Applied Intelligence (2022) 52:1092–1109
https://doi.org/10.1007/s10489-021-02427-x

A novel multi-discriminator deep network for image segmentation

Yi Wang[1] · Hailiang Ye[1] · Feilong Cao[1]

Accepted: 9 April 2021 / Published online: 16 May 2021

Abstract

Several studies have shown the excellent performance of deep learning in image segmentation. Usually, this benefits from a large amount of annotated data. Medical image segmentation is challenging, however, since there is always a scarcity of annotated data. This study constructs a novel deep network for medical image segmentation, referred to as asymmetric U-Net generative adversarial networks with multi-discriminators (AU-MultiGAN). Specifically, the asymmetric U-Net is designed to produce multiple segmentation maps simultaneously and use the dual-dilated blocks in the feature extraction stage only. Further, the multi-discriminator module is embedded into the asymmetric U-Net structure, which can capture the available information of samples sufficiently and thereby promote the information transmission of features. A hybrid loss by the combination of segmentation and discriminator losses is developed, and an adaptive method of selecting the scale factors is devised for this new loss. More importantly, the convergence of the proposed model is proved mathematically. The proposed AU-MultiGAN approach is implemented on some standard medical image benchmarks. Experimental results show that the proposed architecture can be successfully applied to medical image segmentation, and obtain superior performance in comparison with the state-of-the-art baselines.

Keywords Deep learning · Generative adversarial networks (GAN) · Convolution neural network (CNN) · Image segmentation

1 Introduction

Medical image segmentation is one of the most important tasks in biological image processing and analysis. Its purpose is to segment the parts of a medical image with some special implications and extract related features, thereby assisting doctors in diagnosis and pathology research. Previous approaches to medical image segmentation were often based on traditional methods, such as support vector machines (SVMs) [1] and random forests (RF) [2], which generally demanded manual features in advance [3, 4]. These traditional methods often create problems in terms of efficiency and subjectivity. Naturally, it is necessary to explore advanced segmentation algorithms for medical images.

In recent years, deep learning, owing to its powerful and automatic feature extraction capability, has been widely used in image processing and computer vision, such as image reconstruction [5, 6], image classification [7], object detection [8], etc. This technique has also been extensively employed in image segmentation [9, 10]. A fully convolutional network (FCN) in [11] was the first image segmentation approach to perform end-to-end image segmentation. Subsequently, Badrinarayanan et al. [12] improved upon FCN to develop a novel architecture named SegNet. SegNet consists of a 13 layer deep encoder network, which extracts spatial features from the image. A corresponding 13 layer deep decoder network upsamples the feature maps to predict the segmentation masks. And a series of DeepLap model in [13] performed semantic segmentation using dilated convolutions and employed the VGG [14] as a feature extractor to raise the depth of the network.

Despite these approaches have made tremendous successes in image segmentation, a major drawback of the convolutional neural network (CNN) architectures is that

✉ Feilong Cao
icteam@163.com

Yi Wang
kelly_sylvia@163.com

Hailiang Ye
yhl575@163.com

[1] College of Sciences, China Jiliang University, Hangzhou 310018, China

they require massive volumes of training data [15–17]. Unfortunately, in the context of medical images, the situation is always scarcity of labeled images due to the fact that the annotation process is time-consuming and prone to errors. Therefore, developing a novel architecture of medical diagnosis on small samples is of practical significance.

CNN has shown great promise in medical image segmentation recently. This mainly attributes to the development of U-Net [9]. The structure of U-Net is quite similar to SegNet, comprising an encoder and a decoder network. The corresponding layers of the encoder and decoder network are connected by skip connections, which allows efficient information flow and performs well when sufficiently large datasets are not available. Simultaneously, in order to avoid the over-fitting problem caused by the lack of data, the author also proposes a data enhancement method to expand the data in the data pre-processing stage.

Subsequently, several modified versions about skip connections of U-Net have emerged. Drozdzal et al. [18] employed both long and short skip connections to enhance the information flow for biomedical image segmentation. Yu et al. [19] raised a novel combination of residual connections (ConvNet), which can greatly improve the segmentation performance of the proposed network by enhancing the information propagation both locally and globally. Zhou et al. [20] presented nested U-Net structures for medical image segmentation by using short-skip and long-skip connections to link shallow and deep features. Furthermore, Zhuang et al. [21] fused skip connections and residual blocks to acquire more information flow paths on the segmentation task for blood vessels in retinal images. All the approaches for changing the skip connections increase information flow. These tricks perform well when sufficiently large datasets are not available.

Almost all previous works have been designed for a certain kind of medical image model. However, the objects of interest are of irregular and different scales in most cases, which images may originate from various modalities. Therefore, a network should be robust enough to analyze objects at different scales. Various deformable modules on the U-Net have become popular to settle this problem. Oktay et al. [22] proposed a novel attention gate model for two large computed tomography abdominal datasets that automatically learned to focus on target structures of varying shapes and sizes. Gu et al. [23] devised a context extractor in a traditional encoder-decoder structure to capture more high-level information. Moreover, Alom et al. [24] embedded the recurrent convolution module into U-Net. Ibtehaz et al. [17] presented an inception-like block to reconcile the features learned from the image at different scales. In [25], a large kernel encoder-decoder network with deep multiple atrous convolution is proposed, where the use of this network can capture multi-scale contexts by enlarging the valid receptive field. However, image segmentation requires dense pixel-level labeling. A common property across all CNN architectures is that all label variables are predicted independently from each other [26].

Generative adversarial network (GAN) can make the model achieve better results from a distribution perspective by introducing a discriminator, which solves the problem of inconsistent distribution between different data domains [27]. By making the discriminator unable to distinguish data from two different domains, it indirectly leads them to belong to the same distribution. In [26], the image segmentation approach based on GAN has been explored to reinforce spatial contiguity in the output label maps. In medical image segmentation, there have been several types of researches on using U-Net and GAN. These works usually regard medical image segmentation as the process of generating segmentation for samples and introduce a discriminator to fit the generated segmentation distribution to the real segmentation distribution. Dong et al. [28] designed a model called U-Net generative adversarial network (U-Net-GAN), which jointly trained a set of U-Nets as generators and fully convolutional networks as discriminators to implement multimodel segmentation. For segmenting the tumor in breast ultrasound images, Negi et al. [29] used Residual-Dilated-Attention-Gate-UNet as the generator, which serves as a segmentation module. Then the Wasserstein GAN algorithm was employed to stabilize training. However, these approaches involve iterative training between the generator and single discriminator. In fact, it is important for each recovery segmentation in the decoder to increase information flow from high-level semantic information in small sample problems.

To solve the above-mentioned problems, we propose a novel segmentation model for small-sample medical images, referred to as asymmetric U-Net generative adversarial network with multi-discriminators (AU-MultiGAN). More specifically, AU-MultiGAN jointly trains an asymmetric U-Net as generators and multi-discriminators to implement the medical image segmentation tasks. The novelty of the proposed model architecture is twofold. First, the construction of the asymmetric U-Net generates multiple results of different sizes from distinct upsampling levels. Before upsampling, the features of different receptive fields are extracted through multiple proposed dual-dilated block to obtain higher-level semantic information. Second, a multi-discriminator module is designed for improving sample utilization and increasing information flow from the high-level semantic information. The multi-discriminators by employing a discriminator for each upsampling layer of generating the segmentation achieve deep supervision. Furthermore, a hybrid loss is designed for imbalanced sample issues and an adaptive parameter selection method is proposed for this loss. Here, the hybrid loss includes both discriminator and segmentation losses, in which the segmentation loss consists of FocalLoss [30] and the reconstruction loss. The discriminator loss adopts the form of the mean square error. Such a combined loss can possess various functions, such as dealing with the imbalance class, matching the generated segmentation with the real segmentation, and stability training. In addition, we conduct a theoretical and experimental analysis of the proposed method. The experimental results on benchmark datasets indicate that the proposed AU-MultiGAN can be successfully applied to medical image segmentation in small sample cases and it is more effective than the state-of-the-art baselines.

The main contributions of this study can be summarised as follows.

- A multi-discriminator deep network, which mainly embeds the multi-discriminator modules into the asymmetric U-Net in order to utilize the information of samples sufficiently and thereby enhance the information flow of features, is devised to overcome the small sample issue in image segmentation tasks.
- A hybrid loss is proposed that integrates discriminator loss with segmentation loss. This new hybrid loss can not only balance the intra-classes of samples but also keep consistent with the generated and real segmentation maps. Further, an adaptive selection method on the scale factors for this hybrid loss is designed.
- The theoretical convergence of the proposed method is discussed and analyzed rigorously.

The remainder of this paper is organised as follows. Section 2 lists some notations that are used throughout the text. Section 3 details the AU-MultiGAN. Experimental results are presented and analysed in Section 4. The conclusion of the study is provided in Section 5, and the mathematical proofs of the convergence of the proposed model are presented in the final Appendix.

2 Notations

This section lists some notations used in this study. Let $\mathbf{R}^{m\times n}$ be the set of real numbers with $m \times n$ dimensions. For a matrix, $\boldsymbol{X} \in \mathbf{R}^{m\times n}$, we denote its elements as $x_{ij}(i = 1, 2, \dots, m, j = 1, 2, \dots, n)$ and call $\phi_1(x) = \frac{1}{1+\mathrm{e}^{-x}}(x \in \mathbf{R})$ as the logistic sigmoidal function and $\phi_2(x_{ij}) = \frac{\mathrm{e}^{x_{ij}}}{\sum_{i,j}\mathrm{e}^{x_{ij}}}(x_{ij} \in \mathbf{R})$ as the softmax function. $\sigma(x) = \max(0, x)(x \in \mathbf{R})$ denotes as the rectified linear unit (ReLU).

We use $\psi(\cdot)$ to represent max pooling, $\tau(\cdot)$ to denote a random neuron discard operation (dropout), and $\xi(\cdot)$ to express pixelshuffle [31]. Moreover, $[\boldsymbol{X}, \boldsymbol{Y}]$ represents a concatenation operator for $\boldsymbol{X}$ and $\boldsymbol{Y}$; p_{Y_i}, p_X, and p_{g_i} are the distributions of label, sample, and the generated segmentation, respectively; $p_{Y_i}(\cdot)$ and $p_{g_i}(\cdot)$ represents the probability density functions of p_{Y_i} and p_{g_i}, respectively, where i is the discriminator index.

For readability, all the above-mentioned symbols are listed in Table 1.

3 Method

In this section, we first explain the architecture of the proposed network, then describe the training strategy, and finally analyse the convergence for the proposed algorithm.

