



Cross-domain attention-guided generative data augmentation for medical image analysis with limited data

Zhenghua Xu^a, Jiaqi Tang^a, Chang Qi^{a,b,*}, Dan Yao^a, Caihua Liu^c, Yuefu Zhan^d, Thomas Lukasiewicz^{b,e}

^a State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China

^b Institute of Logic and Computation, Vienna University of Technology, Vienna, Austria

^c College of Computer Science and Technology, Civil Aviation University of China, Tianjin, China

^d Department of Radiology, Hainan Women and Children's Medical Center, Haikou, China

^e Department of Computer Science, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:

Computer-aided diagnosis
Data augmentation
Generative Adversarial Networks
Cross-domain medical image generation

ABSTRACT

Data augmentation is widely applied to medical image analysis tasks in limited datasets with imbalanced classes and insufficient annotations. However, traditional augmentation techniques cannot supply extra information, making the performance of diagnosis unsatisfactory. GAN-based generative methods have thus been proposed to obtain additional useful information to realize more effective data augmentation; but existing generative data augmentation techniques mainly encounter two problems: (i) Current generative data augmentation lacks of the capability in using cross-domain differential information to extend limited datasets. (ii) The existing generative methods cannot provide effective supervised information in medical image segmentation tasks. To solve these problems, we propose an attention-guided cross-domain tumor image generation model (CDA-GAN) with an information enhancement strategy. The CDA-GAN can generate diverse samples to expand the scale of datasets, improving the performance of medical image diagnosis and treatment tasks. In particular, we incorporate channel attention into a CycleGAN-based cross-domain generation network that captures inter-domain information and generates positive or negative samples of brain tumors. In addition, we propose a semi-supervised spatial attention strategy to guide spatial information of features at the pixel level in tumor generation. Furthermore, we add spectral normalization to prevent the discriminator from mode collapse and stabilize the training procedure. Finally, to resolve an inapplicability problem in the segmentation task, we further propose an application strategy of using this data augmentation model to achieve more accurate medical image segmentation with limited data. Experimental studies on two public brain tumor datasets (BraTS and TCIA) show that the proposed CDA-GAN model greatly outperforms the state-of-the-art generative data augmentation in both practical medical image classification tasks and segmentation tasks; e.g. CDA-GAN is 0.50%, 1.72%, 2.05%, and 0.21% better than the best SOTA baseline in terms of ACC, AUC, Recall, and F1, respectively, in the classification task of BraTS, while its improvements w.r.t. the best SOTA baseline in terms of Dice, Sens, HD95, and mIOU, in the segmentation task of TCIA are 2.50%, 0.90%, 14.96%, and 4.18%, respectively.

1. Introduction

At present, medical image analysis encounters a common dilemma, requiring a large dataset for model training [1]. However, the collection of medical imaging data faces many constraints. In the context of classification tasks, the model's performance is invariably tethered to the size of the available data. Concurrently, the problem of data class

imbalance also contributes to a reduction in the model's precision [2]. For segmentation tasks [3,4], accurate annotations require a lot of time and energy for experts and are difficult to obtain, which limits the performance improvement of the segmentation model. To achieve accurate medical image analysis with limited data, a data augmentation method is necessary for medical image analysis.

* Corresponding author at: State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, 300130, China.

E-mail address: changqi97@gmail.com (C. Qi).

<https://doi.org/10.1016/j.complbiomed.2023.107744>

Received 5 April 2023; Received in revised form 12 November 2023; Accepted 20 November 2023

Available online 23 November 2023

0010-4825/© 2023 Elsevier Ltd. All rights reserved.

Currently, many traditional and GAN-based data augmentation techniques [5–8] have been proposed to overcome this problem. Although traditional data augmentation solutions (e.g., crop, rotation, and translations) have achieved some successes in the analysis of nature images, it is unreasonable to be used in the medical image analysis tasks, because the images that are synthesized by traditional methods cannot contain related clinical manifestation, which thus makes it unable to achieve satisfactory improvements [9]. Differently, GAN-based data augmentation generates samples with richer and more reasonable information (i.e., contain rational clinical manifestations) to further improve the training process of medical image analysis. GAN-based data augmentation methods in medical imaging applications are divided into single-domain generation and cross-domain generation. In the area of single-domain generation, Frid et al. [10] presented a strategy based on Generative Adversarial Networks (GANs) for Data Augmentation (DA) in classification tasks, utilizing Deep Convolutional Generative Adversarial Networks (DCGAN) [6] to synthesize lung lesion Computed Tomography (CT) images, thereby enhancing the size of the training dataset. In terms of cross-domain generation, which is primarily geared towards modal transitions, a notable approach is proposed by Meng et al. [11]. They introduced a unified multi-modal model that relies on a conditional score-based generative model (SGM) for stochastic sampling from a target probability distribution. In addition, UAGGAN [12,13] and AGGAN [14,15] are two state-of-the-art cross-domain attention-guided image generation models, which are based on CycleGAN and both add an attention module to the model. Although UAGGAN and AGGAN have a certain cross-domain learning ability because of the application of attention mechanism, these cross-domain methods have the following shortcomings in the application of medical imaging tasks: (i) **lack of inter-class difference information problem**: these cross-domain models perform transformations on the same features (e.g. modality transformations) with good generative performance. However, the current SOTA model is weak in learning different information (e.g. with/without lesion), which cannot obtain satisfactory generation results for different samples. Especially, weakening the distinction between different categories of synthetic images will result in unacceptably high false-positive rates in the subsequent medical image analysis tasks. (ii) **inapplicable problem in the segmentation tasks**: In classification tasks applied in tumor diagnosis, generative data augmentation usually combines the generated and original images as a training set, and the class annotations of the synthetic images are known in the generation process. However, generative augmentation cannot provide the ground truth of synthetic images in segmentation tasks applied in tumor treatment. This limits the application of generative data augmentation to image analysis tasks.

Therefore, we first propose an advanced GAN-based generative data augmentation model, **Cross-Domain Attention-guided GAN model (CDA-GAN)**, to resolve the lack of inter-class difference information problem. Then, we further propose an application strategy of using this data augmentation model in the medical image segmentation task, which thus resolves the inapplicable problem in the segmentation task, and achieves more accurate medical image segmentation with limited data. Generally, CDA-GAN is mainly based on CycleGAN with three improvements, the (average-max squeeze-and-excitation) AMSE block, semi-supervised attention-guided generator, and attention-guided discriminator with spectral normalization, which thus can overcome the lack of inter-class difference information problem using CycleGAN-based domain conversion mechanism. Specifically, we first propose to add an AMSE block to CycleGAN, which performs channel-wise feature recalibration to improve the representational ability of the network. Then, we further propose to integrate semi-supervised spatial attention into the generator of CycleGAN to force the generator to focus on the ‘regions of interest’. By adding the attention mechanism and optimization of the adversarial loss, the attention-guided generator can

produce the region of discriminator ‘interest’, and by adding the pixel-wise loss of generated attention map with the region of interest to the training procedure of the attention-guided generator, the generator can produce more precise spatial attention maps, thus enhance the power of spatial attention mechanism. Finally, the attention-guided discriminator is incorporated into CycleGAN to force the discriminator to pay more attention to the changing part and the spectral normalization can prevent the discriminator from mode collapse and stabilize the training procedure.

Furthermore, we propose an application strategy for applying CDA-GAN models to medical image segmentation tasks. Specifically, for classification tasks, the application of CDA-GAN is straightforward: we can use generated images to expand or balance the training dataset. However, when we apply CDA-GAN-generated images to segmentation tasks, we encounter a problem: the generated images lack supervision information, so we need to discover the supervision information for the generated images. Because our synthetic tumor image is generated by adding the tumor to the original real tumor-free image, the straightforward solution is to directly subtract the synthetic tumor image and the original tumor-free image, and then obtain the result (i.e., difference image) as pseudo labels. However, the experiment has proved that this application method is very ineffective because the different image is not clean enough, i.e., it not only has tumor information but also has a lot of other brain tissue information (The generated image is not exactly the same as the original image except for the tumor area). This is the inapplicable problem in the segmentation tasks we mentioned earlier. To resolve this problem, we propose a new application strategy of using the synthesized images and the different images to enhance the feature learning capability of the downstream segmentation model.

Instead of using the different images as the segmentation pseudo-masks, we propose a way of information enhancement in response to its need for accurate pixel information. Specifically, this way is divided into two stages for using the cross-domain information brought by cross-domain. In the first stage, an enhanced map is obtained for the focal part by extracting the difference between output and input. In the second stage, the enhanced map concatenating with raw images as input for pixel-level subsequent tasks weakens the subsequent task’s difficulty. We believe that the information enhancement strategy will enhance the utilization of the generated image, and help the model achieve better segmentation performances. This information enhancement method that takes the difference as part of the input data can not only reduce the influence of other tissue information (i.e., noise information) in the difference image on the model feature learning but also effectively use the information of the lesion contained in it to improve the feature learning ability of the segmentation model. Therefore, with the help of the proposed new application strategy, CDA-GAN can not only be used in medical image classification (mainly for diagnosis) but also medical image segmentation (mainly for treatment), allowing the new generative data augmentation solution to be used to solve the limited data problem in the computer-aided medical diagnosis and treatment in clinical practices.

The contributions of this paper are summarized as follows:

- We first identify the limited data problem in the medical image diagnosis and treatment tasks. To resolve the problem, different from traditional data augmentations, we propose a novel generative data augmentation solution, which can generate images with meaningful clinical manifestation.
- Then we propose a Cross-Domain Attention-Guided GAN (CDA-GAN) model, which is an advanced CycleGAN that utilizes a new AMSE block, a semi-supervised spatial attention module, and the spectral normalization to overcome the lack of inter-domain difference information problem and generate more clinically rational and robust medical images than the existing single-domain and cross-domain generative methods.
- Finally, we identify the inapplicable problem in the segmentation task of the GAN-based generative data augmentation models. To effectively enhance the feature learning capability of the down-

stream segmentation model, we further propose a new application strategy to use the synthesized images and the different images.

- Extensive experimental studies are conducted on two public brain MRI brain tumor datasets, and the results show the following: (i) Using the proposed CDA-GAN for generative data augmentation is much more effective than data augmentation using the traditional DA method and the state-of-the-art GAN-based generative baselines in enhancing the performances of downstream classification and segmentation tasks. (ii) Ablation studies show that the proposed three advancements are all effective and essential for CDA-GAN to achieve superior generative DA performances in both tasks. (iii) We compare the effect of using the proposed application strategy and directly using different images as the segmentation pseudo-masks on the performances of the downstream segmentation tasks; the results show that, due to the noise information in different images, using different images as pseudo-masks will not bring any improvements but reduce the models' segmentation performances, which thus proves the existence of the inapplicable problem in the segmentation task and the effectiveness of the proposed application strategy.

The rest of this paper is organized as shown below. In Section 2, we review the related work. In Section 3, we present the proposed model in detail. In Section 4, we describe the dataset and experimental settings. In Section 5, we show the results in detail. Then, we conclude our work and introduce our feature works in Section 6 and Section 7.

2. Related work

Medical Image Generation for Data Augmentation. In recent years, data augmentation on medical images has received much more attention, and two methods are typically used: (i) affine transformations on given images [16], (ii) generation of new samples from given images [5]. For the first method, although affine translations are easy to implement and quickly increase the number of the training samples, the transformed images are quite similar and have limited contribution over the external test data [17]. Besides, operations such as shearing and translation may distort some important information of the original images, such as the shape features [7,18]. For the second method, inspired by the development of GANs, [7] introduces GAN to generate pneumonia CT images and feed them into the classification model for better learning of lesion features. Then, to generate high-resolution images, Han et al. [5] introduced CPGGAN, an extension of Progressive Growing GAN (PGGAN) [19,20]. This advancement incorporates a conditional module capable of generating medical images with lesions at targeted locations. Similarly, to enhance the quality, detail, and diversity of generated images, Multi-Scale Gradient GAN (MSG-GAN) [21] employs multi-scale features and information transfer mechanisms. However, these studies focus on single-domain transformation generative models. The single-domain limitation constrains the model's ability to acquire diversified information related to the input, thereby reducing its robustness and capacity to learn deep features. This underlines the necessity for further research into multi-domain transformation models to address these limitations. There exists some researches [11,22,23] in medical images applying CycleGAN [24], which learns the mappings between two unpaired image domains in different modalities. Meng et al. [11] propose the unified multi-modal conditional score-based generative model to take advantage of the score-based generative model (SGM) in modeling and stochastically sampling a target probability distribution, and further extend SGM to cross-modal conditional synthesis for various missing-modality configurations in a unified framework. Besides, many works [22,23] related to medical image modality translation demonstrate that adversarial loss with additional cycle-consistence loss proposed by CycleGAN can produce medical images with rich details. Although these studies have addressed the conversion between medical imaging modalities, the key

problem of data limitation due to category imbalance remains. Compared to traditional single-domain generation methods, the CycleGAN solution based on cross-domain image generation performs well on the modality conversion task because CycleGAN assigns a consistent focus to each pixel of the input image. But as mentioned earlier, it has not yet solved the problem of data imbalance and is unable to focus on lesion areas in medical images. Therefore, we add attention mechanisms to the CycleGAN model to better focus on areas of interest and increase the amount of positive and negative samples.

