PCA: Semi-Supervised Segmentation with Patch Confidence Adversarial Training

Zhenghua Xu, Runhe Yang, Zihang Xu, Shuo Zhang, Yuchen Yang, Weipeng Liu, Weichao Xu, Junyang Chen, Thomas Lukasiewicz, Victor C. M. Leung, *(Life Fellow, IEEE)*

Abstract*—Deep-learning-based semi-supervised learning (SSL) methods have achieved a strong performance in medical image segmentation, which can alleviate doctors' expensive annotation by utilizing a large amount of unlabeled data. Unlike most existing semi-supervised learning methods, adversarial training methods distinguish samples from different sources by learning the data distribution of the segmentation map, leading the segmenter to generate more accurate predictions. We argue that the current performance restrictions for such approaches are the problems of feature extraction and learning preferences. In this paper, we propose a new semi-supervised adversarial method called Patch Confidence Adversarial Training (PCA) for medical image segmentation. The PCA method's discriminator penalizes patch-level structures, guiding the generator to optimize different patch areas, by leveraging pixel context, the generator is driven to focus on high-frequency features, making it harder to deceive the discriminator and easy to converge to an ideal state, which more effectively guides the segmenter to generate high-quality pseudo-labels. Furthermore, at the discriminator's input, we supplement image information constraints, making it simpler to fit the expected data distribution. Extensive experiments on the Automated Cardiac Diagnosis Challenge (ACDC) 2017 dataset and the Brain Tumor Segmentation (BraTS) 2019 challenge dataset show that our method outperforms the state-of-the-art semi-supervised methods, which demonstrates its effectiveness for medical image segmentation.

***Index Terms*—Semi-Supervised Learning, Adversarial Learning, Medical Image Segmentation**

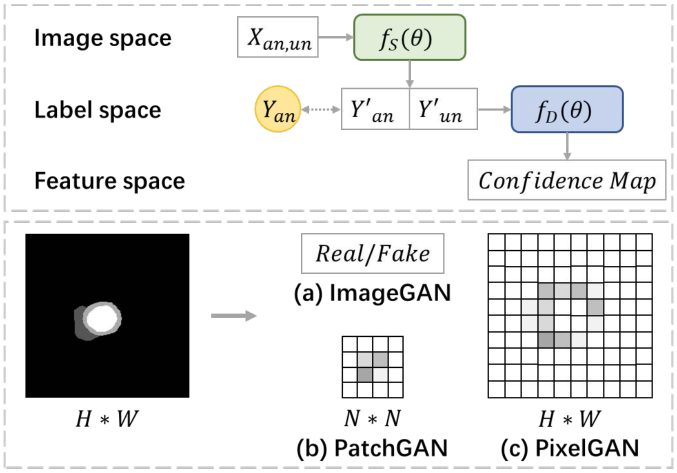


Fig. 1. Different output styles of general adversarial training (ImageGAN), our method (PatchGAN), and confidence-map-based adversarial training (PixelGAN). Where X_{an} and Y'_{an}, X_{un} and Y'_{un} represent the labeled and unlabeled original images and their segmentation predictions, respectively. $f_S(\theta)$ represents the Segmentater, and $f_D(\theta)$ represents the Discriminator.

I. Introduction

Segmentation, identifying interesting regions with anatomical or pathological structures from medical images, is the basic task of medical image analysis, which is of great significance for computer-assisted diagnosis, surgery simulation, and treatment planning. Recently, deep learning methods [1, 2] have achieved excellent results in various medical application fields, which are trained with various typical segmentation networks in a fully supervised way (e.g., FCNs [3], U-Nets [4], and GANs [5]). The success of the deep neural network model is due to its depth and width, which is highly dependent on large-scale and high-quality pixel annotation data. However, the lack of sufficient labeled data has always been a major challenge for medical image segmentation.

To reduce the annotation burden on doctors, many researchers focused on semi-supervised learning method, which can leverage a limited amount of labeled data and a large amount of unlabeled data to improve the accuracy of segmentation. According to the training manner, common semi-supervised segmentation methods can be divided into the following categories: self-training [6]leverages the model's own predictions on unlabeled data as pseudo-labels for further training; consistency regularization [7, 8, 9]encourages the model to produce consistent predictions when small pertur-

This work was supported by the National Natural Science Foundation of China under the grants 62276089, 62027813, 82205314 and 62102265, by the Natural Science Foundation of Hebei Province, China, under the grant F2024202064, by the Ministry of Human Resources and Social Security, China, under the grant RSTH-2023-135-1, by the Shenzhen Talents Special Project - Guangdong Provincial Innovation and Entrepreneurship Team Supporting Project, under the grant 2021344612, by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), under the grant GML-KF-22-29, and by the Natural Science Foundation of Guangdong Province, China, under the grant 2022A1515011474. This work was also supported by the AXA Research Fund. (Corresponding author: Zihang Xu, e-mail: xuzihang@connect.hku.hk; Yuchen Yang, Email: yyc174276@gmail.com; Weipeng Liu, email: liuweipeng@hebut.edu.cn.)

Zhenghua Xu, Runhe Yang and Shuo Zhang are with the State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China.

Zihang Xu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong.

Yuchen Yang is with the Department of Applied Mathematics and Statistic, The Johns Hopkins University, Baltimore, United States.

Weipeng Liu is with the School of Artificial Intelligence, Hebei University of Technology, Tianjin, China.

Weichao Xu is with the Digestive Disease Center, Hebei Provincial Hospital of Traditional Chinese Medicine, Shijiazhuang, China.

Junyang Chen and Victor C. M. Leung are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Thomas Lukasiewicz is with Department of Computer Science, University of Oxford, United Kingdom and the institute of Logic and Computation, Vienna University of Technology, Vienna, Austria.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2025.3548416

bations are applied to the input or the model itself; co-training [10]involves training multiple models simultaneously, with each model learning from the other's predictions on the unlabeled data; and adversarial training [11], which this work focuses on, is particularly advantageous due to its ease of implementation, effectiveness, especially when labeled data is scarce, further introduction are provided below:

Recently, many researchers have focused on semi-supervised segmentation methods based on adversarial training [12, 13, 14, 15, 16, 17, 18], which show a great potential for improving semantic segmentation. Specifically, these methods are inspired by Generative Adversarial Networks (GANs) [5], which consist of a generator and a discriminator. The generator network uses a semantic segmentation network to generate the probability maps of the semantic labels. The discriminator network generates the probability that the input is real by distinguishing generated samples from target ones, as illustrated in the upper figure of Figure 1. Through the discriminator, the quality of the predicted segmentation map can be effectively assessed, the output forms of the general methods can all be summarized in Figure 1 (a). However, those model lack stability due to the insufficient supervision of the discriminator, which brings challenges to semi-supervised adversarial training. Some recent studies have attempted to improve the performance of the discriminator [12, 13, 14]. For example, Son *et al.* [12] connected prediction, inverse prediction, and image, forcing the discriminator to learn the mapping relationship between image and prediction, thus alleviating this problem. Zhang *et al.* [13] multiplied the prediction with the image to obtain an image containing only foreground information. Inspired by the attention mechanism [19], Han *et al.* [14] proposed a dual-attentive-fusion block that has two independent spatial attention paths on the predicted segmentation map and leverages the corresponding original image. Nie *et al.* [16] introduced the concept of a confidence map to supervise the learning of unannotated data. The output forms of the above methods can all be summarized in Figure 1 (c).

However, all of the above methods have the following two shortcomings: (i) *Feature mining problem*: For the discriminator, its primary purpose is to determine whether the generated sample is real or not, with its output representing the probability that the input is genuine. However, when dealing with segmentations generated from unlabeled data, treating all pixels as part of a negative class overlooks the presence of positive class pixels within the unlabeled data, which impedes the discriminator's ability to effectively identify and differentiate features between the two data domains. When the discriminator's output is a pixel-level confidence map rather than a scalar result—i.e., it classifies each pixel individually, this method introduces a new challenge: the generator can manipulate the segmentation maps to reduce information entropy, thereby misleading the discriminator, which prevents the discriminator from extracting sufficient features to generate reliable confidence maps that accurately reflect the probability of the segmentation being correct. This motivates us to consider how to generate more reliable evaluation results. (ii) *Learning preference problem*: The original image directly concatenated or multiplied with the segmentation map as the input of the discriminator not only introduces the relevant information of the segmentation map and the input image but also makes the discriminator generate learning preferences. The reason is that the discriminator may focus more on easy-to-learn features instead of truly distinguishing between real and fake segmentations, which causes the discriminator to change the optimization goal of training and affects the accuracy of the segmentation network. From the results, not only the segmentation accuracy has decreased, but also the problem of under-fitting, which is unacceptable in semi-supervised medical image segmentation and may be difficult to apply to clinical practice.

Inspired by these, we propose a patch confidence adversarial training framework for semi-supervised medical image segmentation. We improve the original adversarial training strategy from the input and output of the discriminator, addressing the two problems mentioned above, respectively. First, to better evaluate segmentations, we introduce the idea of PatchGAN [20], and the output forms of our method can be summarized in Figure 1 (b). The discriminator classifies each patch independently, so that it generates a patch confidence map instead of a scalar classification result or pixel-level confidence map. The discriminator penalizes at the scale of patch-level, compared to a scalar classification result, allows the discriminator to distinguish and mine image features locally, which enhances the generator's ability to capture subtle structures in the image by optimizing each patch, improving the quality of the segmentation maps. Compared to a pixel-level confidence map, and due to the utilization of contextual information, the generator is driven to focus on high-frequency features, making it harder to deceive the discriminator, which aids in its convergence to an optimal state. In addition, to utilize the image information more effectively, we also add the weighted image and the segmentation map. This operation balances the information contained in the image and segmentation maps. In that case, the discriminator will focus on the relationship between the image and the segmentation, learning how to map unlabeled data to an expected distribution. In adversarial training, improving the performance of the discriminator will help the deep model achieve a better segmentation performance.

