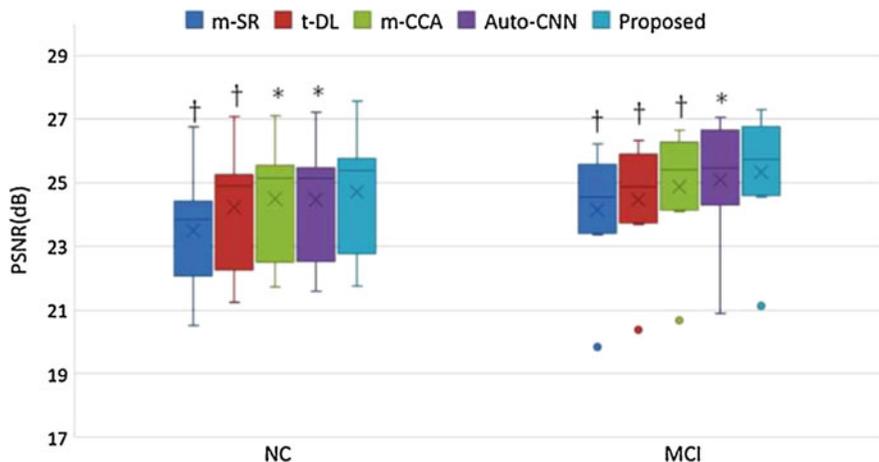
Fig. 11 Qualitative comparison with the state-of-the-art PET estimation methods in terms of PSNR, taken from [47].
† indicates $p < 0.01$ in the paired t-test while ⋆ means $p < 0.05$

(auto-CNN) [51], as shown in Fig. 11. The highest
PSNR values of the LA-GANs indicate that it has
the best performance among all the compared
methods.

Cross-Modality Synthesis

Our two recent cross-modality synthesis works
are presented in this section. They are utilized
in cross-modality MR image synthesis that
aims to better visualize the scanned body parts
from diverse imaging perspectives and facilitate
the following clinical applications, e.g., tumor
segmentation. When setting different scanning
parameters, MRI can generate multiple-modality
images (e.g., T1-weighted, T2-weighted, and
FLAIR) which show the diverse contrasts
of soft tissues. Since each modality image
provides the unique visual representation of
scanned body parts, these multiple modalities
are usually studied together in the subsequent
analysis for disease diagnosis [15] and treatment
planning [30]. However, due to the potential
of modality loss in clinics, the quality of the
analysis will be adversely affected. Therefore,
cross-modality MR image synthesis is highly
desirable to synthesize the unknown target-
modality MR images from the given source-
modality images [26, 53].

3D cGAN with Subject-Specific Local Adaptive Fusion

Our work in [56] proposes a 3D cGAN based
cross-modality MR image synthesis method to
boost brain tumor segmentation performance.
Compared with the single synthesis task, this
is more challenging and requires the higher
quality of synthesized images because of two
main reasons. First, due to the arbitrary location
and appearance of brain tumor, the pathology
involved MR images raise the difficulty of
synthesis in contrast to the healthy subject
images. Second, the source-modality images
may lack some important pathology-related
information which can be seen in the target-
modality. For example, as shown in Fig. 12,
the diffuse changes around tumor parts are only
observed in the FLAIR image. Thus, [56] presents
an additional approach to the 3D cGAN, which is
called subject-specific local adaptive fusion. This
fusion approach aims to polish the local details in
the synthesized target-modality-like images from
the 3D cGAN through a linear combination of the
real target-modality images among the training
set for approximation. During the combination,
the combination weights are estimated from the
synthesized target-modality-like images which
are the outputs of the 3D cGAN model. In this
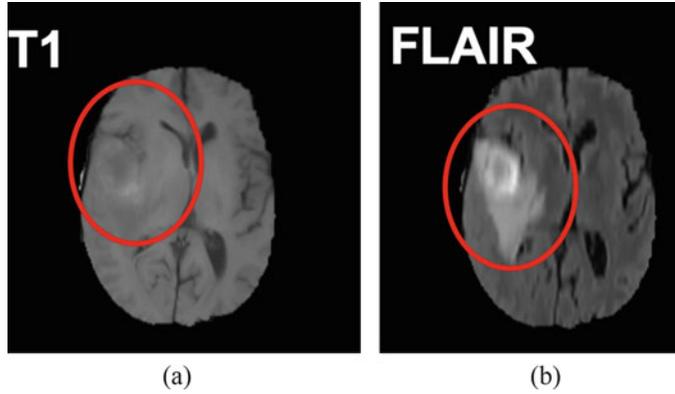way, this local and adaptive approach can improve

