Computer Methods and Programs in Biomedicine 202 (2021) 106019

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Mass Image Synthesis in Mammogram with Contextual Information Based on GANs

Tianyu Shen[b,c,1], Kunkun Hao[d,1], Chao Gou[a,*], Fei-Yue Wang[b,e,f]

[a] School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing, China
[c] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[d] School of Software Engineering, Xi'an Jiaotong University, Xi'an, China
[e] Qingdao Academy of Intelligent Industries, Qingdao, China
[f] Institute of Systems Engineering, Macau University of Science and Technology, Macau, China

ARTICLE INFO

Article history:
Received 15 April 2020
Accepted 16 February 2021

Keywords:
medical image synthesis
generative adversarial network
mammogram
mass detection

ABSTRACT

Background and Objective: In medical imaging, the scarcity of labeled lesion data has hindered the application of many deep learning algorithms. To overcome this problem, the simulation of diverse lesions in medical images is proposed. However, synthesizing labeled mass images in mammograms is still challenging due to the lack of consistent patterns in shape, margin, and contextual information. Therefore, we aim to generate various labeled medical images based on contextual information in mammograms.

Methods: In this paper, we propose a novel approach based on GANs to generate various mass images and then perform contextual infilling by inserting the synthetic lesions into healthy screening mammograms. Through incorporating features of both realistic mass images and corresponding masks into the adversarial learning scheme, the generator can not only learn the distribution of the real mass images but also capture the matching shape, margin and context information.

Results: To demonstrate the effectiveness of our proposed method, we conduct experiments on publicly available mammogram database of DDSM and a private database provided by Nanfang Hospital in China. Qualitative and quantitative evaluations validate the effectiveness of our approach. Additionally, through the data augmentation by image generation of the proposed method, an improvement of 5.03% in detection rate can be achieved over the same model trained on original real lesion images.

Conclusions: The results show that the data augmentation based on our method increases the diversity of dataset. Our method can be viewed as one of the first steps toward generating labeled breast mass images for precise detection and can be extended in other medical imaging domains to solve similar problems.

1. Introduction

Deep learning methods have been widely and successfully applied for natural image analysis with the requirement of large-scale labeled datasets [1]. However, in medical imaging community, collecting a large amount of labeled data is usually challenging, due to the privacy policy in different medical institutions and the scarcity of corresponding diseases [2]. In addition, data annotation of medical images requires specific domain knowledge of experts, which is costly and time consuming [3]. Therefore, medical image synthesis is proposed as a means of data augmentation to address the aforementioned challenges [4–6].

Generative adversarial networks (GANs) [7] have recently made a significant progress in natural image generation. Many researchers have also applied GAN in medical imaging application such as chest X-rays generation [8] and liver lesions synthesis [9]. Nevertheless, the researches on GAN-based medical image synthesis focus on utilizing generated medical samples for augmentation of the original training set to alleviate the issues caused by the insufficiency of data across many classification tasks [8,10,11] rather than detection or segmentation tasks. Conventional GAN-based image synthesis models could only learn the data distribution from a specific class, and thus the generated samples are confined to a specified category. Generating medical images with corresponding

* Corresponding author.
E-mail address: gouchao.cas@gmail.com (C. Gou).
[1] Authors contributed equally to this work

https://doi.org/10.1016/j.cmpb.2021.106019
0169-2607/

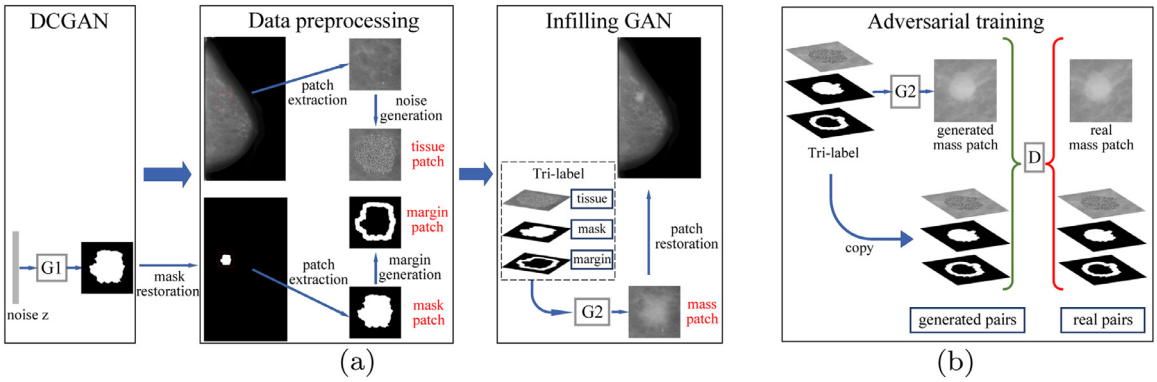

Fig. 1. (a) Overview of the mass image generation process, where G1 is the generator of DCGAN and G2 is proposed infilling generator. We combine the mass segmentation mask with its contextual margin information to synthesize mass image. The input of G2 is the concatenation of three grayscale channels, termed as Tri-label (Triple labels). (b) Adversarial translation scheme of the proposed infilling GAN. G2 is trained to generate mass images as realistic as possible from the three-channel input, the discriminator D is trained to distinguish between the generated pairs and real pairs. In the training stage of infilling GAN, only the mass patch images with precise annotation masks are used. The normal mammogram images are only used for testing and generation in (a).

precise masks is still a challenge. Actually, even for professional experts, the annotations of lesion region are not guaranteed to be completely accurate. And the contextual information surrounding the annotated mass region needs to be considered. Especially for screening mammograms, synthesizing realistic mass images is particularly hard due to the variety of mass in terms of texture and shape as well as the presence of intricate and diverse breast tissue surrounding the masses.

Inspired from the unsupervised anomaly detection with GAN [12] and image-to-image translation technique [13], in this paper, we apply the GAN framework to synthesize high quality mass images with accurate masks by learning the distribution of the real mass images as well as capturing the matching shape, margin and context information. Then the synthetic mass images are inserted into healthy screening mammograms for contextual infilling. The overall framework for mass generation is shown in Fig. 1. Subsequently, the generated labeled mammograms are used as augmented data for mass detection task. In the experiment, improvements for breast mass detection are achieved, as compared to traditional data augmentation methods or training with original data. All of the experiments are carried out on a public mammogram database named DDSM [14] and a private dataset provided by Nanfang Hospital of Southern Medical University. Our work successfully converts a large number of healthy mammorgrams which are readily available to the mammorgrams with lesion regions that are scarce and costly to annotate. This is meaningful for the clinical diagnosis based on medical imaging, since it improves the detection models under small and imbalanced medical image datasets.