3.1 Architectural design

The proposed network includes two major parts: a asymmetric U-Net and a multi-discriminator module, as shown in Fig. 1. The asymmetric U-Net is regarded as a generator for segmentation problems, containing a dual-dilated block, bottleneck block, decoder block, and classifier block. In contrast to the U-Net, the main differences in the asymmetric U-Net are reflected in two aspects. One aspect is that a dual-dilated block is designed for each

Table 1 Some notations used in the paper

Notations	Description	Notations	Description
$\mathbf{R}^{m\times n}$	The set of real number with $m \times n$ dimensions	σ	ReLU $\max(0, x)$
$\boldsymbol{X}$	$m \times n$ matrix	x_{ij}	The elements of matrix $\boldsymbol{X}$
ϕ_1	The logistic sigmoidal function $\frac{1}{1+\mathrm{e}^{-x}}$	ϕ_2	$\frac{\mathrm{e}^{x_{ij}}}{\sum_{i,j}\mathrm{e}^{x_{ij}}}$
ψ	Max pooling	ξ	Pixelshuffle
τ	Dropout	$[\boldsymbol{X}, \boldsymbol{Y}]$	The concatenation operator for $\boldsymbol{X}$ and $\boldsymbol{Y}$
p_{Y_i}	The distribution of label	p_{g_i}	Generated segmentation
p_X	Sample	β_i	The module β in level i of proposed network
$p_{Y_i}(\cdot)$	The probability density function of p_{Y_i}	$p_{g_i}(\cdot)$	The probability density function of p_{g_i}

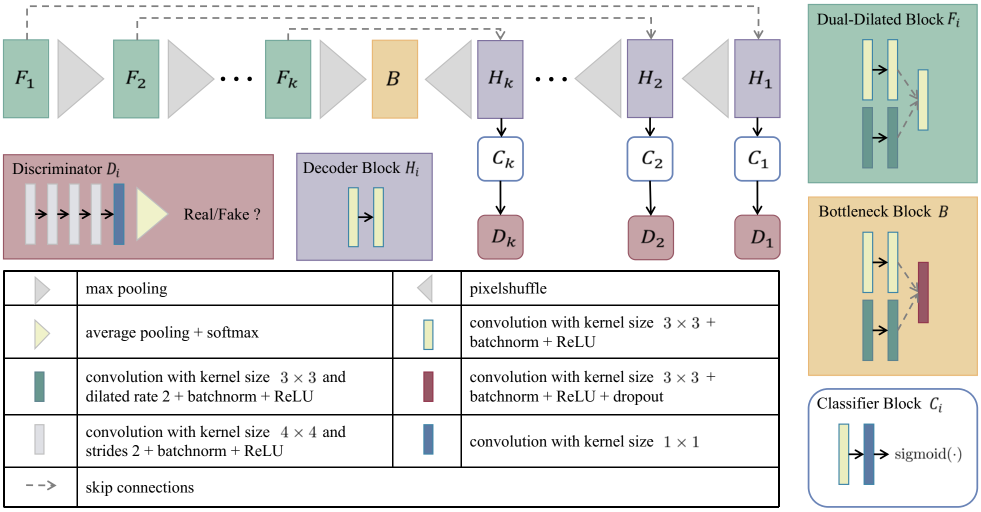

F_1
F_2
F_k
B
H_k
H_2
H_1
C_k
C_2
C_1
D_k
D_2
D_1
Discriminator D_i
Real/Fake ?
Decoder Block H_i
Dual-Dilated Block F_i
Bottleneck Block B
Classifier Block C_i
sigmoid(·)
max pooling
pixelshuffle
average pooling + softmax
convolution with kernel size 3 × 3 + batchnorm + ReLU
convolution with kernel size 3 × 3 and dilated rate 2 + batchnorm + ReLU
convolution with kernel size 3 × 3 + batchnorm + ReLU + dropout
convolution with kernel size 4 × 4 and strides 2 + batchnorm + ReLU
convolution with kernel size 1 × 1
skip connections

Fig. 1 Schematic diagram of AU-MultiGAN

convolutional layer in the encoder. The other is that all level classifier blocks are followed by segmented images for each up-sampling feature extraction stage of the decoder. The multi-discriminator corresponds to the multi-level output of the asymmetric U-Net, and the multiple discriminators are uniform in structure. The description of the architecture is given in Sections 3.2 and 3.3 in further detail.

3.2 Asymmetric U-net

Segmentation tasks can be regarded as a generation of segmented images. The asymmetric U-Net is a generator and employs the dual-dilated blocks in the feature extraction stage to obtain more abundant information, whereas just one convolution branch is used in the process of generating segmentation maps. Concretely, let $\boldsymbol{X} \in \mathbf{R}^{m \times n}$ be the input of the asymmetric U-Net. The image that is segmented at the ith level in the asymmetric U-Net is denoted by $G_i(\boldsymbol{X}; \theta_i)(i = 1, 2 \ldots, k)$, where θ_i is the parameter of generator G_i. For simplicity, we often omit θ_i and write $G_i(\boldsymbol{X}; \theta_i)$ as $G_i(\boldsymbol{X})$ or $\boldsymbol{Y}'_i$. Specifically, we can write $G_i(\boldsymbol{X})$ as a matrix form, $\left(g_{js}^{(i)}\right)_{m_i \times n_i}$, where $g_{js}^{(i)}$ represents the pixel of $G_i(\boldsymbol{X})$ at coordinates (j, s). In addition, $m_i = m/(2^{i-1})$ and $n_i = n/(2^{i-1})$ denote the length and width of $G_i(\boldsymbol{X})$, respectively. Further, $G_i(\boldsymbol{X})$ can be computed as

$$G_i(\boldsymbol{X}) = C_i(H_i(B(F_k(\ldots(F_1(\boldsymbol{X}))\ldots)))), \tag{1}$$

where $B(\cdot)$ is the bottleneck block, and $F_i(\cdot)$, $H_i(\cdot)$, and $C_i(\cdot)(i = 1, 2, \ldots, k)$ represent the dual-dilated, decoder, and classifier blocks in the ith level, respectively. Details of these blocks are provided as follows.

1) **Dual-Dilated Block.** First, the features are extracted from different receptive fields by using two branches of 3 × 3 convolutions with different dilated rates. The dilated rate is set to 1 and 2. The operation is equivalent to adopting a 3 × 3 convolution and a 5 × 5 convolution with fewer parameters. Subsequently, we use a 3 × 3 convolution to further extract features and fuse the distinct features. The encoder in the proposed network consists of k dual-dilated blocks. For each $F_i(i = 1, 2, \ldots, k)$, we employ max pooling for the output features from F_{i-1} in advance, as the input features, $\bar{\boldsymbol{F}}_i$, are expected to be at different scales. Let

$$\bar{\boldsymbol{F}}_i = \psi(F_i(\bar{\boldsymbol{F}}_{i-1})), i = 1, 2, \ldots, k \tag{2}$$

Then, $\bar{\boldsymbol{F}}_{i-1}$ is the input feature of $F_i(\cdot)$; in particular, $\bar{\boldsymbol{F}}_0 = \boldsymbol{X}$ and $\hat{\boldsymbol{F}}_i = F_i(\bar{\boldsymbol{F}}_{i-1})$ are the output features of $F_i(\cdot)$. This strategy can provide abundant information for the discriminator at the lower level.

2) **Bottleneck Block.** The function of the bottleneck block is to process the lower-level features in the asymmetric U-Net. Only $\tau(\cdot)$ is expanded in relation to a dual-dilated block. The remarkable role of this additional operation is to avoid over-fitting. From the block, the high-level features, $\boldsymbol{O}$, are specifically obtained for sample $\boldsymbol{X}$.

3) **Decoder Block.** To correspond with the encoder, the decoder adopts k decoder blocks as well. In parallel, the features in the decoder block require dissimilar scales because the input features $\boldsymbol{Z}_i$ for each $H_i(i = 1, 2, \ldots, k)$ are the concatenation of $\bar{\boldsymbol{H}}_{i+1}$ and $\hat{\boldsymbol{F}}_i$, i.e.,

$$\boldsymbol{Z}_i = [\bar{\boldsymbol{H}}_{i+1}, \hat{\boldsymbol{F}}_i], i = 1, \ldots, k, \tag{3}$$

where

$$\bar{\boldsymbol{H}}_i = \xi(H_i(\boldsymbol{Z}_i))(i = 2, 3, \ldots, k) \text{ and } \bar{\boldsymbol{H}}_{k+1} = \xi(\boldsymbol{O}) \tag{4}$$

denote the upsampling features through pixelshuffle for the output from H_i. Consequently, for the given features $\bar{\boldsymbol{H}}_{i+1}(i = 1, \ldots, k)$, the first phase in $H_i(\cdot)$ involves putting the upsampling results from a low level and the shallow features in the same level dual-dilated block together according to (3). Next, two convolutional layers are exploited to handle the features from the concatenate maps:

$$H_i(\boldsymbol{Z}_i) = \sigma(\boldsymbol{W}^2_{H_i} * \sigma(\boldsymbol{W}^1_{H_i} * \boldsymbol{Z}_i + \boldsymbol{b}^1_{H_i}) + \boldsymbol{b}^2_{H_i}), i = 1, \ldots, k, \tag{5}$$

where $\boldsymbol{W}^j_{H_i}$ $(j = 1, 2)$ expresses a 3×3 convolution kernel with a dilated rate of 1 in module H_i. The decoder block is designed for processing high-abstract features and restoring the size of the image.

4) **Classifier Block.** The classifier block aims to obtain segmented images. To this end, we apply a sigmoidal function ϕ_1 after two convolutional layers to limit the value between 0 and 1, described as

$$C_i(\bar{\boldsymbol{C}}_i) = \phi_1(\sigma(\boldsymbol{W}^2_{C_i} * \sigma(\boldsymbol{W}^1_{C_i} * \bar{\boldsymbol{C}}_i + \boldsymbol{b}^1_{C_i}) + \boldsymbol{b}^2_{C_i})),$$

where $\boldsymbol{W}^1_{C_i}$ and $\boldsymbol{W}^2_{C_i}$ denote the 3×3 convolution kernel and the 1×1 convolution kernel with a dilated rate of 1 in modules $C_i(i = 1, ..., k)$, respectively; $\bar{\boldsymbol{C}}_i = H_i(\boldsymbol{Z}_i)$; the output of $C_i(\cdot)(i = 1, \ldots, k)$ can be regarded as a segmented image defined by (1).

3.3 Multi-discriminator mechanism

The asymmetric U-Net is a generator for producing multi-scale segmentation and the decoder contributes k outputs in different scales, as mentioned earlier. Accordingly, k discriminators are used to train k tasks of the generator. The role of the discriminator is to receive the segmentation generated from the asymmetric U-Net and correspondingly output a confidence value. Here, $D_i(G_i(\boldsymbol{X}); \beta_i)(i = 1, 2 \ldots, k)$ denotes the confidence value of the ith segmentation $G_i(\boldsymbol{X})$, where β_i is the parameter of discriminator D_i. For simplicity, we omit β_i and write $D_i(G_i(\boldsymbol{X}); \beta_i)$ as $D_i(G_i(\boldsymbol{X}))$. The confidence value represents the probability that the input segmentation is a real segmentation label. For segmentation tasks, it is natural that the generated segmentation is expected to be similar to the real segmentation label, and it is similar for the generative tasks. Therefore, it is feasible to transfer generative ideas to segmentation tasks. In applications, the structures of all discriminators $D_i(i = 1, \ldots, k)$ is the same. Specifically, five 4×4 convolutions are followed by average pooling and the softmax function $\phi_2(\cdot)$. Here, the softmax function is used to produce confidence values in the discriminators. A schematic diagram of the dual-dilated block is shown in Fig. 1.

3.4 Hybrid loss and its adaptive scale factor selection

This subsection firstly describes two different losses, the segmentation loss and the discriminator loss, and then builds a combination loss for the proposed AU-MultiGAN model.

