Multi-Model Medical Image Segmentation Using Multi-Stage Generative Adversarial Network

Afifa Khaled, Jian-Jun Han

***Abstract*— Image segmentation is a new challenge problem in medical application. The use of medical imaging has become an integral part of research, as it allows us to see inside the human body without surgical intervention. Many researcher have studied brain segmentation, where one stage method is used to segment the brain tissues.**

In this paper, we propose the multi-stage generative adversarial network(GAN) to solve the problem of information loss in the one-stage. We utilize the coarse-to-fine to improve brain segmentation using multi-stage generative adversarial networks. In the first stage, our model generates a coarse outline for (i) background and (ii) brain tissues. Then, in the second stage, the model generates outline for (i) white matter (WM), (ii) gray matter (GM) and (iii) cerebrospinal fluid (CSF). A good result can be achieved by fusing the coarse outline and refine outline.

We conclude that our model is more efficient and accurate in practice for brain segmentation of both infant and adult. Moreover, we observe that multi-stage model is faster than prior models. To be more specific, the main goal of multi-stage model is to obtain better performance of the model in a few shot learning case where a few labeled data are available. For medical image, this proposed model can work in a wide range of image segmentation where the convolution neural networks and one-stage methods have failed.

***Index Terms*— Brain segmentation, Coarse-to-fine, Generative Adversarial Network, Semi-supervised learning, Multi-stage method.**

I. Introduction

Magnetic resonance imaging (MRI) uses magnetic field to generate detailed images of tissues without using harmful radiation [1] [2]. However, the process of manual segmentation in clinical is time consuming and expensive [3].

Automated segmentation of infant and adult brain has received a substantial research attentions [4] [5]. Training deep models need for large sets of labeled images [6]. Due to the small data sets in the medical application [7] [8], semi supervisor learning approaches solved this problem by using unlabeled image [9] [10]. A good segmentation result can be achieved by adopting unable images [11], or images with weak annotation like image level tag [12]. In the object detection, one-stage method is used to predict the class probability and the position information [13] [14]. To take advantage of the recent success of two-stage method, many models proposed for semantic segmentation. Recently, Xiaohao et al. proposed a two-Stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding. In the computer vision tasks, two-stage method is exploited to generate global information in the first stage and local information in the second stage [15] [16]. A good result can be achieved by fusing the global information and local information [17] [18]. In addition, the adoption of multi-stage generative adversarial network in medical imaging is still in its infancy.

In this paper, we utilize the coarse-to-fine method to improve brain segmentation using multi-stage generative adversarial networks. The main contributions of this paper as follows:

1- In the first stage, our model generates a coarse outline for (i) background, (ii) brain tissues. The main role of the first G is to generate coarse segmentation to be use as guidance information for the third G.

2- In the second generator, the generator takes two input, image x and random vector z. The main idea is to encourage the generator to generate as many different values for each x as there are values of z.

3- The third generator consists of the encoder and decoder. In the encoder and decoder we also used the dense skip connection to combine the features from different scales. The third generator generated outline for (i) white matter (WM), (ii) gray matter (GM) and (iii) cerebrospinal fluid (CSF), which is similar with the process of human learning in clinical practice. To be more specific, the main role of the third G is to generate more detailed results by using the coarse segmentation from the first G.

We evaluate the proposed multi-stage generative adversarial model on two datasets of brain tissues, including infant and adult brain. Empirically, our model achieves a good result compared to the state-of-the-art models.

The rest of this paper is organized as follows. Section 2 presents prior studies related to brain segmentation. Section 3 presents the methodologies used in our paper. Section 4 presents our experimental results. Section 5 discusses threats to the validity of our results. Finally, Section 6 concludes this work and discusses directions for future work.

II. RELATED WORK

The following subsections present the prior studies and technique related to brain segmentation. We start by giving a more detailed description of semi supervised learning. In the subsection B, we introduce generative adversarial network. Finally, we show how developed loss functions for GAN improved the stability of training GAN model.