The contributions of this paper are as follows:

1) We propose a novel GAN-based approach for generating high quality mass images in mammograms. The binary masks are introduced as condition in the adversarial learning scheme and it allows for generating mass images with accurate mask annotations. Thus, the synthesized images can be applied in downstream detection tasks.

2) Compared with conventional GANs for mass generation from random noise, contextual margin feature is incorporated in the proposed mass synthesis approach. It can generate synthetic mass images with richer texture and margin information.

3) Thorough experiments show that our proposed method can generate realistic lesion images with precise masks and further enhance the performance of deep detection network.

The rest of this paper is arranged as follows: In Section 2 we summarize the key theoretical underpinnings and representative applications of GANs as well as related work for medical image synthesis. In Section 3 we outline the proposed overall framework and methodology that builds on previously published literature, and in Section 4 we discuss the results of our experiments in detail. Conclusion and discussion are drawn in Sections 5 and 6.

2. Related work

In this section, we mainly summarize the research on medical image synthesis and introduce the knowledge about GAN-based method.

Medical Image Synthesis: The challenge of medical image scarcity may be partially met by data augmentation based on image synthesis. Recent developments of image synthesis typically fall under two categories including transformation-based and generative model-based approaches. The first and most common method based on transformation can be divided into mathematical affine transformation [1] and feature transformation [15]. Affine transformation is that new samples are created by applying simple mathematical operations such as translation, rotation, flip and scale to the real images [1,16,17], while feature transformation is achieved by applying modifications to feature vector of the existing samples [15,18,19]. Nevertheless, the above image synthesis methods are not suitable for high-quality and sophisticated types of data augmentation because little additional information can be obtained from these small modifications to original images. The second method based on generative models overcomes this shortcoming by producing samples that appear similar to real samples, but whose properties may not have been represented in the existing training set. Generative models leverage on training data to acquire the potential distribution of realistic samples and produce new samples from the same distribution [20–22]. Therefore, the potential path of synthetically generating data has been explored recently. In the work of [23], the authors use Parallel Vision (PV) [24] framework based on generative model to generate eye images, and achieve the state-of-the-art pupil detection results.

In medical imaging community, the approaches based on both transformations and generative models are applied in lots of precedent researches. In transformation approach, new lesions are simulated based on affine transformations of lesion properties in an existing dataset. A few examples of this approach can be found in [25–28] for mammography. In generative approach, new lesions are generated through learning the distribution and property of existing samples based on a generative model [8,9,29,30]. Unlike the first approach, the model learns to generate lesion images directly from data. A typical example is GAN-based method, which has led to a significant breakthrough in medical image synthesis. Many re-

searchers have applied GAN in medical image synthesis [8,9,31], in which artificial medical images are generated directly by random noise, producing results with low quality. There are some methods developed to insert lesions in medical images. For example, Jin et al. introduce a 3D-GAN model in an images-to-image translation setting to synthesize nodules in lung CTs [32]. Baur et al. produce to synthesize highly realistic dermoscopic images of skin lesions with GANs [33]. But the contextual feature are not utilized in these GAN-based frameworks. A data-driven approach has been proposed in [29] to generate CT image from its corresponding MR image. The authors further apply AutoContextModel to implement a context-aware GAN, It is the first work to introduce context-aware GANs in medical image synthesis, but pathological information is not considered in this method. Wu et al. [30] propose conditional infilling GAN to synthesize images of mass area in the mammogram, yet only the lesion image and mask information are utilized while the contextual margin information are neglected.

In this paper, we propose a novel GAN-based approach for generating high quality mass images in mammograms. Different from the aforementioned studies, we incorporate contextual margin feature into the adversarial learning scheme to generate mass images with richer texture and margin information. In addition, most of the above studies focus on synthesizing new samples from a particular class the same as the input data, so the synthetic medical images can be only used for downstream classification tasks. In contrast, our method introduces the binary masks as conditional input of the generator and it allows for generating mass images with precise mask annotations. Therefore, the augmented data can be used in detection or segmentation task.

Generative Adversarial Networks: The framework for training generative models in an adversarial manner was first introduced in the work of Goodfellow et al. [7]. GANs are composed of a generator and a discriminator, which compete with each other over the training data to improve their performance. The aim of GANs is to learn the potential distribution of real data and generate new samples from the same distribution, which guarantees the authenticity of the generated images. Various derivative methods of GANs have gained great popularity in the computer vision community such as high-quality natural images generation [13,22,34], style transformation [13,35], image inpainting [36] and super-resolution [37].

While the original GANs learn a mapping from a random input noise vector z to an output image y, conditional GANs (cGANs) [34] are trained to learn a mapping from an input image x (conditional information) and random noise z to y. One of the famous methods based on cGAN is the image-to-image translation [13]. The loss function of a cGAN can be expressed as:

$$\mathcal{L}_{cGAN}(G,D) = E_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + E_{x\sim p_{data}(x),z\sim p_{data}(z)}[\log(1-D(x,G(x,z)))] \quad (1)$$

where G tries to minimize this objective against an adversarial D that tries to maximize it. The adversarial score in the loss computation of the generator strengthens its capabilities to provide a valid generation. Besides, a traditional L1 loss is added to boost the learning process. As a result, the generator will play a role in not only fooling the discriminator but also being close to the ground truth in an L1 sense.

In medical imaging, the adversarial theory from GAN framework has achieved impressive results in several applications such as image enhancement [38,39], medical images synthesis [8–10,22,29,30], detection [12,40], and segmentation [41–43]. The ways in which GANs are used in the field of medical imaging can generally be divided into two categories[31]. One category pays attention to the generative aspect to generate similar data by learning the data distribution of a certain category, so as to alleviate the problem of data scarcity in medical imaging. The other focuses on the discriminative aspect, and the discriminator can be regarded as a classifier to identify abnormal images or regions.