Fig. 12 A brain T1 image (**a**) and the corresponding FLAIR image (**b**)
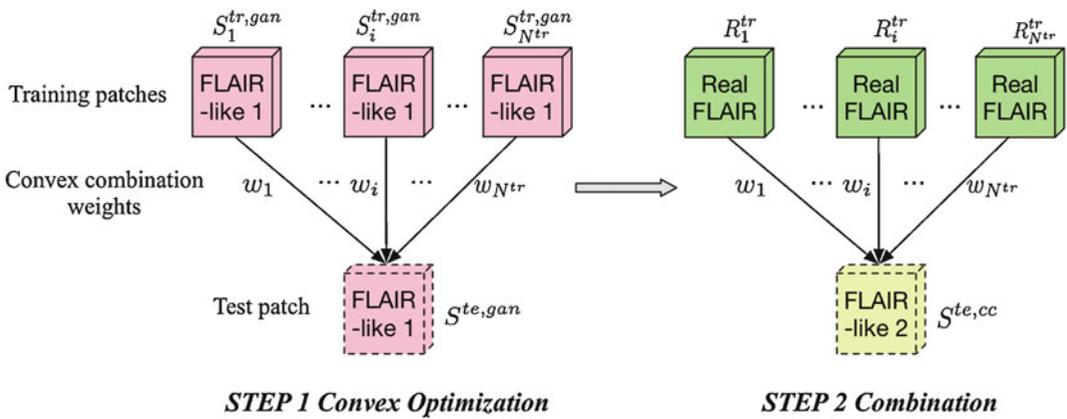

Fig. 13 Framework of subject-specific local adaptive fusion

the quality of synthesized images and further raise the segmentation performance.

Framework

The framework of this subject-specific local adaptive fusion is illustrated in Fig. 13. Here, we take T1-to-FLAIR synthesis task as an example, and the synthesized FLAIR-like image from the 3D cGAN and the final fused FLAIR-like image are called FLAIR-like-1 image and FLAIR-like-2 image, respectively. Before this fusion, for each test subject that only has its real T1 image, its corresponding FLAIR-like-1 image is partitioned into non-overlapped small patches of size $16 \times 16 \times 16$. Each FLAIR-like-1 patch $S^{\text{te,gan}}$ of this test subject is approximated by the convex combination of the patches $S_1^{\text{tr,gan}}, S_2^{\text{tr,gan}} \ldots, S_{N_{\text{tr}}}^{\text{tr,gan}}$ at the

same location from the FLAIR-like-1 images of training subjects. The symbol N_{tr} denotes the number of all training subjects. This approximation is achieved through the following optimization:

$$\min_{w} \| \sum_{i=1}^{N_{\text{tr}}} w_i S_i^{\text{tr,gan}} - S^{\text{te,gan}} \|_2^2 \tag{5}$$

$$s.t. \sum w_i = 1, w_i \geq 0.$$

Therefore, the combination weights w_i ($i = 1, \ldots, N_{\text{tr}}$) are learned via Eq. (5). Since the FLAIR-like-1 image is generated to resemble the corresponding real image, these combination weights could be further applied to linearly combine the real FLAIR training patches $R_1^{\text{tr}}, R_2^{\text{tr}} \ldots, R_{N_{\text{tr}}}^{\text{tr}}$ at the same location to polish

the final FLAIR-like-2 patch $S^{\text{te,cc}}$ through the following equation:

$$S^{\text{te,cc}} = \sum_{i=1}^{N_{\text{tr}}} w_i R_i^{\text{tr}}. \qquad (6)$$

In this way, a better polished target-modality-like image is estimated and used together with its corresponding real source-modality image in the subsequent brain tumor segmentation.

Experimental Results

This work is evaluated on BRATS2015 dataset [33], which includes 274 subjects of four modality images, i.e., T1, T1C, T2, and FLAIR, with size of $240 \times 240 \times 155$, and additional brain tumor labels. In this work, 230 subjects are randomly selected as training data and the rest 44 subjects are test data for the T1-to-FLAIR synthesis task. The brain tumor segmentation model from [27] is utilized to evaluate the segmentation performance of the synthesized FLAIR images. During the synthesis task, 3D large patches of $128 \times 128 \times 128$ are extracted from images to increase the training samples for 3D cGAN. The evaluation measures of PSNR and NMSE on the synthesized whole brains and tumor regions are utilized.

To study the contribution of the subject-specific local adaptive fusion, it (i.e.,3D cGAN (128)+local adaptive fusion) is compared with another two methods for ablation study. They are: (1) 3D cGAN trained on large patches (128^3) and (2) local non-linear mapping (3D cGAN on patches with the size of 32^3) applied after the method (1). The synthesis results are reported in Table 3. As shown, the 3D cGAN

(128)+local adaptive fusion outperforms the other two methods, demonstrating the effectiveness of subject-specific local adaptive fusion in T1-to-FLAIR image synthesis task. The results also indicate that using the linear combination in the local adaptive fusion can obtain better results than the local non-linear mapping of 3D cGAN (32). Table 4 gives the segmentation results on whole tumor parts and tumor core regions by the above three methods, which consistently indicates the better performance of the 3D cGAN (128)+local adaptive fusion approach. The results by using the single modality of T1 are also reported. The paired t-test result verifies that the improvement on the tumor core part is statistically significant. Therefore, the quantitative results of both synthesis and segmentation tasks show the advantage of 3D cGAN (128)+local adaptive fusion approach in synthesizing FLAIR images from T1, and the benefits of using the synthesized FLAIR images to improve the T1-based brain tumor segmentation.