The authors are with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China. E-mail: afifakhaied@tju.edu.cn, asonhan@hust.edu.cn

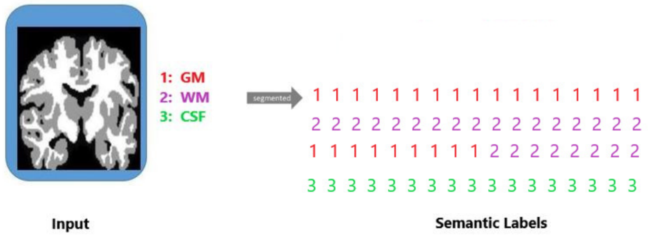

Fig. 1. The Illustration of semantic labels

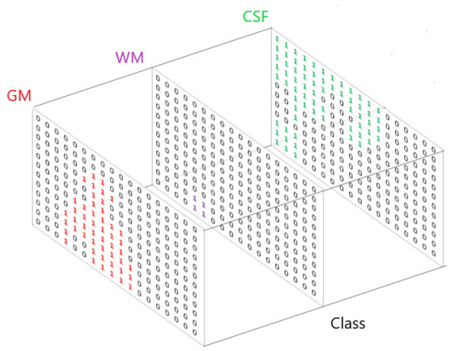

Fig. 2. The Illustration of semantic classes

A. Semi-supervised learning

Training a deep model by using a small datasets may cause over fitting [11] [19]. Preventing Over-fitting can be achieved by exploiting a large amount of unlabeled data with a small amount of labeled data [20] [21]. Training the model using labeled and unlabeled data this encourages the network to have a similar distribution [22] [23].

To be more specific, the idea in the semantic segmentation is to take a image and output a segmentation map [16] [24] [25]. Fig.1 and Fig.2 show the semantic segmentation labeles and semantic segmentation classes respectively.

Up to date, many researchers applied Semantic segmentation for brain, and in particular segmentation of brain tissues. [26] proposed ROAM, a random layer mix up, which allows the network to be less confident for interpolated data points at any selected space. Christina et al. proposed two novel architecture for brain tumor segmentation. Their results have been evaluated using the pinnacle BraTS confront2017 datasets. Similar to the above model, [27] proposed rethinking atrous convolution for semantic image. Different from the above model, rethinking atrous convolution model targets long range context. The model does not require convolution layers. Instead, it utilizes s atrous convolution with up sampled filters to extract dense feature maps. The model evaluate on the PASCAL VOC 2012 semantic image segmentation benchmark including the 3475 finely annotated images and the extra 20000 coarsely annotated images. Their experimental results of the sentiment task show that atrous convolution is necessary when building more blocks cascadedly. The authors also show that the performance improves as more blocks are added.

TABLE I. Some models of generative adversarial network applied to medical applications

Publication	Method
Han et al. [31]	WGAN
Bowles et al. [32]	PGGAN
Andrew et al. [33]	LAPGAN
Mondal et al. [34]	DCGAN
Kang et al. [35]	CycleGAN
Chuquicusma et al. [36]	DCGAN
Frid-Adar et al. [37]	DCGAN/ACGAN

B. Generative Adversarial Network

Generative adversarial network has shown great promise for medical image diagnostics [28]. To be more specific, in brain segmentation [25] [21]. Fig.3 shows an overview of generative adversarial networks.

Up to date, many researchers applied generative adversarial network for brain segmentation. [29] proposed a 3D volume-to-volume generative adversarial network (GAN) for segmentation of brian tumors. Their model achieved a good result when the generator loss is weighted 5 times higher than the discriminator loss. The proposed model has been evaluated on the BraTS 2018 datasets. Their models outperformed previous models with an overall 0.66%. A Super resolution and segmentation using generative adversarial networks is a framework introduced by [30] to neonatal brain MRI. It consists training a generating network that estimates for a given input image to its corresponding HR, and a discriminator network D is designed to distinguish real HR and segmentation images. In Table I, we provide some models of generative adversarial network applied to medical applications.

C. loss functions

To improve the stability of training GAN model, many researchers have developed the loss function [38] [39]. Due to the effectiveness of the loss function for a given model implemention, in this section we summarize five loss functions for GAN.

1) Minimax GAN loss:

The discriminator tries to maximize the loss functions and the generator tries to minimize it.