In our work, we aim to synthesize abnormal lesion regions in mammographic images in terms of the generative aspect. A conditional generative adversarial network is used in an images-to-image translation setting to learn a mapping from segmentation masks with contextual information to the corresponding realistic mass images as well as to learn the distribution of real lesion images. The addition of conditional information and contextual information helps to better utilize the shape/margin feature derived from the masks.

The previous studies on GANs that integrate the contextual information [44,45] focus on applying the modified convolution operation inside the network to extract contextual features. It is inadequate for medical image synthesis different from image inpainting, because we focus on the generation of lesion region with surrounding contextual tissue in medical imaging. Thus, we propose to incorporate contextual margin feature as input conditions in the adversarial learning scheme to generate synthetic mass images with richer texture and margin information. The design of the methods is completely different from above research.

3. Framework and Method

3.1. Overall framework

We propose a united framework for the breast mass generation in mammograms based on GANs. As shown in Fig. 1(a), the framework is composed of three stages.

Firstly, the Deep Convolutional GAN [46] (DCGAN) is trained to learn the underlying distribution of segmentation masks of real mass images and generate various masks from random input noise z. The white region of the masks represents the location of the lesion/mass and black represents the background tissue.

Secondly in the data preprocessing stage, we extract an arbitrary tissue patch from a normal mammogram, where there are no mass or lesion findings. At the same time a generated mask is inserted into a black background with the same size as the normal mammogram image, and the location of mask in the background is the same as the location of the tissue patch in the normal mammogram image. Then the tissue patch is corrupted by noise with the lesion region replaced with uniformly random values between 0 and 255. The addition of random noise is important for the robustness of infilling GAN. Because all the normal tissue patches containing healthy breast tissue are composed of similar pixel features, while the pathological tissue patches containing breast mass are composed of different pixel features. If the input tissue patch is always covered by the same type of pixel information, the deep neural network would only learn features that are robust against that particular type of patch information. The addition of random noise can enhance the generalization of the network and avoid the circumstance that the network only performs well for certain pixel features. Also, we obtain the contextual margin information (termed as margin patch) by applying morphological operation to the generated mask patch.

In the third stage, the corrupted tissue patch, mask patch and margin patch are combined into a concatenated stack of three channels. Then we use the trained generator of infilling GAN, which is trained in an adversarial scheme as shown in Fig. 1(b), to generate a corresponding mass patch. Finally, the generated mass patch is inserted into the original normal mammogram image at the original position.

The architecture of DCGAN and infilling GAN models are demonstrated in Fig. 2 and Fig. 3, respectively. The models are discussed in details below.

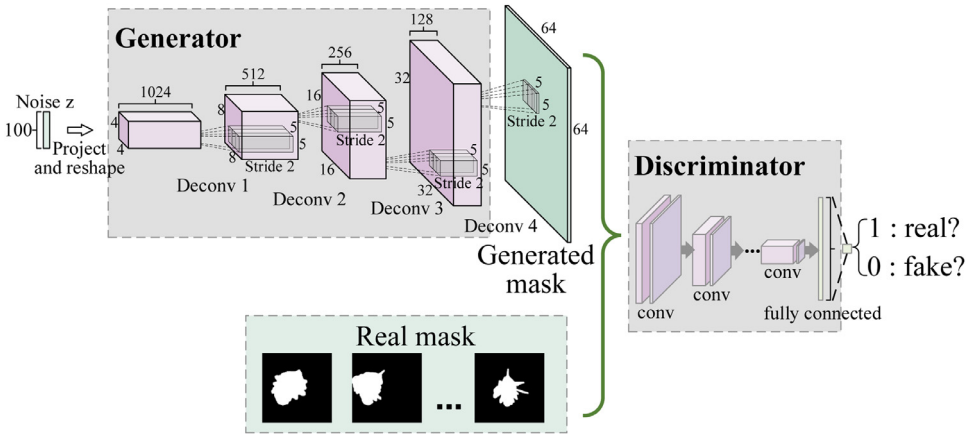

Fig. 2. DCGAN for mask generation from random input noise *z*.

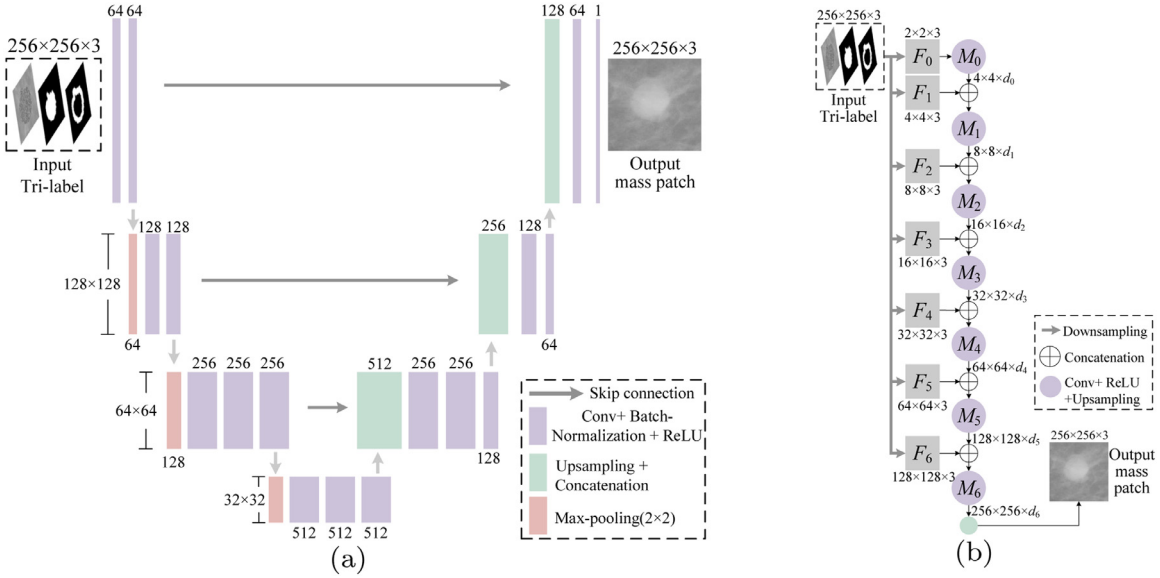

Fig. 3. Two choices for the generator architecture of Infilling GAN. (a) U-Net architecture. The result of the upsampling is connected to the feature map of the corresponding layer in the downsampling path by skip-connection. (b) CRN architecture. M_i denotes the i^{th} single refinement module, which contains two convolutional layers with ReLU activation functions and an upsampled operation at a given resolution. Resolution is doubled between sequential modules. F_i is obtained by applying an downsampled operation on the input stack (Tri-label) at a given resolution. d_i represents the channel number of the output of M_i.