The major contributions of this paper can be summarized as follows:

- We propose a novel and universal semi-supervised method, called PCA, for semi-supervised medical image segmentation tasks. The proposed method effectively addresses two important shortfalls, including the feature mining problem and the learning preference problem.
- Inspired by CGAN [21] and PatchGAN, we propose the patch confidence map and pixel-additive blending, making the discriminator easily converge to a desirable status and guiding the prediction of the unlabeled data to fit the expected data distribution, which can significantly improve the performance of semi-supervised segmentation.
- A large number of experiments on the ACDC 2017 dataset and BraTS 2019 dataset prove that our semi-

supervised method is efficient. We found that the designed patch confidence map is more effective than the most advanced adversarial training strategy. In addition, we achieve the best results compared to other state-of-the-art methods in semi-supervised segmentation.

II. Related Work

As a basic task, medical image segmentation is significant in many biomedical applications [6, 22, 23, 24, 25, 26]. Semi-supervised learning methods have gained widespread attention in the field of medical image processing, demonstrating remarkable performance not only in segmentation tasks [27, 28] but also in object detection [29, 30]. At present, almost all segmentation frameworks are based on deep learning, which has achieved impressive performance improvements on various medical image segmentation tasks and set the new state of the art [31, 32, 33, 34]. However, there is a lack of a large amount of data annotation in the medical field, which limits the performance of the model. We now focus on reviewing related methods and the latest developments in semi-supervised learning, and then discuss semi-supervised methods based on adversarial training, which are most relevant to our work.

Semi-Supervised Segmentation. In semi-supervised image segmentation tasks [27, 28], it is usually assumed that only a small part of the training images have complete pixel-level annotations, but there are also a large number of unlabeled images that can be used to improve the performance of the model. Since unlabeled data do not need to be manually labeled by the doctor, unlabeled data can be used at a low cost to improve performance. The main challenge in this scenario is how to effectively use a large amount of unlabeled data, which is also the main difference between the different methods.

Recently, almost all semi-supervised medical image segmentation frameworks are based on deep learning. For example, Bai et al. [6] proposed an iterative method for heart segmentation of MR images, using pseudo-labels generated by network prediction to update network parameters. Feng *et al.* [22] improved the training strategy based on the work of Bai *et al.* [6]. Only part of the reasonable segmentation maps predicted from the unannotated samples was progressively combined with the annotated samples to improve the training procedure. Zhu *et al.* [35] proposed a fast semi-supervised video object segmentation method which improves the input to the decoder by feature matching and separable structure modeling to improve the segmentation accuracy. Liu *et al.* [36] proposed a guided co-segmentation network for online semi-supervised object segmentation. Inspired by the mean teacher model, Yu *et al.* [37] proposed a semi-supervised framework for uncertainty perception to segment the left atrium from 3D MR images. Cao *et al.* [38] extended the uncertain time model for semi-supervised ABUS quality segmentation. However, such methods do not consider the quality of the predicted segmentation map, which may introduce misinformation into the segmentation network. Adversarial training introduces a discrimination network to evaluate the predicted segmentation map, which is simple and effective for semi-supervised medical image segmentation.

Adversarial Training for Semi-Supervised Segmentation. Generative Adversarial Networks (GANs) were proposed by Goodfellow *et al.* [5]. They generate samples by optimizing the adversarial game between the discriminator and the generator. When applying adversarial learning to semi-supervised segmentation tasks, the model usually uses two networks: a segmentation network (generator) for image segmentation and a discriminant network (discriminator) to identify whether the sample is extracted from real data or generated by the generator. Generally, there are three popular strategies in the medical image analysis community based on adversarial training and semi-supervised methods, including generative models, confidence maps, and segmentation evaluation. For instance, Sedai *et al.* [39] introduced a generative model (VAE [40]) for Optic Cup (OC) segmentation from retinal fundus images. They use a VAE to learn a feature embedding from unlabeled images and then combine the feature embedding with a segmentation autoencoder trained on labeled images to perform pixel segmentation on the cup region. Lahiri *et al.* [17] used unannotated images to generate fake samples to increase the size of the training dataset. However, the model lacks stability due to insufficient supervision of the discriminator.

To improve the performance of the discriminator, a simple strategy is inspired by conditional GANs, where the discriminator is conditioned on the input image to classify whether the segmentation map is real or fake. we discuss several methods for combining the original image enhancement segmentation results. Son *et al.* [12] connected prediction, inverse prediction, and image, and distinguished good or poor segmentation results by finding the mapping relationship between image and prediction. Since the discrimination network has separate model parameters for handling information from the segmentation maps and from the original image, the discriminator may focus on learning the features of the original image to give a judgment score, which leads to the learning preference problem. Zhang *et al.* [13] multiplied the prediction with the image to judge directly whether the target is accurately segmented. The element-wise multiplication method forces the discriminator to learn about the association information between the segmentation maps and the original images, which ensures that the segmentation maps are meaningful in adversarial training. However, element-wise multiplication ignores the correlation between the background and the target. The discriminator will focus on the distinction of some uncomplicated categories, which is not conducive to improving its ability. Han *et al.* [14] introduced a dual attention fusion block based on connection, extracting geometric level and intensity level information, thus digging for more relevant features. However, the attention method does not deal with all the details of the image accurately, introducing a lot of noisy information, which is usually fatal for medical image segmentation tasks.

The above-mentioned works explore the correlations between the segmentation and input image for evaluating the segmentation quality. However, regardless of whether the segmentation and the corresponding image are related or not, it is unreasonable to regard all pixels in the generated samples as negative samples, which is too abstract for the discriminator. As a consequence, the generated evaluation score does not reflect the correct probability of the segmentation result, and the discriminator cannot contribute to instructing the segmentation task.

Nie *et al.* [15] adopted another strategy, using a discriminator to output the confidence map and choose high-confidence regions to obtain the ground truth, then updating the segmentation network in a self-learning manner. The main limitation of this method is that a confidence threshold must be provided, the value of which will affect performance. Furthermore, if the discriminator cannot distinguish between good and bad segmentation, poor confidence maps may reduce the performance of the entire adversarial training. In particular, for medical images, the presence of speckle noise may affect the confidence map. The unstable confidence map may also affect the entire learning process and lead to unsatisfactory segmentation results for unlabeled data.

At the input of the discriminator, we design a new combination of conditional GAN, called pixel additive blending. At the output of the discriminator, differently from the confidence map and single scalar output, we use PatchGAN to output an $N \times N$ array to evaluate the good or bad segmentation results in a region, rather than the entire image or individual pixels. Such improvements effectively solve the feature mining and learning preference problems that exist in other methods.

III. Method

Figure 2 shows a schematic illustration of our patch confidence adversarial training model (PCA) for semi-supervised segmentation. The PCA framework consists of a segmentation network and a discrimination network. First, the segmentation network takes input data and produces the corresponding segmentation probability maps. Then, the segmentation probability map and the image weighted combination are used as the inputs of the conditional discriminator. Subsequently, the discriminator distinguishes the data distribution of new images generated based on different maps, including the ground truth from the annotation data and the segmentation probability maps from all the data. The discriminator tries to classify if each $N \times N$ patch in a segmentation map is real or fake.

A. General Training Strategy

To define the loss function, the symbols utilized are first enlisted. The training set consists of $M+N$ inputs, including M labeled inputs and N unlabeled inputs. Let $A = \{(x_i^a, y_i^a)\}_{i=1}^{M}$ be a labeled set with M samples and $U = \{x_i^u\}_{i=1}^{N}$ be a set with N samples. The segmentation network and the discrimination network are represented by $S(\cdot)$ and $D(\cdot)$.

In our proposed PCA, the segmentation network is trained by minimizing the following loss function L_S:

$$L_S(A, U; \theta_S) = L_{seg} + \lambda_{adv} L_{adv}, \tag{1}$$

where θ_S represents the parameters for the segmentation network, L_{seg} and L_{adv} represent the supervised segmentation and adversarial loss, respectively, and λ_{adv} refers to the weight of adversarial learning.

The loss function L_{seg} is used to determine whether the segmentation probability map generated by the input annotation data is close to ground truth, which is expressed as:

$$L_{seg}(S(x_i^a), y_i^a) = 0.5 * L_{bce} + 0.5 * L_{dice}, \tag{2}$$

$$\begin{aligned} L_{bce}(S(x_i^a), y_i^a) = & -y_i^a \cdot \log(S(x_i^a)) \\ & - (1 - y_i^a) \cdot \log(1 - S(x_i^a))), \end{aligned} \tag{3}$$

$$L_{dice}(S(x_i^a), y_i^a) = 1 - \frac{2 * \sum S(x_i^a) * y_i^a}{\sum S(x_i^a) + \sum y_i^a}, \tag{4}$$

where L_{bce} is constrained by the standard Binary Cross Entropy (BCE) loss function, and L_{dice} is constrained by the standard DICE loss function.