1) **Segmentation Loss.** An intuitive idea for the segmentation task is to minimize the pixel-wise loss between the inputs and the segmented ones, which can be modelled as

$$L_{\text{seg}}(G_i) = L_{\text{FL}}(G_i) + L_{\text{re}}(G_i), \tag{6}$$

where

$$L_{\text{FL}}(G_i) = \mathbb{E}_{X \sim p_X} \sum_{j,s} [-\alpha(1 - g^{(i)}_{js})^{\gamma}] \log(g^{(i)}_{js}), \tag{7}$$

$$L_{\text{re}}(G_i) = \mathbb{E}_{X \sim p_X}[\|\boldsymbol{Y}_i - G_i(\boldsymbol{X})\|_1]. \tag{8}$$

Here, $\boldsymbol{Y}_i(i = 1, 2, \ldots, k)$ denotes the real segmented image; $L_{\text{FL}}(G_i)$ is referred to as FocalLoss, dealing with the imbalance classes; $0 \leq \alpha \leq 1$ and $\gamma \geq 0$ are hyperparameters. In particular, if $\gamma = 1$ and $\alpha = 1$, then $L_{\text{FL}}(G_i)$ is the binary cross-entropy loss. When $\alpha = 0$, the segmentation loss only contains the reconstruction loss $L_{\text{re}}(G_i)$, which ensures that the generated segmentation $G_i(\boldsymbol{X})$ ends up matching closely with $\boldsymbol{Y}_i$. Segmentation loss guides the generator to realize a segmentation task.

2) **Discriminator Loss.** It is well known that the loss (8) may lead to the missing of high-frequency information and blur segmentation results. To generate more realistic results, a least square-based GAN loss is introduced, defined as

$$L_{\text{GAN}}(G_i, D_i) = \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}} \left[\frac{1}{2}(D_i(\boldsymbol{Y}_i) - 1)^2\right] + \mathbb{E}_{X \sim p_X}\left[\frac{1}{2} D^2_i(G_i(\boldsymbol{X}))\right]. \tag{9}$$

3) **Hybrid Loss.** The objective function for the proposed AU-MultiGAN is obtained by combining the segmentation loss (6) and the discriminator loss (9):

$$L(G_i, D_i) = L_{\text{GAN}}(G_i, D_i) + \lambda_1 L_{\text{FL}}(G_i) + \lambda_2 L_{\text{re}}(G_i), \tag{10}$$

where $i = 1, 2, \dots, k$, and $\lambda_1, \lambda_2 > 0$ are scale factors. In the following discussion, the key is the optimisation of the hybrid loss:

$$\min_{G_i, D_i} L(G_i, D_i). \tag{11}$$

We adopt the alternate iteration method to solve this problem. Firstly, we optimize discriminator D_i for a fixed G_i, i.e.

$$D_i^* = \arg\min_{D_i} L_{\text{GAN}}(G_i, D_i), \tag{12}$$

and then, we optimize generator G_i for a fixed D_i^*:

$$G_i^* = \arg\min_{G_i} L_{\text{GAN}}(G_i, D_i^*) + \lambda_1 L_{\text{FL}}(G_i) + \lambda_2 L_{\text{re}}(G_i). \tag{13}$$

4) **Adaptive Scale Factor Selection.** To solve the optimisation problem (13) with scale factor λ_1 and λ_2, the conventional method is to set the parameter values by using manual empirical selection. This manual selection method is always inefficient. Once the selection is inappropriate, it may yield poor results. Hence, we design an adaptive method to select the scale factors. Denote $L(G_i, \lambda_1, \lambda_2) = L_{\text{GAN}}(G_i, D_i^*) + \lambda_1 L_{\text{FL}}(G_i) + \lambda_2 L_{\text{re}}(G_i)$. Then, the function must be optimised as a conditional extremum problem. Fortunately, it can be converted to a Lagrange duality problem [32]:

$$\max_{\lambda_1, \lambda_2 > 0} \min_{G_i} L(G_i, \lambda_1, \lambda_2). \tag{14}$$

Consequently, following (13), the method of gradient ascent can be adopted to update the scale factors λ_1 and λ_2:

$$\begin{cases} \lambda_1 \leftarrow \lambda_1 + \eta \dfrac{\partial L(G_i^*, \lambda_1, \lambda_2)}{\partial \lambda_1} & \text{(15a)} \\ \lambda_2 \leftarrow \lambda_2 + \eta \dfrac{\partial L(G_i^*, \lambda_1, \lambda_2)}{\partial \lambda_2}, & \text{(15b)} \end{cases}$$

where η is the learning rate and is a fixed constant selected empirically, which is discussed in Section 4.1; the partial derivatives can be written as $\frac{\partial L(G_i^*, \lambda_1, \lambda_2)}{\partial \lambda_1} = L_{\text{FL}}(G_i^*)$ and $\frac{\partial L(G_i^*, \lambda_1, \lambda_2)}{\partial \lambda_2} = L_{\text{re}}(G_i^*)$.

Overall, the whole training process of AU-MultiGAN is listed in Algorithm 1.

Algorithm 1 AU-MultiGAN algorithm.

```
Input: Training data, learning rate $\eta > 0$, hyperparameters
    $0 \le \alpha \le 1$, and $\gamma \ge 0$;
Output: Return generator parameters $\{\theta_i\}_{i=1}^k$;
Initialize: Scale factor $\lambda_1 > 0$ and $\lambda_2 > 0$, generator parameters $\{\theta_i\}_{i=1}^k$, and discriminator parameters $\{\beta_i\}_{i=1}^k$;
for epochs do
    for iters do
        Sample minibatch sample pairs from training data;
        for $i = 1$ to $k$ do
            For fixed $G_i$, optimize $D_i$ in (12);
            For fixed $D_i^*$, optimize $G_i$ in (13);
            Update $\lambda_1$ using (15a);
            Update $\lambda_2$ using (15b);
        end for
    end for
end for
```

3.5 Theoretical analysis

This subsection gives the convergence analysis for the proposed method. The main theoretical result is summarised in Theorem 1, and its proof is provided in the Appendix. As a preparation for the analysis, two lemmas are given. Their mathematical proofs are also provided in the Appendix. Specifically, the optimal discriminator D_i for any given generator G_i is considered in Lemma 1, and Lemma 2 indicates that when the discriminator loss achieves the value $\frac{1}{4}$, $p_{g_i} = p_{Y_i}$ holds for all $i (i = 1, 2, \dots, k)$.

Lemma 1 *For a fixed G_i, the optimal discriminator D_i is*

$$D_i^* = \frac{p_{Y_i}(\boldsymbol{Y}_i)}{p_{Y_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)} \tag{16}$$

for all $i = 1, 2, \dots, k$.

Lemma 2 *For all $i = 1, 2, \dots, k$, $p_{g_i} = p_{Y_i}$ if and only if the discriminator loss achieves value $\frac{1}{4}$.*

Theorem 1 *Assume that G_i and D_i have sufficient capacity. If at each step of Algorithm 1, the discriminator loss reaches value $\frac{1}{4}$, and p_{g_i} is updated to improve the criterion*

$$\mathbb{E}_{\boldsymbol{Y}_i \sim p_{Y_i}} \left[\frac{1}{2} (D_i^*(\boldsymbol{Y}_i) - 1)^2 \right] + \mathbb{E}_{\boldsymbol{Y}'_i \sim p_{g_i}} \left[\frac{1}{2} D_i^{*2}(\boldsymbol{Y}'_i) \right], \tag{17}$$

for all $i = 1, 2, \dots, k$, then p_{g_i} converges to p_{Y_i}.

4 Experiments

In this section, a series of experiments are conducted to illustrate the performance of the AU-MultiGAN for small-sample medical image segmentation task. The experiments are carried out in a Python 3.6 environment running on a double NVIDIA GTX 1080 GPU and an Intel(R), Xeon(R) W-2123 CPU @ 3.60 GHz with 64 GB main memory.

4.1 Experiments setup

Datasets We choose the different medical imaging modalities to evaluate the proposed segmentation framework. The datasets are compiled from six databases: ISBI2009 [33], ISBI2012 [34, 35], ISBI2014 [36, 37], DRIVE [38], ISIC [39, 40], and CVC-ClinicDB [41]. These datasets contain various types of medical images, such as cell data, digital eye masks, dermoscopy image and endoscopy image. Meanwhile, these datasets can also be classified into different types of medical images segmentation tasks, such like cell contour segmentation, cell nuclei segmentation, organizational segmentation in several different situations and retinal vessel detection. In these six datasets, there is an underlying commonality for four datasets (ISBI2009, ISBI2012, ISBI2014, and DRIVE) is that the amount of annotated data is small. In order to test and verify the proposed method can be applied to many types of medical images, we randomly select a subset of other two datasets according to the scale of ISBI2012 as new mini-batch datasets in our experiment. To sum up, the details of each dataset are listed in Table 2.

Evaluation Four indices, including dice coefficient (Dice), intersection over union (IoU), accuracy, and sensitivity, are adopted to comprehensively assess the performance of the segmentation. The detail definitions can be described as follows.

- The Dice [42] between two binary pixels can be written as

$$\mathrm{Dice} = \frac{2\sum_{i,j}^{m,n} \hat{y}_{ij} y_{ij}}{\sum_{i,j}^{m,n} \hat{y}_{ij}^2 + \sum_{i,j}^{m,n} y_{ij}^2}, \tag{18}$$

  where $\hat{y}_{ij}$ and y_{ij} denote the pixels of the predicted binary segmented image and the ground truth binary map at coordinates (i, j), respectively.
- A similarity measure related to Dice referred to as the IoU [43] can be defined as

$$\mathrm{IoU} = \frac{\sum_{i,j}^{m,n} \hat{y}_{ij} y_{ij}}{\sum_{i,j}^{m,n} \hat{y}_{ij}^2 + \sum_{i,j}^{m,n} y_{ij}^2 - \sum_{i,j}^{m,n} \hat{y}_{ij} y_{ij}}. \tag{19}$$

- The accuracy (Acc) describes the proportion of correctly classified samples to the total number of samples and can be represented as

$$\mathrm{Acc} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{20}$$

  where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives, respectively.
- The sensitivity (Sen) also called recall calculates the proportion of positives (TP) that are correctly predicted to all positives in the true label image. This metric can be written as

$$\mathrm{Sen} = \frac{TP}{TP + FN}. \tag{21}$$

It is worth mentioning that in the binary image segmentation problem, the Sen metric only considers the proportion of the generated segmentation map that is

Table 2 Image segmentation datasets used in the experiments

Dataset	Images	Input size	Modality	Segmentation task
DRIVE	40	480 × 512	Retina blood vessel[a]	Retinal vessel detection
ISBI2009	97	256 × 256	Fluorescence microscopy[b]	Organizational segmentation
ISBI2012	30	256 × 256	Electron microscopy[c]	Cell contour segmentation
ISBI2014	16	256 × 256	Electron microscopy[d]	Cell nuclei segmentation
ISIC (small)	30	384 × 512	Dermoscopy[e]	Organizational segmentation
CVC-ClinicDB (small)	30	256 × 192	Endoscopy[f]	Organizational segmentation

[a] https://drive.grand-challenge.org/

[b] https://metarabbit.wordpress.com/2013/09/11/nuclear-segmentation-in-microscope-cell-images/