3.2. DCGAN for mass mask generation

Medical datasets are often highly imbalanced with overrepresentation of common medical problems and a paucity of data from rare conditions. For mammogram data, the normal images far exceed those with lesion/mass findings, but such normal data are often not effectively used.

Our goal is to generate mass images with precise annotation masks and transform normal mammograms into the mammograms with mass findings. The generated data can be used to augment the labeled dataset of mammograms with mass findings, which is conducive for deep learning methods that require a lot of training data.

We apply the DCGAN model [46] to generate various segmentation masks. The architecture and learning scheme are shown in Fig. 2. The generator learns the underlying distribution $p_{data}(x)$ from the real mask data x, and then continuously generates masks from random input noise z, where z is obtained from a known simple distribution $p_{data}(z)$, usually a uniform distribution. The loss function can be expressed as:

$$\min_G \max_D \mathcal{L}_{DCGAN}(G, D) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_{data}(z)}[\log(1 - D(G(z)))] \quad (2)$$

where $G(z)$ is the mapping function from the input space to output image space of distribution p_g. The goal of G is to achieve $p_g = p_{data}(x)$.

3.3. Infilling GAN for adversarial translation from Tri-label to mass image

Considering the mass annotation problem and the importance of mass contextual information, the proposed infilling GAN method is based on image-to-image translation technique.

Traning Scheme: The proposed adversarial translation framework from labels to mass image is shown in Fig. 1(b). The input of the generator is composed of three grayscale channels, termed as Tri-label (Triple labels). The first channel is the tissue patch with

segmentation region filled by using uniformly random values between 0 and 255, termed as tissue patch. The second channel is the generated mask from DCGAN, termed as mask patch. It represents the mass region of mammogram. The third channel is the margin information obtained from the mask patch, termed as margin patch. The mask patch and margin patch restrict the generated area and imply the shape/margin features. The tissue patch provides the contextual information of the normal tissue.

The generator is trained to output a mass patch image matching to the input three-channel labels and generate images as realistic as possible. The discriminator is trained to distinguish between generated pairs, which is composed of the three-channel input of generator and generated mass patch image, and real pairs, which is composed of the three-channel input and real mass patch image. It is worth noticing that in the training stage of infilling GAN, only the mass patch images with precise annotation masks are used. The normal mammogram images without lesion findings are only used for testing and generation process in Fig. 1(a).

Therefore, the goal of G is to maximize the classification error of D, while D aims to minimize the classification error until the Nash equilibrium is reached. This training process is called adversarial learning. Similar as in [34], the adversarial loss function is as follows:

$$\min_G \max_D \mathcal{L}_{adv}(G, D) = \begin{array}{l} \mathbb{E}_{(x,R)\sim p_{data}(x,R)}[\log D(x, R)]+ \\ \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D(x, G(x)))] \end{array} \quad (3)$$

where R is the real mass images. x represents the Tri-label denoting the concatenation of tissue patch, mask patch and margin patch obtained from corresponding R. In the training stage of infilling GAN, only the mammogram images with mass findings and annotation masks are used. The method for obtaining corresponding Tri-label is the same as the data preprocessing stage in Fig. 1(a). The Tri-label information is added into both the generator and discriminator so that all the features and matching information can be better learned.

In order to overcome the instability of GAN training, we introduce perceptual loss or feature matching loss [47], termed as $\mathcal{L}_{fm}$. The basic idea is to calculate the convolutional layer activation error between real images and generated images using a visual perception network. In this work we use VGG-19 [48] pre-trained on the ImageNet dataset. Let $\{\phi_l\}$ represent the collection of layers in the VGG-19 network ϕ, the feature matching loss is defined by:

$$\mathcal{L}_{fm}(G) = \sum_{l=0}^{5} \lambda_l \|\phi_l(R) - \phi_l(G(x; \omega_G))\|_1 \quad (4)$$

where R denotes the real image and $G(x; \omega_G)$ denotes the generated image. ϕ_0 represents an identity map of the input image. For $\phi_l (l \geq 1)$, we use the l^{th} convolutional layer in VGG-19. The hyperparameters $\{\lambda_l\}$ are initialized to the inverse of the number of elements in each layer. The addition of $\mathcal{L}_{fm}$ can guide the generation model to cover fine-grained details and more global context information by simultaneously matching low-level and high-level activations in a perceptual network.

Finally, the overall loss function is

$$\mathcal{L}_{loss} = \arg\min_G \max_D [\mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_{fm}(G)] \quad (5)$$

where λ is used to balance the contribution of the two losses, here we set $\lambda = 100$.

Network Architectures: For the design of generator architecture, we apply two improved generator including Cascaded Refinement Network (CRN) [49] and U-net [50]. Overview of the two architectures is shown in Fig. 3.

CRN is a serial cascade of refinement modules. Each refinement module M_i contains two convolutional layers with ReLU activation functions and an upsampled operation at a given resolution. In our implementation, the resolution of the first module (M_0) is $2 \times 2 \times 3$. And the resolution of the last module (M_n) is $256 \times 256 \times 3$. Resolution is doubled between sequential modules. This structure enables the network to cover the low-level fine-grained features (shape/margin feature) and the high-level general features such as objects and tissue categories.

U-Net architecture is an encoder-decoder network in general, except that skip connections are added between mirrored layers in the downsampling and upsampling block. The result of the upsampling is connected to the feature map of the corresponding layer in the downsampling path by skip connection. In our implementation, the input size of U-Net is $256 \times 256 \times 3$, and the output size is $256 \times 256 \times 1$.

The discriminator network has a typical CNN architecture similar as DCGAN. By comparing the generated pairs with real pairs, the discriminator outputs the decision that whether this mass region is real or not. Specifically, a PatchGAN proposed in [13] is used in the discriminator. That is, the discriminator is required to classify if each patch is real or fake and output a matrix with the size of 30×30. Then it takes the average value as a final result.