The general semi-supervised adversarial training objective loss function L_{adv} is defined as:

$$\begin{aligned} \min_{\theta_S} \max_{\theta_D} L(\theta_S, \theta_D) = & \mathrm{E}_{x_i \sim P_A}[-\mathrm{log} D(S(x_i))] \\ & + \mathrm{E}_{x_i \sim P_U}[-\mathrm{log}(1 - D(S(x_i)))], \end{aligned} \tag{5}$$

where x_i is the input image from the data distribution P_{A+U}, and θ_D represents the parameters for the discrimination network. As in GANs [5], when updating the segmentation network, we replace $\log(1 - D(x_i))$ by $-\mathrm{log} D(x_i)$.

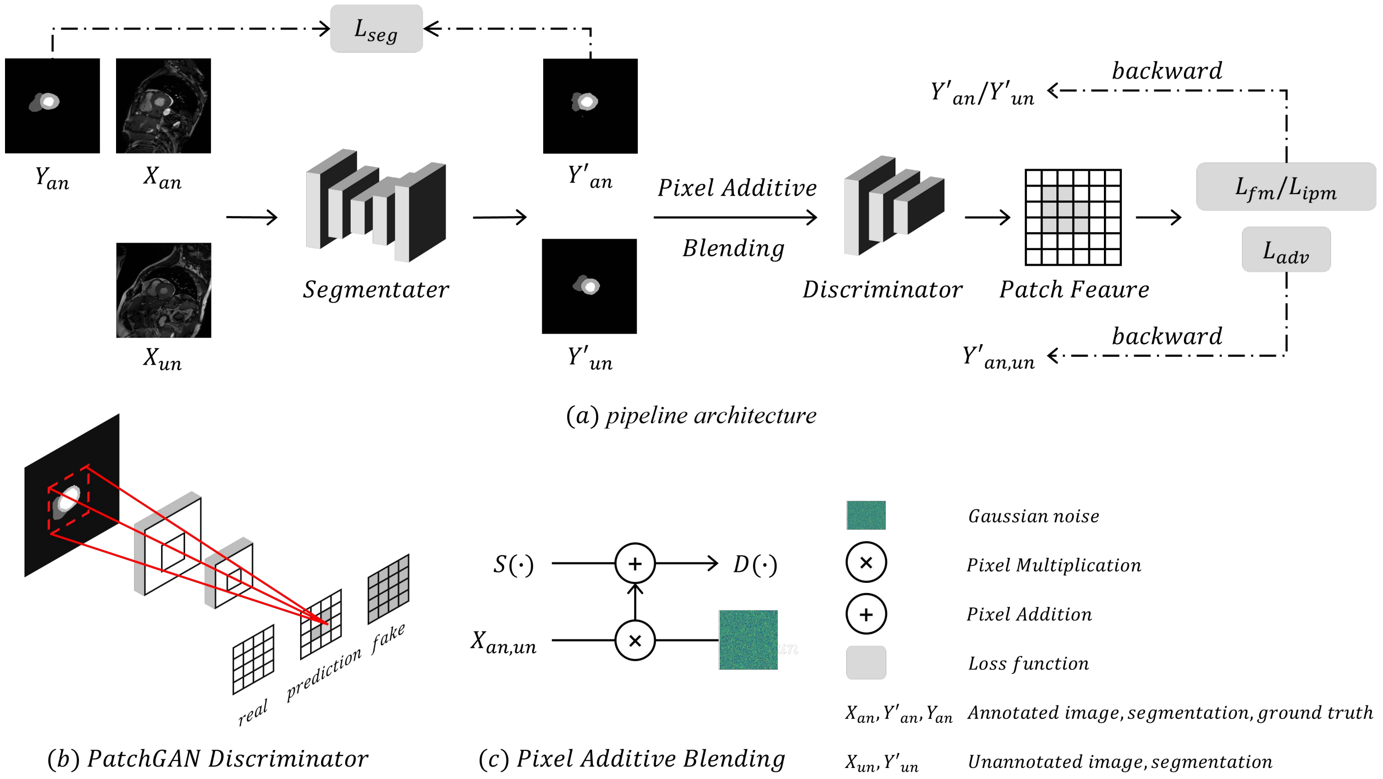
B. Patch Confidence Map

The choice of the discriminator differs due to the different output sizes at which the decision is made. In this work, we explored several models for the discriminators with various output sizes, as done in the previous work, corresponding to several adversarial training strategies. Notably, our $D(\cdot)$ produces an $N \times N$ matrix as a result.

According to the general adversarial training strategy, the image-level discriminator's purpose is to determine whether the input sample is real. The last layer of the general discriminator outputs a scalar value, which is a weighted value of the whole image but cannot reflect the local features of the image. It is hard to train for tasks requiring high precision, such as image segmentation. Especially in semi-supervised segmentation tasks, if the discriminator treats predictions from unlabeled data as negative samples and only learns global features, it will not be easy to mine enough features to guide the segmentator to generate more accurate predictions.

The pixel-level discriminator generates the confidence map to determine whether each pixel in the input sample is real or not. Afterwards, unlabeled data are trained by generating reliable pseudo-label masks from the confidence map. This requires the confidence map to infer sufficiently close regions from the ground-truth distribution to find more confident pixels. However, the confidence map generated by the current method cannot reflect the correct probability of the segmentation result, which will affect the performance of the segmentation network and even cause a negative transfer.

Between the extremes, it is also possible to set the receptive field to a $K \times K$ patch where the decision can be given at the patch-level (PatchGAN). The final output of the patch-based discriminator is a matrix of size $N \times N$. Each of the items in the matrix represents a local region in the image. The patch-based discriminator tries to classify whether each



(a) pipeline architecture

(b) PatchGAN Discriminator

(c) Pixel Additive Blending

Fig. 2. (a) Schematic view of our proposed patch confidence adversarial training model (with a cardiac image as an example). (b) Each value of the output matrix from the PatchGAN discriminator represents the probability of whether the corresponding segmentation patch is real or fake. (c) The process of pixel additive blending calculation, where $S(\cdot)$ represents the Segmentater, $D(\cdot)$ represents the Discriminator.

$N \times N$ patch in the image is real or fake, which achieves the extraction and characterization of local image features and is conducive to the generation of high-precision images. Compared with image-level (ImageGAN) or pixel-level (PixelGAN) adversarial learning, PatchGAN has the ability to capture the local statistics of the output space and guide the segmentation network to focus on the local structure similarity in the image patches. We achieve this variation in patch size by adjusting the depth of the GAN discriminator.

In the early stages of training, the distribution of positive and negative samples may not overlap, and the discriminator network can easily distinguish between them. To stabilize the training of the semi-supervised framework, we directly align the features learned by the discriminator network to the segmentation network and introduce the feature matching [41] loss L_{fm}. It aims to minimize the discrepancy in feature statistics between the generating samples and the ground truth. It is calculated by a mean squared error (MSE) loss:

$$L_{fm}(x_i, y_i) = \mathrm{E}_{x_i, y_i \sim P_A} ||D(S(x_i)) - D(y_i)||^2, \quad (6)$$

where x_i and y_i are sampled from the labeled set A.

For unlabeled data, it is impossible to align ground-truth features directly. Inspired by the integrated probability metric [42], we choose the intermediate layer of the discriminative network as the mapping function and calculate the mean value to measure the distance between the distributions of positive and negative samples. We assume that the two distributions are the same when the feature centres match (i.e., the difference in means is minimal). L_{ipm} is used for unlabeled prediction, forcing the segmentation network to use a reasonable solution without label information:

$$L_{ipm}(x_i, y_i) = ||\mathrm{E}_{x_i \sim P_U} D(S(x_i)) - \mathrm{E}_{y_i \sim P_A} D(y_i)||^2, \quad (7)$$

where x_i represents unlabeled images sampled from unlabeled set U, and y_i represents the ground truth sampled from the labeled set A. The final training objective L_S is as follows:

$$L_S = L_{seg} + \lambda_{adv} L_{adv} + \lambda_{fm1} L_{fm} + \lambda_{ipm} L_{ipm}, \quad (8)$$

where λ_{fm} and λ_{ipm} are the corresponding weights.

C. Pixel-additive Blending

The simplest input form can be just the segmentation probability maps [43, 44, 45], which allow the discriminator to learn useful shape properties of the objects, thereby evaluating the segmentation result quality. However, in this form, the discriminator has weaker discriminability, which can easily cause an overfitting problem.