Attention-Guided Image Generation Solution. The current research on GAN-based attention mechanisms is developed in two directions. The first direction is to use extra data to provide attention. For example, Roy et al. propose a semantics-aware translation model [25], which uses the object mask annotations from each dataset as extra input data. Sun et al. [26] generate a facial mask by using FCN for face attribute manipulation. Moreover, Lai et al. [27] propose CWT-GAN, which can generate diverse and higher-quality images with the aid of the weight transfer mechanism, since features learned by discriminator tend to be more expressive than those learned by generator trained via maximum likelihood. The second direction is to train another segmentation or attention model to generate attention maps and fit them into the system. For example, Liu et al. [28] propose an implicit style function (ISF) to straightforwardly achieve multi-modal and multi-domain image-to-image translation from pre-trained unconditional generators. Kearney et al. [29] suggest an attention-aware, cycle-consistent generative adversarial network (A-CycleGAN) enhanced with variational autoencoding as a superior alternative to current state-of-the-art MR-to-CT image translation methods. Yang et al. [30] propose to add an attention module to predict an attention map to guide the image translation process. Xu et al. [31] propose SAGAN for the image generation task. However, all these methods are suitable for single-domain transformation-generative attention mechanisms that do not acquire stable and desired attention characteristics in cross-domain transformation. Unsupervised Attention-Guided Image to Image Generation (UAGGAN) applied in [12,13], and Attention-Guided Generative Adversarial Networks (AGGAN) applied in [14,15] are two state-of-the-art cross-domain attention-guided image generation models, which are also based on CycleGAN and are the most similar methods to our work. They both add an attention module to the model, output the attention map, and later multiply the output of the transformation function with the attention map pixel by pixel. They differ in that the attention module and the conversion function of UAGGAN are two separate and independent modules, while the attention module and the conversion function of AGGAN share the encoder part. But they all have limits. Firstly, the problem of AGGAN and UAGGAN is that in the discriminator part, both models focus on the learning of the attention part and lack the learning of the overall information of the image. This results in the model being less resilient to the tissue parts of the image of interest. Additionally, both models only improve the ability of CycleGAN to adjust image details from the perspective of spatial attention. Therefore, we propose the CDA-GAN model, which uses different attention mechanisms from AGGAN and UAGGAN. Specifically, to capture channel attention and improve the ability of the model to express features, we add an AMSE block. To address the issue of the inability of the above methods (UAGGAN, AGGAN) to generate more accurate attention maps, a semi-supervised spatial attention module is used to guide the training process of the attention module to obtain a more precise attention map. Furthermore, we added spectral normalization to the discriminator. It can prevent the discriminator from mode collapse and stabilize the training procedure due to over-focusing on the changing part.

In summary, to enhance the generative capability of CycleGAN, the stability of the model when performing the generative task, and the attention of the model to the region of interest, we propose a Cross-Domain Attention-Guided CycleGAN (CDA-GAN), which is an improvement on our previous work [31] on the attention mechanism.

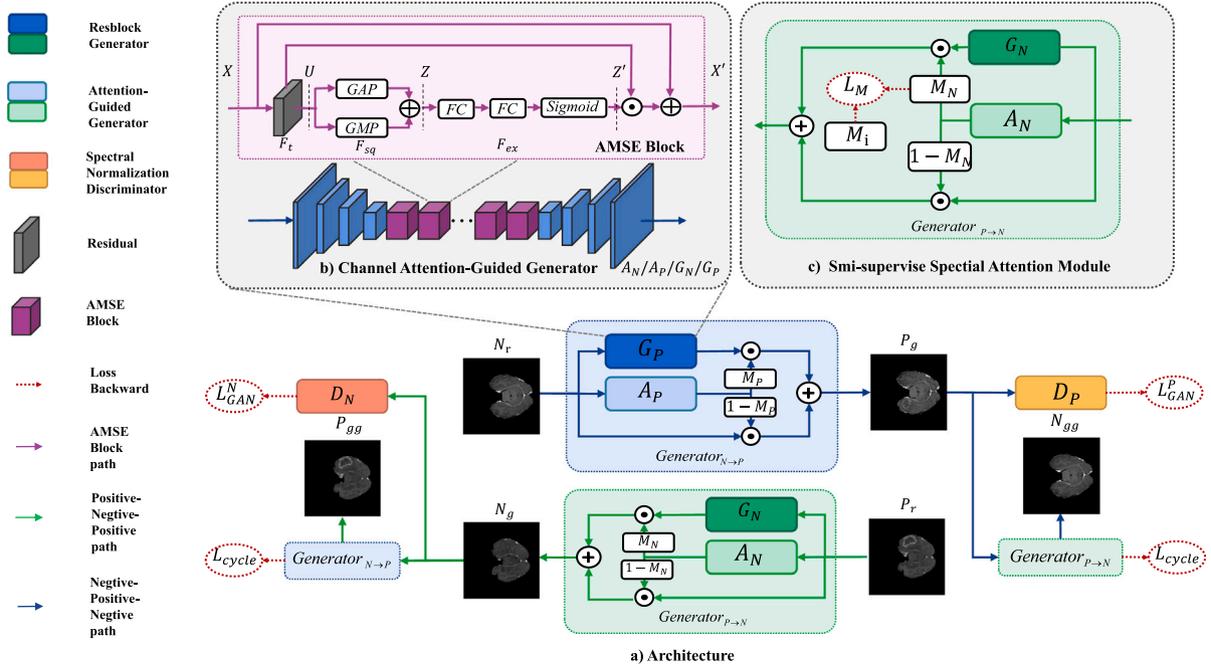


Fig. 1. Illustration of our proposed Cross-Domain Attention-Guided GAN (CDA-GAN). The symbols with a subscript P denote abnormal images that contain target objects originating from the positive sample domain, while symbols with a subscript N represent normal images from the negative sample domain. Here, GMP stands for Global Max Pooling and GAP represents Global Average Pooling. G_P / G_N : The Resblock generators within the model. A_P / A_N : The spatial attention modules. D_P / D_N : The attention-guided discriminators. N_r : The real normal images. N_g : The generated normal images. N_{gg} : The reconstructed normal images used for calculating cycle-consistency loss. P_r : The real abnormal images with the target object. P_g : The generated abnormal images with the target object. P_{gg} : The reconstructed abnormal images with the target object, which are used to calculate cycle-consistency loss. The term F_t denotes the convolutional component of the original residual module, which incorporates non-linear transformations into the model. These transformations can be optimized through the backpropagation algorithm.

The results of classification and segmentation experiments, which compare the two methods of two SOTA attention-guided image generation solutions: UAGGAN [12,13], and AGGAN [14,15] with CDA-GAN, show that our model generates data augmentation images can improve the accuracy of medical image analysis.

3. Methodology

The proposal of the Cross-Domain Attention-Guided GAN (CDA-GAN) for generative medical image data augmentation is driven by several key observations. Primarily, we identified that single-domain generation methods frequently encounter difficulties in generating high-resolution medical images that are interpretable by radiologists [5,32]. The generation of high-resolution images necessitates larger models equipped with a greater number of parameters. However, the restricted size of available datasets poses significant challenges to training these models effectively, often leading to convergence difficulties. Concurrently, we observed the extensive utilization of modality conversion in data augmentation methodologies within the medical domain [30,33]. This approach facilitates the stable generation of high-resolution imaging data, even when constrained by limited dataset sizes. Given these considerations, we devised a novel strategy: treating positive samples (those with lesions) and negative samples (those without lesions) as two distinct domains for transfer generation. This approach enables the production of meaningful and high-resolution medical images, which can significantly enhance the effectiveness of downstream tasks. Our method, therefore, aims to address the challenges encountered by previous single-domain generation methods, promising superior performance in the generation and augmentation of medical imaging data.

The scheme is mainly divided into two parts: in the first part, we use a CycleGAN-based cross-domain generative network integrating attention mechanism, called CDA-GAN in Fig. 1, to generate positive samples (with lesions) and negative samples (without lesions). Moreover, the two domains involved in the term cross-domain are the positive sample

domain and the negative sample domain, which are labeled P and N in the following text respectively. Logical implementation of the network is described in Algorithm 1. And the workflows of downstream classification and segmentation tasks are shown in Algorithm 2 and Algorithm 3, respectively. In the second part, we propose an information enhancement strategy for the segmentation task in Fig. 3. Sections 3.1 and 3.2 describe the technical details of the overall scheme.

3.1. Cross domain attention-guided GAN (CDA-GAN) model

Our goal is to construct a cross-domain generative model that implements data augmentation for medical image analysis tasks, such as the expansion of positive and negative sample data for disease diagnosis. Denote $X = \{N, P\}_i$ as a dataset consisting of i patients, where N represents negative samples and P is positive samples. However, we found the generative model lacks learning of domain commonality and characteristics in generating images, which leads to limited data augmentation in tumor classification tasks. To solve this problem, we consider CycleGAN as a backbone network to compare and learn the correlations and differences between the two domains. We define $G_N: P \rightarrow N$ as the mapping function from positive to negative samples, and $G_P: N \rightarrow P$ as the mapping function from negative to positive samples. Then, the goal of the system becomes $\mathbb{P}(N, P) \sim \mathbb{P}'(N, P, G_N(N, P), G_P(N, P))$, where $\mathbb{P}(N, P)$ stands for a medical image analysis system and $\mathbb{P}'(N, P, G_N(N, P), G_P(N, P))$ stands for a medical image analysis system with generative data augmentation.

However, while CycleGAN's cross-domain generative model is capable of effectively transforming modalities, it falls short in generating and eliminating the content of lesions. We propose a CycleGAN-based cross-domain attention generation model, called CDA-GAN, to build a medical image analysis system with data augmentation.

As illustrated in Fig. 1, this model includes three major components: the average-max squeeze-and-excitation (AMSE) block, semi-supervised attention-guided generator, and attention-guided discriminator with

spectral normalization, which thus can overcome the lack of inter-class difference information problem using CycleGAN-based domain conversion mechanism. Our improvements can encourage models to generate specific regions of brain images in cross-domain samples, leading to improvements in medical image analysis in small and imbalanced datasets.

3.1.1. Channel attention module

Specific brain disease is often highly relevant to particular regions [34,35]. However, in CycleGAN, the ability of the residual convolution generator module and the discriminator is weak to capture hierarchical patterns. Inspired by the SENet [36], a type of convolutional neural network that introduces a mechanism to allow the network to perform dynamic channel-wise feature recalibration. To achieve this end, the model adaptively adjusts the importance of each channel based on the information aggregated globally across the spatial dimensions of the input feature map. Therefore, the information aggregated globally across the spatial dimensions plays a crucial role in the subsequent re-weighting of feature importance in the channel dimension. However, SENet's use of either max pooling or mean pooling alone to handle spatial information is not optimal. As a result, we propose the Average-Max Squeeze-and-Excitation (AMSE) module to better handle the information aggregated globally across the spatial dimensions thus improving the quality of the channel-wise feature recalibration, which allows the model to accurately capture the differences between the two domains and correctly translate them. There is another work that proposed a SE block with both average pooling and max pooling, called CBAM block [37]. The differences between our proposed AMSE-Residual block and the CBAM block lie in the squeeze part and attention mechanism. In the squeeze part, for the proposed AMSE block, the features from max-pooling and mean-pooling are added together before being passed through a network consisting of two fully connected (FC) layers with a ReLU activation function in between, while in the channel attention part of CBAM block, the features passed through a shared network then added together to re-denoting the weights of the channels. In multi-task models, a shared network is typically employed to enhance feature extraction capabilities, thereby improving the performance across multiple tasks. However, models using a shared network structure across multiple inputs generally aim to achieve approximate outcomes while reducing the number of model parameters. By placing the addition operation before the fully connected (FC) layer, we can better integrate the features from max-pooling and mean-pooling through the backpropagation algorithm. This approach allows the model to effectively combine the features of both pooling methods, rather than simply finding a compromise in optimization direction between mean-pooling and max-pooling feature input. As for the attention mechanism, the CBAM block integrates both channel attention and spatial attention within a single module. In contrast, we separate these two types of attention. We believe that such a separation can better enhance the model's ability to learn representations and improve the flexibility of model training since we can use different labels to guide the learning process of the model's attention mechanism.

The detailed architecture of this module is presented in Fig. 1, the proposed AMSE-Residual module consists of two blocks: the transformation functions F_i , and the AMSE block. F_i represents the convolution part of the original residual module. This introduces non-linear transformations to the model, which can be optimized through the backpropagation algorithm. F_i takes the image features $X \in R^{C \times H \times W}$ captured in the last AMSE-Residual block as input, and outputs transformed image features $U \in R^{C \times H \times W}$. At the squeeze stage, a statistic $Z \in R^{C \times 1 \times 1}$ is generated by ignoring U 's spatial dimensions $H \times W$. The p_{th} element of Z is calculated by:

$$Z_p = F_{sq}(U) = \frac{1}{H \times W} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} U(p, i, j) + \max U(p, i, j) \quad (1)$$

To learn the nonlinear interactions between channels and maintain nonreciprocal relations, a gating mechanism consisting of two fully connected (FC) layers and a sigmoid layer is used in the excitation phase. Denoting the two FC layers and the final sigmoid layer cascaded in Fig. 1 by F_{ex} , given the output of the squeeze stage Z , the output activates $Z' \in R^{C \times 1 \times 1}$ of the excitation stage is:

$$Z' = F_{ex}(Z) \quad (2)$$

The final outputs $X' \in R^{C \times H \times W}$ of this AMSE-Residual module is the reassignment of the value on the channel of U plus the origin input X , $F_{rescale}$ denoting the function to reassignment the channel of U , it is known that the shape of U is C, H, W and the shape of Z' is C , so the reassignment of the channel of U is calculated as the value of each item of U on the same channel multiplied by the value of the corresponding layer of Z' , which is calculated as:

$$X' = F_{rescale}(Z', U) + X = Z'U + X \quad (3)$$

This module allows the network to automatically determine the importance of each feature channel and adjust its weights by the proposed Average-Max pooling-based squeeze operation and the following FC layer-based excitation operation. This dynamic feature recalibration enhances the representational power of the model, improving its understanding of the input data.