Edge-Aware GANs

The aforementioned cGAN models enforce the pixel/voxel-wise intensity similarity between the real and the synthesized images through using an L1-norm penalty during training. However, the structure of image content, like the textural information in MRI [6], is not sufficiently captured by these models. The edge information in an image provides the details about the textural structure of image content through capturing the local intensity changes and the boundaries

Table 3 Quantitative evaluation results of the synthesized images

Methods	Synthesis quality (PSNR/NMSE%)	
	Whole brain	Tumor
3D cGAN (128)	20.45/<u>25.08</u>	<u>19.13</u>/<u>12.68</u>
3D cGAN (128)+3D cGAN (32)	<u>19.94</u>/<u>24.99</u>	<u>18.73</u>/<u>13.45</u>
3D cGAN (128) + local adaptive fusion	**20.68/22.67**	**19.27/11.86**

Values with underline indicate they are statistically significantly different from 3D cGAN (128)+local adaptive fusion, according to a two-sided, paired t-test (solid line $p < 0.05$). t-Test values are given as follows: (1) proposed method over 3D cGAN: (a) whole brain ($1.17e-1/4.59e-2$); (b) tumor ($1.17e-2/8.13e-4$). (2) Proposed method over 3D cGAN(128) + (32): (a) whole brain ($8.75e-7/5.59e-4$); (b) tumor ($2.1e-3/6.12e-5$)

Table 4 Quantitative evaluation results of segmentation

Methods	Segmentation (dice ratio%)	
	Whole tumor	Core
3D cGAN (128)	66.35	72.09
3D cGAN (128)+3D cGAN (32)	66.61	72.14
T1	67.18	63.00
T1+real FLAIR (ideal scenario)	*82.17*	*85.49*
3D cGAN (128) + local adaptive fusion	**68.23**	**72.28**

Values with underline indicate they are statistically significantly different from 3D cGAN (128)+local adaptive fusion, according to a two-sided, paired t-test (solid line $p < 0.05$, dotted line $p < 0.1$). t-Test values are given as follows: (1) proposed method over 3D cGAN: (a) whole tumor (0.0643); (b) core (0.886). (2) Proposed method over 3D cGAN(128) + (32): (a) whole tumor (0.0672); (b) tumor (0.912). (3) Proposed method over T1: (a) whole tumor (0.262); (b) tumor ($9.44e − 5$)
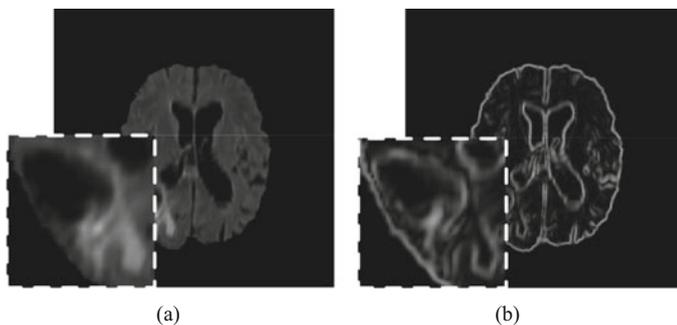

(a)   (b)

Fig. 14 A brain FLAIR image (**a**), and its corresponding edge map (**b**) after the 3D Sobel edge detection, taken from [57]. The contour of abnormal tissues can be depicted clearer by the edge map, which is shown as the zoomed regions

between different tissues. Thus, maintaining the edges during the synthesis can help to sharpen the synthesized target-modality MR images. Especially, for a pathology involved MR image, the edge details benefit to distinguish between the normal and the abnormal tissues, which is important to depict the contour of the arbitrary pathological regions. For example, Fig. 14 shows that the zoomed gliomas tumor is very clear in the edge map of a brain MR image. Therefore, our work in [57] proposes new cGAN models to enforce edge preservation for cross-modality MR image synthesis. This work adds an extra constraint to 3D cGAN models to realize edge-aware generative adversarial networks (Ea-GANs) by ensuring the similarity of the edge maps extracted from the real and the synthesized images during training. These edge maps are calculated via the commonly applied Sobel filters as shown in Fig. 15. These three $3 \times 3 \times 3$ Sobel filters, i.e., F_i, F_j, and F_k, are applied to convolve a given

image A to produce its three edge maps which correspond to the intensity gradients along i, j, and k directions, respectively. After that, a final edge map $S(A)$ of A is obtained by merging the three-direction edge maps through the following equation:

$$S(A) = \sqrt{(F_i * A)^2 + (F_j * A)^2 + (F_k * A)^2},$$
(7)

where $*$ denotes the convolution operation.