In conditional GANs (CGANs), the discrimination model explores the relationship between the segmentation probability map and the image, to enhance the discriminative ability. In that case, the discriminator might generate the probability values by distinguishing between labeled and unlabeled images without learning segmentation information. So, the key is how to encode the correlation information to construct the effective input of the discriminator. The methods for combining the original image enhancement segmentation results such as concatenated method [12], Element-wise multiplication method [13] and Element-wise multiplication method[13] is

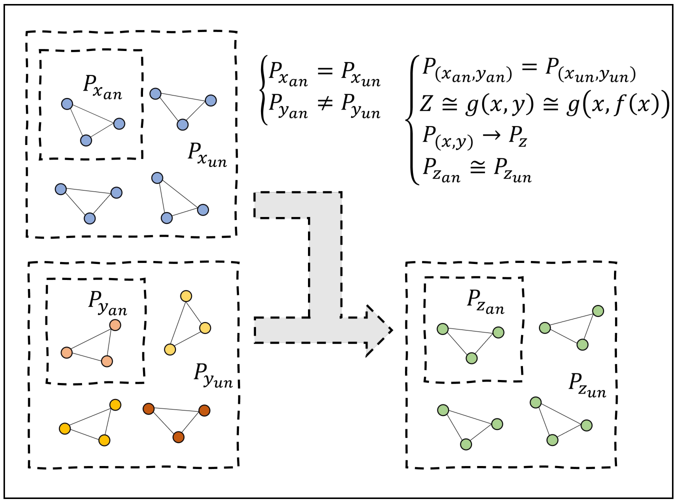

$P_{x_{an}}$
$P_{x_{un}}$
$P_{y_{an}}$
$P_{y_{un}}$
$P_{z_{an}}$
$P_{z_{un}}$
$\begin{cases} P_{x_{an}} = P_{x_{un}} \\ P_{y_{an}} \neq P_{y_{un}} \end{cases}$
$\begin{cases} P_{(x_{an}, y_{an})} = P_{(x_{un}, y_{un})} \\ Z \cong g(x,y) \cong g(x, f(x)) \\ P_{(x,y)} \to P_Z \\ P_{z_{an}} \cong P_{z_{un}} \end{cases}$

Fig. 3. Using the condition that the labeled data and unlabeled data are identically distributed, the input and output are coupled, and the obtained joint probability is approximately identically distributed. Where P represents the joint probability distribution of annotated and unannotated data, Z represents pixel additive mixture (refer to equation 10), and $g(\cdot)$ represents the generator.

is unreasonable, because they regard all pixels in the generated samples as negative samples, which is abstract for the discriminator.

To alleviate the influence of learning preferences on network optimization and to improve the feature mining ability, we introduce a novel conditional GAN, called pixel additive blending, where the discriminator is conditioned on the input image x. The main idea is to fit the joint probability of the image and the segmentation to narrow the gap between the labeled data pair and the unlabeled data pair, thereby helping the discriminator to better learn the correlation between the image and the segmentation. This is because semi-supervised learning assumes that the labeled data and unlabeled data samples belong to the same distribution, but there is a gap between their corresponding segmentations. If the joint probability distribution of the image and the segmentation can be fitted, then the joint probability distribution of the unlabeled data will be the same as that of the labeled data. Figure 3 shows the process information.

Our solution is to assign Gaussian noise with a small value to the image and add it to the segmentation map pixel by pixel. Note that here we use the segmentator's predictions as negative samples and the ground truth of the labeled data as positive samples. It is to avoid the discriminator to complete the classification task by distinguishing between labeled and unlabeled images, which will deviate from the original purpose. Thus, the objective function becomes:

$$\begin{aligned}\min_{\theta_S}\max_{\theta_D} L(\theta_S, \theta_D) &= \mathrm{E}_{x_i, y_i \sim P_A}[-\log D(Z(x_i, y_i))] \\ &+ \mathrm{E}_{x_i \sim P_{A+U}}[-\log(1 - D(Z(x_i, S(x_i))))].\end{aligned} \tag{9}$$

Particularly, the pixel additive blending $Z(\cdot)$ is expressed as follows:

$$Z(x_i, y_i) = y_i + \lambda_{noise} x_i \cdot noise, \tag{10}$$

where λ_{noise} and $noise$ are the weight coefficient of the image and Gaussian noise, respectively. All $D(x)$ are replaced by $D(Z(x_i, y_i))$ when introducing the pixel-additive blending.

There are three reasons for this: Firstly, pixel-addition ensures both the segmentation map and the image are used in the discriminator's decision-making and in the adversarial training process. Also, the randomness of Gaussian noise reduces the correlation between image pixels, forcing the discriminator to learn the correlation between segmentation and image. In addition, we believe that the output is less complex than the input in the semi-supervised segmentation task. That means the segmentation map may have very little influence on both the decision-making process of the discriminator and the parameter updates of the segmentator. To reduce this bias, we must limit the influence of image information. Here, we assign a small weight to the image.

D. Network Architecture

For the segmentation network, we follow the spirit of U-Net, where skip connections are added between the down-sampling path and the up-sampling path to save low-level information. These skip connections are crucial to segmentation tasks, as the initial feature maps maintain low-level features such as edges and blobs that can be properly exploited for accurate segmentation. In addition, the U-Net model utilizes bi-linear interpolation to expand the feature maps.

The proposed discrimination network generally follows the architecture of PatchGAN [20]. The network contains three convolutional layers with a kernel size of 4×4 and a stride of 2×2. The channel numbers of the three convolutional layers are 32, 64, and 1, respectively. The activation function following each convolutional layer is LeakyReLU with an alpha value of 0.2, except for the last one using the sigmoid function. The output size $m \times n$ of the patch-based discriminator is 32×32, in which one pixel corresponds to a patch of size 22×22 in the input probability maps. Each patch is classified into real (1) or fake (0) through the discriminator. We employ this adversarial learning strategy to force each generated patch in the prediction of unlabeled data to be similar to the patch of labeled data.

IV. Experiments

A. Datasets

We evaluate our proposed semi-supervised segmentation method on two typical medical images, including cardiac images and brain tumor images.

Cardiac image dataset. This dataset consists of 200 MRI scans from 100 patients [52] for training and 50 patients for testing. Three cardiac regions are labeled in the ground truth: Left Ventricle (LV), Right Ventricle (RV), and Myocardium (Myo). For a fair comparison, we only select the training set in our experiments and divide sets of 70, 10, and 20 for training, validation, and testing. Slices within 3DMRI scans are considered 2D images, which are fed as input to the network.

Brain tumor image dataset. The brain tumor image dataset comes from the Brain Tumor Segmentation (BraTS) 2019 challenge. The released BraTS dataset contains 335 3D cases. The experiment of whole brain tumor segmentation was performed using the multi-modal MRI data from the BraTS 2019

TABLE I
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE ACDC2017 DATASET.

Method	Scans used		Area Overlap				Boundary Error			
	An.	Un.	DSC(%)↑	Imp.	JA(%)↑	Imp.	95HD(voxel)↓	Imp.	ASD(voxel)↓	Imp.
Baseline	7(10%)	0	77.34	-	66.20	-	9.18	-	2.45	-
GAN [12]	7(10%)	63(90%)	83.28	5.94%	72.84	6.64%	9.84	-0.66	2.78	-0.33
DAN [13]			79.33	1.99%	67.93	1.73%	11.85	-2.67	3.08	-0.63
BUS-GAN [14]			81.80	4.46%	71.20	5.00%	6.11	3.07	2.21	0.24
ASDNet [16]			82.89	5.55%	71.98	5.78%	16.50	-7.32	4.13	-1.68
SASSNet [15]			84.14	6.80%	74.09	7.89%	5.03	4.15	1.40	1.05
MT [9]			80.40	3.06%	69.28	3.08%	10.05	-0.87	2.65	-0.20
ICT [8]			83.54	6.20%	72.84	6.64%	7.58	1.60	2.27	0.18
UAMT [37]			81.58	4.24%	70.48	4.28%	12.35	-3.17	3.62	-1.17
CCT [46]			83.34	6.00%	72.84	6.64%	7.07	2.11	2.18	0.27
DTC [47]			82.71	5.37%	72.14	5.94%	11.31	-2.13	2.99	-0.54
CPS [48]			85.32	7.98%	75.42	9.22%	6.64	2.54	1.98	0.47
UPRC [49]			81.77	4.43%	70.85	4.65%	5.04	4.14	1.41	1.04
CLB [50]			82.12	4.78%	71.26	5.06%	9.19	-0.01	2.80	-0.35
BCP [51]			85.79	8.45%	75.97	9.77%	10.18	-1.00	2.46	-0.01
PCA(w/o CGAN)			86.39	9.05%	76.91	10.71%	4.63	4.55	**1.20**	**1.25**
PCA(w/ CGAN)			**87.33**	**9.99%**	**78.31**	**12.11%**	**3.10**	**6.08**	1.35	1.10
Baseline	14(20%)	0	83.69	-	74.00	-	6.63	-	1.74	-
GAN [12]	14(20%)	56(80%)	84.52	0.83%	74.92	0.92%	10.51	-3.88	2.64	-0.90
DAN [13]			86.18	2.49%	76.71	2.71%	9.23	-2.60	2.38	-0.64
BUS-GAN [14]			85.01	1.32%	75.83	1.83%	6.64	-0.01	1.76	-0.02
ASDNet [16]			84.18	0.49%	74.01	0.01%	6.68	-0.05	2.10	-0.36
SASSNet [15]			85.99	2.30%	76.63	2.63%	5.32	1.31	1.47	0.27
MT [9]			85.58	1.89%	76.38	2.38%	4.89	1.74	1.60	0.14
ICT [8]			85.25	1.56%	75.71	1.71%	7.66	-1.03	2.32	-0.58
UAMT [37]			85.87	2.18%	76.78	2.78%	5.06	1.57	1.54	0.20
CCT [46]			86.09	2.40%	77.05	3.05%	7.01	-0.38	1.98	-0.24
DTC [47]			86.28	2.59%	77.03	3.03%	6.14	0.49	2.11	-0.37
CPS [48]			87.38	3.69%	78.61	4.61%	6.06	0.57	1.69	0.05
UPRC [49]			85.07	1.38%	75.61	1.61%	6.26	0.37	1.77	-0.03
CLB [50]			85.98	2.29%	76.68	2.68%	5.96	0.67	1.69	0.05
BCP [51]			87.81	4.12%	78.93	4.93%	5.31	1.32	1.55	0.19
PCA(w/o CGAN)			87.86	4.17%	79.13	5.13%	5.10	1.53	1.51	0.23
PCA(w/ CGAN)			**88.09**	**4.40%**	**79.44**	**5.44%**	**2.91**	**3.72**	**0.98**	**0.76**
Upper bound	70(100%)	0	91.65	-	84.93	-	1.89	-	0.56	-

challenge [53, 54, 55]. The entire dataset contains multi-institutional preoperative MRI of 335 glioma patients (259 HGG and 76 LGG), where each patient has four modalities of MRI scans with neuroradiologist-examined pixel-wise labels. Here, we use the T1-CE modal of HGG for whole tumor segmentation, since this modality can better manifest malignant tumors. In our experiments, the MRI scans are normalized to zero mean and unit variance. We randomly selecte 207 samples for training and 52 for validation. We slice 3D images into 2D images and crop them to 160×160.