Generator loss function:

$$\mathrm{l}_{\mathrm{D}}^{GAN} = -E_{\mathrm{z}\sim p_d}[logD(x)] - E_{\mathrm{x}\sim p_g}[log(1-d(x))][41], \quad (1)$$

Discriminator loss function:

$$\mathrm{l}_{\mathrm{G}}^{GAN} = -E_{\mathrm{z}\sim pg}[log(1-D(X))][41], \quad (2)$$

In this function:

$D(x)$ denotes the discriminator's estimate of the probability that real data x is real.

$E(x)$ denotes the expected value over all real data.

$G(z)$ denotes the generator's output when given noise z.

$D(G(z))$ denotes the discriminator's estimate of the probability that a fake data is real.

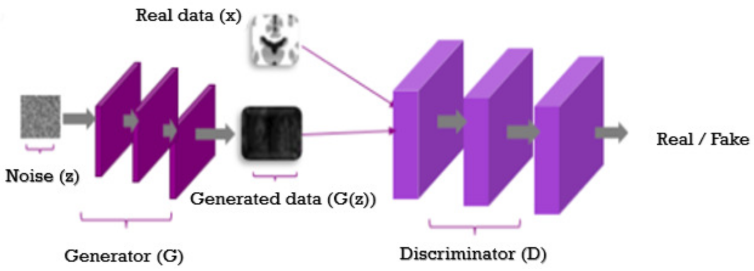

Fig. 3. The Illustration of Generative Adversarial Networks(GAN)

$E(z)$ denotes the expected value over all generated fake data $G(z)$.

2) Non-saturating loss (NSGAN):

Non-saturating loss is used to solve the saturation problem.

Generator loss function:

$$l_{\mathrm{D}}^{NSGAN} = -E_{\mathrm{z}\sim p_d}[logD(x)] - E_{\mathrm{x}\sim p_g}[log(1-D(x))][45], \tag{3}$$

Discriminator loss function:

$$l_{\mathrm{G}}^{NSGAN} = -E_{\mathrm{z}\sim p_g}[logD(x)][45], \tag{4}$$

3) Wassersstein loss (WGAN):

Generative adversarial networks (GANs) are useful in the area of computer vision [40] [41], but the main problem is suffering from training instability [27]. Many researchers have developed loss functions toward stable training of GAN [34]. Wassersstein ($WGAN$) achieves a good progress for training stability of GAN, but still suffer from poor result. Many researchers have argue that the problems of poor result due to the use of weight clipping. Hence, [42] proposed a way to clipping weights. This model is modification of the standard GAN. The output of the discriminator is a number. The discriminator training tries to make the output bigger for real data than for fake data.

Discriminator Loss:

$$l_{\mathrm{D}}^{WGAN} = -E_{\mathrm{z}\sim p_d}[D(x)] - E_{\mathrm{x}\sim p_g}[D(x)][46], \tag{5}$$

Generator Loss:

$$l_{\mathrm{G}}^{NSGAN} = -E_{\mathrm{z}\sim p_g}[D(x)][46], \tag{6}$$

In these functions:
$D(x)$ denotes the discriminator's output for a real data.
$G(z)$ denotes the generator's output when given noise z.
$D(G(z))$ denotes the discriminator's output for a fake data.
To be more specific, the output of the discriminator does not have to be between 0 and 1. for more information we refer reader to read [42].

4) Least-squares loss (LSGAN):

This model proposed $a - b$ coding scheme for the discriminator where a, b denote to the labels of fake and real data.

Discriminator Loss:

$$l_{\mathrm{D}}^{LSGAN} = -E_{\mathrm{z}\sim p_d}[D(x)-1]^2 - E_{\mathrm{x}\sim p_g}[D(x^2)][47], \tag{7}$$

Generator Loss:

$$l_{\mathrm{G}}^{LSGAN} = -E_{\mathrm{z}\sim p_g}[D(x-1)^2][47], \tag{8}$$

5) Wassersstein Gradient Penalty loss (AC-GAN):

AC-GAN uses the noise z and sample has class label $c \sim p$. This model is modification of the standard GAN. In the standard GAN, X_{fake}= $G(Z)$, but in $AC - GAN$ X_{fake}=$G(c, z)$. Moreover, the outputs of standard GAN is a probability distribution $P(s, x) = D(x)$, but in $AC - GAN$ the output is two probability distribution.