3.1.2. Semi-supervised spatial attention module

In this paper, the generation task focuses on the perfect restoration of tumor regions and generating positive samples with clear structural details and matching tumor image features. Therefore, we consider that the attention mechanism should be dynamically distributed in each image region to enhance the capability of the CycleGAN in the proposed medical image generation task. The above purpose is achieved in two steps: (i) locating the area for manipulation and (ii) taking the proper translation in the located area. Both unsupervised attention-guided image-to-image generative adversarial networks (UAGGAN) [12,13] and attention-guided generative adversarial networks (AGGAN) [14, 15] are performing image generation following the above approach. However, both UAGGAN and AGGAN have some issues. The first is the insufficient generative ability of the models. Although they can generally eliminate the original tumor lesions and restore negative images, even UAGGAN, which is the better performer in both UAGGAN and AGGAN, still has the problem of incomplete tumor elimination. This problem is due to the weak feature representation capability of the model, which leads to incomplete tumor elimination in the localized region. This problem can be solved by adding the channel attention method introduced in the previous section. Secondly, the model cannot generate attention maps accurately, resulting in incomplete localization of the operation area. Therefore, we propose a semi-supervised attention module to use a tumor mask to co-guide the training procedure of the spatial attention module, a semi-supervised strategy that is different from the traditional method [38]. Especially, the loss function in the training procedure of the spatial attention module comes from two parts: (1) adversarial Loss from the discriminator: The part represents an unsupervised loss as it does not require labeled data. It tries to make the distribution of the model's outputs match the distribution of the training data and (2) L2 Loss computed with tumor masks: The part constitutes a supervised loss as it relies on the labeled tumor mask for the module to learn from. The combination of these supervised and unsupervised losses is what gives our module its "semi-supervised". We use mask information to supervise the generation process of positive samples to negative samples, and the generation process of negative samples to positive samples without mask supervision information. In addition, attention-guided discriminators lead to the difficulty of the model to learn the tissue boundary information of the negative image, resulting in the difficulty of the model to recover the tissue boundaries invaded by the tumor when it generates negative samples that do not contain tumor lesions. Therefore, the spatial attention module and

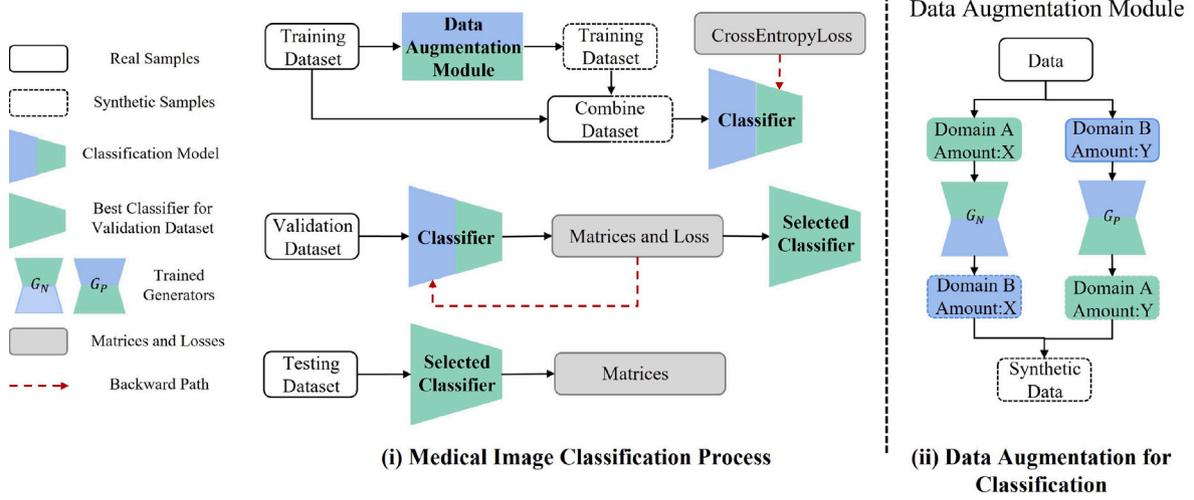


Fig. 2. The flowchart of data augmentation in classification.

the generated attention map do not guide the training procedure of discriminators.

Specifically, as shown in Fig. 1, in the forward processing, the generated image is a combination of two parts: the foreground from the generator and the background from the input image. Take the translation from positive samples to negative samples as an example. Firstly, the positive image $\{p_i\} \in P$ is fed into the generator G_N , which maps $\{p_i\}$ to the target domain N , generating the negative image $n'_i = G_N(p_i)$. Secondly, the same input $\{p_i\}$ is fed into the attention module A_N , resulting in the attention map $M_i^p = A_N(p_i)$. Thirdly, to create the 'foreground' object $\{n_i^f\} \in n'_i$, we apply M_i^p to n'_i via an element-wise product: $\{n_i^f\} = M_i^p \odot n'_i$. Finally, the inverse of attention map $M_i^{p'} = 1 - M_i^p$ will be applied to the input image via an element-wise product as the background. Thus, the mapped image $\{n_i^o\}$ is obtained by:

$$\begin{aligned} n_i^o &= G_N^A(p_i) = \underbrace{M_i^p \odot n'_i}_{\text{Foreground}} + \underbrace{M_i^{p'} \odot p_i}_{\text{Background}} \\ &= A_N(p_i) \odot G_N(p_i) + (1 - A_N(p_i)) \odot p_i \end{aligned} \quad (4)$$

We only described the mapping function G_N^A ; the inverse mapping function G_P^A is defined similarly:

$$\begin{aligned} p_i^o &= G_P^A(n_i) = \underbrace{M_i^n \odot p'_i}_{\text{Foreground}} + \underbrace{M_i^{n'} \odot n_i}_{\text{Background}} \\ &= A_P(n_i) \odot G_P(n_i) + (1 - A_P(n_i)) \odot n_i \end{aligned} \quad (5)$$

The attention map of $P \rightarrow N$ translation is exactly the whole lesion region. Therefore, we supervised the training process of attention network G_N^A by segmentation mask. Given a training set $\{(p_1, M_1), \dots, (p_k, M_k)\}$ of k examples, where M_i is the tumor mask of segmentation and $\{p_i\} \in P$ is the given positive image. To reduce changes and constrain generators, we design pixel loss between the tumor mask M_i and the generated attention map M_i^p . We express this loss as:

$$\mathcal{L}_M(M_i, M_i^p) = \|M_i - M_i^p\|_1 \quad (6)$$

3.1.3. Spectral normalization

It is well known that the difficulty of training GANs is caused by the fact that the objective function of vanilla GAN is equivalent to optimizing the J-S divergence between the distribution p_g of the generated data and the distribution p_r of the real data. Then, WGAN [39] is proposed to solve the problem of the J-S divergence in the vanilla GAN with Wasserstein distance. Specifically, the Wasserstein distance of WGAN can continuously represent the distance of two distributions

without intersection and interruption, so it can eliminate the convergence problem in the training process of the original generative adversarial network and make the training process stable. However, the implementation of Wasserstein distance is conditional: the parameter matrix of the discriminator needs to satisfy the 1-Lipschitz continuity, i.e., it achieves the 1-Lipschitz continuity in the whole definition domain of the function.

To address this problem, Lin et al. [40] propose the spectral normalization that achieves the 1-Lipschitz continuity without destroying the matrix structure. Specifically, they implement the 1-Lipschitz continuity constraint by dividing the network parameters of each layer of the network by the spectral norm of that layer's parameter matrix. The spectral normalization method makes the discriminator achieve the 1-Lipschitz continuity in a way without destroying the proportionality between the parameters and allowing the model to express the distribution distance between the generated and real images in Wasserstein distance. We apply spectral normalization to the image-to-image generation, which brings a more elegant way to prevent the discriminator from mode collapse and stabilize the training procedure.

3.1.4. Objective functions

I. Attention-guided Adversarial Loss. The attention-guided adversarial loss is proposed to train the attention-guided generators and discriminators. We update the adversarial loss:

$$\begin{aligned} \mathcal{L}_{GAN}^P(G_P^A, D_P, P, N) &= \mathbb{E}_{p \sim p_{\text{data}}(P)} [\log D_P(M^n, p)] \\ &\quad + \mathbb{E}_{n \sim p_{\text{data}}(N)} [\log(1 - D_P(M^p, G_P^A(n)))] \end{aligned} \quad (7)$$

where G_P^A aims to translate the negative image to the positive image and maximize the probability that the discriminator makes a mistake, while D_P is trained to distinguish between the generated positive images and its corresponding attention mask in generation produce (M^p, p^o) with real positive images and its attention mask (M^n, p) . This means G_P^A tries to minimize the attention-guided adversarial loss $\mathcal{L}_{GAN}^P(G_P^A, D_P, P, N)$, while D_P tries to maximize it. There are also another loss for the discriminator D_N and the generator G_N^A update:

$$\begin{aligned} \mathcal{L}_{GAN}^N(G_N^A, D_N, N, P) &= \mathbb{E}_{n \sim p_{\text{data}}(N)} [\log D_N(M^p, n)] \\ &\quad + \mathbb{E}_{p \sim p_{\text{data}}(P)} [\log(1 - D_N(M^n, G_N^A(p)))] \end{aligned} \quad (8)$$

II. Pixel-level Attention Loss. The pixel-level attention loss is used to reduce the difference between the region of interest M_i and the generated attention map M_i^p . The above-mentioned adversarial loss and pixel-wise loss jointly optimize the spatial attention mechanism. Pixel-wise loss narrows the convergence space of the attention network, forcing the attention map to converge to the region of interest, while the adversarial loss brings uncertainty to the attention map generated

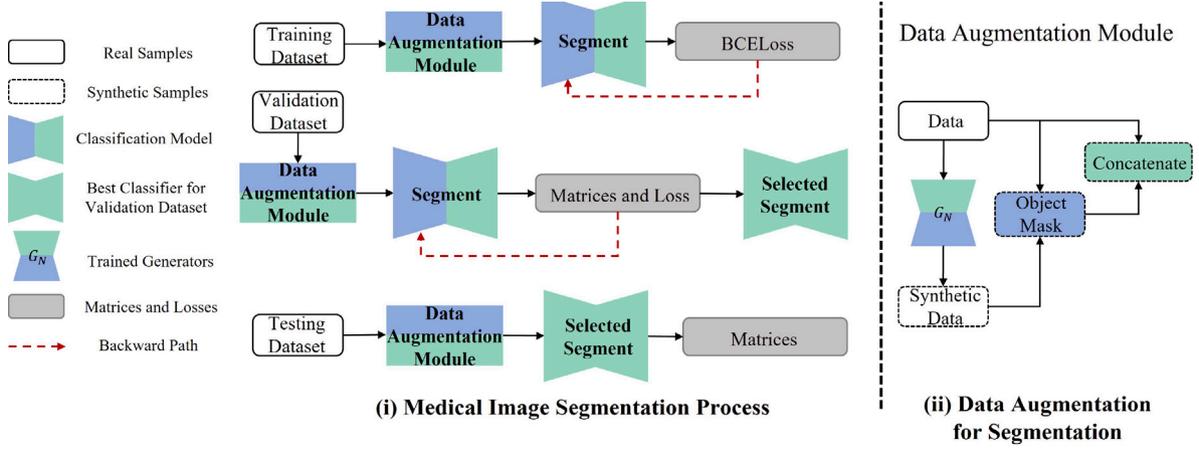


Fig. 3. The flowchart of data augmentation in segmentation.

by the model so that the attention map converges to the part that the discriminator cares about, preventing overfitting problems that may be caused by pixel-wise loss.

Thus, the loss function of Pixel-level Attention Loss is defined as:

$$\mathcal{L}_M(M_i, M_i^p) = \|M_i - M_i^p\|_1 \quad (9)$$

III. Cycle-consistency Loss. The cycle-consistency loss is used to enforce forward and backward consistency. Thus, the loss function of cycle-consistency is defined as:

$$\mathcal{L}_{cycle}(G_P^A, G_N^A, N, P) = \mathbb{E}_{n \sim p_{data}(N)} \left[\|G_N^A(G_P^A(n)) - n\|_1 \right] + \mathbb{E}_{p \sim p_{data}(P)} \left[\|G_P^A(G_N^A(p)) - p\|_1 \right] \quad (10)$$

In Algorithm 1, we introduce a data augmentation method named Cross-Domain Attention-Guided GAN (CDA-GAN). This method is specifically designed for image translation between a source domain dataset X and a target domain dataset Y , thus effectively executing data augmentation. The algorithm primarily consists of two key components: the Attention-based Generator (G_A) and the Discriminator (D). During each training iteration, we randomly sample mini-batches from the source domain dataset X and the target domain dataset Y . Subsequently, we perform the generator update step, mapping the source domain data X to the target domain Y , and compute a cycle consistency loss to ensure image consistency. We also introduce a pixel-level attention loss to enhance the generator's performance. Simultaneously, we update the discriminator by computing the discriminator loss and training it to differentiate between real target domain data and data generated by the generator. Weight parameters in the loss function, denoted as λ_{cycle} , and λ_M , are used to control the importance of different loss terms. By iteratively executing the generator and discriminator update steps, we eventually achieve the training of the generator G_A and the discriminator D , enabling them to map source domain images to the target domain and generate high-quality results.

This algorithm finds widespread applications in domain adaptation and image translation tasks, effectively addressing image translation challenges across different domains.

3.2. Information enhancement strategy for medical image analyze

For medical imaging tumor diagnosis task flow is shown in Fig. 2. In the data augmentation stage, all positive samples containing tumor lesions used for training are input to the generator G_N to generate negative samples, and all negative samples are input to the generator G_P to generate positive samples.

Firstly, in the stage of tumor classification, the combination of synthetic samples with real samples is used to train the classification model. Our proposed inter-class cross-domain translation model inputs positive images to the generator G_N and outputs the same number of

Algorithm 1 CDA-GAN Training

Description: Training the CDA-GAN model to map from source domain X to target domain Y while maintaining cycle consistency and pixel-level attention.

Input:

- X : Source domain dataset;
- Y : Target domain dataset;
- G_A : Generator mapping function from X to Y ;
- A : Attention-based mapping function from X to M ;
- M : Attention map of Generator A ;
- \hat{M} : Segmentation mask;
- Loss function weights λ_{cycle} , λ_M ;
- Number of training iterations N .

Output: Trained generator G_A and discriminator D .