4. Experimental Evalution

4.1. Data and preprocessing

The DDSM (Digital Database for Screening Mammography) dataset [14] is a public mammogram dataset containing 10,480 total images. We select 765 mammogram images with malignant mass findings and accurate masks in this work. To prevent the overfitting problem, the original samples are augmented by rotating, flipping, and scaling. The resulted dataset is in a total of 3825 images.

The private dataset provided by Nanfang Hospital in China has a total of 549 cases from Chinese patients with both breasts affected (four images per case) containing 376 mass regions. We select 275 mammogram images with malignant mass findings and accurate masks in this work. Similarly, the original samples are augmented into 1375 images.

The following preprocessing procedures are performed on the two datasets: (1) the tape marks and white noises are removed from the mammogram images; (2) the mass patches containing the malignant mass from mammograms are extracted and then scaled to 256×256, and the same procedure is conducted on the mask; (3) the original segmentation mask region is resized and scaling up to 296×296 to obtain an enlarged segmentation mask. Then the original segmentation mask region is subtracted from the enlarged segmentation region to obtain the margin patch. In this step, we conduct experimental analysis on how many pixels are needed for obtaining an adequate margin. Three sizes that are 276×276, 296×296 and 316×316 are tested and the size of 296×296 are observed to be the most reasonable; (4) the obtained margin patch and mask patch are concatenated with the tissue patch into three-channel labels (Tri-label).

4.2. Details of training

For infillingGAN model, it is trained using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.5$, and the learning rate is 0.0002, the mini-batch size is 2. The number of iterations of training is about 100,000. The hyperparameters $\{\lambda_l\}$ are initialized to the inverse of the number of elements in each layer and the balance item λ in the loss function is 100. For DCGAN model, Adam optimizer is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$. The learning rate is 0.0002, and the mini-batch size is 32. The number of iterations of training is 3,000.

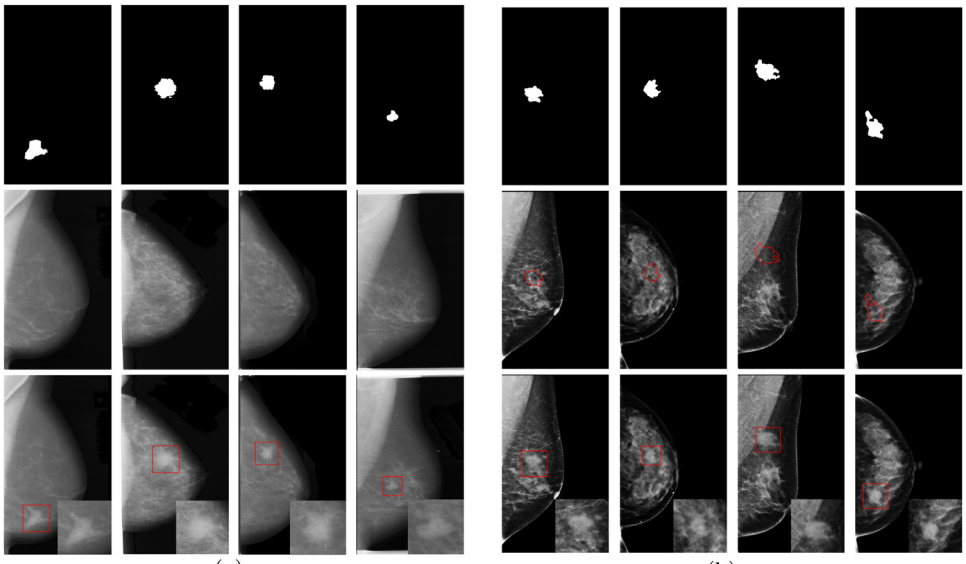

Fig. 4. Generated mammogram images with mass findings (the third row) transformed from normal mammogram images (the second row) and given mask (the first row). The corresponding annotation including bounding-box label and segmentation mask. Zoom in for details. (a) DDSM dataset. (b) Private dataset.

4.3. Mass synthesis in mammograms

Firstly, we simulate the size of generated mass images. We calculate the mean and variance of the height or width of original unscaled mass patches (padded 100 pixels around mass). We perform a pre-processing procedure to remove the white noise from the synthetic mask generated by DCGAN. Then the Gaussian distribution is applied to simulate the size (height and width) of the generated mass images.

Subsequently, we simulate the position of generated mass patch in the normal full mammogram images. We perform the mass generation experiment on normal mammograms using the trained generator of infilling GAN. In order to simulate the position of the mass patch in the full mammogram, the mammogram images with the breast on the right side are flipped. Let w and h denote the width and height of the mammogram, and we use a uniform distribution to select the abscissa $X \sim U(w/10, w/2)$ and ordinate $Y \sim U(h/4, 3 \times h/4)$ of the breast tissue patch on the normal mammogram images. Then the generated mass patch is restored to the corresponding mammogram according to the recorded coordinate position (i.e. the breast tissue patch at original position is replaced by the generated mass patch). At the same time, an automatic annotation procedure including bounding-box location for detection and segmentation mask for segmentation is finished.

As shown in Fig. 4, some normal mammogram images are transformed into the mammogram images with mass findings. The corresponding annotations including bounding-box label and segmentation mask are also shown.

The binary masks are applied as indispensable precondition in the adversarial learning scheme for generating mass images with accurate mask annotations. The mask patch only restricts the generated area and imply the shape features. In contrast, the margin patch provides the contextual margin feature better. When the margin patch is added, some of the parameters of the deep neural network focus on learning the information of the global block, and some parameters focus on learning about the local edge information. Therefore, we conduct the qualitative and quantitative experiments with the margin patches removed to demonstrate the importance of contextual margin feature in the mass generation process. Some qualitative comparisons of generated mass patch are shown in Fig. 5. The first row shows the generated mass patches without the contextual margin information, that is, the margin patch is not contained in the input Tri-label (only two-channel input). The second row shows the generated results with the context margin information, and the third row shows the real mass patch images. The results show that the masses generated by the proposed method have richer texture and margin information.

4.4. Quantitative quality evaluation

Recently, a quantitative quality analysis method, Learned Perceptual Image Patch Similarity (LPIPS) [51] metric, is proposed. The basic idea is to calculate the distance between two similar images through a network $\mathcal{F}$ such as AlexNet.