$P(s, x), P(c, x) = D(x)$.

Arnab et al. introduce a model that use generative adversarial network to brain segmentation [34]. The authors use a dataset of 43 subjects. First, the authors generat fake images by using generator. Second, the authors use the labeled data, unlabled data and fake data to train the discriminator to distinguish between generated data and true data. While the encoder is used to compute the predicted noise mean and log-varaince. The approach proposed by Arnab et al. supports only one-stage, while our model supports multi-stage. Our paper aims to solve the problem of loss informations in the one-stage. The first generator generated coarse outline to be use in the third generator. Moreover, the encoder and decoder generated fine outline. We also used dense skip connection to combine the features from different scales. To validate the idea of multi-stage, we used Dice coefficient.

III. Methodology

In this section, we present the design of multi-stage model. In subsection A, a more detailed description of the generative adversarial network approach that is used in our work. In the subsection B, a more detailed description of the loss functions for the discriminator and the generator that is used in our work. Table II shows the list of defined symbols in the paper.

TABLE II. List of defined symbols in the paper

Symbols	Definition
G	Generator
D	Discriminator
z	Noise
$G(z)$	Generated data
x	Real data
WM	White matter
GM	Gray matter
CSF	Cerebrospinal fluid
$Conv$	Convolutional
$LeReLU$	Activation function
GAN	Generative adversarial network
E	Expected value
DC	Dice Coefficient
MRI	Magnetic resonance imaging
$T1$	subject-1-to-subject-10
T	subject-11-to-subject-23
$NSGAN$	Non-saturating loss
$WGAN$	Wassersstein loss
$lSGAN$	Least-squares loss
$ACGAN$	Wassersstein Gradient Penalty loss
V_auto	Automated segmentation
V_ref	Reference segmentation