- 1: **for** n in 1 to N **do**
- 2: Randomly sample minibatches x and y from source domain dataset X and target domain dataset Y .
- 3: **Generator Update:**
- 4: Compute generator's generated outputs $\hat{y} = G_A(x)$ and $\hat{x} = G_A(y)$.
- 5: Compute cycle consistency loss:
- 6: $\mathcal{L}_{cycle} = \|x - G_A(\hat{y})\|_1 + \|y - G_A(\hat{x})\|_1$.
- 7: Compute pixel-level attention loss (from positive samples to negative samples):
- 8: Compute attention-based generator's generated attention map $M = A(x)$ and $M' = A(y)$.
- 9: $\mathcal{L}_M = \|M - \hat{M}\|_1$.
- 10: Compute total generator loss:
- 11: $\mathcal{L}_{G_A} = \mathcal{L}_{cycle} + \lambda_M \mathcal{L}_M$.
- 12: Update generator parameters using Adam optimizer:
- 13: $\theta_{G_A} \leftarrow \theta_{G_A} - \text{Adam}(\nabla_{\theta_{G_A}} \mathcal{L}_{G_A})$.
- 14: **Discriminator Update:**
- 15: Compute discriminator loss:
- 16: $\mathcal{L}_D = -\log(D(y)) - \log(1 - D(\hat{y})) - \log(D(x)) - \log(1 - D(\hat{x}))$.
- 17: Update discriminator parameters using Adam optimizer:
- 18: $\theta_D \leftarrow \theta_D - \text{Adam}(\nabla_{\theta_D} \mathcal{L}_D)$.
- 19: **end for**

corresponding negative images. Meanwhile, inputting negative images to the generator G_P will output the same number of corresponding positive images. With an equal number of each class, CDA-GAN doubles the original training dataset for the classification task and balances the number of classes. Finally, we select the model parameters with optimal

Algorithm 2 Classification Task Data Augmentation Flow

Input: Positive samples with tumor lesions X_p , Negative samples X_n , Generator G_N , Generator G_P , Classification Model M

- 1: Initialize empty dataset X_{aug}
- 2: **for** $x \in X_p$ **do**
- 3: Generate corresponding negative sample $\hat{x} = G_N(x)$
- 4: Add $(x, 1)$ and $(\hat{x}, 0)$ to X_{aug}
- 5: **end for**
- 6: **for** $x \in X_n$ **do**
- 7: Generate corresponding positive sample $\hat{x} = G_P(x)$
- 8: Add $(x, 0)$ and $(\hat{x}, 1)$ to X_{aug}
- 9: **end for**
- 10: Train classification model M on X_{aug}

return Trained classification model M

Algorithm 3 Segmentation Task Information Enhancement Flow

Input: Positive samples with tumor lesions X_p , Negative samples X_n , Cross-Domain GAN-based Generator G_N , Segmentation Model S

- 1: Initialize empty dataset X_e
- 2: **for** $x \in X_p$ **do**
- 3: Generate corresponding negative sample $\hat{x} = G_N(x)$
- 4: Obtain coarse mask of tumor area $M_{coarse} = |x - \hat{x}|$
- 5: Add (x, M_{coarse}) and annotated tumor mask to X_e
- 6: **end for**
- 7: **for** $x \in X_n$ **do**
- 8: Create a zero vector M_{zero} of the same size as x
- 9: Add (x, M_{zero}) and annotated tumor mask to X_e
- 10: **end for**

return Train segmentation model S on X_e

evaluation metrics by the validation dataset and obtain the final classification results in the test dataset. The specific process is described in Algorithm 2. Specifically, in this data augmentation pipeline for classification tasks, we use a two-step process: data generation and model tasks. First, in the data generation stage, we use a generator named G_N to convert positive samples (samples containing tumor lesions) into corresponding negative samples. At the same time, we use another generator G_P to convert negative samples into corresponding positive samples. Next, we create an empty dataset X_{aug} to store the generated sample pairs. For each positive sample x , we generate a corresponding negative sample $\hat{x} = G_N(x)$, and combine $(x, 1)$ and $(\hat{x}, 0)$ is added to X_{aug} , where 1 represents a positive sample and 0 represents a negative sample. For each negative sample x , we generate a corresponding positive sample $\hat{x} = G_P(x)$, and combine $(x, 0)$ and $(\hat{x}, 1)$ is added to X_{aug} . Finally, we use the data on X_{aug} to train the classification model M . This model will utilize the generated samples to improve its performance. After training, we get the trained classification model M . This algorithmic process makes full use of data augmentation methods to improve the performance of classification tasks by generating sample pairs related to the original data.

Secondly, in the process of tumor treatment, precise segmentation of lesions is required, and information enhancement of the segmentation task is shown in Fig. 3. At the training stage, the segmentation model can be divided into two parts. The first stage is to train the cross-domain GANs-based generation model. The second stage is to train the segmentation model. Initially, we input all positive samples into the positive-to-negative generator from the trained cross-domain GANs-based generation model to generate negative samples. By subtracting the generated negative samples from the input ones at the pixel level, we obtain a coarse mask of the tumor area. We then input both the generated masks and the original slices into the segmentation model and use the human-annotated tumor mask as the label. We believe this

strategy can reduce the difficulty of the segmentation task, thus yielding better segmentation results. The training process of the information enhancement strategy applied in the segmentation task is shown in the Algorithm 3. In this information augmentation pipeline, we aim to improve the performance of image segmentation models, especially in the task of tumor segmentation in medical images. We have a set of positive sample images X_p and negative sample images X_n with tumor lesions, a cross-domain generator named G_N , and a Segmentation model S . First, we create an empty dataset X_e , which will be used to store information-enhanced samples. Then, we iterate over each image x in the positive sample set X_p . For each positive sample x , we use the generator G_N to generate a corresponding negative sample \hat{x} . Next, we calculate a coarse mask M_{coarse} of the tumor area, where $M_{coarse} = |x - \hat{x}|$. Then, we concat x and M_{coarse} on the channel and add them to the information augmentation dataset X_e together with the annotated tumor mask. Next, we iterate over each image x in the negative sample set X_n . For each negative sample x , we create a zero vector M_{zero} with the same size as x . Then, we concat x and M_{zero} on the channel and add them to the information augmentation dataset X_e together with the annotated tumor mask. Finally, we return the trained segmentation model S , which was trained using the information-augmented dataset X_e . In the testing stage, there is no category label for the testing dataset. Luckily, there is a loss function called identity loss in the Improved CycleGAN model. The loss is proposed to force the generator to learn the category information in the input image. If a negative image feeds into the positive-to-negative generator, which should feed the positive image, the generator will output an image that is similar to the input image. Therefore, in the testing stage, all the images will feed into the positive-to-negative generator. As the same, the trained segmentation model will take two inputs: the original image and the pixel-wise difference between the original image and the corresponding synthetic image, and output the segmentation results. For medical image analysis, the process of information enhancement is similar to the segmentation task, where the pixel-level difference is used as a rough detection map to train the segmentation model. In addition to the segmentation task, the classification task can also be used with the two-stage method as described above.

4. Datasets and experiments

4.1. Datasets and pre-processing

To evaluate the performances of our proposed CDA-GAN and information enhancement strategy in brain tumor classification and segmentation tasks, we conduct extensive experiments on two different magnetic resonance imaging (MRI) brain tumor datasets (BraTS 2020¹ and TCIA²). Please note that although only brain tumor MR images are used here, CDA-GAN is a generic method that can be adapted to other types of medical images, such as CT scans, ultrasound images, and gastrointestinal endoscopy images.

BraTS 2020, as a public real-world multimodal magnetic resonance imaging (MRI) dataset provided by Mehta et al. [41], aims to segment the glioma tumor. This dataset has two types of brain tumor data, namely high-grade glioblastoma (HGG) and low-grade glioma (LGG). The MRI of each sample contains four modalities: fluid attenuation inversion recovery (FLAIR), T1 weighting (T1), T1-weighted contrast-enhanced (T1-CE), and T2 weighting (T2). The ground truth masks are tagged by expert board-certified neuroradiologists. To further verify the model's effect with different sample magnitudes, we divided the public dataset into two different datasets ($BraTS$ and $BraTS_S$). The details of the division will be shown later in this subsection. The Cancer Imaging Archive (TCIA) contains brain MR images and manual anomaly

¹ <https://www.med.upenn.edu/cbica/brats-2020/>

² <https://www.kaggle.com/datasets/mateuszbudala/lgg-mri-segmentation>

Table 1
The information of three public datasets.

Datasets	Training set		Validation set		Testing set	
	Positive	Negative	Positive	Negative	Positive	Negative
BraTS	6645	11 913	1016	1729	1972	3418
BraTS _s	897	1448	132	215	1972	3418
TCIA	998	1888	94	230	281	438

Datasets are divided by patients rather than the number of slices.

segmentation masks of 110 patients. Among them, the brain MR images were low-grade glioma (LGG) with fluid-attenuated inversion recovery (FLAIR) sequence imaging. The patient's brain MR images are 2D slices ranging in number from 20 to 88. The statistical information about datasets is shown in Table 1.

BraTS: There are 322 cases in the Brats 2020. We chose HGG samples as our experimental dataset, which is 20% of all cases are assigned as the testing data, and 10% of all the cases are assigned as the validation data. Besides, for each patient sample, there are four 3D volumes with different modalities and ground truth of brain tumors corresponding to them. For data preprocessing, each 3D MRI volume with 240*240*155 sizes is sliced into 155 2D slices. Since each case has four modal MR images, four modal images for each slice are concatenated into four-channel data with 240*240 size. Due to the poor image quality of the first 30 slices and the last 30 slices, we choose the middle slices between 30 to 125.

BraTS_s: A subset of BraTS to verify the performance of the proposed data augmentation scheme with limited data. The amount of the BraTS_s dataset is one-eighth of the training and validation amount of the BraTS dataset. To fairly evaluate the performance of the model on different scales of the same dataset, our testing set is consistent with the BraTS dataset. For data preprocessing, BraTS_s and BraTS have the same data pre-processing method.

TCIA: The TCIA dataset contains 110 LGG cases, and the MRI images in the dataset are sliced 256*256 2D images. The number of slices per patient varies between 20 and 88. We choose 20% of all cases as the testing data and 10% of all the cases as the validation data. Besides, for each patient sample, images for each slice are one-channel data with 256*256 size.

4.2. Evaluation metrics

We demonstrate the effectiveness of cross-domain data augmentation solutions for medical image classification tasks. The classification performance is evaluated by two metrics, classification accuracy (ACC), Recall, F1 score (F1), and area under the curve (AUC). Then, we elaborate on the meaning and function of each metric and follow the equations in detail. The equation of ACC, Rec, and F1 are calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

$$Recall = \frac{TP}{TP + FN}, \quad (12)$$

$$F1 = \frac{2 * TP}{T + P}, \quad (13)$$

where TP , TN , FP , and FN are the number of true positive points, true negative points, false positive points, and false negative points, respectively. T is the number of ground truth points of that class, and P is the number of predicted positive points. Accuracy measures the number of correctly predicted samples (both true positives and true negatives) as a percentage of the total number of samples. The performance of the classification is evaluated with TP, TN, FP, and FN, describing the number of samples predicted correctly. Recall has evaluated the probability that positive samples are correctly classified as positive. F1 score is the harmonic mean of Precision and Recall, which thus can evaluate the model's performances more comprehensively from the perspectives of both Precision and Recall.

The equation of AUC is defined as:

$$AUC = \frac{\sum_{i \in \text{positive class}} \text{rank}_i - \frac{M \times (M+1)}{2}}{M \times N} \quad (14)$$

AUC is defined as the area under the ROC curve enclosed by the coordinate axes. To calculate AUC, we need to sort the output of the classification model from largest to smallest. Then, the rank of the largest output sample is n , the rank of the second largest output sample is $n-1$, and so on. Then we add the rankings of the entire positive samples and subtract the $M-1$ combination of the two positive samples. The final result is to output the number of positive class samples that is greater than the number of negative class samples and then divide by $M \times N$.

To evaluate the segmentation performances of our proposed cross-domain data augmentation solutions for medical image segmentation tasks, two widely used segmentation evaluation metrics, sensitivity (Sens), dice similarity coefficient (Dice), mean intersection over union (mIoU) and 95% Hausdorff Distance (HD95) are adopted. The formal definitions of Dice and Sens are as follows:

$$Sens = \frac{TP}{TP + FN} \quad (15)$$

Sensitivity, also known as Recall, is the proportion of positive pixels that are correctly segmented to all the pixels that are annotated as positive in the ground truths.

$$Dice = \frac{2 * TP}{2 * TP + FP + FN}. \quad (16)$$

The Dice Similarity Coefficient is a statistical measure utilized to evaluate the similarity between two samples. The terms True Positive (TP), False Positive (FP), and False Negative (FN) retain their standard meanings within this context. However, in the case of segmentation tasks, these metrics are computed in relation to the real and predicted mask regions.

$$mIoU = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i + FN_i} \quad (17)$$

Where n represents the number of categories, TP_i represents the number of pixels correctly classified as positive samples in the i_{th} category, FP_i represents the number of pixels misclassified as positive samples in the i_{th} category, and FN_i represents the misclassified pixels in the i_{th} category. It measures the ratio of the intersection area between the predicted segmentation mask and the ground truth mask to the union area of the two masks across all classes.

$$HD95 = \max_{k95\%} [d(P, G), d(G, P)], \quad (18)$$

$$d(P, G) = \sum_{p \in P} \min_{g \in G} d(p, g), \quad (19)$$

$$d(G, P) = \sum_{g \in G} \min_{p \in P} d(g, p) \quad (20)$$

Where G represents the real label, and P represents the segmentation result. Hausdorff Distance 95%, is a metric used to evaluate the boundary accuracy of a segmentation model. It measures the maximum distance between the predicted segmentation boundary and the ground truth boundary for which 95% of the distances are smaller or equal to the measured value.

4.3. Implementation details

All models are implemented using PyTorch and conducted over a server equipped with 8 Nvidia GeForce 2080 Ti GPUs. Each graphics card has 11019M of memory, and the server has an Intel(R) Xeon(R) Silver 4110 CPU with 2.10 GHz and 16G of RAM. In terms of software configuration, the CUDA version of the server is 10.2, and all the code is implemented in Python language, based on the PyTorch framework.