In this work, we apply the LPIPS metric to evaluate the quality of generated images quantitatively. 100 real mass images are randomly selected and denoted as R. 100 mass images generated by the CRN generator are randomly selected and denoted as G_{crn}. And 100 mass images generated by the U-Net generator are randomly selected and denoted as G_{unet}. In order to verify the effectiveness of the proposed contextual margin information, we also perform contrast experiments between using only the first two channels and all the three channels (Tri-label) for the input. G_{crn-} denotes the generated images by the CRN using two channels as input. G_{unet-} denotes the generated images by the U-Net generator using two channels. Then we calculate the LPIPS distance mean between 4 pairs of image groups: R and G_{crn-}, R and G_{crn}, R and G_{unet-}, R and G_{unet}. The quantitative quality evaluation results for generated images in terms of mean LPIPS distance on DDSM and Private datasets are shown in Table 1. Moreover, we apply the generated images by all above different designs to supplyment the original real training dataset as augmented data in the detection tasks as shown in section 4.5.

4.5. Mass detection using generated data augmentation

Our proposed method not only achieves mass generation but also provides an automatic annotation procedure, which further alleviates the problem of data scarcity in the lesion detection or segmentation tasks. In this work, we use the detection model

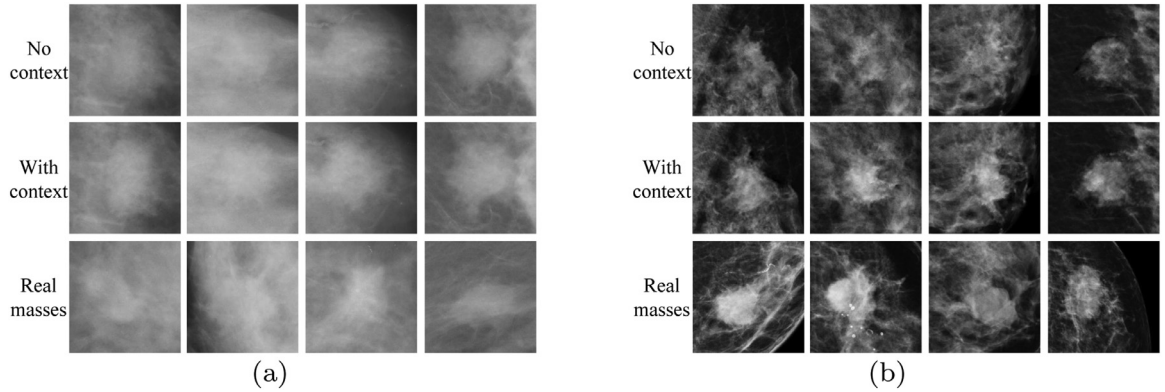

Fig. 5. Qualitative comparison of generated mass patch with/without contextual margin feature in the mass generation process (using CRN generator). (a) DDSM dataset. (b) Private dataset.

Table 1
Quantitative quality evaluation for generated images in terms of mean LPIPS distance and detection accuracy evaluation in terms of Recall(TPR) on DDSM and Private datasets.

		DDSM Dataset	Private Dataset
LPIPS Distance	G_{crn-} and R	0.02549	0.17134
	G_{crn} and R	0.02436	0.16768
	G_{unet-} and R	0.02170	0.15839
	G_{unet} and R	**0.02021**	**0.15342**
Recall (TPR)	CRN + two-channel input	0.8165	0.8100
	CRN + Tri-label input	0.8188	0.8128
	U-Net + two-channel input	0.8259	0.8177
	U-Net + Tri-label input	**0.8273**	**0.8182**

Light-Head R-CNN [52] to verify the effectiveness of the proposed generated data augmentation. Light-Head R-CNN constructs a lightweight head R-CNN network that not only has high accuracy, but also maintains high time efficiency, which is suitable for high-resolution medical images such as mammograms.

For Light-Head R-CNN model, we use a batch size of 4 with a learning rate of 0.0002 for 10 epochs. The input images are cropped and resized to 800×1333 while ensuring that the image does not deform.

We apply the traditional data augmentation method as baseline to compare with the proposed data augmentation method. In the same way as [10], we record the Recall (equal to True Positive Rate (TPR)) of mass detection by continuously adding data with data augmentation methods to the original training set. The metric is defined in the following equation:

$$\text{Recall(TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

where TP and FN are defined to represent the number of true positive detection boxes (i.e. mass regions that are successfully detected) and false negative detection boxes (i.e. mass regions that are not detected), respectively. In our experiments, a potential detection box is considered to be successfully detected when the IOU exceeds 0.5.

Taking DDSM dataset as example, we use the traditional data augmentation method (rotate, flip, and scale) to obtain 4000 images, and the number of images added to training set each time is denoted as $tra_1 = 300$, $tra_2 = 300$, $tra_3 = 400$, $tra_4 = 500$..., $tra_7 = 500$, $tra_8 = 1000$. We notice that the detection performance is not improved when the training set size is 2066. Then we continue to supplement 2800 generated images on the basis of 2066 real images, and the number of images added each time is denoted as $gan_1 = 300$, $gan_2 = 500$..., $gan_6 = 500$. To further validate and compare our proposed methods, we apply the generated images by different designs, which are U-Net generator using Tri-label as input, U-Net generator using two channels as input, CRN generator using Tri-label as input and CRN generator using two channels as input, to increase the training samples. The processes on DDSM and Private datasets are shown in Fig. 6 and Fig. 7, respectively. We note that as the amount of generated data increases, the corresponding TPR value increases largely, indicating that the masses generated by proposed method increase the diversity of the data. Also, the final detection accuracy in terms of Recall(TPR) is shown in Table1.

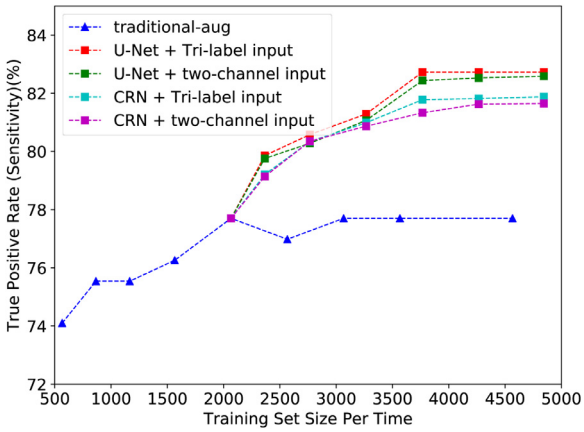

Fig. 6. Detection results with the increase of training set on DDSM dataset.