Framework

As shown in Fig. 16, [57] proposes two different frameworks, i.e., a generator-induced Ea-GAN (gEa-GAN) and a discriminator-induced Ea-GAN (dEa-GAN), according to the different strategies of using the edge maps. Both of these two Ea-GANs are composed of three modules: (1) a generator G, (2) a discriminator D, and (3) a Sobel edge detector S.
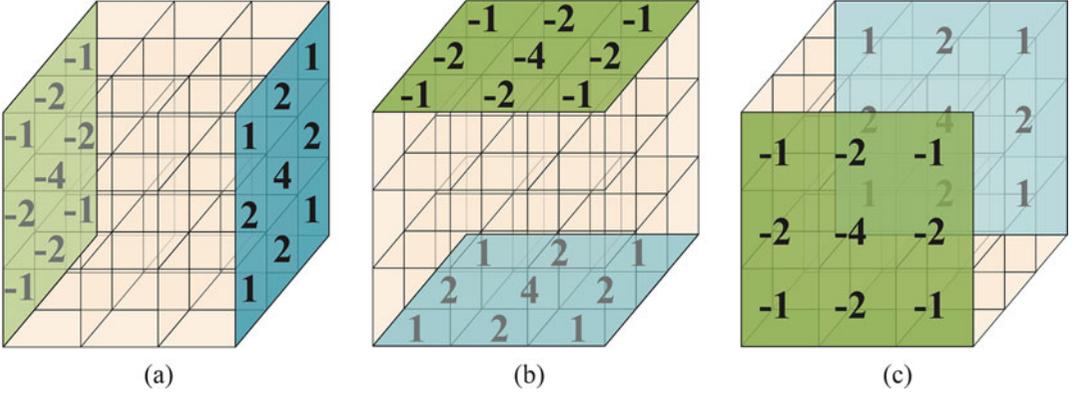
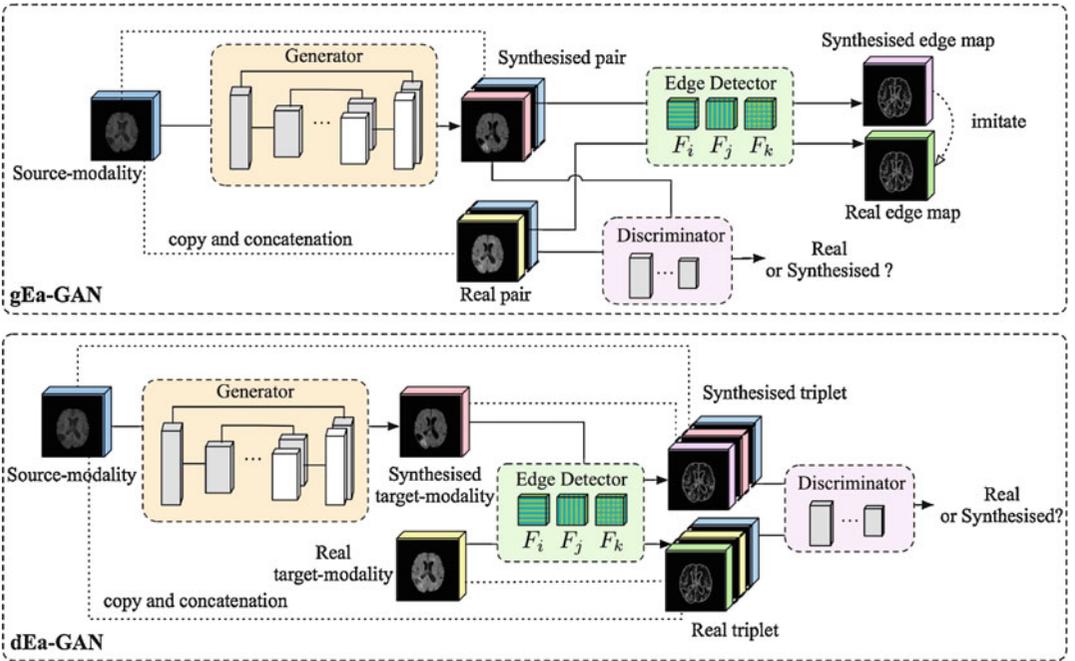**Fig. 15** Three-dimensional Sobel operator, taken from [57]. (**a**) F_i. (**b**) F_j. (**c**) F_k

Fig. 16 Frameworks of Ea-GANs, taken from [57]

For the gEa-GAN model, when given a source-modality image **x** and its target-modality counterpart **y** as the groundtruth for the cross-modality MR image synthesis task, its generator G tries to synthesize target-modality-like images $G(\mathbf{x})$ that can be misclassified by its discriminator D through the adversarial learning. The L1-norm penalty through G is applied to ensure the voxel-wise intensity similarity between the real and the synthesized images, similar to the 3D cGAN model. Additionally, another L1-norm penalty is used to discourage the difference between their

corresponding Sobel edge maps which are extracted from S during the training of gEa-GAN. Therefore, the loss function of the generator G is formulated as follows:

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

$$(8)$$

where the hyper-parameters, λ_{l1} and λ_{edge}, are used to balance the three terms in Eq. (8).

Similar to the case of 3D cGAN model, the loss function of its discriminator D is defined as follows:

$$\mathcal{L}_{\text{gEa}-\text{GAN}}^{D} = -\mathbb{E}_{\mathbf{x},y \sim p_{\text{data}}(\mathbf{x},\mathbf{y})}[\log D(\mathbf{x}, \mathbf{y})]$$
$$- \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}[\log (1 - D(\mathbf{x}, G(\mathbf{x})))]. \quad (9)$$

Thus, the final objective function of gEa-GAN integrates the above two loss functions of the generator and the discriminator as follows:

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

Different from the gEa-GAN that maintains the edge similarity only by its generator during training, the dEa-GAN model additionally brings the edge information into the adversarial learning between its generator and discriminator. In this way, its discriminator could also perceive the edge details of the synthesized images and further benefit the synthesis processing. Specifically, the edge maps of the real and the synthesized target-modality images are correspondingly concatenated with the real and the synthesized image pairs as the real triplet $(\mathbf{x}, y, S(\mathbf{y}))$ and the synthesized triplet $(\mathbf{x}, G(\mathbf{x}), \text{and } S(G(\mathbf{x})))$. The discriminator D of dEa-GAN tries to distinguish between these two kinds of triplets, and this in turn enforces its generator G to estimate the better edge details for synthesis.