Table 2

Comparison with the state-of-the-art GAN-based data augmentation in the classification task, where *O-s* represents oversampling in traditional data enhancement, and *U-s* represents undersampling in traditional data enhancement.

Methods	BraTS				BraTS _s				TCIA			
	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1
w/o DA	0.9081	0.9158	0.8839	0.9052	0.8243	0.8339	0.7942	0.8489	0.9221	0.9131	0.8719	0.8974
O-s	0.9194	0.9282	0.8867	0.9134	0.8313	0.8389	0.8076	0.8561	0.9248	0.9192	0.8732	0.9029
U-s	0.9208	0.9213	0.8891	0.9175	0.8419	0.8451	0.8319	0.8674	0.9249	0.9160	0.8754	0.9011
Mixup [43]	0.9345	0.9369	0.8921	0.9194	0.8438	0.8476	0.8625	0.8769	0.9263	0.9197	0.8897	0.9042
CutMix [44]	0.9268	0.9402	0.8803	0.9256	0.8431	0.8321	0.8457	0.8652	0.9235	0.9245	0.8818	0.9047
CutOut [45]	0.9328	0.9329	0.8868	0.9137	0.8346	0.8309	0.8258	0.8547	0.9266	0.9054	0.9066	0.8989
PGGAN [19,20]	0.9336	0.9427	0.8909	0.9215	0.8440	0.8334	0.8771	0.8749	0.9249	0.9173	0.8826	0.9018
MSG-GAN [21]	0.9354	0.9468	0.8915	0.9266	0.8487	0.8546	0.8302	0.8721	0.9291	0.9201	0.8790	0.9064
CycleGAN [24]	0.9404	0.9512	0.8970	0.9256	0.8500	0.8422	0.8744	0.8787	0.9332	0.9242	0.8826	0.9118
UAGGAN [12,13]	0.9433	0.9515	0.8953	0.9311	0.8672	0.8603	0.8787	0.8927	0.9471	0.9375	0.8932	0.9296
AGGAN [14,15]	0.9486	0.9527	0.9021	0.9391	0.8577	0.8549	0.8661	0.8832	0.9346	0.9304	0.9110	0.9159
CDA-GAN (Ours)	0.9534	0.9691	0.9206	0.9411	0.8755	0.8770	0.8809	0.8969	0.9485	0.9418	0.9120	0.9326

The main Python libraries involved in the experiments are Numpy (for matrix operations), PIL, Nibabel (for reading, processing, and saving medical images), visdom (for tracking and analysis of the experimental process), and torchvision (a PyTorch-related image processing library). We adopt the Adam algorithm to optimize the networks. For generating tasks, the learning rate is initialized to $1e-4$, and after 100 epochs, the learning rate starts decaying to zero linearly. For classification tasks, the learning rate is initialized to $5e-4$, followed by decreasing the learning rate 2 times every 10 epoch during the training. For segmentation tasks, the learning rate is initialized to $1e-4$, followed by the warmupCosine learning rate scheduler. Both two tasks are trained using the Adam optimizer with a mini-batch size of 2, where the weight decay parameter in Adam is set to $1e-4$.

4.4. The selection of data augmentation baselines

The CDA-GAN model aims to accomplish image generation tasks by transforming positive and negative samples. However, the transformed positive and negative samples, as a solution for data augmentation, are more difficult in technology. Positive and negative sample transformation is required to determine the presence of tumors at the pixel level during the generation process. Therefore, to verify the validity of this method, we compare the data augmentation method with (i) classic traditional data augmentation methods [16] and (ii) GAN-based data augmentation solutions on medical images in single-domain [19–21]. Besides, the proposed CDA-GAN is further compared with (iii) several state-of-the-art image-to-image translation models in cross-domain [12, 13, 24, 42].

(i) **Traditional data augmentation method:** Following the previous data augmentation method in medical images [16], we use the traditional data augmentation scheme (e.g., crop, rotation, flip) followed by a sampling-based strategy to cope with datasets.

(ii) **Mixup-based data augmentation method:** The mixup-based data augmentation method is also widely used in medical imaging tasks [46], and we process datasets through the Mixup, CutMix, and CutOut data augmentation methods.

(iii) **GAN-based data augmentation methods on medical images:** PGGAN [19,20] is the first model introducing a progressive growth strategy to break the limitation in the resolution of GANs-based image generation tasks. Similar to PGGAN, MSG-GAN [21], which proposed a different growth strategy to reach high-resolution synthetic images, is the latest method state-of-the-art in image generation. In this experimental study, they are trained to generate images on datasets. After that, the synthetic samples were combined with the real samples to train the classification model.

(iv) **Unpaired image-to-image translation state-of-the-art methods:** Unsupervised image translation is more and more popular inspired by recent progress in GANs. AGGAN [14,15] is the latest state-of-the-art

in attention-guided image-to-image translation. We also compare with several known translation models, CycleGAN [24], which introduces the cycle consistency to the unpaired image translation for the first time, and UAG-GAN [12,13], which first introduces the spatial attention mechanism into cycle-consistence, loss-based models. We combine the above translation models with the proposed data augmentation scheme to improve the performance of subsequent medical image analysis tasks. Specifically, the above translation models are trained to model the translation link between classes, and then, the training data is fed into translation models to generate images. Finally, the synthetic images are combined with original training data to train the medical image analysis tasks.

5. Results

In this section, we conduct extensive experiments to investigate the performance of the CDA-GAN model across a range of datasets, module architectures, and tasks. Firstly, we compare our proposed CDA-GAN medical image generation method with several state-of-the-art data augmentation methods in the brain tumor classification tasks. Then, we further validate the performance of the important components in our model, including the AMSE module, the semi-supervised spatial attention module, and the spectral normalization in discriminators. Further, we experimented with and analyzed the impact of the ratio of synthetic to original images in the training data on the medical image classification task. Finally, we validate the performance of the proposed data augmentation strategy on a small dataset with few positive samples in classification tasks, which is common in real-world medical practice.

5.1. Data augmentation methods in classification

In this group of experiments, we compare the proposed cross-domain medical image generation scheme with several state-of-the-art GANs-based data augmentation schemes for the medical image classification task in Section 3.2.

The classification results show that the CDA-GAN achieved the best performance. The finding observations are in Fig. 4 and Table 2: (i) All data augmentation methods show better classification results than models trained without data augmentation methods, proving that data augmentation plays a key role in modeling feature distributions. (ii) Mixup-based data augmentation methods are superior to traditional data augmentation methods. Mixup-based can improve data diversity, mitigate overfitting, and enhance the robustness of the model by introducing sample blending to improve the generalization ability of the model. (iii) GANs-based data augmentation solutions outperform traditional data augmentation solutions and obtain relatively more additional information gained from the training set. It proves that traditional data augmentation obtains little additional information limiting

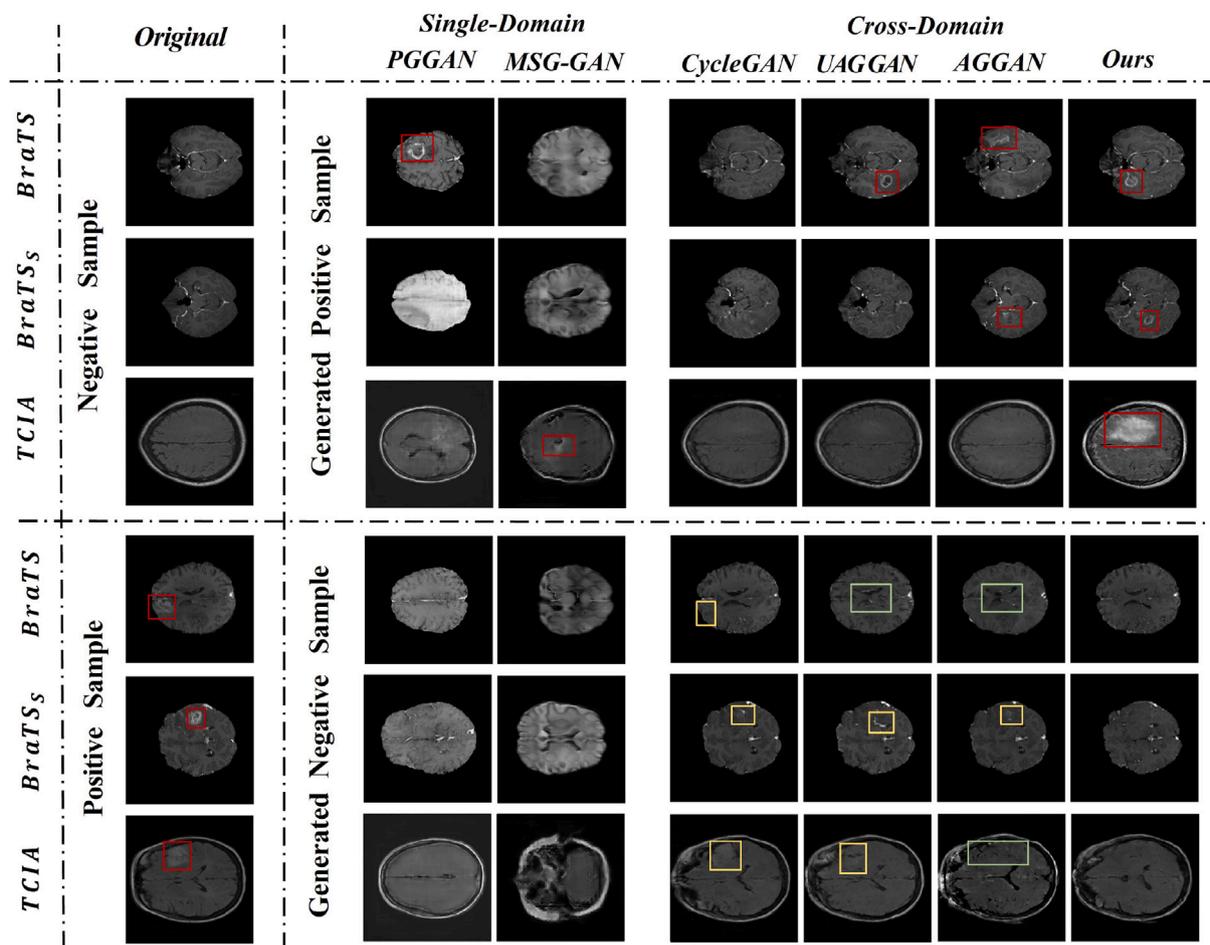


Fig. 4. The synthetic images of positive and negative samples in different methods. The red box indicates the obvious tumor area, the yellow box indicates the incomplete area of tumor generation, and the green box indicates the area where the brain tissue was over-modified except the tumor during the generation process.

the improvement of the model and GANs improving the scale of data feature distributions learned by the classifier from the dataset. (iv) Focusing on the results from *BraTS* to *BraTS_s* and *TCIA*, in the ACC and AUC of classification results, traditional sampling-based data augmentation methods are lower than GAN-based data augmentation methods, which indicates the fragility of traditional methods when encountering small datasets. However, GANs-based models show advantages on small datasets. (v) The cross-domain GANs-based models have competitive performance compared with single-domain GANs-based models, representing that our proposed cross-domain medical image generation scheme can generate images with more diverse and obvious distinctions between classes, further improving performance on subsequent classification tasks.

From the synthetic image of the cross-domain translation model in Fig. 4, we can find several obvious conclusions: (i) Mixup-based data augmentation methods are suitable for pixel-level labeling tasks (segmentation tasks). However, compared with the information augmentation method based on cross-domain GAN, it cannot produce an advantage in the amount of effective information, so the overall data enhancement effect is not as good as that of the cross-domain GAN-based information augmentation method. (ii) When compared with cross-domain generative models that effectively preserve and manipulate relevant features through cycle-consistent loss, single-domain generative models, such as Progressive Growing of GANs (PGGAN) and Multi-Scale Gradients GANs (MSG-GAN), often fall short. The images these single-domain models generate not only fail to ensure the depiction of tumor tissue but also significantly alter the inherent characteristics of various tissues impacted by brain tumors. This underscores

the superiority of cross-domain models in producing more accurate and clinically relevant representations. (iii) Beyond the recovery of tissue structure, the medical images generated by attention-guided GANs, including models such as UAGGAN, AGGAN, and our proposed CDA-GAN, exhibit superior performance compared to those built on CycleGAN. This enhancement in performance underscores the crucial role that attention mechanisms play in medical image generation tasks. These models achieve their improved performance by integrating attention mechanisms into the CycleGAN framework. This integration effectively directs the model's learning focus towards areas of greater significance, thereby enhancing the quality of the generated images and their utility in medical diagnosis. (iv) The proposed Cross-Domain Attention GAN (CDA-GAN) demonstrates exceptional performance in maintaining generation stability, even when handling smaller datasets such as *BraTS_s* and *TCIA*. This is a clear advantage over other models, which may only perform incomplete generation tasks. Specifically, CDA-GAN is capable of generating tumor foci in negative samples and eradicating tumors from positive samples, exhibiting its strength in handling complex medical image generation tasks. While all these models are attention-guided GANs, the attention mechanism in our CDA-GAN outperforms the other two (UAGGAN, AGGAN). Firstly, we have introduced channel attention in addition to spatial attention. Secondly, we employ a semi-supervised approach by using tumor labels to guide the training of the attention module. Although AGGAN also employs supervised loss to guide the training of its attention module, its supervision applies to the entire image. In contrast, we believe that the attention mask generated by guiding our attention module is more inclined to output the tumor part, providing a supervision method more

Table 3
Comparison with the state-of-the-art GAN-based data augmentation in the segmentation task.