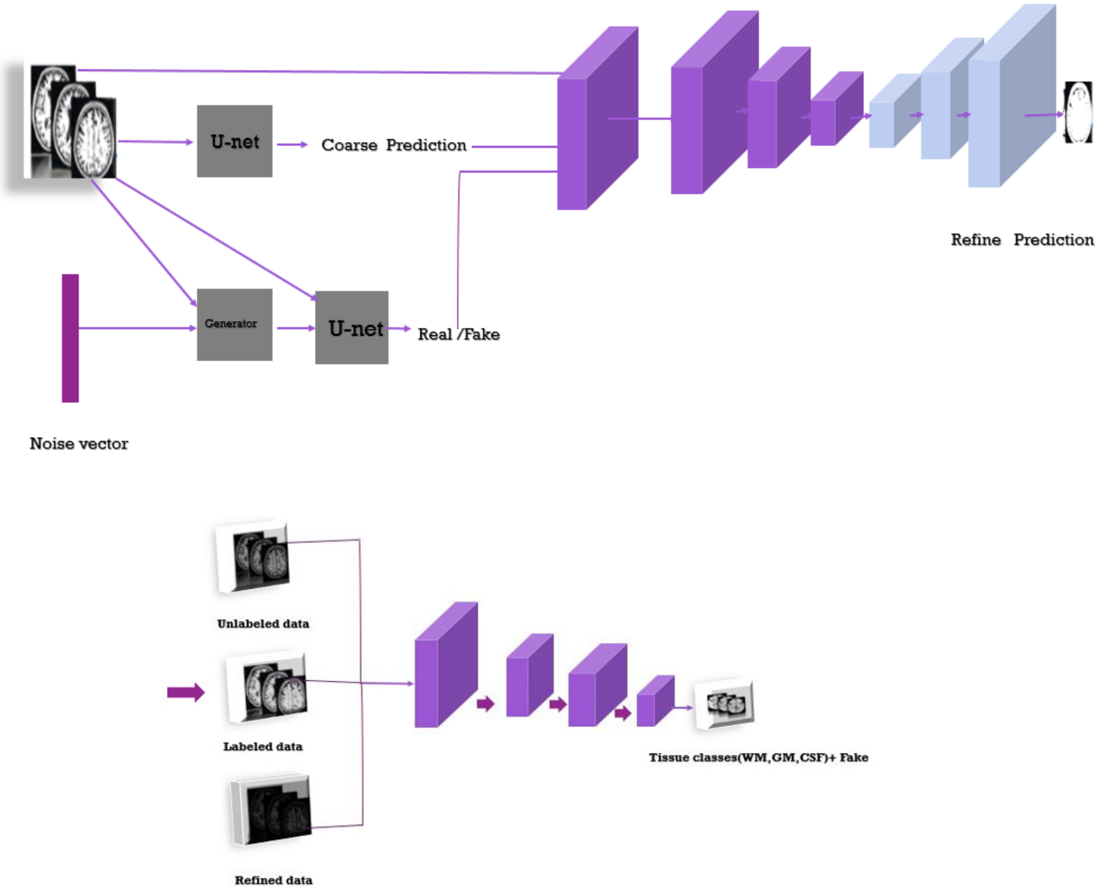

Fig. 4. The proposed Multi-Stage Generative Adversarial Network

A. Generative Adversarial Network (GAN)

Generative adversarial network refers to a network composed of two networks. A generator is used to generate a fake images from a noise vector. A discriminator is used to distinguish between generated data and true data. To be more specific, G is trained to map a noise vector $z \in R$ to a fake

image. D is trained to differential between true data x and generated data $G(z)$. The core idea of a GAN is to play a two player $min - max$ game. $min_G max_D E_{x \sim pdata}[logD(x)] + E_{z \sim noise}[1 - D(G(z))]$ Fig.4 shows an overview of our proposed network, which mainly consists of three stages: the three stages generator networks and the discriminator network. In this work, the first generator G is mainly used to generate outline for the background and brain tissues from the input images. The second generator takes two input, image x and random vector z. The idea is to encourage the generator to generate as many different values for each x as there are values of z. To be more specific, training network with random vector z and image x encourage network to act better in the output. While the third G is used to generate outline for (i) white matter (WM), (ii) gray matter (GM) and (iii) cerebrospinal fluid (CSF). On the other hand, the discriminator is used to distinguish between true and generated data. The main role of the first G is to generate coarse segmentation to be used as guidance information for the third G. The main role of the third G is to generate more detailed results by using the coarse segmentation from the first G. The third G consists of the encoder and decoder. In the encoder and decoder we also used the dense skip connection to combine the features from different scales. Fig.5 shows the network architecture of encoder decoder and the dense skip connection. Zihang et al. is used to implement the multi-stage model. In order to use this model in the multi-stage model, we changed this model as follows:
1-K classes were changed to $(K + 1)$ classes.
2-The segmentation network changed to be fully-convolutions.
Özgün et al. is used to implement the discriminator network. In order to use this model in the multi-stage model, we changed this model as follows:
1- ReLUs were changed to leaky ReLUs.
2- Max pooling was changed to average pooling.
For implement the encoder and decoder, in our proposed model, we use four blocks to implement the encoder where each block as follows:
The 1st block: consists of conv, LeReLU and concatenation.
The 2nd block: consists of conv, LeReLU and concatenation.
The 3rd block: consists of conv, LeReLU and concatenation.
The 4th block: consists of conv, LeReLU and concatenation.
While four blocks use to implement the decoder where each block as follows:
The 1st block: consists of deconv, conv, LeReLU and concatenation.
The 2nd block: consists of deconv, conv, LeReLU and concatenation.
The 3rd block: consists of deconv, conv, LeReLU and concatenation.
The 4th block: consists of deconv, conv, LeReLU and concatenation.
Furthermore, the dense skip connection is used to combine the features from different layers and used to help each other.