[c] http://brainiac2.mit.edu/isbi_challenge/

[d] https://cs.adelaide.edu.au/~carneiro/isbi14_challenge/dataset.html

[e] https://challenge2018.isic-archive.com/

[f] http://www.cvc.uab.es/CVC-Colon/index.php/databases/

Table 3 The effectiveness of different methods on the ISBI2009

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	87.0393 ± 9.8339	86.9861 ± 1.9753	74.5218 ± 6.4906	59.4572 ± 10.8662
RF	94.9879 ± 0.0104	93.7949 ± 0.0287	86.3585 ± 0.0697	75.9930 ± 0.1680
U-Net	91.0421 ± 0.0021	97.1333 ± 0.0020	94.1552 ± 0.3072	90.0799 ± 0.3732
Unet++	94.4628 ± 0.0614	97.7075 ± 0.0561	95.7605 ± 0.2429	91.9275 ± 0.7691
LadderNet	90.9953 ± 0.2254	95.8987 ± 0.2361	92.0341 ± 1.9388	85.5846 ± 0.0280
Attention U-Net	94.1523 ± 0.3412	97.6124 ± 0.1907	95.2568 ± 0.6614	91.2092 ± 1.4424
R2U-Net	96.1461 ± 0.2304	97.4231 ± 0.2297	94.4617 ± 0.5645	89.5773 ± 1.7483
CE-Net	92.2508 ± 0.0053	97.0112 ± 0.0043	94.0018 ± 0.0091	88.6982 ± 0.0309
MultiResUnet	94.8939 ± 0.5401	97.5798 ± 0.2104	94.6682 ± 2.6850	90.5479 ± 2.9552
Ours	**98.8662 ± 0.0012**	**99.1551 ± 0.0010**	**97.6750 ± 0.0054**	**95.4697 ± 0.0195**

correctly predicted in the real segmentation map. Even if there are many noises or outliers in the generated segmentation map, it does not affect the value of this metric. The main reason is that these noises or outliers are always false positives (FP), which is not included in the denominator part of the formula (21). Therefore, it is generally unreasonable. In contrast, Acc, Dice, and IoU have fully considered the true and the false prediction in the real segmentation map. Based on theses analyses, we mainly focus on the three indices Acc, Dice, and IoU when comparing with other methods. The index Sen is given as a reference item.

Implementation details All methods are implemented without data augmentation. The 5-fold cross-validation is adopted in all the experiments. Furthermore, the learnable weight parameters of the asymmetric U-Net and the multi-discriminator are optimised by using the adaptive moment estimation (Adam) method with a learning rate of 4×10^{-3}. Also, we set the hyperparameters $\alpha = 0.25$, $\gamma = 1$, and $k = 2$, respectively. A discussion of these parameters is presented concretely in Section 4.3.

4.2 Comparison results

In this subsection, we compare the proposed method with seven methods based on the U-Net architecture as the baseline models, including U-Net [9], Unet++ [20], CE-Net [23], LadderNet [21], R2U-Net [24], Attention U-Net [22], and MultiResUnet [17]. Simultaneously, two traditional methods, SVM [1] and RF [2], are also used in this paper. Next, we will analyze the experimental results from both quantisation and vision.

The quantitative results for different datasets are listed in Tables 3, 4, 5, 6, 7 and 8. Bold represents the best performance. It can be observed that the proposed method has a great improvement on the Dice, IoU, and Acc for almost all types of datasets when comparing with other methods, although it is not the best performance on the index Sen. But this is reasonable as explained in Section 4.1. Specially, for ISBI2009, ISBI2012, ISBI2014, DRIVE, and CVC-ClinicDB(small), the proposed method achieves the best performances on the Dice, IoU, and Acc over all the baseline models. Further, for ISIC(small) dataset, the effectiveness of the proposed AU-MultiGAN method surpasses those of

Table 4 The effectiveness of different methods on the ISBI2012

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	49.8983 ± 397.2313	79.5333 ± 3.9409	88.6679 ± 0.2824	79.6469 ± 0.7369
RF	59.5372 ± 2.9798	86.2458 ± 0.0409	91.3377 ± 0.0048	84.0566 ± 0.0138
U-Net	90.7071 ± 0.0324	87.9354 ± 0.0357	92.8368 ± 0.3052	86.6740 ± 0.8967
Unet++	91.6146 ± 0.0143	90.8084 ± 0.0124	93.9688 ± 0.2174	88.6465 ± 0.6798
LadderNet	95.2380 ± 0.0025	90.2888 ± 0.0013	93.8664 ± 0.0442	88.4516 ± 0.1392
Attention U-Net	94.3214 ± 0.0022	90.1131 ± 0.0021	93.7055 ± 0.0596	88.1820 ± 0.1718
R2U-Net	96.4009 ± 0.0198	90.0257 ± 0.0212	93.2589 ± 0.1453	87.3972 ± 0.4360
CE-Net	93.9183 ± 0.0053	90.2682 ± 0.0044	93.8656 ± 0.0662	88.4520 ± 0.2045
MultiResUnet	94.2747 ± 0.0034	90.0523 ± 0.0033	93.4554 ± 0.0376	87.7331 ± 0.1178
Ours	**96.7074 ± 0.0133**	**94.9392 ± 0.0165**	**95.1314 ± 0.0428**	**90.7281 ± 0.1404**

Table 5 The effectiveness of different methods on the ISBI2014

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	43.6242 ± 0.9453	98.7000 ± 0.0704	58.7660 ± 8.0891	41.6654 ± 7.8531
RF	**99.6613 ± 0.0157**	99.0417 ± 0.0160	74.4123 ± 8.8632	59.3405 ± 14.1639
U-Net	64.2248 ± 0.0221	99.0075 ± 0.0259	70.3776 ± 1.6603	54.4548 ± 2.1660
Unet++	64.6362 ± 0.0255	99.3102 ± 0.0287	77.1738 ± 0.7480	63.0308 ± 1.2586
LadderNet	79.1710 ± 0.0211	99.4174 ± 0.0289	79.1248 ± 1.0990	65.7082 ± 1.9227
Attention U-Net	74.2248 ± 0.0214	99.3575 ± 0.0266	78.7765 ± 1.0274	65.0959 ± 1.7670
R2U-Net	83.1930 ± 0.0316	99.4765 ± 0.0322	85.2426 ± 3.1881	74.4710 ± 7.0570
CE-Net	69.7540 ± 0.0376	99.3556 ± 0.0378	79.7766 ± 0.7561	66.4709 ± 1.4315
MultiResUnet	79.1831 ± 0.2301	99.4263 ± 0.0011	79.1696 ± 6.9267	65.7778 ± 13.1382
Ours	95.4894 ± 0.0023	**99.8548 ± 0.0010**	**87.5352 ± 0.3273**	**77.9360 ± 0.7037**

Table 6 The effectiveness of different methods on the DRIVE

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	25.7824 ± 1.1102	91.1600 ± 1.1307	48.0032 ± 11.3300	31.6460 ± 8.4280
RF	37.8543 ± 0.1035	93.6950 ± 0.0926	56.6126 ± 2.8500	39.5019 ± 2.7741
U-Net	61.3522 ± 0.3146	95.2815 ± 0.3226	69.3849 ± 0.1229	53.1825 ± 0.1693
Unet++	74.0815 ± 0.2306	96.0579 ± 0.2239	76.1274 ± 0.0178	61.4999 ± 0.0309
LadderNet	60.4607 ± 0.0401	95.5301 ± 0.0339	70.0946 ± 0.0659	54.0112 ± 0.0914
Attention U-Net	72.8634 ± 0.3319	96.1265 ± 0.3127	77.1398 ± 0.0244	62.8265 ± 0.0430
R2U-Net	58.8284 ± 0.4123	95.2693 ± 0.3786	68.1106 ± 1.7839	52.0015 ± 2.1230
CE-Net	**77.3591 ± 0.7798**	95.6206 ± 0.7315	70.5756 ± 0.2766	54.6102 ± 0.3845
MultiResUnet	75.5403 ± 1.1938	94.1571 ± 1.6532	65.4252 ± 30.4548	49.7060 ± 31.8315
Ours	74.8853 ± 0.0228	**96.2081 ± 0.0193**	**78.5965 ± 0.1187**	**64.8962 ± 0.1451**

Table 7 The effectiveness of different methods on the CVC-ClinicDB(small)

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	28.3722 ± 255.0145	79.9018 ± 75.2226	44.7483 ± 5.8467	28.8542 ± 3.9968
RF	34.2975 ± 0.1609	80.8078 ± 0.2513	52.4319 ± 3.7327	35.5540 ± 3.1792
U-Net	86.3685 ± 0.0943	89.2632 ± 0.0088	81.9984 ± 12.9208	70.1967 ± 20.3121
Unet++	77.5476 ± 0.2620	90.7097 ± 0.0099	83.9432 ± 8.6908	74.2916 ± 14.7084
LadderNet	64.2563 ± 0.3570	89.0171 ± 0.0653	77.0148 ± 52.6166	64.7222 ± 63.3766
Attention U-Net	87.7648 ± 0.2818	91.2043 ± 0.0709	85.7505 ± 20.2002	75.8469 ± 29.7161
R2U-Net	76.8619 ± 1.4909	90.5436 ± 0.1160	83.3855 ± 26.6117	73.6234 ± 18.8510
CE-Net	**89.3081 ± 0.1145**	91.1891 ± 0.0391	85.5977 ± 4.7637	75.3646 ± 4.3133
MultiResUnet	81.3246 ± 0.0939	91.1840 ± 0.0158	85.0758 ± 3.7045	74.9459 ± 8.3660
Ours	85.4425 ± 0.2853	**91.2479 ± 0.0571**	**85.8079 ± 2.3763**	**76.8933 ± 2.9567**

Table 8 The effectiveness of different methods on the ISIC(small)

Method	Sen(%)	Acc(%)	Dice(%)	IoU(%)
SVM	97.7707 ± 2.2295	94.2875 ± 0.1060	88.9775 ± 1.1507	80.1604 ± 2.9734
RF	**98.3949 ± 0.0010**	95.3000 ± 0.1225	90.8312 ± 0.2879	83.2070 ± 0.8142
U-Net	87.4052 ± 0.0127	96.9820 ± 0.0141	94.0334 ± 0.7591	89.0805 ± 2.3496
Unet++	92.9950 ± 0.0111	96.9028 ± 0.0051	93.9671 ± 1.5560	89.1188 ± 4.9427
LadderNet	84.6721 ± 0.0054	95.1094 ± 0.0059	91.6466 ± 1.0944	85.1468 ± 2.8768
Attention U-Net	85.9732 ± 0.0216	94.9678 ± 0.0319	91.0932 ± 0.5476	84.2856 ± 1.1908
R2U-Net	78.0633 ± 0.0277	96.1950 ± 0.0292	92.8048 ± 2.3718	86.9698 ± 6.2732
CE-Net	86.6122 ± 0.0274	95.0386 ± 0.0225	91.4771 ± 4.3291	84.5272 ± 12.1439
MultiResUnet	91.6595 ± 0.1209	**97.4704 ± 0.1114**	**94.8670 ± 1.7577**	**90.4979 ± 5.0583**
Ours	90.8895 ± 0.0013	97.3184 ± 0.0024	94.4686 ± 1.5205	89.7871 ± 2.8238

the existing methods except for MultiResUnet according to Table 8. One possible reason for failure is that the parameter number of MultiResUnet are more larger than our framework. This is generally unfair. To this end, we reduce the parameter number for MultiResUnet to a situation similar to the proposed method. The related results are recorded in Table 9. Obviously, the proposed AU-MultiGAN method can obtain the improvement with roughly 0.75%, 1.1%, and 2.1% for Dice, IoU, and Acc in comparison with MultiResUnet with the same magnitude. These good performances largely benefit from the fact that the proposed multi-discriminator module and the use of dual-dilated block can guide the generator to capture more favorable information and thereby improve the segmentation performance. To sum up, the above-mentioned analyses indicate the superiority of the proposed method.