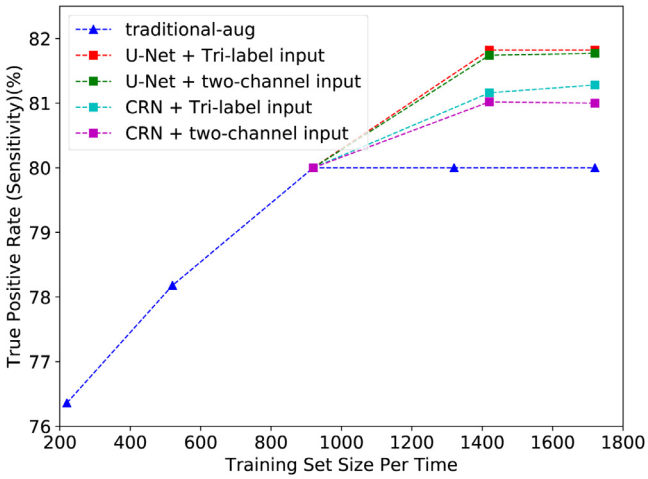

Fig. 7. Detection results with the increase of training set on Private dataset.

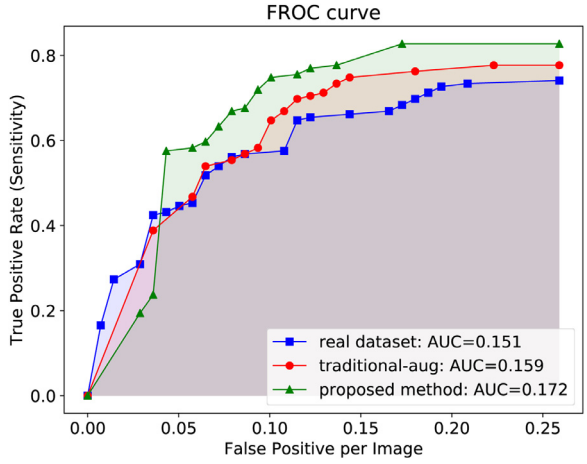

Fig. 8. FROC curves of mass detection on DDSM dataset. Proposed method denotes the data augmentation method through the generated images by the U-Net generator using Tri-label input.

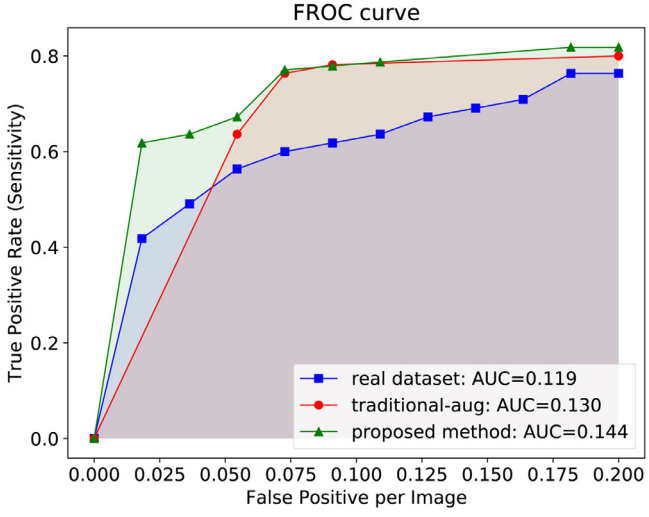

Fig. 9. FROC curves of mass detection on Private dataset. Proposed method denotes the data augmentation method through the generated images by the U-Net generator using Tri-label input.

To better illustrate the improvement in mass detection performance, free-response receiver operating characteristic (FROC) curves [53] and their AUCs (Area Under the Curve) are used to evaluate the detection performance using different confidence thresholds of the detection bounding-box. The horizontal axis is for false positive per image (FFPI) and vertical axis is for true positive rate (TPR). Larger AUC represents better performance [54]. The FFPI metric is defined to represent the number of false positive detection boxes (i.e. the normal regions that are detected as mass regions) per image. The FROC curve is shown in Fig. 8 and Fig. 9. Although our proposed method does not perform the best at certain specific thresholds, it does the best overall with an AUC of 0.172 for DDSM dataset in Fig. 8 and 0.144 for Private dataset in Fig. 9.

5. Conclusion

In conclusion, we propose a novel GAN-based approach for generating high quality mass images, which are usually scarce and costly to annotate, from plenty of normal mammogram images. We introduce the binary masks combining with contextual information based on pathological features, such as shape, margin, and density, in order to generate mass patch with richer texture imformation. Also, our method achieves automatic annotation of generated mass patch with different locations and patterns. Furthermore, we use the generated mass images for data augmentation in the detection task and the performance is significantly improved. The result shows that data augmentation using GAN-based generated medical images increases the diversity of dataset. Our method is meaningful for improving the detection models under small and imbalanced medical image datasets and potentially, can be extended in other medical imaging domains to solve similar problems. In the future, we will further investigate the improvement of model architecture, such as the application of spatially adaptive normalization [55] in the Infilling GAN, and the utilization of different adversarial loss.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest to this work entitled “Mass Image Synthesis in Mammogram with Contextual Information Based on GANs”.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (2020YFB1600400) and National Natural Science Foundation of China under Grant 61806198, 61533019 . The authors would like to thank Nanfang Hospital in China.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cmpb.2021.106019

References

[1] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.
[2] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique, IEEE Transactions on Medical Imaging 35 (5) (2016) 1153–1159.
[3] T. Shen, J. Wang, C. Gou, F.-Y. Wang, Hierarchical Fused Model With Deep Learning and Type-2 Fuzzy Learning for Breast Cancer Diagnosis, IEEE Transactions on Fuzzy Systems 28 (12) (2020) 3204–3218.
[4] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al., Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: International conference on information processing in medical imaging, Springer, 2017, pp. 597–609.
[5] P. Costa, A. Galdran, M.I. Meyer, M.D. Abràmoff, M. Niemeijer, A.M. Mendonça, A. Campilho, Towards adversarial retinal image synthesis, arXiv preprint: 1701.08974(2017).
[6] T. Shen, C. Gou, F.-Y. Wang, Z. He, W. Chen, Learning from adversarial medical images for x-ray breast mass segmentation, Comput. Methods Progr. Biomed. 180 (2019) 105012.
[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.
[8] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, J. Barfett, Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 990–994.
[9] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using gan for improved liver lesion classification, in: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 289–293.
[10] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.
[11] A. Pezeshk, N. Petrick, W. Chen, B. Sahiner, Seamless lesion insertion for data augmentation in cad training, IEEE transactions on medical imaging 36 (4) (2017) 1005–1015.
[12] T. Schlegl, P. Seeböck, S.M. Waldstein, U. Schmidt-Erfurth, G. Langs, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 146–157.