B. Implementation Details

We implemente our framework in PyTorch with 2 Nvidia 2080Ti GPUs. On the ACDC dataset, all settings follow the public benchmark Luo *et al.* [56] for fair comparisons. We normalize the samples as a 0–1 range and used data augmentation for the random rotation and flip operations. All the 2D patches are interpolated to a size of 256×256 and randomly extracted. The batch size is set at 24, including 12 labeled samples and 12 unlabeled samples. The model is trained via 30K iterations. To train the segmentation network and discrimination network, an SGD and an Adam optimizer are employed to minimize l_s and l_d. The initial learning rates are set to 1e-2 and 1e-4, respectively, and decayed according to the equation $lr = (1 - \frac{iterations}{iterartions_{total}})^{0.9}$. During the testing time, we also interpolate the output results to 256×256 as input and then restore them to their original size. For the brain tumor image dataset, we normalize the samples as zero mean and unit variance. The batch size is set at 30. The initial learning rates were set at 2e-2 and 1e-4. The model is trained over 100 epochs. After obtaining the segmentation probability map from the segmentation network, we apply thresholding with 0.5 to generate a binary segmentation result. λ_{adv},λ_{fea},λ_{ipm}, and λ_{noise} were set to 0.1, 1, 0.1, and 0.001.

C. Evaluation Metrics

Following Luo *et al.* [56], we adopt four metrics, including Dice Similarity Coefficient (DSC), Jaccard (JA), the 95% Hausdorff Distance (95HD), and the average surface distance (ASD). DSC and JA are employed to examine the overlap areas between the two comparisons. 95HD and ASD are exploited to measure the Euclidean distance between a computer-identified lesion boundary and the boundary determined by physicians. Higher DSC and JA, along with lower 95HD and ASD, correspond to the higher similarity between the two compared regions.

The definitions of DSC and Jaccard are as follows: DSC $= 2 * TP/(FP + 2 * TP + FN)$, Jaccard $= TP/(TP + FN + FP)$.

TP is the number of true positives, where the label is positive and the prediction is also positive. TN is the number of true negatives, where the label is negative and the prediction is also negative. FP is the number of false positives, where the label is negative but the prediction is positive. FN is the number of false negatives, where the label is positive but the prediction is negative.

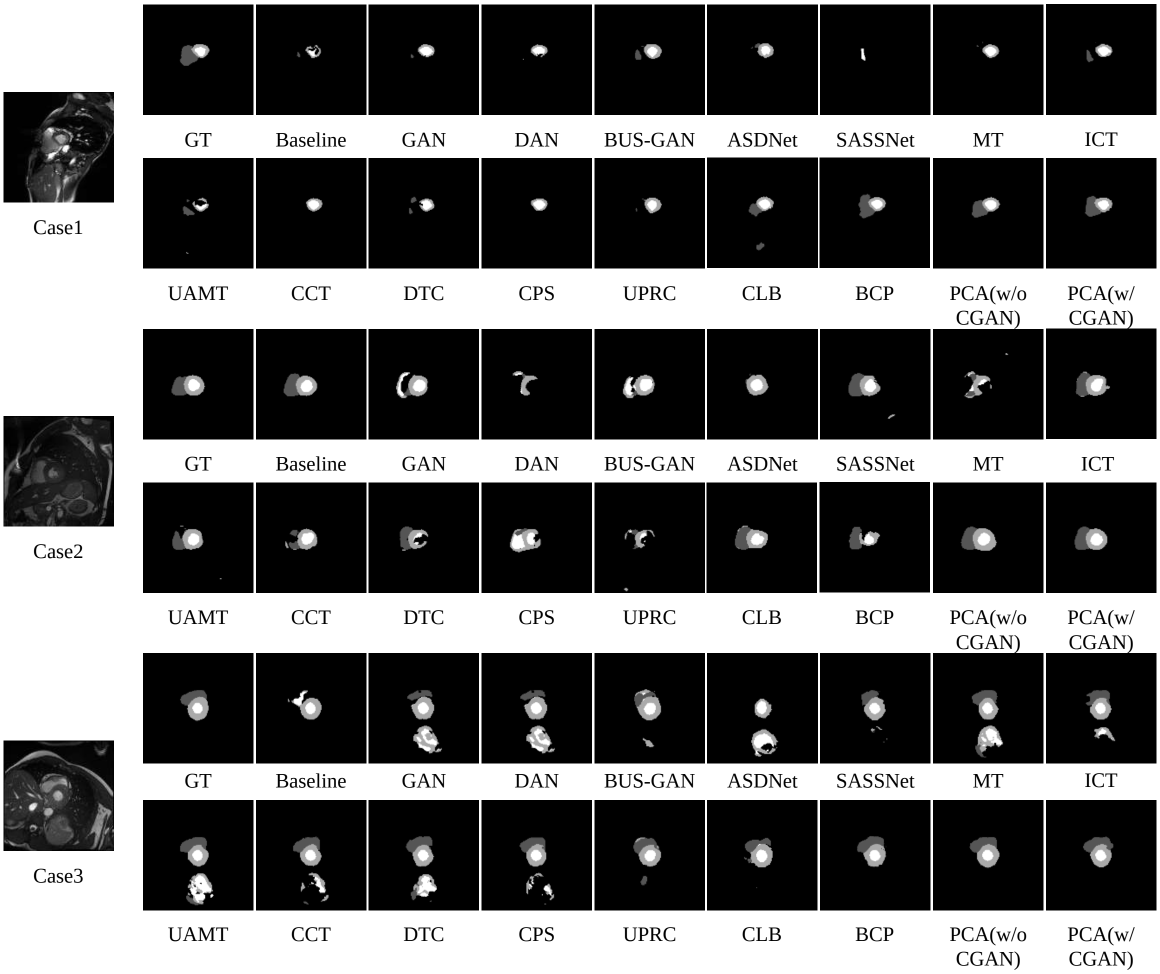


Fig. 4. Visualization of different semi-supervised segmentation methods under 10% labeled data on the ACDC2017 dataset, where GT is the ground truth.

The 95th percentile Hausdorff Distance (HD95) is a statistical measure of the maximum minimum distance between two sets of points. It is defined as:

$$\begin{aligned} HD_{95}(A,B) = \text{percentile}_{95}(\max\Big\{ &\sup_{a\in A}\inf_{b\in B}\|a-b\|, \\ &\sup_{b\in B}\inf_{a\in A}\|b-a\|\Big\}) \end{aligned} \tag{11}$$

where A and B represent the predicted results and the ground truth labels, respectively. The 95th percentile of the Hausdorff Distance measures the value within which 95% of the maximum errors fall, making it less sensitive to outliers.

The formula for Average Surface Distance (ASD) is as follows:

$$\text{ASD} = \frac{1}{|S_A|+|S_B|}\left(\sum_{x\in S_A} d(x,S_B) + \sum_{y\in S_B} d(y,S_A)\right)$$

Where S_A and S_B are the sets of boundary points of the segmentation result and the ground truth, respectively, and $d(x,S_B)$ denotes the shortest distance from point x to S_B.

Those valuation metrics are particularly useful in evaluating results in tasks such as medical image segmentation. Additionally, in all the experimental tables, we present the improvements relative to the baseline, denoted by ”imp.”

D. Experiments on Cardiac Image Data Set

1) Comparison with other semi-supervised methods: We compare our method with the latest semi-supervised learning methods, including GAN-based methods [12, 13, 14, 16, 15] and current state-of-the-art semi-supervised methods [37, 46, 47, 48, 49, 8, 9]. We implement these methods and conduct comparative experiments on the public dataset ACDC 2017. The reported results in Table I are the average performance of four classes on the test set. In each semi-supervised setting, we list the performance of the fully-supervised baseline, adversarial training methods, the latest semi-supervised methods, and our methods in turn. We train U-Net with 100%, 20%, and 10% of training data as upper bounds and the two baselines. The GAN-based methods perform better compared to the baseline overall, showing the effectiveness of GAN-based methods for semi-supervised segmentation. However, we can notice that the performance of ASDNet and BUS-GAN is similar to that of GAN, indicating that the application of the confidence map and segmentation evaluation are still challenging for semi-supervised segmentation. Compared to GAN-based methods, consistency-based methods achieve a

TABLE II
DISCUSSION OF DIFFERENT PROPORTIONS OF DATA TO TRAIN OUR MODEL ON THE ACDC2017 DATASET.