Moreover, from the view of visual effects, we present segmentation results of some representative images for these different datasets. The corresponding visual images are displayed in Figs. 2, 3, 4, 5, 6 and 7. It can be observed that our method performs well on various modalities of datasets. For example, Fig. 2 shows the result of ISBI2009, where some images of this dataset are with bright objects that are almost indistinguishable from the actual nuclei. Specially, the input image is polluted by small particles that are not actual cell nuclei. But the AU-MultiGAN method can segment the images reliably and acquire perfect segmentation in the regions of interest in comparison with other approaches. Similarly, we can see that the proposed method have obtain clearer boundaries or textures compared with other methods from Figs. 3–5, corresponding to the datasets ISBI2012, ISBI2014 and DRIVE, respectively. Moreover, for the ISIC and CVC-ClinicDB datasets, the segmentation tasks in Figs. 6 and 7 are more difficult. It seems that all the baselines have unsatisfactory results. However, most regions of the proposed framework are segmented accurately when comparing with other methods, although it still has few wrong segmentation part. These effective visual results are likely to depend on the proposed multi-discriminator module and dual-dilated block, which contribute to produce useful features in the deep model.

In summary, the proposed method can achieve superior performance in comparison with the baselines from both quantitative and visual results.

4.3 Discussion

1) **The impact of hyperparameters α and γ.** Adaptive selecting the scale factors α and γ is generally a challenging work. Here, we take the ISBI2012 dataset as an example. Table 10 depicts the effects of different values of α and γ in a relatively wide. As can be seen, the best results are in the case of $\alpha = 0.25$ and $\gamma = 1$. Consequently, we set them empirically.
2) **Choice of the discriminators number k.** We will discuss the influence of different number of discriminators

Table 9 The effects of MultiResUnet(small) and AU-MultiGAN with the similar parameter number on the ISIC(small)

Method	Parameters	Sen(%)	Acc(%)	Dice(%)	IoU(%)
MultiResUnet(samll)	4,413,212	**92.2344 ± 0.2409**	95.2134 ± 0.2113	93.7192 ± 2.5669	88.6647 ± 6.2637
AU-MultiGAN	4,249,156	90.8895 ± 0.0013	**97.3184 ± 0.0024**	**94.4686 ± 1.5205**	**89.7871 ± 2.8238**

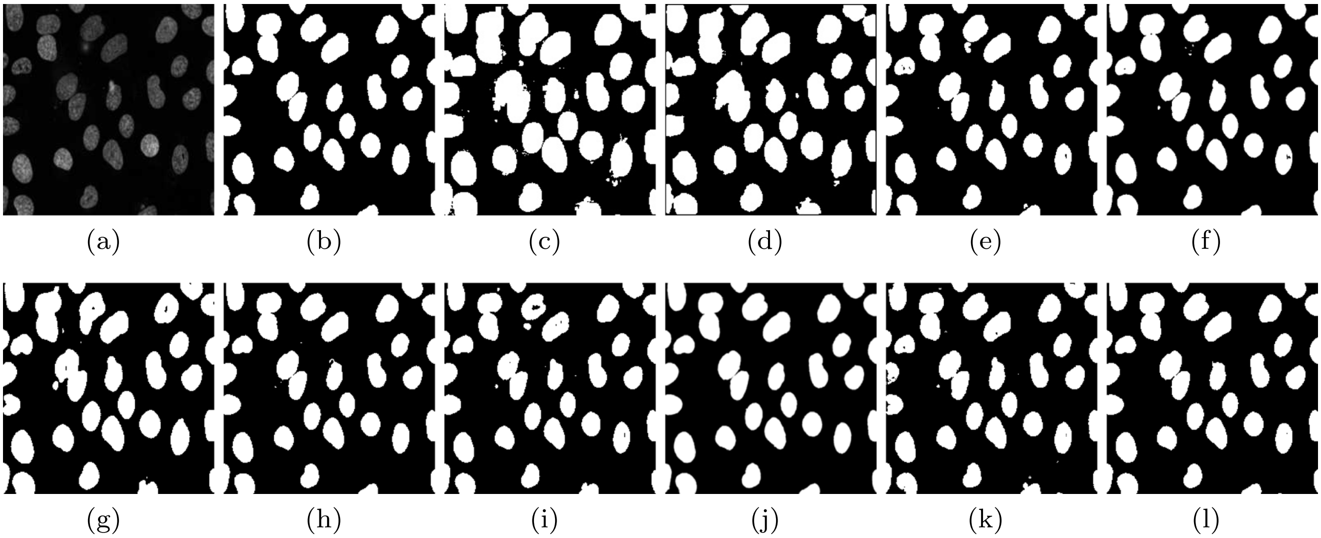


Fig. 2 Segmented images of different methods on ISBI2009. **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

k to determine the sizes in the overall network. To this end, we complete a set of comparative experiments with $k = \{1, 2, 3, 4\}$ on the ISBI2012 dataset, as shown in Table 11. It is easy to see that the performance of our framework is optimal when $k = 2$. Consequently, we select this value empirically.

3) **Effectiveness of the multi-discriminator mechanism and the dual-dilated block.** We consider the influence of the multi-discriminator mechanism and the dual-dilated block. To this end, some comparison experiments are carried out, taking the ISBI2012 dataset as an example. The baselines include four cases, i.e. only the asymmetric U-Net called MultiUnet, only the single discriminator model referred to as SingleGAN, the single-scale U-Net with a multi-discriminator named UGAN, and the proposed AU-MultiGAN. The results are presented in Table 12. Clearly, we can see that the multi-discriminator mechanism is effective when compared with SingleGAN and MultiUnet. Further, the dual-dilated block of the proposed method is also useful in comparison with UGAN, which use a single-scale block. These demonstrate that the multi-discriminator

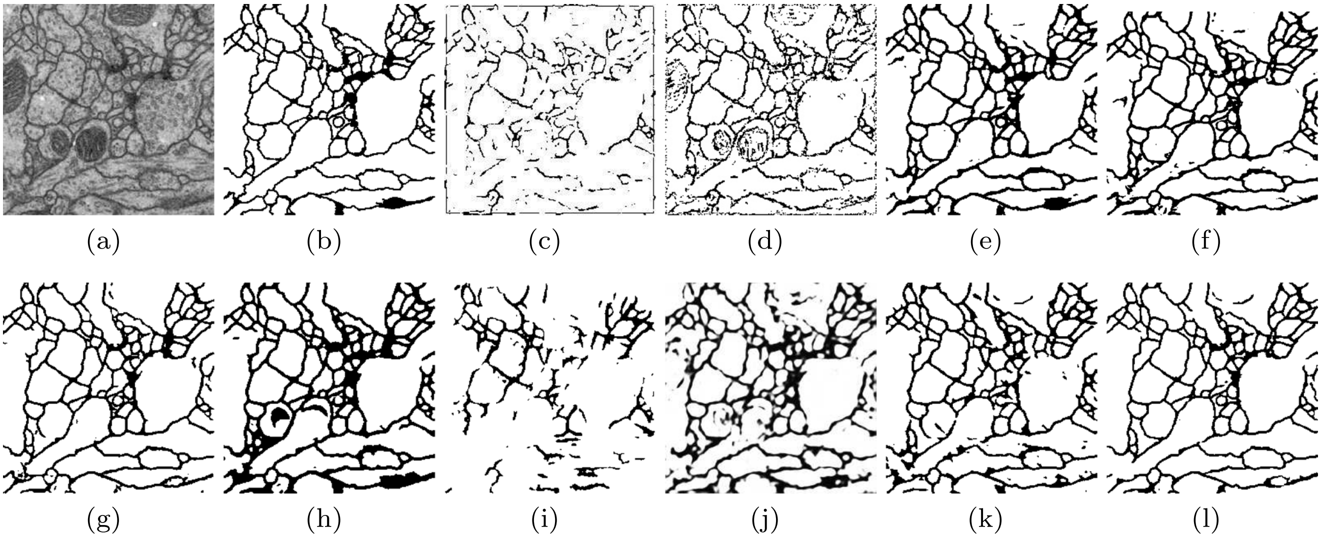


Fig. 3 Segmented images of different methods on ISBI2012. **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

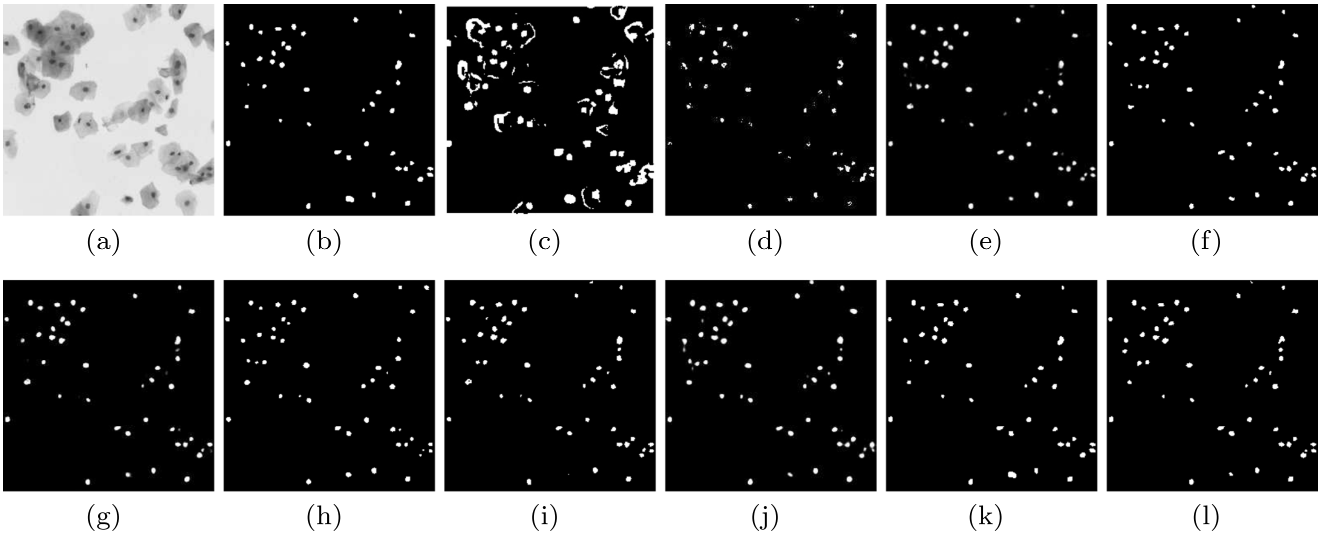


Fig. 4 Segmented images of different methods on ISBI2014. **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

and dual-dilated block are of importance in the proposed method.