For dEa-GAN, its generator G is also trained by the adversarial loss, the voxel-wise intensity difference loss, and the edge difference loss for synthesis, following the designed objective:

$$\mathcal{L}_{\text{dEa}-\text{GAN}}^{G}$$
$$= \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}[\log (1 - D(\mathbf{x}, G(\mathbf{x}), S(G(\mathbf{x}))))$$
$$+ \lambda_{l1} \mathbb{E}_{\mathbf{x},\mathbf{y} \sim p_{\text{data}}(\mathbf{x},\mathbf{y})}[\|y - G(\mathbf{x})\|_1]$$
$$+ \lambda_{\text{edge}} \mathbb{E}_{\mathbf{x},\mathbf{y} \sim p_{\text{data}}(\mathbf{x},\mathbf{y})}[\|S(\mathbf{y}) - S(G(\mathbf{x}))\|_1]. \quad (11)$$

Different from the gEa-GAN model, the edge map $S(G(\mathbf{x}))$ in dEa-GAN is implicitly utilized in the first term of Eq. (11) through calculating the loss error of the outputs from its discriminator D.

The objective function of the discriminator D is accordingly designed as:

$$\mathcal{L}_{\text{dEa}-\text{GAN}}^{D} = -\mathbb{E}_{\mathbf{x},\mathbf{y} \sim p_{\text{data}}(\mathbf{x},\mathbf{y})}[\log D(\mathbf{x}, \mathbf{y}, S(\mathbf{y}))]$$
$$- \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}$$
$$[\log (1 - D(\mathbf{x}, G(\mathbf{x}), S(G(\mathbf{x}))))]. \quad (12)$$

Finally, the objective for training the entire dEa-GAN model is

$$\mathcal{L}_{\text{dEa}-\text{GAN}} = \mathcal{L}_{\text{dEa}-\text{GAN}}^{G} + \mathcal{L}_{\text{dEa}-\text{GAN}}^{D}. \quad (13)$$

Experimental Results

The Ea-GANs are evaluated on BRATS2015 dataset by the way of 5-fold cross validation. They are compared with five methods: (1) handcrafted feature used replica [26], (2) common CNN based multimodal [8], (3) 2D cGAN based pix2pix [25], (4) 3D cGAN, and (5) gradient loss utilized gradient cGAN. The evaluation measures of PSNR, NMSE, and SSIM are separately applied on the synthesized whole images including the background and the brain part. Two synthesis tasks, T1-to-FLAIR and T1-to-T2, are conducted to show the performance of Ea-GANs for cross-modality MR image synthesis. Their quantitative and visual results are presented in Table 5, Fig. 17, Table 6, and Fig. 18.

When comparing two Ea-GANs with the 3D cGAN model through the given quantitative results, the Ea-GANs produce higher-quality images than 3D cGAN with the significant improvements of PSNR from 29.26dB (3D cGAN) to 30.11dB (dEa-GAN), SSIM from 0.958 (3D cGAN) to 0.963 (dEa-GAN), and NMSE from 0.119 (3D cGAN) to 0.105 (dEa-GAN), respectively, in the T1-to-FLAIR task. Similarly, these two Ea-GANs also outperform the 3D cGAN in the T2 image synthesis task. These results demonstrate the effectiveness of preserving edge information in the synthesized images. Additionally, the dEa-GAN model performs better than the gEa-GAN model in both of two synthesis

Table 5 Quantitative evaluation results of the synthesized FLAIR-like from T1 on the BRATS2015 dataset (mean $\pm$ standard deviation)

Methods	PSNR	NMSE	SSIM
Replica [26]	$\underline{27.17 \pm 2.60}$	$\underline{0.171 \pm 0.267}$	$\underline{0.939 \pm 0.013}$
Multimodal [8]	$\underline{27.26 \pm 2.82}$	$\underline{0.184 \pm 0.284}$	$\underline{0.950 \pm 0.014}$
Pix2pix [25]	$\underline{27.46 \pm 2.55}$	$\underline{0.144 \pm 0.189}$	$\underline{0.940 \pm 0.015}$
3D cGAN	$\underline{29.26 \pm 3.21}$	$\underline{0.119 \pm 0.205}$	$\underline{0.958 \pm 0.016}$
Gradient cGAN	$\underline{29.38 \pm 3.25}$	$\underline{0.116 \pm 0.204}$	0.960 ± 0.017
gEa-GAN	$\underline{29.55 \pm 3.24}$	$\underline{0.115 \pm 0.199}$	0.960 ± 0.017
dEa-GAN	$\mathbf{30.11 \pm 3.22}$	$\mathbf{0.105 \pm 0.174}$	$\mathbf{0.963 \pm 0.016}$