B. Loss function

1) Discriminator loss function: The discriminator in our model has unlabeled data loss, labeled data loss and refined prediction loss. And the overall loss function is as follows:

$$\mathrm{l}_{\mathrm{discriminator}} = \lambda_{\mathrm{labeled}} l_{\mathrm{labeled}} + \lambda_{\mathrm{unlabeled}} l_{\mathrm{unlabeled}} + \lambda_{\mathrm{fake}} l_{\mathrm{fake}}, \quad (9)$$

Where $\lambda_{\mathrm{labeled}}, \lambda_{\mathrm{unlabeled}} and \lambda_{\mathrm{fake}}$ are hyper-parameters. We set the hyper-parameters in Equation (9) to
$\lambda_{\mathrm{labeled}} = 1.0, \lambda_{\mathrm{unlabeled}} = 1.0 and \lambda_{\mathrm{fake}}$= 2.0.
For labeled data, we use the same loss function in the standard segmentation network. [34] it was shown that by using $l_{i,k+1}$ as substracted function, the softmax function changed as follows:

$$l_{\mathrm{labeled}} = -E_{\mathrm{x,y} \,\sim \mathrm{pdata(x,y)}} \sum_{\mathrm{i=1}}^{H \times W \times D} \log(Pmodel(y_{\mathrm{i}}|X), \quad (10)$$

$$l_{\mathrm{unlabeled}} = -E_{\mathrm{x} \sim \mathrm{pdata(x)}} \sum_{\mathrm{i=1}}^{H \times W \times D} \log((Z_{\mathrm{i}}(x)/Z_{\mathrm{i}}(x)) + 1), \quad (11)$$

$$l_{\mathrm{fake}} = -E_{\mathrm{z} \,\sim \mathrm{noise}} \sum_{\mathrm{i=1}}^{H \times W \times D} \log[((1/Z_{\mathrm{i}}(G_{\Theta G}(z) + 1)], \quad (12)$$

$$Z_{\mathrm{i}} = \sum_{\mathrm{k=1}}^{K} exp[l_{\mathrm{i,k}}(x)], \quad (13)$$

To be more specific, the idea is to introduce unlabeled loss and fake loss, which have analogues to the two components of discriminator loss in the standard GAN and the labeled loss is the cross-entropy. for more information we refer reader to read [34].

2) Generator loss function: We introduce a novel generated loss to encourage G to generate real data. Let x and z denote to the real data and noise, respectively.

$$C = E_{\mathrm{x} \sim \mathrm{pdata(x)}} f(x) - log(1 - D(G(z))), \quad (14)$$

In our paper, $f(x)$ contains the activation of the last layer.

$$L(G) = ||C - x||_2^2, \quad (15)$$

By minimizing this loss, we force the generator to generate real data in order to match our data and the correspond K classes of real data, which can be define as: $Classes = 1, ..., K$.

IV. Experiments

The following subsections present our experimental design and evaluation. In the subsection A, we present the evaluations and discussions. We start by giving a more detailed description of the datasets. Then, we show the experimental setup of our work. Finally, we explain the Dice coefficient of the segmentation evaluation.

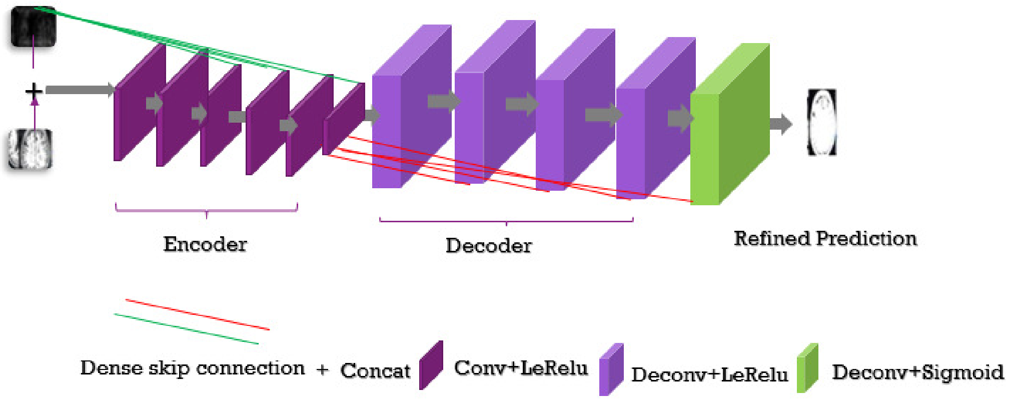

Fig. 5. The Illustration of Encoder Decoder Network

A. Evaluations and Discussions

1) Datasets:

In our work, we use two different datasets of brain images: the $MICCAIiSEG$ dataset and MRBrains dataset. We describe each of these datasets in the following.