4) **Ablation study of the hybrid loss.** We discuss the effectiveness of the proposed hybrid loss through three groups of experiments on the ISBI2012 dataset as an example. The first one only consider the discriminator loss L_{GAN} denoted by AU-MultiGAN (L_{GAN}). The second case embeds the Focal Loss into the discriminator loss L_{GAN}, referred to as AU-MultiGAN($L_{\mathrm{GAN}} + \lambda_1 L_{\mathrm{FL}}$). The last one is the proposed method. The corresponding results are listed in Table 13. It can be seen that the proposed method outperforms than other two cases. This verifies the effectiveness of the hybrid loss and may balance the intra-classes of samples so that keeping consistent with the generated and real segmentation maps.

5) **Convergence of AU-MultiGAN.** In Fig. 8, we plot the curves of the segmentation loss in each epoch on the first four datasets in Table 2. It can be seen that for all cases, the proposed model attains convergence quickly. This can be attributed to the synergy between the multi-discriminator mechanism and batch normalisation. These imply that the proposed AU-MultiGAN method are likely to obtain superior results in fewer training epochs and are consistent with the theoretical results.

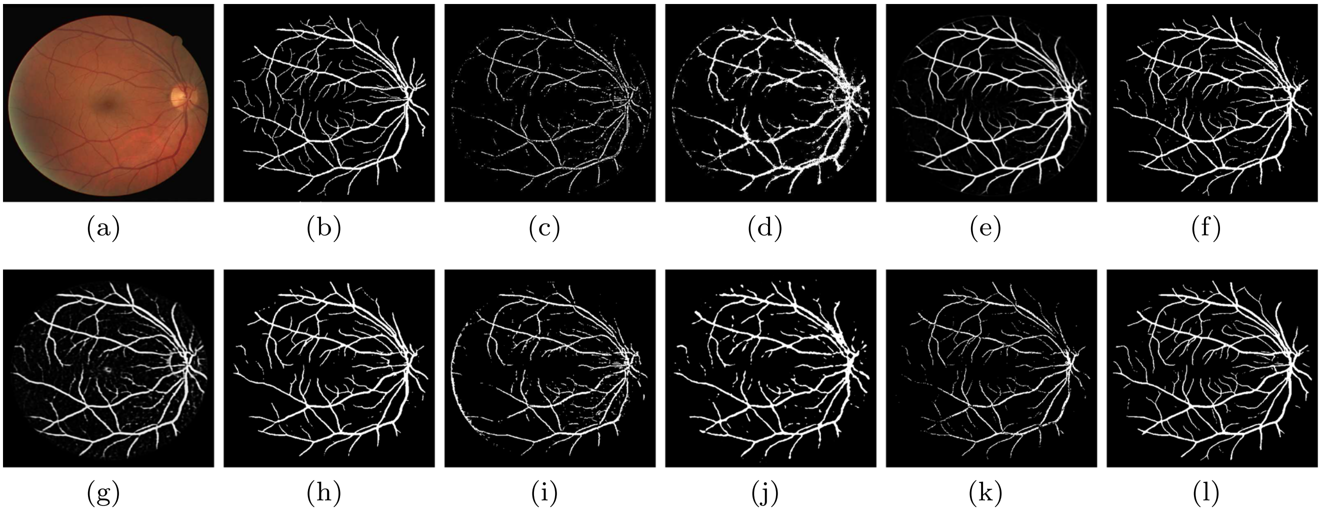


Fig. 5 Segmented images of different methods on DRIVE. **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

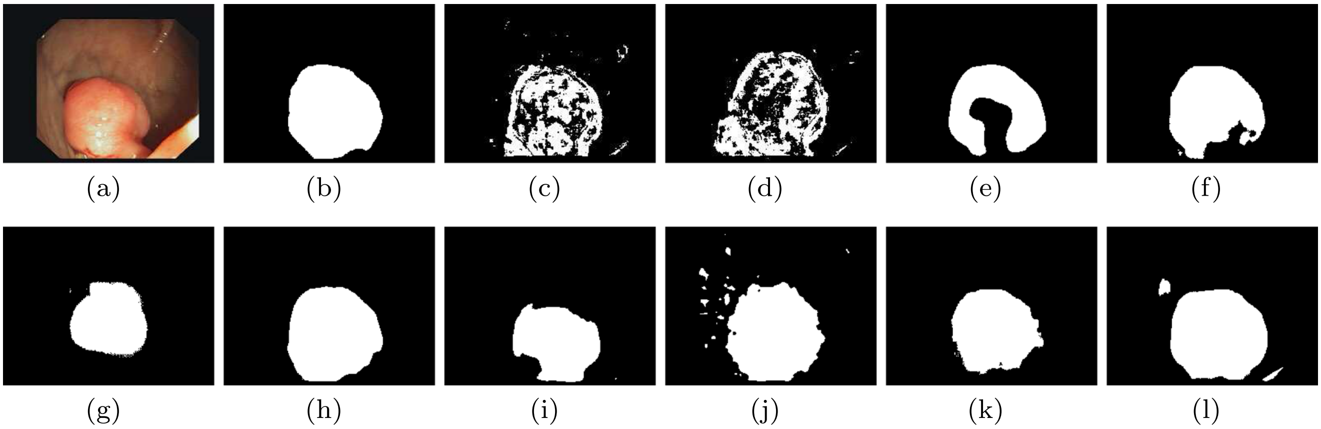


Fig. 6 Segmented images of different methods on CVC-ClinicDB(small). **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

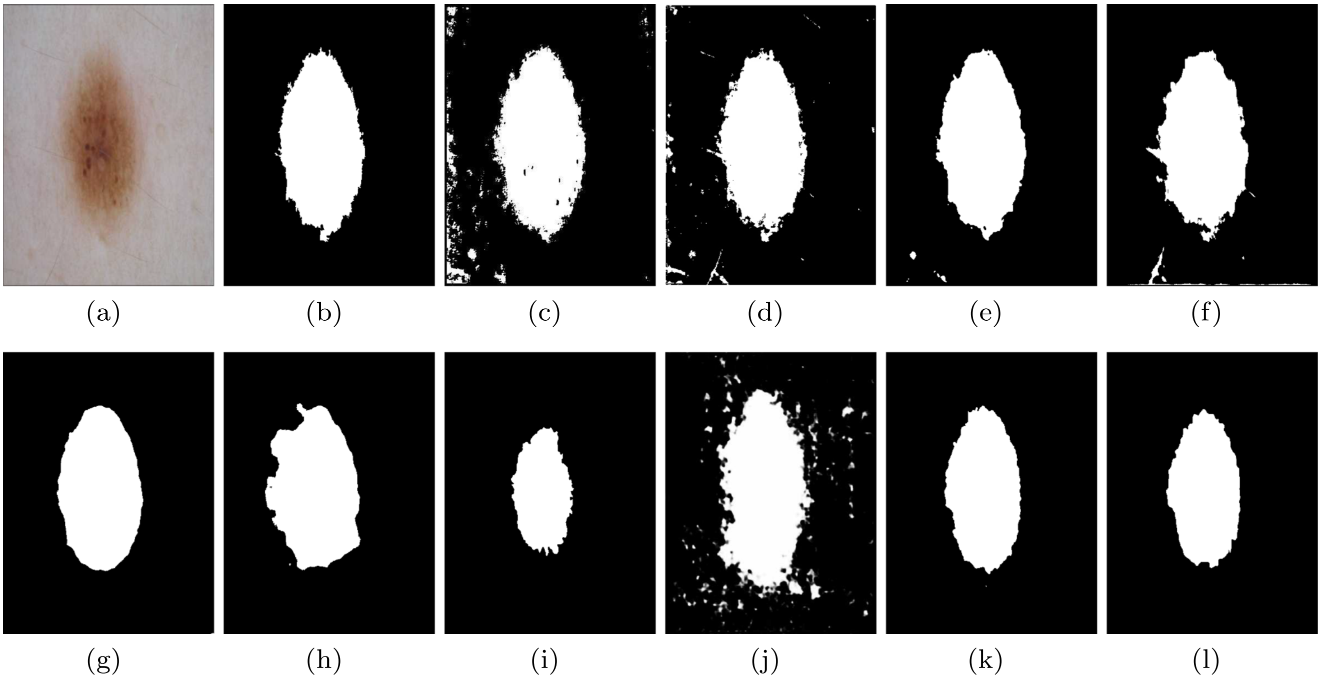


Fig. 7 Segmented images of different methods on ISIC(small). **a** Input image, **b** ground truth, **c** SVM, **d** RF, **e** U-Net, **f** Unet++, **g** LadderNet, **h** Attention U-Net, **i** R2U-Net, **j** CE-Net, **k** MultiResUnet, and **l** AU-MultiGAN

Table 10 Results on different scale factors for the ISBI2012 dataset

	$\alpha = 0$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
$\gamma = 0$	94.8267 ± 0.0214	95.0773 ± 0.0069	94.9994 ± 0.1538	94.9994 ± 0.1538	94.6308 ± 0.0611
$\gamma = 1$	94.8308 ± 0.0768	**95.1314 ± 0.0428**	94.9045 ± 0.0799	94.6978 ± 0.0437	94.8244 ± 0.0354
$\gamma = 2$	94.9313 ± 0.0142	95.0532 ± 0.0688	94.9869 ± 0.1140	93.1068 ± 0.1046	94.8652 ± 0.0468
$\gamma = 3$	94.9000 ± 0.0721	95.0828 ± 0.0205	94.8006 ± 0.0154	94.9894 ± 0.0542	94.9899 ± 0.0072