Methods	BraTS				BraTS _S				TCIA			
	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU
w/o DA	0.6787	0.6507	8.5598	0.5762	0.5566	0.5442	15.4252	0.4382	0.6474	0.7234	38.7380	0.5364
Pseudo-label method	0.6729	0.6492	8.7826	0.5715	0.5215	0.5150	19.5641	0.4053	0.5968	0.7670	44.1802	0.4692
Mixup [43]	0.6895	0.6690	8.2624	0.5793	0.5685	0.5848	15.2345	0.4326	0.6650	0.7571	32.1632	0.5436
CutMix [44]	0.6989	0.6834	7.8994	0.5828	0.5668	0.5627	16.7834	0.4387	0.6459	0.7474	43.8078	0.5096
CutOut [45]	0.6969	0.6753	8.0696	0.5762	0.5577	0.5438	16.4386	0.4256	0.6508	0.7738	36.5420	0.5427
CycleGAN [24]	0.6931	0.6651	7.6409	0.5801	0.5661	0.5554	14.5276	0.4492	0.6772	0.7492	35.5067	0.5583
UAGGAN [12,13]	0.6964	0.6717	8.1684	0.5922	0.5921	0.5874	16.1591	0.4730	0.6956	0.8035	38.0205	0.5668
AGGAN [14,15]	0.6991	0.6756	7.3038	0.5852	0.5876	0.5821	19.7026	0.4685	0.6897	0.7843	35.0214	0.5604
CDA-GAN (Ours)	0.7030	0.6866	6.8097	0.5993	0.5941	0.5989	14.0390	0.4760	0.7130	0.8107	29.7807	0.5905

tailored to the requirements of generative tasks. This nuanced approach allows CDA-GAN to focus more precisely on areas of interest, enhancing its performance in medical image generation tasks. In summary, we can find that our proposed CDA-GAN model achieves the best performance among the whole models on datasets. CDA-GAN is more accomplished in the task of recovering positive samples. Comparing synthetic positive images in all models, the CDA-GAN stabilizes the translation process and achieves successful negative-to-positive translation in as many cases as possible. However, other models can only achieve incomplete translation in certain situations.

The results shown in Table 2 can confirm our conclusions in Fig. 4 that GAN-based data augmentation methods are significantly more effective than traditional data augmentation methods for brain tumor classification tasks. Meanwhile, the cross-domain generation model has some improvements compared to the single-domain generation model. In addition, in cross-domain generation methods, CDA-GAN achieves the best outcome. Numerically, UAGGAN, AGGAN, and CDA-GAN on brain tumor classification in three datasets achieve certain improvements, and this is consistent with the performance in the generated images. This illustrates that although the methods of adding spatial attention mechanisms are different for each model, the addition of a spatial attention module is an enhancement to the generation capacity of the model, which also proves the necessity of adding a spatial attention module.

In summary, the robustness of translation models in the image generation procedure plays a crucial role in the subsequent image analysis task. Our proposed model outperforms other models because of its stable ability in the image generation task. Meanwhile, CycleGAN achieves the worst performance among cross-domain models, which proves that the attention module plays an essential role in our image generation task.

5.2. Information enhancement methods in segmentation

In this group of experiments, firstly, we validate the performance of the pseudo-labels method for the segmentation task. Specifically, we subtract the tumor image generated by the CDA-GAN model from the original tumor-free image to obtain a pseudo-label for the generated tumor image. Then, we add the generated tumor images and their pseudo-labels to the training set and train the model with the original data. From Table 3, we find that segmentation performance with generated differences as pseudo-labels is significantly weaker than that of the base model without pseudo-labels in three datasets. In particular, when the dataset is smaller (*BraTS_S/TCIA*), the segmentation performance using pseudo-label methods is worse. The reason for this phenomenon may be that the poor generation effect in small datasets leads to poorer quality of pseudo-labels, which leads to a sharp drop in segmentation performance. In Fig. 5, we can also see that the tumor difference is more obvious in the larger dataset (*BraTS*), which also proves our explanation for the poor segmentation performance of the pseudo-label method under the small dataset. Secondly, we use an information enhancement strategy in Section 3.2 and compare the proposed CDA-GAN model with several state-of-the-art cross-domain

image-to-image translation models. The conclusion of the experiment is as follows: (i) The segmentation results of the cross-domain model are better than the pseudo-label method. This proves that directly using the difference image as the segmentation pseudo-label will introduce a lot of noise, resulting in performance degradation. The information enhancement strategy is a way to significantly improve segmentation performance. (ii) UAGGAN, AGGAN, and CDA-GAN outperform CycleGAN when using an information augmentation strategy. This is consistent with the performance in the classification task, which proves that adding attention mechanisms to CycleGAN is beneficial for medical image generation tasks. At the same time, in Fig. 5, the aforementioned conclusion can be clearly observed in generated tumor-free images, difference images, or segmented tumor regions. The results also correspond to the experimentally derived tabular data. (iii) Our proposed CDA-GAN model has state-of-the-art performance on information-augmented segmentation tasks. In particular, the effect of model information enhancement is more pronounced on smaller datasets, both in the table and figure. Especially, as shown in Fig. 5, it can be seen that the boundary of the brain tumor enhanced by the information is not clear enough, and in addition to the distinct tumor tissue, it also contains brain tissue. This is because the pixel-level generation task based on the attention mechanism is likely to notice other abnormal regions in non-tumor tissues. However, according to the segmentation results, this information enhancement strategy can significantly improve the segmentation accuracy of tumors.

Overall, compared with other state-of-the-art generative models, the medical images generated by CDA-GAN can not only improve classification tasks but also achieve remarkable results in pixel-level segmentation tasks. Specifically, CDA-GAN outperforms PGGAN and MSG-GAN because it is a cross-domain model that leverages cycle-consistent loss. This type of loss ensures that our model can maintain consistent information in the generation, which is a significant advantage when dealing with medical images. When it comes to CycleGAN, it is the base model for our work. While CycleGAN has made significant strides in image translation tasks, our proposal furthers its capabilities by introducing AMSE Block, Semi-supervised Spatial Attention Module and Spectral Normalization, which have been shown to improve the performance of generative models in both classification and segmentation tasks. While UAGGAN uses an attention mechanism similar to ours, it does not incorporate the same constraints on attention as ours. Our model incorporates the use of a segmentation label as a prior to guide the attention module. This strategic use of prior knowledge makes the attention mechanism in our model more effective and discriminative. Finally, the difference between AGGAN and CDA-GAN is the attention constraint strategy. By integrating a segmentation label into our model, our attention module can be trained more effectively, leading to better performance in capturing cross-domain features. Our proposed information augmentation strategy can further improve the accuracy of the segmentation task by making full use of the generated images.

5.3. Comparison of the computational costs

In this section, we compare the computational costs of cross-domain GAN methods, which can enrich the evaluation dimensions of the

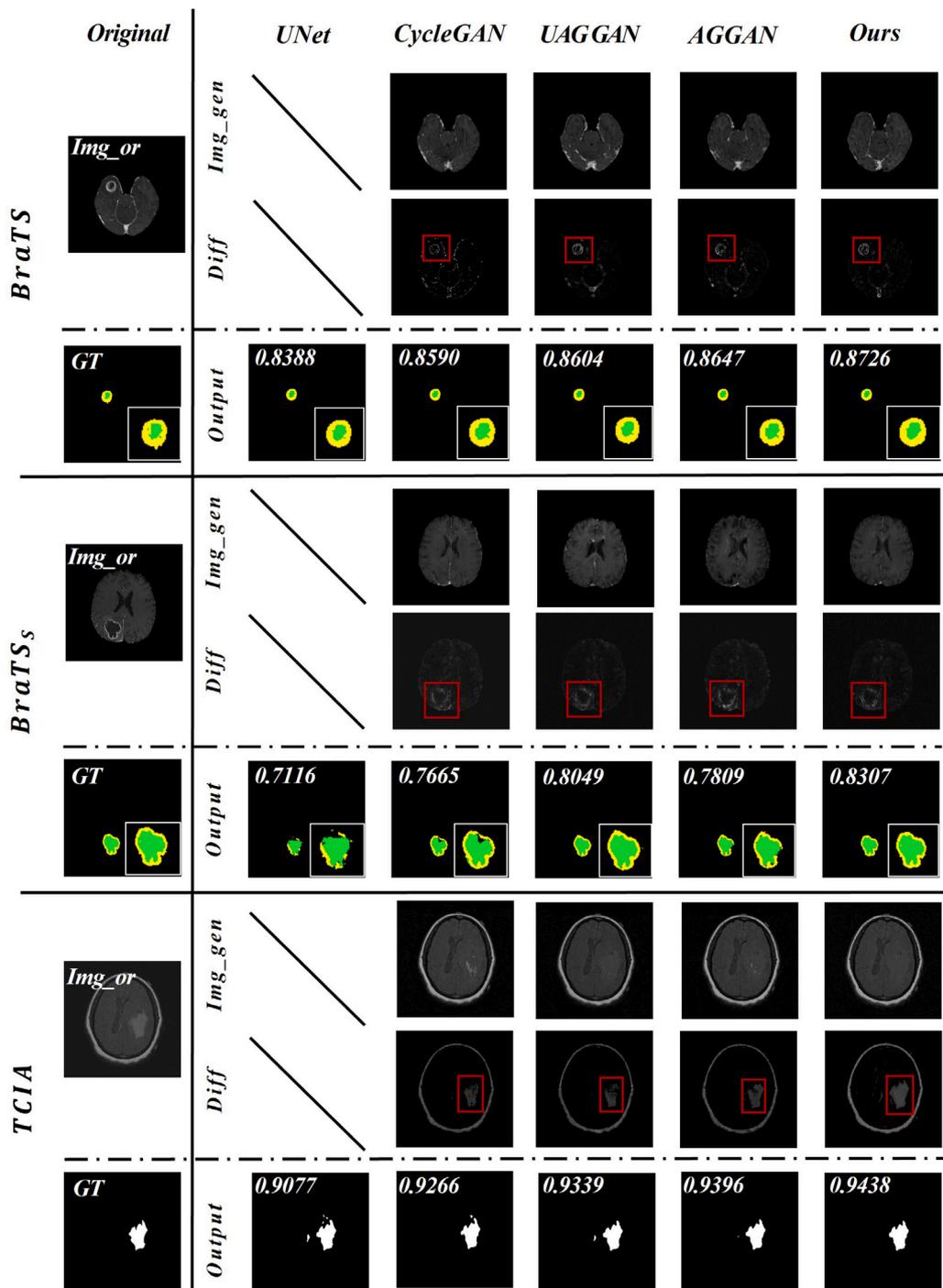


Fig. 5. Input and output of the segmentation task applying the information augmentation strategy. The original image and the different images are combined on the channel as the input of the segmentation, and this input method is applied to the training, validation, and testing processes.

Table 4
Comparison with the GAN-based data augmentation methods in training cost, which is measured by the average time (in hour) spent per epoch.

Methods	BraTS	BraTS _s	TCIA
CycleGAN	0.6774	0.0906	0.0556
UAGGAN	0.6752	0.0834	0.0547
AGGAN	0.7181	0.0869	0.0588
CDA-GAN	0.7932	0.0963	0.0629

CDA-GAN method. Our work involves a two-step process: data generation and model task. During the data generation phase, our proposed

scheme does incur higher computational costs due to the additional modules and steps involved. However, during the model task phase, our scheme does not impose any additional computational cost compared to other schemes. This is because we use the same task model as other scenarios. Therefore, it is necessary to compare GAN-based data augmentation methods with additional modules and steps. In addition, the single-domain GAN method trains corresponding models for positive samples and negative samples respectively, and the cross-domain GAN method completes the conversion between positive and negative samples at one time, the computational cost between them cannot be directly quantified, we focus is on comparing the computational cost of cross-domain GAN models associated with our proposed CDA-GAN method.

Table 5
Ablation studies in the classification and segmentation tasks.

Classification task												
Methods	BraTS				BraTS _s				TCIA			
	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1
CDA-GAN w/o SE	0.9477	0.9686	0.8777	0.9055	0.8623	0.8542	0.7399	0.7675	0.9443	0.9358	0.8968	0.9265
CDA-GAN w/o SA	0.9447	0.9681	0.8824	0.9083	0.8691	0.8665	0.8774	0.8929	0.9401	0.9299	0.8826	0.9202
CDA-GAN w/o SS	0.9471	0.9688	0.8761	0.9135	0.8626	0.8596	0.8716	0.8874	0.9420	0.9324	0.8968	0.9214
CDA-GAN w/o SN	0.9489	0.9691	0.8835	0.9107	0.8726	0.8729	0.8559	0.8930	0.9444	0.9416	0.9028	0.9288
CDA-GAN (Ours)	0.9534	0.9691	0.9206	0.9411	0.8755	0.8770	0.8809	0.8969	0.9485	0.9418	0.9120	0.9326
Segmentation task												
Methods	BraTS				BraTS _s				TCIA			
	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU
CDA-GAN w/o SE	0.6927	0.6786	9.3287	0.5886	0.5857	0.5647	15.8306	0.4699	0.6671	0.7701	32.9830	0.5445
CDA-GAN w/o SA	0.6903	0.6769	8.1433	0.5881	0.5686	0.5606	14.5530	0.4548	0.6629	0.7640	32.3488	0.5398
CDA-GAN w/o SS	0.6826	0.6606	9.1328	0.5896	0.5769	0.5784	17.1248	0.4588	0.6733	0.7975	32.7705	0.5478
CDA-GAN w/o SN	0.6895	0.6690	8.0340	0.5857	0.5821	0.5704	14.5375	0.4676	0.6820	0.7600	32.7416	0.5579
CDA-GAN (Ours)	0.7030	0.6866	6.8097	0.5993	0.5941	0.5989	14.0390	0.4760	0.7130	0.8107	29.7807	0.5905

The computational cost results of the cross-domain GAN method are in the Table 4. We can obtain the following conclusions: (i) The size of the training data significantly affects the training time of each method. For example, the average time spent per epoch in a four-modal, large-data-volume dataset like BraTS is much higher than that of a single-modal, small-data-volume dataset like TCIA. (ii) The UAGGAN method has the shortest time among the three datasets. The average time per epoch is similar to the CycleGAN method (UAGGAN: 0.6752 h vs CycleGAN: 0.6774 h in BraTS). This also shows that the spatial attention method of UAGGAN can achieve better data enhancement effects without extra time expenditure. (iii) The computational cost of the CDA-GAN method we proposed is relatively the largest, but the model computational cost does not increase significantly. Secondly, in the brain tumor classification and segmentation tasks, the data enhancement effect of CDA-GAN is significantly improved compared to AGGAN and UAGGAN. We believe that this is the possible price for achieving better generation quality.