2) MICCAI iSEG Dataset:

The aim of the evaluation framework introduced by the $MICCAIiSEG$ organizers is to compare segmentation models of WM, GM and CSF on $T1$ and $T2$. The MICCAI iSEG dataset contains 10 images, named subject-1 through subject-10, subject $T1$: $T1$ -weighted image, subject $T2$: $T2$-weighted, and a manual segmentation label used as a training set. The dataset also contains 13 images, named subject-11 through subject-23, used as a testing set. An example of the $MICCAIiSEG$ dataset ($T1$, $T2$, and manual reference contour) is shown in Fig.6. On the other hand, Table II shows the parameters used to generate $T1$ and $T2$. The dataset has two different times (i.e., longitudinal relaxation time and transverse relaxation time), which are used to generate $T1$ and $T2$. The dataset has been interpolated, registered, and skull-removed by the $MICCAIiSEG$ organizers. We present the evaluation equations in subsection B.

3) `MRBrains` Dataset:

The MRBrainsdataset contains 20 subjects for adults for segmentation of (a) cortical gray matter, (b) basal ganglia, (c) white matter, (d) white matter lesions, (e) peripheral cerebrospinal fluid, (f) lateral ventricles, (g) cerebellum, and (h) brain stem on $T1$, $T2$, and FLAIR. Five (i.e., 2 male and 3 female) subjects are provided to the training set and 15 subjects are provided for the testing set. On the evaluation of the segmentation, these structures merged into gray matter ($a-b$), white matter ($c-d$), and cerebrospinal fluid ($e-f$). The cerebellum and brainstem were excluded from the evaluation.

4) Experimental Setup:

The proposed model was performed Python on a PC with $NVIDIAGPU$ with a Ubuntu 16.04 operating system. Training our model took 30 hours in total. Testing our model took 5 minutes for the images employed for the testing set.

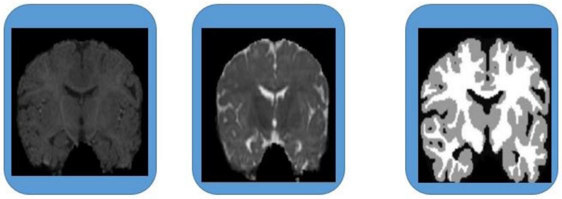

Fig. 6. Example of the MICCAIiSEG dataset ($T1$,$T2$, manual reference contour)

TABLE III. The parameters used to generate $T1$ and $T2$.

Parameter	TR/TE	Flip angle	Resolution
$T1$	1900/4.38 ms	7	1×1×1
$T2$	7380/119 ms	150	1.25×1.25×1.25

B. Segmentation Evaluation

1) Dice Coefficient (DC):

To better demonstrate the significance of our model, we have used Dice Coefficient (DC) metric to evaluate our model. Dice Coefficient (DC) has been considered as a baseline (i.e., benchmark) to compare segmentation models in the literature to compare brain segmentation models. We use V_{ref} for the reference segmentation and V_{auto} for the automated segmentation. The DC is given by the following equation:

$$DC(V_{\text{ref}}, V_{\text{auto}}) = \frac{2V_{\text{ref}} \bigcap V_{\text{auto}}|}{|V_{\text{ref}}| + |V_{\text{auto}}|}[48], \qquad (16)$$

where DC values are given in the range of [0, 1]. 1 corresponding to the perfect overlap and 0 indicating the total mismatch.

TABLE IV. Experiments on Training epoch obtained on the $MRBRAINS$ datasets. The best performance for each tissue class is highlighted in bold.