Table 11 AU-MultiGAN with various k values for the ISBI2012 dataset

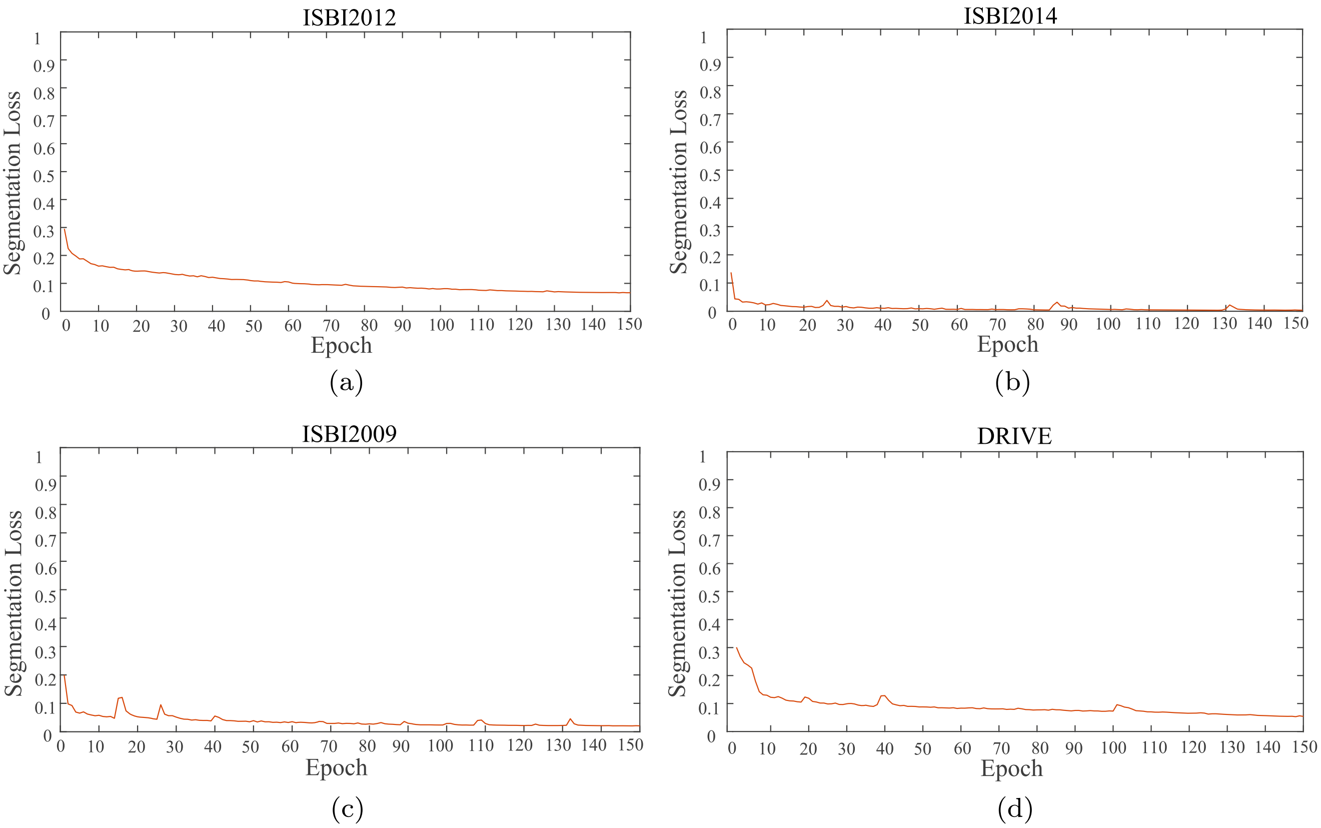
k	Parameters	Dice(%)	IoU(%)
1	1,341,410	94.9610 ± 0.2225	90.4131 ± 0.7053
2	4,249,156	**95.1314** ± **0.0428**	**90.7281** ± **0.1404**
3	13,795,110	94.9443 ± 0.0679	90.3861 ± 0.2164
4	49,888,520	94.9308 ± 0.0405	90.3613 ± 0.1331

Table 12 Effectiveness of the multi-discriminator mechanism and the dual-dilated block

Method	Parameters	Dice(%)	IoU(%)
MultiUnet	2,862,722	95.0370 ± 0.0408	90.5572 ± 0.1306
SingleGAN	3,555,939	94.9241 ± 0.0621	90.3430 ± 0.2022
UGAN	2,532,100	95.0595 ± 0.0256	90.5884 ± 0.0843
AU-MultiGAN	4,249,156	**95.1314** ± **0.0428**	**90.7281** ± **0.1404**

Table 13 Ablation study of the hybrid loss

Method	Dice(%)	IoU(%)
AU-MultiGAN(L_{GAN})	68.6179 ± 156.8310	53.6859 ± 229.5927
AU-MultiGAN($L_{\text{GAN}} + \lambda_1 L_{\text{FL}}$)	93.0980 ± 0.0612	87.1012 ± 0.1809
AU-MultiGAN($L_{\text{GAN}} + \lambda_1 L_{\text{FL}} + \lambda_2 L_{\text{re}}$)	**95.1314** ± **0.0428**	**90.7281** ± **0.1404**



(a)

(b)

(c)

(d)

Fig. 8 Convergence analysis of the proposed method for the former four datasets. **a** ISBI2012. **b** SBI2014. **c** ISBI2009. **d** DRIVE

Table 14 The model sizes of the proposed method and all the baselines

Model	U-Net	Unet++	LadderNet	Attention U-Net
Parameters	7,784,577	9,045,668	14,155,041	34,877,421
Model	R2U-Net	CE-Net	MultiResUnet	Ours
Parameters	6,214,209	38,969,176	7,251,322	4,249,156

6) **Analysis of model size.** We give the parameters for different methods. The results are recorded in Table 14. As can be seen, the parameter number of the proposed method is smaller than all the baselines, demonstrating that the proposed method is a lightweight framework.

5 Conclusion

We present a novel method based on GAN for medical image segmentation with small samples by designing multiple adversarial networks, referred to as AU-MultiGAN. This framework mainly contains an asymmetric U-Net module and a multi-discriminator module. The former is designed to produce multiple segmentation maps. Further, the multi-discriminator module is embedded into the asymmetric U-Net structure, capturing the available information of samples sufficiently and thereby promote the information transmission. Also, a hybrid loss is developed and an adaptive method of selecting the scale factors is designed. Simultaneously, the convergence of the proposed method is proved mathematically. Experimental results demonstrate that the effectiveness of proposed method surpasses the existing baselines.

There is scope for further research in this task. For example, it is feasible to introduce a more sophisticated architecture that can adapt to few-shot segmentation. In addition, to extend our work for the selection of hyperparameters, meta-learning [44] may be a viable approach.

Appendix A

A.1 Proof of Lemma 1

For a given generator G_i, the training criterion for the sub-discriminator $D_i (i = 1, 2, \ldots, k)$ is to minimize the discriminator loss, $L_{\mathrm{GAN}}(G_i, D_i)$. Let

$$\begin{aligned}
&L_{\mathrm{GAN}}(G_i, D_i)\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}\left[\frac{1}{2}(D_i(\boldsymbol{Y}_i) - 1)^2\right] + \mathbb{E}_{\boldsymbol{X} \sim p_{\boldsymbol{X}}}\left[\frac{1}{2}D_i^2(G_i(\boldsymbol{X}))\right]\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}[\frac{1}{2}(D_i(\boldsymbol{Y}_i) - 1)^2] + \mathbb{E}_{\boldsymbol{Y}_i' \sim p_{g_i}}\left[\frac{1}{2}D_i^2(\boldsymbol{Y}_i')\right]\\
&= \frac{1}{2}\int_{\boldsymbol{Y}_i} p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)(D_i(\boldsymbol{Y}_i) - 1)^2 + p_{g_i}(\boldsymbol{Y}_i)D_i^2(\boldsymbol{Y}_i)d\boldsymbol{Y}_i. \qquad (22)
\end{aligned}$$

By the formula in (22), the optimisation problem of the sub-discriminator can be transformed into a least squares problem:

$$\min_{D_i}(D_i(\boldsymbol{Y}_i) - 1)^2 p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + D_i^2(\boldsymbol{Y}_i)p_{g_i}(\boldsymbol{Y}_i). \qquad (23)$$

It achieves the minimum at $\frac{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)}{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)+p_{g_i}(\boldsymbol{Y}_i)}(i = 1, 2, \ldots, k)$ in [0, 1]. This completes the proof.

A.2 Proof of Lemma 2

For all $i(i = 1, 2, \ldots, k)$, if the relation $p_{g_i} = p_{\boldsymbol{Y}_i}$ is satisfied, then $D_i^* = \frac{1}{2}$ is calculated by (16). Hence,

$$\begin{aligned}
&\min_{D_i} L_{\mathrm{GAN}}(G_i, D_i)\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}[\frac{1}{2}(D_i^*(\boldsymbol{Y}_i) - 1)^2] + \mathbb{E}_{\boldsymbol{X} \sim p_{\boldsymbol{X}}}[\frac{1}{2}D_i^{*2}(G_i(\boldsymbol{X}))]\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}[\frac{1}{2}(D_i^*(\boldsymbol{Y}_i) - 1)^2] + \mathbb{E}_{\boldsymbol{Y}_i' \sim p_{g_i}}[\frac{1}{2}D_i^{*2}(\boldsymbol{Y}_i')]\\
&= \frac{1}{4}. \qquad (24)
\end{aligned}$$

To see that $\frac{1}{4}$ is the best possible value of $\min_{D_i} L_{\mathrm{GAN}}(G_i, D_i)$, reached only for $p_{g_i} = p_{\boldsymbol{Y}_i}$, we observe

$$\begin{aligned}
&\min_{D_i} L_{\mathrm{GAN}}(G_i, D_i)\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}[\frac{1}{2}(D_i^*(\boldsymbol{Y}_i) - 1)^2] + \mathbb{E}_{\boldsymbol{X} \sim p_{\boldsymbol{X}}}[\frac{1}{2}D_i^{*2}(G_i(\boldsymbol{X}))]\\
&= \mathbb{E}_{\boldsymbol{Y}_i \sim p_{\boldsymbol{Y}_i}}[\frac{1}{2}(D_i^*(\boldsymbol{Y}_i) - 1)^2] + \mathbb{E}_{\boldsymbol{Y}_i' \sim p_{g_i}}[\frac{1}{2}D_i^{*2}(\boldsymbol{Y}_i')]\\
&= \frac{1}{2}\int_{\boldsymbol{Y}_i}[p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)(D_i^*(\boldsymbol{Y}_i) - 1)^2 + p_{g_i}(\boldsymbol{Y}_i)D_i^{*2}(\boldsymbol{Y}_i)]d\boldsymbol{Y}_i,
\end{aligned} \qquad (25)$$

where the relationship between $D_i^{*2}(\boldsymbol{Y}_i)$, the label distribution $p_{\boldsymbol{Y}_i}$ and the generated segmentation distribution p_{g_i} are obtained in Lemma 1. Here, we introduce this relationship into the above formula.

$$\begin{aligned}
&\frac{1}{2}\int_{\boldsymbol{Y}_i}\left[\frac{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)p_{g_i}^2(\boldsymbol{Y}_i)}{(p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i))^2} + \frac{p_{g_i}(\boldsymbol{Y}_i)p_{\boldsymbol{Y}_i}^2(\boldsymbol{Y}_i)}{(p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i))^2}\right]d\boldsymbol{Y}_i\\
&= \frac{1}{4} + \frac{1}{8}\int_{\boldsymbol{Y}_i}\left[\frac{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)p_{g_i}(\boldsymbol{Y}_i) - p_{\boldsymbol{Y}_i}^2(\boldsymbol{Y}_i)}{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)} + \frac{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i)p_{g_i}(\boldsymbol{Y}_i) - p_{g_i}^2(\boldsymbol{Y}_i)}{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)}\right]d\boldsymbol{Y}_i\\
&= \frac{1}{4} - \frac{1}{8}\int_{\boldsymbol{Y}_i}\left[\frac{(2p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) - (p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)))^2}{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)}\right]d\boldsymbol{Y}_i\\
&= \frac{1}{4} - \frac{1}{8}\chi^2(p_{\boldsymbol{Y}_i} + p_{g_i} \| 2p_{\boldsymbol{Y}_i}), \qquad (26)
\end{aligned}$$

where χ^2 is the Pearson χ^2 divergence and $\chi^2(p_{\boldsymbol{Y}_i} + p_{g_i} \| 2p_{\boldsymbol{Y}_i})$ denotes the simplified representation of $\int_{\boldsymbol{Y}_i}\left[\frac{(2p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) - (p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)))^2}{p_{\boldsymbol{Y}_i}(\boldsymbol{Y}_i) + p_{g_i}(\boldsymbol{Y}_i)}\right]d\boldsymbol{Y}_i$. Thus, the results of

(26) achieve the value $\frac{1}{4}$ when p_{Y_i} and p_{g_i} are equal. We have shown that $L_{\text{GAN}} = \frac{1}{4}$ and that the only solution is $p_{g_i} = p_{Y_i}$. Thus, the asymmetric U-Net can perfectly replicate the distribution of the real segmented image. This completes the proof.