Label/Unlabel	Model	Area Overlap				Boundary Error			
		DSC(%)↑	Imp.	JA(%)↑	Imp.	95HD(voxel)↓	Imp.	ASD(voxel)↓	Imp.
1/67	Baseline	38.26	-	27.66	-	62.76	-	31.32	-
	PCA(w/o CGAN)	47.36	9.10%	35.42	7.76%	49.90	12.86	23.02	8.30
	PCA(w/ CGAN)	**56.11**	**17.84%**	**43.4**	**15.74%**	**45.47**	**17.29**	**17.20**	**14.12**
2/67	Baseline	44.66	-	33.16	-	48.94	-	23.37	-
	PCA(w/o CGAN)	52.18	7.58%	40.72	7.56%	15.98	32.96	8.92	14.45
	PCA(w/ CGAN)	**60.66**	**16.06%**	**48.12**	**14.96%**	**14.58**	**34.36**	**7.02**	**16.35**
3/67	Baseline	48.09	-	36.54	-	50.04	-	21.03	-
	PCA(w/o CGAN)	56.98	8.89%	46.43	9.89%	**14.15**	**35.89**	7.51	13.52
	PCA(w/ CGAN)	**66.15**	**18.06%**	**54.70**	**18.16%**	15.22	22.74	**5.42**	**15.61**
7/63	Baseline	77.34	-	66.20	-	9.18	-	2.45	-
	PCA(w/o CGAN)	86.39	9.05%	76.91	10.71%	4.63	4.55	**1.20**	**1.25**
	PCA(w/ CGAN)	**87.33**	**9.99%**	**78.31**	**12.11%**	**3.10**	**6.08**	1.35	1.10
14/56	Baseline	83.69	-	74.00	-	7.31	-	2.35	-
	PCA(w/o CGAN)	87.86	4.17%	79.13	5.13%	5.10	2.21	1.51	0.84
	PCA(w/ CGAN)	**88.09**	**4.40%**	**79.44**	**5.44%**	**2.91**	**4.40**	**0.98**	**1.37**
70/0	Baseline	91.65	-	84.93	-	1.89	-	0.56	-

TABLE III
ABLATION STUDIES OF OUR PROPOSED METHODS ON THE ACDC2017 DATASET.

Setting					Scans used		Metrics	
UNet	Im.	Pa.	Pi.	CGAN	An.	Un.	DSC_Mean	Imp.
✓					3(4%)	67(96%)	48.09	-
✓	✓						54.43	6.34%
✓		✓					56.98	8.89%
✓			✓				54.49	6.40%
✓	✓			✓			63.81	15.72%
✓		✓		✓			**66.15**	**18.06%**

comparable performance, demonstrating them effectively utilizing unlabeled data. For example, under the 10% setting, UAMT achieved a 4.24% improvement in the DSC indicator. Obviously, compared with other methods, the model that we proposed obtained the best quantitative results for DSC and JA performance in each semi-supervised setting. Specifically, we achieve an average DSC and JA improvement of 9.99% and 12.11% or 4.40% and 5.44% than the fully-supervised baseline trained with 10% or 20% labeled data. In addition, by exploiting the unlabeled data effectively, our model almost always obtains the lowest standard values of boundary error (i.e., 95HD and ASD). Figure 4 shows the visualization of all methods. We can see that these methods often perform well for lesions. However, for blurry lesions, our method can predict them better than the other methods. Compared with the most advanced semi-supervised methods, our results can describe the target boundary more accurately and retain more details, which also proves the effectiveness of our method.

2) Different ratio of labeled data: To further verify the influence of different percentages of labeled data on performance for our method, a different number of labeled images were selected from the training set. We choose five settings of 1, 2, 3, 7, and 14 cases, which are 1%, 3%, 4%, 10%, and 20% of labeled training data. Table II shows the segmentation results using different numbers of labeled and unlabeled images. All PCA models perform better than the corresponding supervision methods, which shows that our method effectively utilizes unlabeled data and promotes performance. Our method can achieve a substantial surpass of 2.4% DSC with 10% labeled scans compared with the baseline with 20% labeled scans, demonstrating the significant advantage of our method under a small-scale labeled dataset. When the number of labeled data is small (i.e., labeled data = 3), our method obtains a substantial increase (18.06% DSC and 18.16% JA) in accuracy over the fully supervised method (48.09% DSC and 36.54% JA). Furthermore, our method achieves a large reduction (about 70% 95HD and 75% ASD) in boundary error. Both improvements indicate that our proposed method has a broad potential for further clinical applications. It can also be noticed that the performance of all methods increases slowly with the increase of labeled data, which illustrates that the performance of models tends to converge as labeled data increases.

3) Ablation study: We propose an ablation study to measure the contribution of different components of the method, and the results are shown in Table III. The abbreviations "Im.", "Pa.", and "Pi." stand for ImageGAN, PatchGAN, and PixelGAN. Our framework contains two main components: CGAN and PatchGAN. To investigate the effectiveness of each component, we performed an ablation study by adding the two components to the baseline one by one. In addition, we also explore the impact of a variety of discriminators, including PixelGAN and ImageGAN. The experiments are conducted in the setting of three labeled datasets. The results of different settings are presented in Table III. The first line is the supervised baseline model, which was trained with only 3 labels. First, we add the unlabeled data and the adversarial loss. Our method achieves the best results (56.98% DSC) among the three, surpassing the baseline by 8.89%. Other adversarial methods are also better than the baseline, showing the effectiveness of the GAN framework. Then, we explore the influence of the CGAN strategy. From the results, we can observe that conditional image constraints significantly improve segmentation performance by 63.81%. We believe that this is because the image-segmentation association plays a key role in the few annotations. Finally, with the joint learning of CGAN and PatchGAN, the performance of our framework is further promoted to the state of the art, surpassing the supervised baseline by 18.06% DSC. It is observed that applying conditional image constraints alone contributes more to the model's performance.

TABLE IV
COMPARISONS WITH STATE-OF-THE-ART METHODS ON THE BRATS2019 DATASET.

Method	Scans used		Metrics							
	An.	Un.	DSC_WT(%)	Imp.	DSC_TC(%)	Imp.	DSC_ET(%)	Imp.	DSC_Mean(%)	Imp.
Baseline	11(5%)	0	38.84	-	35.69	-	37.47	-	37.33	-
GAN[12]	11(5%)	196(95%)	46.50	7.66%	52.85	17.16%	50.77	13.30%	50.04	12.71%
DAN [13]			42.50	3.66%	47.92	12.23%	47.07	9.60%	45.83	8.50%
BUS-GAN [14]			40.25	1.41%	44.64	8.95%	44.95	7.48%	43.28	5.95%
ASDNet [16]			39.22	0.38%	42.61	6.92%	41.98	4.51%	41.27	3.94%
SSASNet [15]			39.45	0.61%	49.12	4.49%	42.37	4.90%	40.67	3.33%
MT[9]			40.27	1.43%	43.43	7.74%	44.10	6.63%	42.60	5.27%
ICT [8]			41.64	2.80%	45.32	9.63%	46.13	8.66%	44.36	7.03%
UAMT [37]			39.62	0.78%	43.77	8.08%	44.53	7.06%	42.64	5.31%
CCT [46]			42.29	3.45%	46.50	10.81%	46.22	8.75%	45.00	7.67%
DTC [47]			36.72	-2.12%	42.73	7.04%	42.98	5.51%	40.81	3.48%
CPS [48]			40.03	1.19%	47.00	11.31%	48.00	10.53%	45.01	7.68%
UPRC [49]			37.81	-1.03%	42.30	6.61%	43.17	5.70%	41.09	3.76%
CLB [50]			45.57	6.73%	53.43	17.74%	51.32	13.85%	50.11	12.78%
BCP [51]			45.62	6.78%	55.77	20.08%	52.89	15.42%	51.43	14.10%
PCA(w/o CGAN)			**49.80**	**10.96%**	57.61	21.92%	**55.40**	**17.93%**	**54.27**	**16.94%**
PCA(w/ CGAN)			44.59	5.75%	**57.71**	**22.02%**	55.20	17.73%	52.50	15.17%