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. t-Test values of proposed dEa-GAN over the following methods: (a) Replica: $7.96e - 53$; $3.94e - 13$; $3.67e - 106$. (b) Multimodal: $9.10e - 39$; $1.75e - 12$; $2.11e - 41$. (c) Pix2pix: $7.72e - 67$; $8.50e - 21$; $8.71e - 131$. (d) 3D cGAN: $1.05e - 63$; $6.62e - 6$; $6.84e - 42$. (e) Gradient cGAN: $5.29e - 30$; $1.08e - 2$; $1.80e - 22$. (f) gEa-GAN: $3.41e - 25$; $9.15e - 4$; $1.67e - 15$
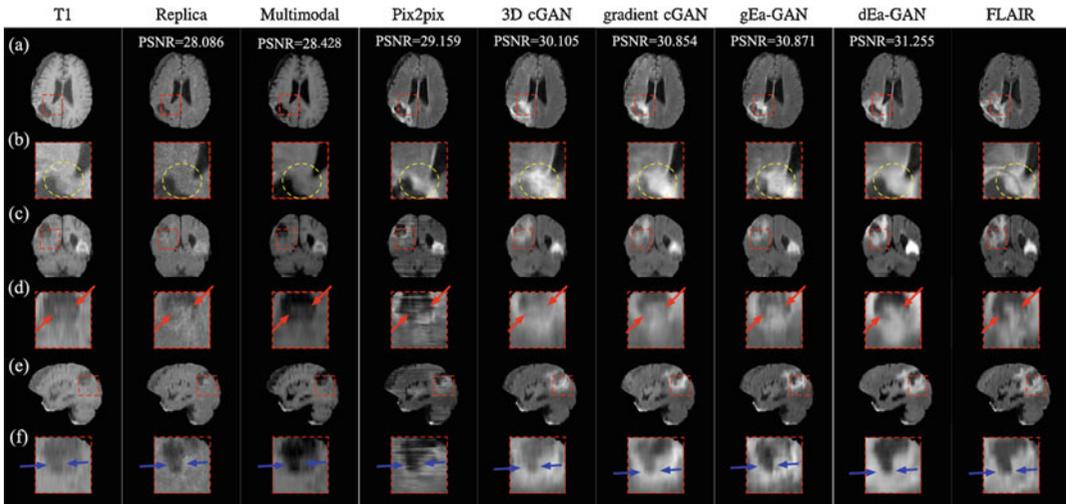

Fig. 17 Visual comparison of the synthesized FLAIR images between Ea-GANs and other state-of-the-art methods taken from [57]: (**a**) axial slices, (**b**) zoomed parts of axial slices, (**c**) coronal slices, (**d**) zoomed parts of coronal slices, and (**e**) sagittal slices, (**f**) zoomed parts of sagittal slices

tasks, showing the necessity of bringing the edge information into the training of the discriminator. Furthermore, the superiority of the Ea-GANs is consistently shown when compared with the other four state-of-the-art methods in terms of all the three measures. When looking into the visual comparison examples in Figs. 17 and 18, from all the three views, it can be seen that the Ea-GANs synthesize sharper edges and more local details than the compared methods as indicated. Therefore, both the quantitative and visual results

demonstrate that the Ea-GANs synthesize better MR images by using edge maps via two different strategies in cGAN models than the compared methods.

Conclusion

In this chapter, we focus on presenting deep learning approaches for medical image synthesis. Through the experimental results in our four

Table 6 Quantitative evaluation results of the synthesized T2-like images from T1 on the BRATS2015 dataset (mean ± standard deviation)

Methods	PSNR	NMSE	SSIM
Replica [26]	26.92 ± 2.36	0.158 ± 0.324	0.946 ± 0.015
Multimodal [8]	27.31 ± 2.39	0.140 ± 0.229	0.951 ± 0.016
Pix2pix [25]	28.12 ± 2.45	0.110 ± 0.220	0.953 ± 0.014
3D cGAN	29.34 ± 3.23	0.095 ± 0.199	0.964 ± 0.017
Gradient cGAN	29.43 ± 3.28	0.097 ± 0.210	0.966 ± 0.017
Proposed gEa-GAN	29.58 ± 3.29	0.093 ± 0.218	0.966 ± 0.018
Proposed dEa-GAN	$\mathbf{29.98 \pm 3.37}$	$\mathbf{0.088 \pm 0.223}$	$\mathbf{0.967 \pm 0.016}$

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. t-Test values of proposed dEa-GAN over the following methods: (a) Replica: $4.03e-41$; $4.21e-7$; $1.85e-72$. (b) Multimodal: $4.25e-48$; $4.32e-23$; $1.49e-78$. (c) Pix2pix: $1.90e-42$; $2.19e-9$; $2.44e-106$. (d) 3D cGAN: $4.25e-30$; $1.54e-4$; $3.43e-40$. (e) Gradient cGAN: $2.24e-33$; $3.77e-10$; $2.28e-16$; (f) gEa-GAN: $8.59e-18$; $1.82e-7$; $3.72e-5$

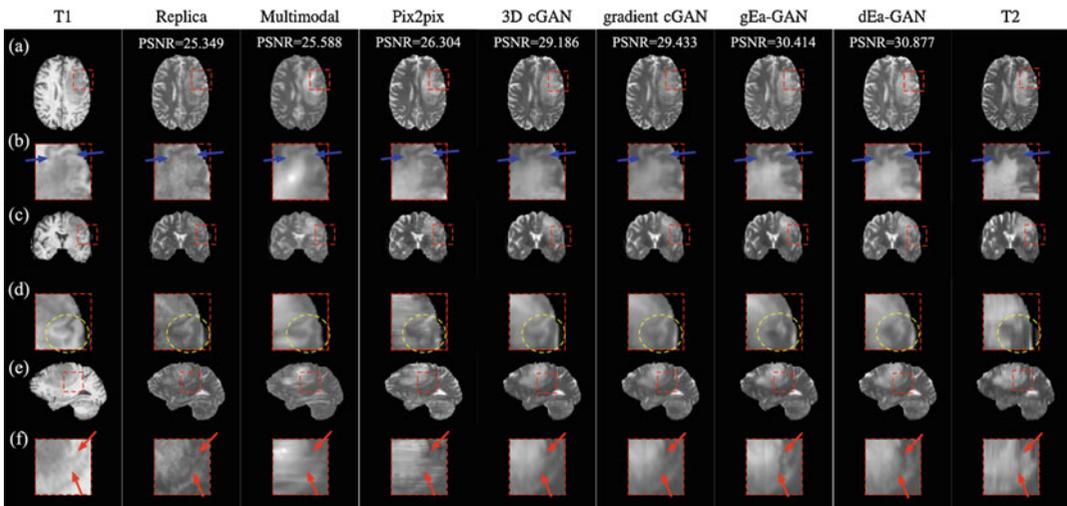**Fig. 18** Visual comparison of the synthesized T2 images between Ea-GANs and other state-of-the-art methods taken from [57]: (**a**) axial slices, (**b**) zoomed parts of axial slices, (**c**) coronal slices, (**d**) zoomed parts of coronal slices, and (**e**) sagittal slices, (**f**) zoomed parts of sagittal slices