[13] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.
[14] M. Heath, K. Bowyer, D. Kopans, R. Moore, W.P. Kegelmeyer, The digital database for screening mammography, in: Proceedings of the 5th international workshop on digital mammography, Medical Physics Publishing, 2000, pp. 212–218.
[15] L. Maaten, M. Chen, S. Tyree, K. Weinberger, Learning with marginalized corrupted features, in: International Conference on Machine Learning, 2013, pp. 410–418.
[16] C. Papageorgiou, T. Poggio, A trainable system for object detection, International journal of computer vision 38 (1) (2000) 15–33.
[17] I. Laptev, Improving object detection with boosted histograms, Image and Vision Computing 27 (5) (2009) 535–544.
[18] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.
[19] P. Melville, R.J. Mooney, Constructing diverse classifier ensembles using artificial training examples, in: IJCAI, 3, Citeseer, 2003, pp. 505–510.
[20] T.K. Ho, H.S. Baird, Large-scale simulation studies in image pattern recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (10) (1997) 1067–1079.
[21] A. Pezeshk, R.L. Tutwiler, Automatic feature extraction and text recognition from scanned topographic maps, IEEE transactions on Geoscience and Remote sensing 49 (12) (2011) 5047–5063.
[22] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, arXiv preprint: 1610.09585(2016).
[23] C. Gou, H. Zhang, K. Wang, F.-Y. Wang, Q. Ji, Cascade learning from adversarial synthetic images for accurate pupil detection, Pattern Recognition 88 (2019) 584–594.
[24] K. Wang, C. Gou, N. Zheng, J.M. Rehg, F.-Y. Wang, Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives, Artificial Intelligence Review 48 (3) (2017) 299–329.
[25] T. Kooi, B. van Ginneken, N. Karssemeijer, A. den Heeten, Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network, Medical physics 44 (3) (2017) 1017–1027.
[26] R. Saunders, E. Samei, J. Baker, D. Delong, Simulation of mammographic lesions, Academic radiology 13 (7) (2006) 860–870.
[27] A. Rashidnasab, P. Elangovan, M. Yip, O. Diaz, D. Dance, K. Young, K. Wells, Simulation and assessment of realistic breast lesions using fractal growth models, Physics in Medicine & Biology 58 (16) (2013) 5613.
[28] I.S. Sahli, H. Bettaieb, A.B. Abdallah, I. Bhouri, M.H. Bedoui, Synthesis of mammographic images based on the fractional brownian motion, International Journal of Medical Imaging 6 (1) (2018) 1.
[29] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 417–425.
[30] E. Wu, K. Wu, D. Cox, W. Lotter, Conditional infilling gans for data augmentation in mammogram classification, in: Image Analysis for Moving Organ, Breast, and Thoracic Images, Springer, 2018, pp. 98–106.
[31] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Medical image analysis 58 (2019) 101552.
[32] D. Jin, Z. Xu, Y. Tang, A.P. Harrison, D.J. Mollura, Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 732–740.
[33] C. Baur, S. Albarqouni, N. Navab, Generating highly realistic images of skin lesions with gans, in: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer, 2018, pp. 260–267.
[34] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint: 1411.1784(2014).
[35] D. Ulyanov, A. Vedaldi, V.S. Lempitsky, Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis., in: CVPR, 1, 2017, p. 3.
[36] R. Yeh, C. Chen, T.Y. Lim, M. Hasegawa-Johnson, M.N. Do, Semantic image inpainting with perceptual and contextual losses. arxiv preprint, arXiv preprint: 1607.07539 2 (2016).
[37] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A.P. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network., in: CVPR, 2, 2017, p. 4.
[38] J.M. Wolterink, T. Leiner, M.A. Viergever, I. Išgum, Generative adversarial networks for noise reduction in low-dose ct, IEEE transactions on medical imaging 36 (12) (2017) 2536–2545.
[39] X. Yi, P. Babyn, Sharpness-aware low-dose ct denoising using conditional generative adversarial network, Journal of digital imaging (2018) 1–15.
[40] V. Alex, M.S. KP, S.S. Chennamsetty, G. Krishnamurthi, Generative adversarial networks for brain lesion detection, in: Medical Imaging 2017: Image Processing, 10133, International Society for Optics and Photonics, 2017, p. 101330G.
[41] Y. Xue, T. Xu, H. Zhang, L.R. Long, X. Huang, Segan: Adversarial network with multi-scale l1 loss for medical image segmentation, Neuroinformatics (2018) 1–10.
[42] W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, E.P. Xing, Scan: Structure correcting adversarial network for chest x-rays organ segmentation. arxiv preprint, arXiv preprint: 1703.08770(2017).
[43] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al., Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 597–609.
[44] G. Liu, F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 85–100.
[45] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, Free-form image inpainting with gated convolution, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, pp. 4471–4480.
[46] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint: 1511.06434(2015).
[47] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European conference on computer vision, Springer, 2016, pp. 694–711.
[48] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint: 1409.1556(2014).
[49] Q. Chen, V. Koltun, Photographic image synthesis with cascaded refinement networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1511–1520.
[50] O. Ronneberger, P. Fischer, 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.
[51] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.
[52] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, J. Sun, Light-head r-cnn: In defense of two-stage object detector, arXiv preprint: 1711.07264(2017).
[53] J.P. Egan, G.Z. Greenberg, A.I. Schulman, Operating characteristics, signal detectability, and the method of free response, The Journal of the Acoustical Society of America 33 (8) (1961) 993–1007.
[54] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, D. Feng, Dermoscopic image segmentation via multistage fully convolutional networks, IEEE Transactions on Biomedical Engineering 64 (9) (2017) 2065–2074.
[55] T. Park, M. Liu, T. Wang, J. Zhu, Semantic image synthesis with spatially-adaptive normalization, CoRR abs/1903.07291 (2019) 1903.07291.