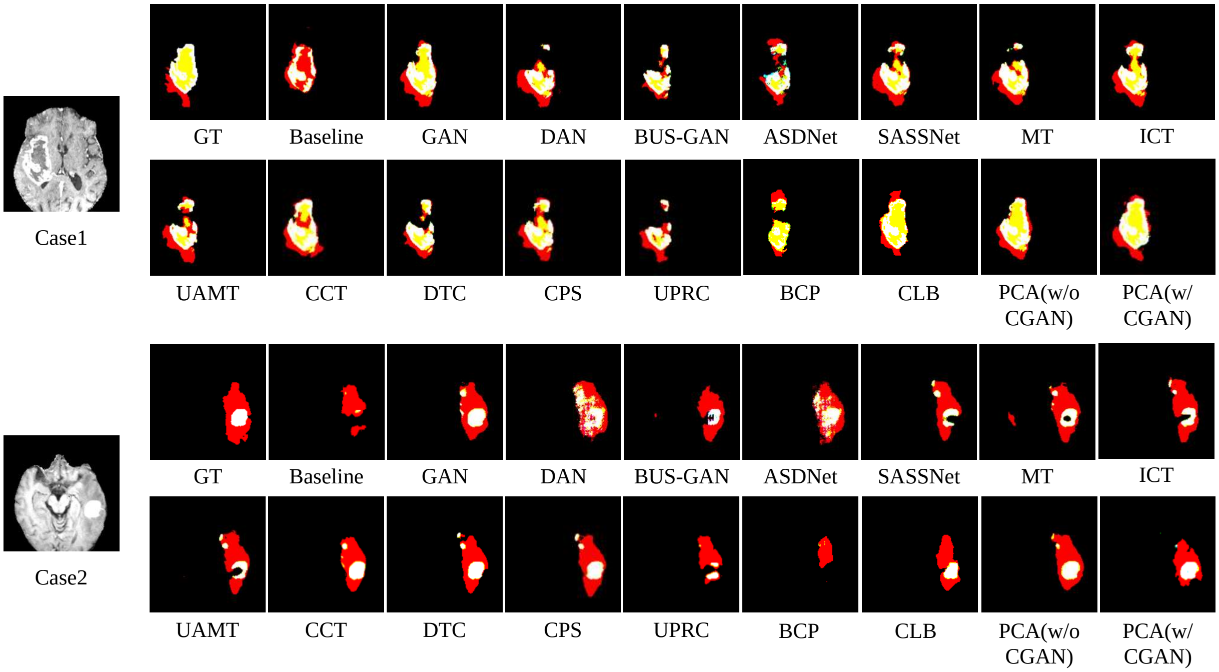


Fig. 5. Visualization of different semi-supervised segmentation methods under 10% labeled data on the BraTS2019 dataset, where GT is the ground truth, the black area is the background, the green area is the whole tumor region (wt), the yellow area represents the core tumor region (ct), and the red area represents the enhancing tumor region (et).

E. Experiments on Brain Tumor Image Data Set

We further tested our proposed method on the brain tumor dataset from the public 2019 Brain Tumor Segmentation Challenge. Since the test and validation sets are not publicly available, we randomly divided the training set into training and test set at a ratio of 8:2. The input are T1-CE modal brain images, and the output result is the binarized segmentation of the three target regions. We performed a quantitative comparison of all semi-supervised segmentation methods on the test set and trained the model with 10% labeled data.

The results in Table IV show that among these semi-supervised segmentation methods, our PCA (without CGAN) model achieves the best performance (Mean DSC = 54.27%) in both settings. Compared with the supervised segmentation method trained on only 10% of labeled data, the improvement is 16.94%. Furthermore, we find that unconditional GANs perform much better than conditional GANs. A possible reason could be that small-amplitude jitter produces significant changes in the grey value of brain tumor images. With the introduction of the constraints of the original image information, our model also received noisy information. However, the robustness and stability of the w/ CGAN method are superior. We have conducted stability experiments comparing the w/o CGAN and w/ CGAN methods, where each method is run

TABLE V
MEAN AND VARIANCE OF PCA (W/O CGAN) AND PCA (W/ CGAN) ON THE BRATS DATASET WITH 5% LABEL RATIO.

	DSC_WT(%)	DSC_TC(%)	DSC_ET(%)	DSC_Mean(%)
PCA(w/o CGAN)	**49.13(8.49)**	54.04(22.5)	51.12(22.8)	51.43(17.0)
PCA(w/ CGAN)	48.63(0.67)	**55.22(3.76)**	**54.62(9.03)**	**52.83(2.61)**

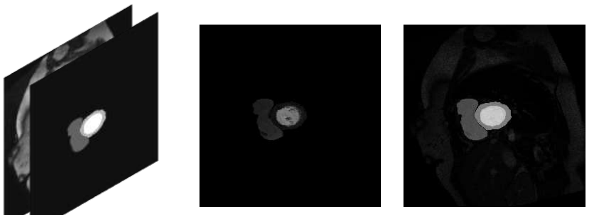

(a) Concate (b) Element-wise multiplication (c) Pixel-additive blending

Fig. 6. Visualization of different input blending methods for the discriminator.

10 times under the same settings, and the mean and variance are calculated, as shown in Table V. It can be observed that under multiple trials, the w/ CGAN method exhibits a smaller variance and overall better segmentation performance. We visualized the segmentation results in Figure 5. Compared with other methods, our results have a higher overlap rate with the ground truth, produce fewer false positives, and remain more detailed. This part of the experiment further demonstrates the effectiveness of our method.

F. Hyperparameter Experiments

1) Experiments on λ_{noise}: On the ACDC dataset with 10% labeled data, we present the experimental results of the selection of hyperparameter λ_{noise} in our method CPCA in Table VI. From the analysis in Table VI, it can be observed that when λ_{noise} is around 0.001, the model achieves better segmentation performance.

TABLE VI
DSC VALUES OF THE CPCA METHOD USING 10% LABELED DATA ON THE ACDC DATASET.

λ_{noise}	0.0002	0.0004	0.0006	0.0008	**0.001**	0.002	0.003
DSC	85.48	86.34	86.76	86.86	**87.33**	86.86	86.62
λ_{noise}	0.004	0.005	0.006	0.007	0.008	0.009	0.01
DSC	86.54	86.38	85.54	85.90	85.76	85.56	85.08

We provide the visualizations of the concatenated-way, element-wise multiplication, and pixel-additive blending methods in Figure 6. Additionally, on the ACDC dataset with 10% labeled data, we conduct quantitative analysis experiments comparing these methods under the same settings in Table VII. The results show that the pixel-additive blending method outperforms the other approaches.

TABLE VII
COMPARISON OF DIFFERENT BLENDING METHODS ON MEDICAL IMAGE SEGMENTATION PERFORMANCE.

blend method	DSC(%)↑	JA(%)↑	95HD(voxel)↓	ASD(voxel)↓
Concate	85.34	76.08	4.12	1.92
Element-wise multiplication	86.14	76.93	4.31	1.83
Pixel-additive blending	**87.33**	**78.31**	**3.10**	**1.35**

2) Experiments on the N×N Size of Patch Confidence Map: The size of our discriminator network is designed based on previous successful work [20]. Specifically, the N × N size is related to the number of convolutional layers in the discriminator. We use a network that contains three convolutional layers with a kernel size of 4 × 4 and a stride of 2 × 2. The number of channels for the three convolutional layers are 32, 64, and 1, respectively. On the ACDC dataset, the output size m × n of the patch-based discriminator is 32 × 32.

To further validate the applicability of this hyperparameter selection in our medical image segmentation task, we conducted hyperparameter experiments under the same settings. On the ACDC dataset with 10% labeled data, we varied the number of convolutional layers in the discriminator to 2, 3, 4, and 5, corresponding to N × N sizes of 64 × 64, 32 × 32, 16 × 16, and 8 × 8, respectively. The experimental results are shown in Table VIII. From the analysis of the table, we can observe that the segmentation results of the overall model are better when the number of convolutional layers in the discriminator is 3, and the size of N × N is 32 × 32. Because N × N is too small is not conducive to the discriminator capturing local structures and detailed features, while an N × N is too large does not contain contextual information. The appropriate size of N × N at the output of the appropriate convolutional layer can improve the discriminator's ability to distinguish and guide the generator for better segmentation performance.

TABLE VIII
PERFORMANCE COMPARISON OF DIFFERENT PATCH SIZES (N × N) FOR THE DISCRIMINATOR ON THE ACDC DATASET WITH 10% LABEL RATIO.

N*N	DSC(%)↑	JA(%)↑	95HD(voxel)↓	ASD(voxel)↓
64*64	76.17	65.11	33.9	10.8
32*32	**87.33**	**78.31**	**3.10**	**1.35**
16*16	85.72	76.54	5.46	1.88
8*8	85.04	75.59	5.92	1.53

V. CONCLUSION

In this paper, we proposed a novel semi-supervised learning model (PCA) for medical image segmentation. Specifically, we built a PatchGAN and CGAN framework based on GAN to mine segmentation contextual information and stabilize discriminator training, thereby improving the accuracy of the segmentation model. In addition, we proposed a novel conditional input for the GAN framework to alleviate the overfitting of unlabeled data prediction. Extensive experiments on two public data sets prove the effectiveness of the method. Compared with the current state-of-the-art methods, our proposed model shows a superior performance. In the future, we will extend the segmentation scheme to other medical image segmentation tasks that lack enough annotated data and explore the potential of PCA in more visual tasks.

REFERENCES

[1] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2015.

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

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of International Conference on Neural Information Processing Systems*, 2014, pp. 2672–2680.

[6] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, "Semi-supervised learning for network-based cardiac MR image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 253–260.

[7] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne, "Semi-supervised medical image segmentation via learning consistency under transformations," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 810–818.

[8] V. Verma, A. Lamb, J. Kannala, Y. Bengio, and D. Lopez-Paz, "Interpolation consistency training for semi-supervised learning," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2019, pp. 3635–3641.

[9] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proceedings of International Conference on Neural Information Processing Systems*, 2017, pp. 1195–1204.

[10] Y. Xia, F. Liu, D. Yang, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, "3D semi-supervised learning with uncertainty-aware multi-view co-training," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3646–3655.

[11] W. C. Hung, Y. H. Tsai, Y. T. Liou, Y.-Y. Lin, and M. H. Yang, "Adversarial learning for semi-supervised semantic segmentation," in *Proceedings of British Machine Vision Conference*, 2018.

[12] J. Son, S. J. Park, and K.-H. Jung, "Retinal vessel segmentation in fundoscopic images with generative adversarial networks," *arXiv preprint arXiv:1706.09318*, 2017.

[13] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 408–416.

[14] L. Han, Y. Huang, H. Dou, S. Wang, S. Ahamad, H. Luo, Q. Liu, J. Fan, and J. Zhang, "Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network," *Computer Methods and Programs in Biomedicine*, vol. 189, p. 105275, 2020.