5.4. Ablation study

To show the effectiveness and necessity of the proposed advanced modules, ablation studies are further conducted, where several intermediate models that only use one or two advanced modules are introduced and evaluated. Table 5 summarizes our experiments on applying ablated versions of the CDA-GAN with/without different modules.

5.4.1. Effectiveness of channel attention module

As introduced in Section 3.1.1 in the CDA-GAN model, we employ AMSE-residual blocks to capture inter-dependencies between channels. This design choice is instrumental in enhancing the quality of the image representations generated by our model. By exploiting the inter-channel dependencies, these blocks allow for a more nuanced understanding of the input data, thus contributing to the generation of higher-quality, more accurate medical images. To evaluate the effectiveness of this module, we designed another version (i.e., CDA-GAN w/o SE) of our proposed model (CDA-GAN) for comparison. In Table 5, the variants of our proposed model are compared on the task of brain tumor classification and segmentation. According to Tables 2 and 5, several observations can be found: Firstly, compared with state-of-the-art data augmentation methods, the above two models (i.e., CDA-GAN, CDA-GAN w/o SE) lead to competitive performance (e.g., ACC: 0.9477 and **0.9534**, Dice: 0.6927 and **0.7030** in BraTS) on both tasks and the results of our model are the best in the four indicators of the three datasets. It reflects the effectiveness of the proposed method. Secondly, our proposed CDA-GAN outperforms CDA-GAN w/o SE on three datasets. It indicates that channel-level inter-dependencies are related to the procedure of modeling feature distribution.

5.4.2. Effectiveness of semi-supervised spatial attention module

In experiments, we evaluate the effectiveness of the spatial attention module as well as the semi-supervised strategy. Specifically, we perform two variants of our model (i.e. CDA-GAN w/o SA, CDA-GAN w/o SS). First, we compare the CDA-GAN model with the CDA-GAN model without the spatial attention module and semi-supervised training strategy (CDA-GAN w/o SA). Second, we compare the CDA-GAN model with a CDA-GAN model without a semi-supervised training strategy (CDA-GAN w/o SS).

The experimental results of classification and segmentation are presented in Table 5. From the results we can analyze the following conclusions: (i) Our proposed CDA-GAN outperforms CDA-GAN w/o SA and CDA-GAN w/o SS in each dataset, e.g., in the BraTS_s, the ACC for tumor classification is **0.8755** versus 0.8691 versus 0.8626, and the Dice for tumor segmentation is **0.5941** versus 0.5686 versus 0.5769. It indicates that spatial attention can force the generator to translate relevant regions, which improves the robustness of the image generation procedure. This proves that both semi-supervised methods and spatial attention can improve the generator's acquisition of lesion features, reliably generate positive and negative samples, and perform effective data enhancement in downstream medical image analysis tasks. (ii) The results of CDA-GAN w/o SA are mostly lower than those of CDA-GAN w/o SS on classification and segmentation tasks. This finding illustrates that the application of spatial attention can effectively guide the generator to focus on translating relevant regions, thus improving the robustness of the image generation process. Therefore, the proposed semi-supervised strategy is essential in our model.

5.4.3. Effectiveness of spectral normalization for stable training procedure

As introduced in Section 3.1.3, we used the spectral normalization [40] in discriminators to constrain the Lipschitz constant. This improvement prevents the proposed CDA-GAN from mode collapse and stabilizes the training procedure. In this group of experiments, we aim to evaluate the performance of the spectral normalization (SN) module for the data augmentation task. We perform a variant of the CDA-GAN (i.e., CDA-GAN w/o SN). The results presented in Table 5, we find that all metrics in classification and segmentation for three datasets have improvements by restricting spectral norm to each layer of discriminators. It indicates that the proposed CDA-GAN stabilizes the training process by introducing spectral normalization and improves the classification and segmentation performance with a small computational cost.

5.5. Different attention-based blocks in cross-domain GAN

Although the above ablation studies have proved the necessity and effectiveness of the components in attention modules. In this subsec-

Table 6
The performances on datasets with different attention blocks of CDA-GAN.

Methods	BraTS				BraTS _s				TCIA			
	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1
CycleGAN+SE	0.9402	0.9428	0.8996	0.9348	0.8563	0.8608	0.8724	0.8692	0.9375	0.9305	0.8970	0.9147
CycleGAN+CABM	0.9469	0.9526	0.9087	0.9324	0.8732	0.8685	0.8706	0.8757	0.9338	0.9358	0.9036	0.9317
CDA-GAN(AMSE)	0.9534	0.9691	0.9206	0.9411	0.8755	0.8770	0.8809	0.8969	0.9485	0.9418	0.9120	0.9326
Methods	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU	Dice	Sens	HD95	mIoU
CycleGAN+SE	0.6935	0.6783	7.0997	0.5890	0.5796	0.5760	44.1802	0.4692	0.6753	0.7925	33.6505	0.5421
CycleGAN+CABM	0.6944	0.6821	6.8134	0.5905	0.5828	0.5759	41.0231	0.4756	0.6803	0.7526	32.7346	0.5575
CDA-GAN(AMSE)	0.7030	0.6866	6.8097	0.5993	0.5941	0.5989	14.0390	0.4760	0.7130	0.8107	29.7807	0.5905

Table 7
The performances on datasets with different amounts of synthetic samples.

Ratios	BraTS				BraTS _s				TCIA			
	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1	ACC	AUC	Recall	F1
20%	0.9215	0.9320	0.8935	0.9261	0.8504	0.8554	0.8350	0.8740	0.9263	0.9197	0.8897	0.9042
40%	0.9256	0.9399	0.8976	0.9298	0.8509	0.8487	0.8576	0.8773	0.9305	0.9244	0.8968	0.9097
60%	0.9324	0.9456	0.9148	0.9317	0.8645	0.8658	0.8603	0.8875	0.9346	0.9259	0.8861	0.9138
80%	0.9487	0.9497	0.9157	0.9358	0.8726	0.8729	0.8559	0.8930	0.9332	0.9280	0.9039	0.9208
100%	0.9534	0.9691	0.9206	0.9411	0.8755	0.8770	0.8809	0.8969	0.9485	0.9418	0.9120	0.9326

tion, we further compare the proposed CDA-GAN with two similar models, CycleGAN+SE and CycleGAN+CABM, to show that AMSE with the Semi-supervised Attention Module is a better attention choice than the existing SOTA attention baselines, e.g., SE and CABM, in the medical image argumentation tasks.

We can obtain the following conclusions from the results in [Table 6](#): (i) In the classification and segmentation task, the results of the CABM module are mostly better than those of the SE module. This shows that the combined action of channels and spaces in the data enhancement method based on cross-domain GAN can better capture image features and further improve the data augmentation effect. (ii) Compared with the SE block and CBAM block, the attention block we proposed has significant improvements in both classification and segmentation tasks. The attention block in the CDA-GAN method of separating the two types of attention, channel attention and spatial attention, allows for more flexibility and specificity in the learning process of the model. In addition, the AMSE residue block combines average and maximum pooling features and can capture more global and local attention features than the SE block.

5.6. Influence of the amount of synthetic samples

In this group of experiments, we conduct experiments to verify the value of generated images. 20%, 40%, 60%, and 80% of the generated samples were mixed into the real samples, and then, four classification models were trained with four combined datasets, respectively. However, in the segmentation task, we found that training UNet according to the percentage of added information enhancement does not enable positive samples to obtain corresponding augmentation information but degrades the segmentation results. This indicates that the information enhancement strategy is for the entire data and cannot be augmented by percentage. [Table 7](#) showed the classification results with four datasets. The results show that the classification performance of the model steadily improves as the number of generated samples increases. This demonstrated that our proposed model could increase the feature space uncovered by original data, provide extra feature information to the combined datasets, and serve as additional information to enhance the capability of classification and segmentation.

5.7. Performances of CDA-GAN in extremely imbalanced small dataset

In the clinic, abnormal findings of disease are rare, especially in brain tumors, so training models on small datasets with extreme class imbalances are common in real-world medical practice. To verify the

capability of the proposed model in coping with the real-world situation, we collected a small dataset with only 70 positive samples and 730 negative samples randomly chosen from the *BraTS* dataset. In imbalanced small datasets, the classification performance of data augmentation models is shown in [Fig. 6](#). We can obtain from this analysis that the CDA-GAN outperforms other models, which proves the stability of our model in dealing with extremely unbalanced small datasets and in the face of real-world medical practice situations. The single-domain GANs (PGGAN and MSG-GAN) are unable to model the feature distributions when dealing with a small dataset, thus failing to generate meaningful new data. This demonstrates their fragility in augmenting small datasets. Comparing the results of CycleGAN with attention-guided UAGGAN and AGGAN shows that in the face of small datasets and very few positive samples, although the attention mechanism helps to generate new samples, the improvement is small. Because the training stability of the attention module with adversarial loss is extremely susceptible to the size of the training samples, that is, the more unbalanced and the smaller the dataset, the more likely it will reach model instability.

6. Discussion and future work

In this section, we first summarize the main differences between the proposed CDA-GAN data augmentation model and previous studies in medical image data augmentation. We also point out the limitations of our proposed model as well as potential solutions to deal with these limitations in the feature.

6.1. Comparison with previous work

Compared with the existing data augmentation models in medical imaging, our proposed CDA-GAN is different from existing image-to-image translation models [[12](#),[13](#),[24](#),[42](#)]. In contrast to the model [[24](#)], our proposed model integrates a spatial-attention module, so the model can pay more attention to the region of interest. In contrast to translation models that also use spatial-attention modules to guide the training procedure of generators, our train mechanism can force the attention module to produce a more precise spatial attention map, thus strengthening the quality of spatial encodings. Furthermore, CDA-GAN substitutes the conventional residual block with an AMSE-residual block, a move designed to capture channel-wise dependencies. This replacement serves to enhance the representational power of Convolutional Neural Networks (CNNs), and to stabilize the training procedure, thereby improving the overall effectiveness of the model. Notably, we

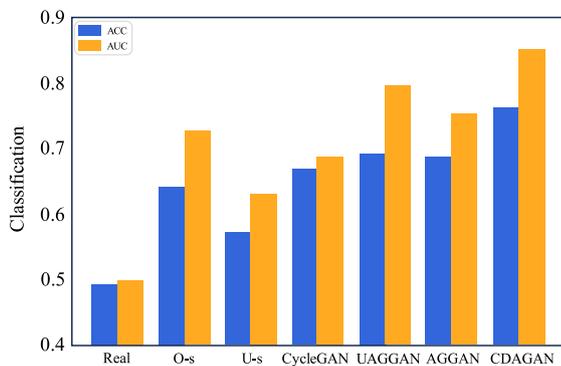


Fig. 6. Performances on the dataset with extreme class imbalance, where *O-s* represents oversampling in traditional data enhancement, and *U-s* represents undersampling in traditional data enhancement.

incorporate spectral normalization into our model. This addition is crucial in preventing mode collapse in the discriminator. Mode collapse is a common challenge in training GANs, where the generator starts producing a limited diversity of samples, and the discriminator cannot effectively distinguish between real and generated data. By introducing spectral normalization, we can mitigate this issue, enhancing the stability and robustness of our CDA-GAN model.

6.2. Limitations and future work

Although the CDA-GAN data augmentation model achieves good performance in classification and segmentation tasks, in the future, its performance and image generation capacity can be further improved by carefully dealing with the following limitations or challenges. Firstly, in current research, the image-to-image translation is the one-to-one generation. The model generates individual outputs based on given inputs, limiting the diversity of generated samples. Inspired by StarGAN [47], we will try to solve this problem by introducing multi-domain transformation to the proposed model in our future work. The complex training process of GAN limits its generation ability. Secondly, our current work is based on 2D slices. We plan to extend our model to incorporate 3D imaging, fully leveraging the model's training stability and the authenticity of the generated images. By generating 3D medical images, we aim to enhance the performance of 3D medical imaging tasks, including classification, segmentation, and detection. Furthermore, the Diffusion-based model has been drawing significant interest in the field of image generation lately. We found that there are still no unpaired image-to-image generative models in the field of medical image generation. Meanwhile, in the domain of natural images, there are two works on unpaired image-to-image generation, a common feature of these studies is the requirement to pre-train a DDPM [48] model on data from both modalities [49,50]. We then conducted experiments with DDPM. However, the model did not perform well on our dataset. We were unable to train a DDPM model capable of generating reasonable brain MRI images. As a result, we also could not train the above-mentioned two unpaired image-to-image translation models based on diffusion. We hypothesize that the underperformance could be due to the extensive diffusion involved, which demands a larger dataset for effective learning. In future work, we plan to either expand our dataset to accommodate the requirements of diffusion models or to integrate diffusion models into our existing strategy. Besides, We hope that generative data augmentation can be applied in the detection tasks of medical imaging [51,52], and achieve better segmentation results than the method of optimizing segmentation models [53] with limited data.

7. Conclusion

In this work, we propose a novel cross-domain generative data augmentation model called CDA-GAN to enlarge the number of sam-

ples and balance the number of samples across categories to improve performance on brain tumor classification tasks. Further, we propose an information enhancement strategy for the segmentation task to the attention-generated tumor feature map. In CDA-GAN, we apply the AMSE-residual module, semi-supervised spatial attention, and spectral normalization to capture the differences between different classes more precisely, bridge the gap between local and global feature representations and stabilize the training process. In three public datasets based on the BraTS and TCIA datasets, the effectiveness of our proposed data augmentation model on medical image analysis tasks is extensively evaluated.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China under the grants 62276089 and 61906063, by the Natural Science Foundation of Hebei Province, China, under the grant F2021202064, by the Key Research and Development Project of Hainan Province, China, under the grant ZDYF2022SHFZ015, and by the Natural Science Foundation of Hainan Province, China, under the grant 821RC1131. This work was also supported by the AXA Research Fund.