works, we can see that using the recent GAN based models achieves better medical image synthesis performance than the conventional CNN based models. Besides, we can also conclude two main factors that benefit the successful application of the presented GAN models in within-modality and cross-modality synthesis. First, due to the 3D structure of medical images, the 3D architecture of GAN models can preserve continuous contextual information along all the three directions and therefore

improve the synthesis results. Second, in order to synthesize more realistic images, additionally exploiting the spatially local details in the source or the target images for different subjects can further boost the synthesis performance of the GAN based methods, since the subtle visual difference in medical images is essential in clinical applications. In summary, deep learning based medical image synthesis, especially the recent GAN based one, has become an active research topic. With the participation of more

researchers, it is expected that new synthesis approaches and methods will be developed in the coming years to further boost its performance and efficiency.

References

1. An L, Zhang P, Adeli E, Wang Y, Ma G, Shi F, Lalush DS, Lin W, Shen D (2016) Multi-level canonical correlation analysis for standard-dose PET image estimation. IEEE Trans Image Process 25(7):3303–3315

2. Armanious K, Jiang C, Fischer M, Küstner T, Nikolaou K, Gatidis S, Yang B (2018) MedGAN: medical image translation using GANs. arXiv preprint. arXiv:1806.06397

3. Ben-Cohen A, Klang E, Raskin SP, Amitai MM, Greenspan H (2017) Virtual PET images from CT data using deep convolutional networks: initial results. In: International workshop on simulation and synthesis in medical imaging. Springer, Cham, pp 49–57

4. Bi L, Kim J, Kumar A, Feng D, Fulham M (2017) 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, Cham, pp 43–51

5. Bullitt E, Zeng D, Gerig G, Aylward S, Joshi S, Smith JK, Lin W, Ewend MG (2005) Vessel tortuosity and brain tumor malignancy: a blinded study1. Acad Radiol 12(10):1232–1240

6. Bustin A, Voilliot D, Menini A, Felblinger J, de Chillou C, Burschka D, Bonnemains L, Odille F (2018) Isotropic reconstruction of MR images using 3D patch-based self-similarity learning. IEEE Trans Med Imaging 37:1932–1942

7. Chartsias A, Joyce T, Dharmakumar R, Tsaftaris SA (2017) Adversarial image synthesis for unpaired multi-modal cardiac data. In: International workshop on simulation and synthesis in medical imaging. Springer, Cham, pp 3–13

8. Chartsias A, Joyce T, Giuffrida MV, Tsaftaris SA (2018) Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans Med imaging 37(3):803–814

9. Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, Lee JH, Gold GE, Hargreaves BA (2018) Super-resolution musculoskeletal MRI using deep learning. Magn Reson Med 80(5):2139–2154

10. Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, Zhou J, Wang G (2017) Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Trans Med Imaging 36(12):2524–2535

11. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, Wang G (2017) Low-dose CT via convolutional neural network. Biomed Opt Express 8(2):679–694

12. Chen W (2007) Clinical applications of PET in brain tumors. J Nucl Med 48(9):1468–1481

13. Choi H, Lee DS (2018) Generation of structural MR images from amyloid PET: application to MR-less quantification. J Nucl Med 59(7):1111–1117

14. Costa P, Galdran A, Meyer MI, Niemeijer M, Abràmoff M, Mendonça AM, Campilho A (2018) End-to-end adversarial retinal image synthesis. IEEE Trans Med Imaging 37(3):781–791

15. Dadar M, Pascoal TA, Manitsirikul S, Misquitta K, Fonov VS, Tartaglia MC, Breitner J, Rosa-Neto P, Carmichael OT, Decarli C et al (2017) Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease. IEEE Trans Med Imaging 36:1758–1768

16. Dar SUH, Yurt M, Karacan L, Erdem A, Erdem E, Çukur T (2018) Image synthesis in multi-contrast MRI with conditional generative adversarial networks. arXiv preprint. arXiv:1802.01221

17. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK (2018) Generating synthetic CTS from magnetic resonance images using generative adversarial networks. Med Phys 45(8):3627–3636

18. Frangi AF, Tsaftaris SA, Prince JL (2018) Simulation and synthesis in medical imaging. IEEE Trans Med Imaging 37(3):673

19. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680

20. Han X (2017) MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 44(4):1408–1419

21. Hiasa Y, Otake Y, Takao M, Matsuoka T, Takashima K, Carass A, Prince JL, Sugano N, Sato Y (2018) Cross-modality image synthesis from unpaired data using cycleGAN. In: International workshop on simulation and synthesis in medical imaging. Springer, Cham, pp 31–41

22. Hsieh J et al (2009) Computed tomography: principles, design, artifacts, and recent advances. SPIE, Bellingham, WA

23. Huang Y, Shao L, Frangi AF (2017) Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning. IEEE Trans Med Imaging 37(3):815–827

24. Huang Y, Shao L, Frangi AF (2017) Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. arXiv preprint. arXiv:1705.02596

25. Isola P, Zhu JY, Zhou T, Efros AA (2016) Image-to-image translation with conditional adversarial networks. arXiv preprint. arXiv:1611.07004

26. Jog A, Carass A, Roy S, Pham DL, Prince JL (2017) Random forest regression for magnetic resonance image synthesis. Med Image Anal 35:475–488

27. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3D CNN with fully connected

CRF for accurate brain lesion segmentation. Med Image Anal 36:61–78

28. Kang E, Min J, Ye JC (2017) A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction. Med Phys 44(10):e360–e375

29. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105

30. Lê M, Delingette H, Kalpathy-Cramer J, Gerstner ER, Batchelor T, Unkelbach J, Ayache N (2017) Personalized radiotherapy planning based on a computational tumor growth model. IEEE Trans Med Imaging 36(3):815–825

31. Leynes AP, Yang J, Wiesinger F, Kaushik SS, Shanbhag DD, Seo Y, Hope TA, Larson PE (2017) Direct pseudoCT generation for pelvis PET/MRI attenuation correction using deep convolutional neural networks with multi-parametric MRI: zero echo-time and dixon deep pseudoCT (ZeDD-CT). J Nucl Med 59:852–858 (2017)

32. Liu F, Jang H, Kijowski R, Bradshaw T, McMillan AB (2017) Deep learning MR imaging–based attenuation correction for PET/MR imaging. Radiology 286(2):676–684

33. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R et al (2015) The multimodal brain tumor image segmentation benchmark (brats). IEEE Trans Med Imaging 34(10):1993–2024

34. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint. arXiv:1411.1784

35. Mok TC, Chung AC (2018) Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: International MICCAI brainlesion workshop. Springer, Cham, pp 70–80

36. Nie D, Cao X, Gao Y, Wang L, Shen D (2016) Estimating CT image from MRI data using 3D fully convolutional networks. In: Deep learning and data labeling for medical applications. Springer, Cham, pp 170–178

37. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q, Shen D (2018) Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng 65:2720–2730

38. Olut S, Sahin YH, Demir U, Unal G (2018) Generative adversarial training for MRA image synthesis using multi-contrast MRI. In: International workshop on predictive intelligence in medicine. Springer, Cham, pp 147–154

39. Peterfy C, Schneider E, Nevitt M (2008) The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthr Cartil 16(12):1433–1441

40. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241

41. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML (2019) Deep learning in medical imaging and radiation therapy. Med Phys 46(1):e1–e36

42. Van Nguyen H, Zhou K, Vemulapalli R (2015) Cross-domain synthesis of medical images using efficient location-sensitive deep network. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham. pp 677–684

43. van Tulder G, de Bruijne M (2015) Why does synthesized data improve multi-sequence classification? In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 531–538

44. Wang Y, Zhang P, An L, Ma G, Kang J, Shi F, Wu X, Zhou J, Lalush DS, Lin W et al (2016) Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. Phys Med Biol 61(2):791

45. Wang Y, Ma G, An L, Shi F, Zhang P, Lalush DS, Wu X, Pu Y, Zhou J, Shen D (2016) Semisupervised tripled dictionary learning for standard-dose PET image prediction using low-dose PET and multimodal MRI. IEEE Trans Biomed Eng 64(3): 569–579

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

47. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, Shen D (2019) 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans Med Imaging 38(6):1328

48. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

49. Wei W, Poirion E, Bodini B, Durrleman S, Ayache N, Stankoff B, Colliot O (2018) Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 514–522

50. Welander P, Karlsson S, Eklund A (2018) Generative adversarial networks for image-to-image translation on multi-contrast MR images-a comparison of cycle-GAN and unit. arXiv preprint. arXiv:1806.07777

51. Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, Shen D (2017) Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 267:406–416

52. Yang Q, Li N, Zhao Z, Fan X, Chang EI, Xu Y et al (2018) MRI cross-modality neuroimage-to-neuroimage translation. arXiv preprint. arXiv:1801.06940

53. Ye DH, Zikic D, Glocker B, Criminisi A, Konukoglu E (2013) Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 606–613

54. Yi X, Babyn P (2017) Sharpness-aware low dose CT denoising using conditional generative adversarial network. arXiv preprint. arXiv:1708.06453

55. Yi X, Walia E, Babyn P (2018) Generative adversarial network in medical imaging: a review. arXiv preprint. arXiv:1809.07294

56. Yu B, Zhou L, Wang L, Fripp J, Bourgeat P (2018) 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, Washington, DC, pp 626–630

57. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P (2019) Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans Med Imaging 38:1750–1762

58. Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z (2018) Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. Comput Biol Med 99:133–141

59. Zhang Z, Yang L, Zheng Y (2018) Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 9242–9251

Part II

Applications: Screening and Diagnosis