[15] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 552–561.

[16] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 370–378.

[17] A. Lahiri, K. Ayush, P. Kumar Biswas, and P. Mitra, "Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: Automated vessel segmentation in retinal fundus image as test case," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 42–48.

[18] C. Cao, T. Lin, D. He, F. Li, H. Yue, J. Yang, and E. Ding, "Adversarial dual-student with differentiable spatial warping for semi-supervised semantic segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 2, pp. 793–803, 2022.

[19] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proceedings of International Conference on Machine Learning*, 2019, pp. 7354–7363.

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

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

[22] Z. Feng, D. Nie, L. Wang, and D. Shen, "Semi-supervised learning for pelvic MR image segmentation based on multi-task residual fully convolutional networks," in *Proceedings of International Symposium on Biomedical Imaging*, 2018, pp. 885–888.

[23] D. Yuan, Y. Liu, Z. Xu, Y. Zhan, J. Chen, and T. Lukasiewicz, "Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing," *Computers in Biology and Medicine*, vol. 153, p. 106487, 2023.

[24] Z. Xu, X. Zhang, H. Zhang, Y. Liu, Y. Zhan, and T. Lukasiewicz, "Efpn: Effective medical image detection using feature pyramid fusion enhancement," *Computers in Biology and Medicine*, vol. 163, p. 107149, 2023.

[25] M. Yu, M. Guo, S. Zhang, Y. Zhan, M. Zhao, T. Lukasiewicz, and Z. Xu, "Rirgan: An end-to-end lightweight multi-task learning method for brain mri super-resolution and denoising," *Computers in Biology and Medicine*, vol. 167, p. 107632, 2023.

[26] Z. Xu, Y. Liu, G. Xu, and T. Lukasiewicz, "Self-supervised medical image segmentation using deep reinforced adaptive masking," *IEEE Transactions on Medical Imaging*, vol. Early Access, pp. 1–14, 2024.

[27] J. Zhang, S. Zhang, X. Shen, T. Lukasiewicz, and Z. Xu, "Multi-condos: Multimodal contrastive domain sharing generative adversarial networks for self-supervised medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 76–95, 2024.

[28] S. Zhang, J. Zhang, B. Tian, T. Lukasiewicz, and Z. Xu, "Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 83, p. 102656, 2023.

[29] L. Su, Z. Wang, X. Zhu, G. Meng, M. Wang, and A. Li, "Dual consistency semi-supervised nuclei detection via global regularization and local adversarial learning," *Neurocomputing*, vol. 529, pp. 204–213, 2023.

[30] L. Zhang, X. Xie, K. Xiao, W. Bai, K. Liu, and P. Dong, "Manomaly: Mutual adversarial networks for semi-supervised anomaly detection," *Information Sciences*, vol. 611, pp. 65–80, 2022.

[31] Z. Xu, B. Tian, S. Liu, X. Wang, D. Yuan, J. Gu, J. Chen, T. Lukasiewicz, and V. C. M. Leung, "Collaborative attention guided multi-scale feature fusion network for medical image segmentation," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 2, pp. 1857–1871, 2024.

[32] Z. Xu, S. Liu, D. Yuan, L. Wang, J. Chen, T. Lukasiewicz, Z. Fu, and R. Zhang, "ω-net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution," *Neurocomputing*, vol. 500, pp. 177–190, 2022.

[33] D. Yuan, Z. Xu, B. Tian, H. Wang, Y. Zhan, and T. Lukasiewicz, "μ-net: Medical image segmentation using efficient and effective deep supervision," *Computers in Biology and Medicine*, vol. 160, p. 106963, 2023.

[34] Z. Xu, W. Xu, R. Wang, J. Chen, C. Qi, and T. Lukasiewicz,

"Hybrid reinforced medical report generation with m-linear attention and repetition penalty," *IEEE Transactions on Neural Networks and Learning Systems*, vol. Early Access, pp. 1–15, 2023.

[35] W. Zhu, J. Li, J. Lu, and J. Zhou, "Separable structure modeling for semi-supervised video object segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 1, pp. 330–344, 2021.

[36] W. Liu, G. Lin, T. Zhang, and Z. Liu, "Guided co-segmentation network for fast video object segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1607–1617, 2020.

[37] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 605–613.

[38] X. Cao, H. Chen, Y. Li, Y. Peng, S. Wang, and L. Cheng, "Uncertainty aware temporal-ensembling model for semi-supervised abus mass segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 431–443, 2020.

[39] S. Sedai, D. Mahapatra, S. Hewavitharanage, S. Maetschke, and R. Garnavi, "Semi-supervised segmentation of optic cup in retinal fundus images using variational autoencoder," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 75–82.

[40] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of International Conference on Learning Representations*, 2014.

[41] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proceedings of International Conference on Neural Information Processing Systems*, 2016, pp. 2234–2242.

[42] A. Jolicoeur-Martineau, "The relativistic discriminator: A key element missing from standard GAN," in *Proceedings of International Conference on Learning Representations*, 2018.

[43] Y. Dong, D. Xu, S. K. Zhou, B. Georgescu, M. Chen, S. Grbic, D. Metaxas, and D. Comaniciu, "Automatic liver segmentation using an adversarial image-to-image network," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 507–515.

[44] Z. Han, B. Wei, M. Ashley, L. Stephanie, and S. Li, "Spine-GAN: Semantic segmentation of multiple spinal structures," *Medical Image Analysis*, vol. 50, pp. 23–35, 2018.

[45] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, "Joint optic disc and cup segmentation using fully convolutional and adversarial networks," in *Proceedings of Fetal, Infant and Ophthalmic Medical Image Analysis*, 2017, pp. 168–176.

[46] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 674–12 684.

[47] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, 2021, pp. 8801–8809.

[48] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2613–2622.

[49] X. Luo, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, N. Chen, G. Wang, and S. Zhang, "Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2021, pp. 318–329.

[50] J. Liu, C. Desrosiers, D. Yu, and Y. Zhou, "Semi-supervised medical image segmentation using cross-style consistency with shape-aware and local context constraints," *IEEE Transactions on Medical Imaging*, 2023.

[51] Y. Bai, D. Chen, Q. Li, W. Shen, and Y. Wang, "Bidirectional copy-paste for semi-supervised medical image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 11 514–11 524.

[52] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[53] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[54] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, no. 1, pp. 1–13, 2017.

[55] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[56] X. Luo, "SSL4MIS," https://github.com/HiLab-git/SSL4MIS, 2020.

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.

Runhe Yang is currently a PhD student in the State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, China. His research interests lie in medical image processing using deep learning methods.

This article has been accepted for publication in IEEE Transactions on Network Science and Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2025.3548416

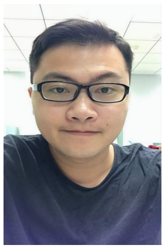
Zihang Xu is going to be a master student in the Department of Electrical and Electronic Engineering (EEE) of HKU. He received a B.Eng. degree in biomedical engineering from Hebei University of Technology, China, in 2022. His research interests lie in medical image analysis and computer vision.

Shuo Zhang is currently a master student in the State Key Laboratory of Reliability and Intelligence of Electrical Equipment, Hebei University of Technology, China. He received B.Eng. degree in Internet of Things Engineering from Jilin Agricultural University, China, in 2020. His research interests lie in medical image processing using deep learning methods. He has published high-quality papers in top journals, e.g., Medical Image Analysis.

Yuchen Yang is currently a master's student at Johns Hopkins University, USA. He received a B.S. degree in Data Science from Duke University. His research interests include deep learning in image reconstruction in CT and applications of CT.

Weipeng Liu received his Ph.D. degree from Hebei University of Technology. Currently, he works as a professor at Hebei University of Technology. His research interests include robot vision, artificial intelligence, and deep learning. He has published more than 20 papers and received several awards in robot control.

Weichao Xu obtained a Master's degree in Medicine from Hebei University of Traditional Chinese Medicine in China in 2013. Since 2013, he has been serving as the Deputy Director of the Digestive Disease Center at Hebei Provincial Hospital of Traditional Chinese Medicine. He has published over 30 papers in top artificial intelligence or traditional Chinese medicine conferences and journals, such as the Journal of Traditional Chinese Medicine and the Chinese Journal of Integrated Traditional Chinese and Western Medicine. His current research focuses on traditional Chinese medicine, intelligent medical image analysis, and computer vision.



Junyang Chen received the Ph.D. degree in computer and information science from the University of Macau, Macau, China, in 2020. He is currently an Assistant Professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include graph neural networks, text mining, deep learning, and recommender systems.

Thomas Lukasiewicz is a Professor of Computer Science at the Department of Computer Science, University of Oxford, UK, heading the Intelligent Systems Lab within the Artificial Intelligence and Machine Learning Theme. 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.

Victor C. M. Leung (Life Fellow, IEEE) is currently a Distinguished Professor of computer science and software engineering with Shenzhen University, Shenzhen, China. He is also an Emeritus Professor of electrical and computer engineering and the Director of the Laboratory for Wireless Networks and Mobile Systems, University of British Columbia (UBC), Vancouver, BC, Canada. He is a Fellow of Royal Society of Canada, Canadian Academy of Engineering, and Engineering Institute of Canada. He is named in the current Clarivate Analytics list of "Highly Cited Researcher". His research interests include wireless networks and mobile systems.