References

- [1] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, O. Tabona, A survey on missing data in machine learning, *J. Big Data* 8 (1) (2021) 1–37.
- [2] D. Yuan, Y. Liu, Z. Xu, Y. Zhan, J. Chen, T. Lukasiewicz, Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing, *Comput. Biol. Med.* (2022) 106487.
- [3] D. Yuan, Z. Xu, B. Tian, H. Wang, Y. Zhan, T. Lukasiewicz, μ -Net: Medical image segmentation using efficient and effective deep supervision, *Comput. Biol. Med.* 160 (2023) 106963.
- [4] J. Zhang, S. Zhang, X. Shen, T. Lukasiewicz, Z. Xu, Multi-ConDoS: Multimodal contrastive domain sharing generative adversarial networks for self-supervised medical image segmentation, *IEEE Trans. Med. Imaging Early Access* (2023) 1–20, <http://dx.doi.org/10.1109/TMI.2023.3290356>.
- [5] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, S. Satoh, Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images, in: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 119–127.
- [6] I. Lorencin, S. Baressi Šegota, N. Anđelić, V. Mrzljak, T. Čabov, J. Španjol, Z. Car, On urinary bladder cancer diagnosis: Utilization of deep convolutional generative adversarial networks for data augmentation, *Biology* 10 (3) (2021) 175.
- [7] D. Srivastav, A. Bajpai, P. Srivastava, Improved classification for pneumonia detection using transfer learning with GAN based synthetic image augmentation, in: *Proceedings of the International Conference on Cloud Computing, Data Science & Engineering*, 2021, pp. 433–437.
- [8] J. Wang, T. Lukasiewicz, X. Hu, J. Cai, Z. Xu, RSG: A simple but effective module for learning imbalanced datasets, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3784–3793.
- [9] K. Santosh, S. Ghosh, Covid-19 imaging tools: How big data is big? *J. Med. Syst.* 45 (7) (2021) 71.
- [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] X. Meng, Y. Gu, Y. Pan, N. Wang, P. Xue, M. Lu, X. He, Y. Zhan, D. Shen, A novel unified conditional score-based generative framework for multi-modal medical image completion, 2022, *ArXiv preprint*, [arXiv:2207.03430](https://arxiv.org/abs/2207.03430), URL <https://arxiv.org/abs/2207.03430>.
- [12] A. Abu-Srhan, I. Almallahi, M.A. Abushariah, W. Mahafza, O.S. Al-Kadi, Paired-unpaired unsupervised attention guided GAN with transfer learning for bidirectional brain MR-CT synthesis, *Comput. Biol. Med.* 136 (2021) 104763.
- [13] Y. Alami Mejjati, C. Richardt, J. Tompkin, D. Cosker, K.I. Kim, Unsupervised attention-guided image-to-image translation, in: *Advances in Neural Information Processing Systems*, Vol. 31, 2018.

- [14] L. Gao, K. Xie, X. Wu, Z. Lu, C. Li, J. Sun, T. Lin, J. Sui, X. Ni, Generating synthetic CT from low-dose cone-beam CT by using generative adversarial networks for adaptive radiotherapy, *Radiat. Oncol.* 16 (1) (2021) 1–16.
- [15] H. Tang, H. Liu, D. Xu, P.H. Torr, N. Sebe, Attentiongan: Unpaired image-to-image translation using attention-guided generative adversarial networks, *IEEE Trans. Neural Netw. Learn. Syst.* (2021).
- [16] T. Rejusha, V.K. KS, Artificial MRI image generation using deep convolutional GAN and its comparison with other augmentation methods, in: *Proceedings of the International Conference on Communication, Control and Information Sciences*, Vol. 1, 2021, pp. 1–6.
- [17] S. Dhivya, S. Mohanavalli, S. Karthika, S. Shivani, R. Mageswari, GAN based data augmentation for enhanced tumor classification, in: *Proceedings of the International Conference on Computer, Communication and Signal Processing*, 2020, pp. 1–5.
- [18] Y. Li, J. Zhao, Z. Lv, J. Li, Medical image fusion method by deep learning, *Int. J. Cogn. Comput. Eng.* 2 (2021) 21–29.
- [19] C. Han, L. Rundo, R. Araki, Y. Furukawa, G. Mauri, H. Nakayama, H. Hayashi, Infinite brain MR images: PGGAN-based data augmentation for tumor detection, in: *Neural Approaches to Dynamics of Signal Exchanges*, 2020, pp. 291–303.
- [20] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of gans for improved quality, stability, and variation, 2017, arXiv preprint arXiv:1710.10196.
- [21] A. Karnewar, O. Wang, Msg-gan: Multi-scale gradients for generative adversarial networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7799–7808.
- [22] R. Oulbacha, S. Kadoury, MRI to CT synthesis of the lumbar spine from a pseudo-3D cycle GAN, in: *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2020, pp. 1784–1787.
- [23] G. Zeng, T.D. Lerch, F. Schmaranzer, G. Zheng, J. Burger, K. Gerber, M. Tannast, K. Siebenrock, N. Gerber, Semantic consistent unsupervised domain adaptation for cross-modality medical image segmentation, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021, pp. 201–210.
- [24] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [25] P. Roy, N. Häni, V. Isler, Semantics-aware image to image translation and domain transfer, 2019, ArXiv preprint, arXiv:1904.02203, URL <https://arxiv.org/abs/1904.02203>.
- [26] R. Sun, C. Huang, H. Zhu, L. Ma, Mask-aware photorealistic facial attribute manipulation, *Comput. Vis. Media* 7 (2021) 363–374.
- [27] X. Lai, X. Bai, Y. Hao, Unsupervised generative adversarial networks with cross-model weight transfer mechanism for image-to-image translation, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1814–1822.
- [28] Y. Liu, Y. Chen, L. Bao, N. Sebe, B. Lepri, M. De Nadai, ISF-GAN: An implicit style function for high-resolution image-to-image translation, 2021, ArXiv preprint, arXiv:2109.12492, URL <https://arxiv.org/abs/2109.12492>.
- [29] V. Kearney, B.P. Ziemer, A. Perry, T. Wang, J.W. Chan, L. Ma, O. Morin, S.S. Yom, T.D. Solberg, Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks, *Radiol. Artif. Intell.* 2 (2) (2020).
- [30] C. Yang, T. Kim, R. Wang, H. Peng, C.-C.J. Kuo, Show, attend, and translate: Unsupervised image translation with self-regularization and attention, *IEEE Trans. Image Process.* 28 (10) (2019) 4845–4856.
- [31] Z. Xu, C. Qi, G. Xu, Semi-supervised attention-guided cyclegan for data augmentation on medical images, in: *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine*, 2019, pp. 563–568.
- [32] Y. Skandarani, P.M. Jodoin, A. Lalonde, Gans for medical image synthesis: An empirical study, *J. Imaging* 9 (3) (2023) 69.
- [33] H. Choi, D.S. Lee, Generation of structural MR images from amyloid PET: application to MR-less quantification, *J. Nucl. Med.* 59 (7) (2018) 1111–1117.
- [34] B. Cheng, B. Zhu, S. Pu, Multi-auxiliary domain transfer learning for diagnosis of MCI conversion, *Neuro. Sci.* 43 (2021) 1721–1739.
- [35] N. Deepa, S. Chokkalingam, Optimization of VGG16 utilizing the arithmetic optimization algorithm for early detection of alzheimer's disease, *Biomed. Signal Process. Control* 74 (2022).
- [36] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [37] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: *Proceedings of the European Conference on Computer Vision*, (ECCV), 2018, pp. 3–19.
- [38] S. Zhang, J. Zhang, B. Tian, T. Lukasiewicz, Z. Xu, Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation, *Med. Image Anal.* 83 (2023) 102656.
- [39] D. Ramegowda, I. Gondra, Wasserstein-based feature map knowledge transfer to improve the performance of small deep neural networks, in: *Proceedings of the International Conference on Pattern Recognition and Machine Learning*, 2022, pp. 409–415.
- [40] Z. Lin, V. Sekar, G. Fanti, Why spectral normalization stabilizes GANs: Analysis and improvements, *Adv. Neural Inf. Process. Syst.* 34 (2021) 9625–9638.
- [41] R. Mehta, A. Filos, U. Baid, C. Sako, R. McKinley, M. Rebsamen, K. Dätwyler, R. Meier, P. Radojewski, G.K. Murugesan, et al., QU-brats: MICCAI brats 2020 challenge on quantifying uncertainty in brain tumor segmentation—analysis of ranking metrics and benchmarking results, 2021, ArXiv preprint, arXiv:2112.10074, URL <https://arxiv.org/abs/2112.10074>.
- [42] H. Tang, D. Xu, N. Sebe, Y. Yan, Attention-guided generative adversarial networks for unsupervised image-to-image translation, in: *Proceedings of the International Joint Conference on Neural Networks*, 2019, pp. 1–8.
- [43] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond empirical risk minimization, 2017, arXiv preprint arXiv:1710.09412.
- [44] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization technique to train strong classifiers with localizable features, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6023–6032.
- [45] T. DeVries, G.W. Taylor, Improved regularization of convolutional neural networks with cutout, 2017, arXiv preprint arXiv:1708.04552.
- [46] D. Yao, Z. Xu, Y. Lin, Y. Zhan, Accurate and intelligent diagnosis of pediatric pneumonia using X-ray images and blood testing data, *Front. Bioeng. Biotechnol.* 11 (2023) 1058888.
- [47] S. Fuhai, X. Tang, Multi-modal MRI synthesization based on stargan, in: *The Fourth International Symposium on Image Computing and Digital Medicine*, 2020, pp. 19–22.
- [48] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: *Advances in Neural Information Processing Systems*, Vol. 33, 2020, pp. 6840–6851.
- [49] H. Sasaki, C.G. Willcocks, T.P. Breckon, Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models, 2021, arXiv preprint arXiv:2104.05358.
- [50] M. Zhao, F. Bao, C. Li, J. Zhu, Egsde: Unpaired image-to-image translation via energy-guided stochastic differential equations, *Adv. Neural Inf. Process. Syst.* 35 (2022) 3609–3623.
- [51] Z. Xu, T. Li, Y. Liu, Y. Zhan, J. Chen, T. Lukasiewicz, PAC-net: Multi-pathway FPN with position attention guided connections and vertex distance IoU for 3D medical image detection, *Front. Bioeng. Biotechnol.* 11, 106.
- [52] Z. Xu, X. Zhang, H. Zhang, Y. Liu, Y. Zhan, T. Lukasiewicz, EFPN: Effective medical image detection using feature pyramid fusion enhancement, *Comput. Biol. Med.* 163 (2023) 107149, <http://dx.doi.org/10.1016/j.combiomed.2023.107149>.
- [53] Z. Xu, S. Liu, D. Yuan, L. Wang, J. Chen, T. Lukasiewicz, Z. Fu, R. Zhang, ω -Net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution, *Neurocomputing* 500 (2022) 177–190.



Zhenghua Xu received a M.Phil. in Computer Science from The University of Melbourne, Australia, in 2012, and a D.Phil in computer Science from University of Oxford, United Kingdom, in 2018. From 2017 to 2018, he worked as a research associate at the Department of Computer Science, University of Oxford. He is now a professor at the Hebei University of Technology, China, and a awardee of “100 Talents Plan” of Hebei Province. He has published more than thirty papers in top AI or database conferences and journals, e.g., NeurIPS, AAAI, IJCAI, ICDE, IEEE TNNLS, Medical Image Analysis, etc. His current research focuses on intelligent medical image analysis, deep learning, reinforcement learning and computer vision.



Jiaqi Tang is currently a master's student at Hebei University of Technology, China. She received B.Eng. degree in computer science and B.Sci. degree in information science from Hebei University of Technology, China, in 2021. Her research interests lie in medical image generation using deep learning methods.



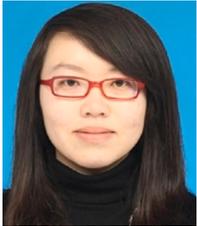
Chang Qi obtained a master's degree in biomedical engineering from Hebei University of Technology, China, in 2021. She received B.Eng. degree in biomedical engineering from Southern Medical University, China, in 2018. Her research interests lie in medical image generation using deep learning.



Dan Yao obtained a master's degree in electronic information engineering from Hebei University of Technology, China, in 2023. She received B.Eng. degree in communication engineering from Henan Institute of Engineering, China, in 2019. Her research interests lie in medical image generation using deep learning.



Yuefu Zhan is currently an Associate Chief Physician at Department of Radiology, Hainan Women and Children's Medical Center. He received a Doctor of Medicine degree from the West China Medical School, Sichuan University, China, in 2022, and a Master of Medicine degree from the Xiangya School of Medicine, Central South University, China, in 2010. His current research interests mainly focus on imaging diagnosis, minimally invasive intervention, and AI-based medical image analysis.



Caihua Liu received the B.S. degree in software engineering and the Ph.D. degree in computer application technology from Nankai University, China, in 2010 and 2017, respectively. She is currently a Lecturer with the College of Computer Science and Technology, Civil Aviation University of China. Her research interests include computer vision and machine learning.



Thomas Lukasiewicz is a Professor at Institute of Logic and Computation, Vienna University of Technology (TU Wien), Vienna, Austria, and Department of Computer Science, University of Oxford, UK. He currently holds an AXA Chair grant on "Explainable Artificial Intelligence in Healthcare" and a Turing Fellowship at the Alan Turing Institute, London, UK, which is the UK's National Institute for Data Science and Artificial Intelligence. He received the IJCAI-01 Distinguished Paper Award, the AIJ Prominent Paper Award 2013, the RuleML 2015 Best Paper Award, and the ACM PODS Alberto O. Mendelzon Test-of-Time Award 2019. He is a Fellow of the European Association for Artificial Intelligence (EurAI) since 2020. His research interests are especially in artificial intelligence and machine learning.