A.3 Proof of Theorem 1

Consider $L_{\text{GAN}}(G_i, D_i) = \text{V}(p_{g_i}, D_i)$ as a function of p_{g_i}, as done in the above criterion (17), in which $\text{V}(p_{g_i}, D_i)$ is the criterion. Note that $\text{V}(p_{g_i}, D_i)$ is convex on p_{g_i}. The inf-derivatives of an infimum of convex functions are the derivative of the function at the point where the minimum is attained. This is equivalent to computing a gradient descent update for p_{g_i} at the optimal D_i, given the corresponding G_i. From [45], $\inf_{D_i} L_{\text{GAN}}(G_i, D_i)$ is convex on p_{g_i}. Moreover, $\inf_{D_i} L_{\text{GAN}}(G_i, D_i)$ takes the value $\frac{1}{4}$ as proven in Lemma 2; therefore, with sufficiently small updates of p_{g_i}, p_{g_i} converges to p_{Y_i} $(i = 1, 2, \ldots, k)$. This completes the proof.

Acknowledgments This work was supported by the National Natural Science Foundation of China under grant 61672477, and the Zhejiang Provincial Natural Science Foundation of China under grant LZ20F030001.

References

1. Barkana BD, Saricicek I, Yildirim B (2017) Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ANN, SVM, and classifier fusion. Knowledge-Based Syst 118:165–176. https://doi.org/10.1016/j.knosys.2016.11.022
2. Mitra J, Bourgeat P, Fripp J, Ghose S, Rose S, Salvado O, Connelly A, Campbell B, Palmer S, Sharma G et al (2014) Lesion segmentation from multimodal MRI using random forest following ischemic stroke. Neuroimage 98:324–335. https://doi.org/10.1016/j.neuroimage.2014.04.056
3. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88. https://doi.org/10.1016/j.media.2017.07.005
4. Yu H, Yang Z, Tan L, Wang Y, Sun W, Sun M, Tang Y (2018) Methods and datasets on semantic segmentation: a review. Neurocomputing 304:82–103. https://doi.org/10.1016/j.neucom.2018.03.037
5. Cao F, Liu H (2019) Single image super-resolution via multi-scale residual channel attention network. Neurocomputing 358:424–436. https://doi.org/10.1016/j.neucom.2019.05.066
6. Zhang J, Gu Y, Tang H, Wang X, Kong Y, Chen Y, Shu H, Coatrieux J (2020) Compressed sensing MR image reconstruction via a deep frequency-division network. Neurocomputing 384:346–355. https://doi.org/10.1016/j.neucom.2019.12.011
7. Cao F, Guo W (2020) Cascaded dual-scale crossover network for hyperspectral image classification. Knowledge-Based Syst 189:105122. https://doi.org/10.1016/j.knosys.2019.105122
8. Li Z, Dong M, Wen S, Hu X, Zhou P, Zeng Z (2019) Clu-cnns: object detection for medical images. Neurocomputing 350:53–59. https://doi.org/10.1016/j.neucom.2019.04.028
9. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: Proceedings of international conference on medical image computing and computer-assisted intervention. Springer, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
10. Fang L, Wang X, Wang L (2020) Multi-modal medical image segmentation based on vector-valued active contour models. Inf Sci 513:504–518. https://doi.org/10.1016/j.ins.2019.10.051
11. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of IEEE conference on computer vision and pattern recognition. IEEE, pp 3431–3440. https://doi.org/10.1109/TPAMI.2016.2572683
12. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
13. Wang N, Zhang Z, Xiao J, Cui L (2019) DeepLap: a deep learning based non-specific low back pain symptomatic muscles recognition system. In: 2019 16th annual IEEE international conference on sensing, communication, and networking (SECON). IEEE, pp 1–9. https://doi.org/10.1109/SAHCN.2019.8824868
14. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations
15. Roy AG, Siddiqui S, Pölsterl S, Navab N, Wachinger C (2020) 'Squeeze & excite' guided few-shot segmentation of volumetric images. Med Image Anal 59:101587. https://doi.org/10.1016/j.media.2019.101587
16. Cui H, Wei D, Ma K, Gu S, Zheng Y (2020) A unified framework for generalized low-shot medical image segmentation with scarce data. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2020.3045775
17. Ibtehaz N, Rahman MS (2020) MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks 121:74–87. https://doi.org/10.1016/j.neunet.2019.08.025
18. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C (2016) The importance of skip connections in biomedical image segmentation: 179–187. https://doi.org/10.1007/978-3-319-46976-8_19
19. Yu L, Yang X, Chen H, Qin J, Heng PA (2017) Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. In: Proceedings of thirty-first AAAI conference on artificial intelligence. ACM, pp 66–72
20. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet++: a nested U-Net architecture for medical image segmentation. In: Proceedings of deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
21. Zhuang J (2018) LadderNet: multi-path networks based on U-Net for medical image segmentation. arXiv:1810.07810
22. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B et al (2018) Attention U-Net: learning where to look for the pancreas. In: Proceedings of medical imaging with deep learning
23. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans Med Imaging 38(10):2281–2292. https://doi.org/10.1109/TMI.2019.2903562

24. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-Net for medical image segmentation. J Med Imaging 6(1):014006. https://doi.org/10.1117/1.JMI.6.1.014006
25. Zhang C, Shu H, Yang G, Li F, Wen Y, Zhang Q, Dillenseger JL, Coatrieux JL (2020) HIFUNet: multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning. IEEE Trans Med Imaging 39(11):3309–3320. https://doi.org/10.1109/TMI.2020.2991266
26. Luc P, Couprie C, Chintala S, Verbeek J (2016) Semantic segmentation using adversarial networks. In: NIPS workshop on adversarial training
27. Xie H, Lei H, Zeng X, He Y, Chen G, Elazab A, Yue G, Wang J, Zhang G, Lei B (2020) Amd-gan: attention encoder and multi-branch structure based generative adversarial networks for fundus disease detection from scanning laser ophthalmoscopy images. Neural Networks 132:477–490. https://doi.org/10.1016/j.neunet.2020.09.005
28. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, Liu T, Yang X (2019) Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Medical Physics 46(5):2157–2168. https://doi.org/10.1002/mp.13458
29. Negi A, Raj ANJ, Nersisson R, Zhuang Z, Murugappan M (2020) RDA-UNET-WGAN: an accurate breast ultrasound lesion segmentation using wasserstein generative adversarial networks. Arabian Journal for Science and Engineering 45:6399–6410
30. Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of ieee international conference on computer vision, pp 2980–2988. https://doi.org/10.1109/iccv.2017.324
31. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of IEEE conference on computer vision and pattern recognition. pp 1874–1883. https://doi.org/10.1109/CVPR.2016.207
32. De Haan L, Ferreira A (2007) Extreme value theory: an introduction. Springer Science & Business Media
33. Coelho LP, Shariff A, Murphy RF (2009) Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In: Proceedings of IEEE international symposium on biomedical imaging: from nano to macro. IEEE, pp 518–521. https://doi.org/10.1109/ISBI.2009.5193098
34. Arganda-Carreras I, Turaga SC, Berger DR, Cireşan D, Giusti A, Gambardella LM, Schmidhuber J, Laptev D, Dwivedi S, Buhmann JM et al (2015) Crowdsourcing the creation of image segmentation algorithms for connectomics. Front Neuroanat 9:142. https://doi.org/10.3389/fnana.2015.00142
35. Cardona A, Saalfeld S, Preibisch S, Schmid B, Cheng A, Pulokas J, Tomancak P, Hartenstein V (2010) An integrated micro-and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol 8(10):1000502. https://doi.org/10.1371/journal.pbio.1000502
36. Lu Z, Carneiro G, Bradley AP, Ushizima D, Nosrati MS, Bianchi AG, Carneiro CM, Hamarneh G (2016) Evaluation of three algorithms for the segmentation of overlapping cervical cells. IEEE J Biomed Health Inform 21(2):441–450. https://doi.org/10.1109/JBHI.2016.2519686
37. Lu Z, Carneiro G, Bradley AP (2015) An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. IEEE Trans Image Process 24(4):1261–1272. https://doi.org/10.1109/TIP.2015.2389619
38. Staal JJ, Abramoff M, Niemeijer M, Viergever M, van Ginneken B (2004) Drive: digital retinal images for vessel extraction. IEEE Trans Med Imaging 23(4):501–509
39. Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H et al (2018) Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: Proceedings of IEEE 15th international symposium on biomedical imaging, pp 168–172. https://doi.org/10.1109/ISBI.2018.8363547
40. Tschandl P, Rosendahl C, Kittler H (2018) The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 5(1):1–9. https://doi.org/10.1038/sdata.2018.161
41. Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F (2015) Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics 43:99–111. https://doi.org/10.1016/j.compmedimag.2015.02.007
42. Fausto M, Nassir N, Seyed-Ahmad A (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of 2016 fourth international conference on 3D vision. IEEE, pp 565–571. https://doi.org/10.1109/3DV.2016.79
43. Hassan T, Akram MU, Werghi N, Nazir N (2020) Rag-fw: A hybrid convolutional framework for the automated extraction of retinal lesions and lesion-influenced grading of human retinal pathology. IEEE J Biomed Health Inform 24(99):1–1. https://doi.org/10.36227/techrxiv.11877879.v1
44. Shu J, Xie Q, Yi L, Zhao Q, Zhou S, Xu Z, Meng D (2019) Meta-weight-net: learning an explicit mapping for sample weighting. In: Advances in neural information processing systems, pp 1919–1930
45. Boyd S, Boyd SP, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Yi Wang received the B.Sc. degree in mathematics and applied mathematics from Leshan Normal University, Leshan, China, in 2018. She is currently pursuing the M.Sc. degree in the College of Sciences, China Jiliang University, Hangzhou, China. Her research interests include deep learning and image processing.

Hailiang Ye received the B.Sc. and M.Sc. degree in applied mathematics from China Jiliang University, Hangzhou, China, in 2012 and 2015, respectively. In 2019, he received the Ph.D. degree in computational mathematics from Huazhong University of Science and Technology, Wuhan, China. He is currently a lecturer of the College of Sciences, China Jiliang University, Hangzhou, China. His research interests include deep learning, pattern recognition, and image processing.

Feilong Cao received the Ph.D. degree in Applied Mathematics from Xi'an Jiaotong University, China in 2003. He was a Research Fellow with the Center of Basic Sciences, Xi'an Jiaotong University, China, from 2003 to 2004. From 2004 to 2006, he was a Post-Doctoral Research Fellow with the School of Aerospace, Xi'an Jiaotong University, China. He is currently a Professor of the College of Sciences, China Jiliang University, Hangzhou, China. He has authored or co-authored over 230 scientific papers in refereed journals. His current research interests include neural networks, pattern recognition, and approximation theory.