Training epoch	Dice Coefficient (DC) Accuracy		
	CSF	GM	WM
20	67%	69%	60%
40	86%	84%	82%
60	87%	88%	85%
80	**95%**	**94%**	**92%**

TABLE V. Experiments on Learning Rate obtained on the $MRBRAINS$ datasets. The best performance for each tissue class is highlighted in bold.

Learning Rate	Dice Coefficient (DC) Accuracy		
	CSF	GM	WM
1×10^-1	80%	80%	82%
1×10^-2	84%	83%	82.4%
1×10^-3	87%	88%	87%
1×10^-4	**95%**	**94%**	**92%**

C. Evaluating the hyper-parameters in multi-stage

To evaluate the effectiveness of our model, we evaluated different hyper parameters. The model has different hyper parameters, e.g., batch size, learning Rate, etc. Table III, Table IV and Table V show training epochs, learning rate and batch size, respectively. We find that the batch size of 30 is 95%, 94% and 92% for CSF, GM and WM respectively. A large training epochs caused the over fitting problem and a small training epochs caused under fitting problem. To mitigate this problem and validate whether the training epochs will be significantly impacted the network performance. Training epoch involving 20, 40, 60, 80 epoch is conducted. In the 80 epochs, we find that the network performance was good. We followed the same approach to choose the learning rate values. A large learning rate caused the parameters of network are updated quickly. A small learning rate caused the parameters are updated slowly. First, we randomly start with a learning rate is 1×10^1. Second, we use multiple runs by changing the learning rate value. Finally, Our experimental results show that multi-stage model achieves a good result in the learning rate 1×10^-4.

V. Result and Discussion

To better demonstrate the significance of our model, We train and test multi-stage GAN model on two datasets of different ages (i.e., infants and adults). To train multi-stage GAN model, we used the 13 unlabeled of the testing dataset to train the GAN and for the 10 labeled subjects of the $iSEG - 2017$ dataset, we used two labeled subjects for training set, one labeled subject used for validation set and 7 labeled subjects used for testing set. Similarly, for the 5 labeled subjects of the MR Brains 2013 dataset, we used one labeled subject for training set, one labeled subject used for validation set and 3 labeled subjects used for testing set. And 15 unlabeled of the testing dataset used to train the multi-stage GAN model. The main goal of multi-stage GAN model is to see the performance in a few-shot learning case. Table VI presents the results of our model to segment CSF ,GM, and WM using the $MICCAIiSEG$ dataset. Our model obtains a DC values of 95% in CSF segmentation. The DC values obtained from segmenting CSF by state-of-the-art models range between 86% and 91%. In addition, our model obtains a DC values of 94% and 92% in segmenting GM and WM , respectively. The state-of-the-art models, on the other hand, obtain DC values in the ranges of 80%- 93% for GM segmentation and 81%- 90% for WM segmentation. Such results highlight the remarkable efficiency gained by using multi-stage GAN. Table VII compares the results obtained using the MRBrains dataset. We observe that our model achieves a DC value of 93% on CSF segmentation, 93% on GM segmentation, and 88% on WM segmentation. Such results are superior to the results obtained using the state-of-the-art models. Therefore, we argue that our mode can perform better in a few-shot learning case. Fig. 7 shows Visualization results of our models on the subject used as a validation set. We observe that segmentation results obtained by multi-stage model is fairly close to the manual reference contour provided by the MICCAI iSEG organizers.

TABLE VI. Experiments on batch size obtained on the $MRBRAINS$ datasets. The best performance for each tissue class is highlighted in bold.

Batch size	Dice Coefficient (DC) Accuracy		
	CSF	GM	WM
10	77%	76%	76%
20	82%	83.2%	82.4%
30	**95%**	**94%**	**92%**
40	89%	88%	87%

TABLE VII. Segmentation performance in Dice Coefficient (DC) obtained on the $MICCAIiSEG$ dataset. The best performance for each tissue class is highlighted in bold.

Model	Dice Coefficient (DC) Accuracy		
	CSF	GM	WM
U-net	86.2%	80.1%	81.1%
Standard-GAN	87.5%	89.2%	82.4%
One-stage	91.3%	93.8%	90.3%
Multi-stage	**95%**	**94%**	**92%**

TABLE VIII. Segmentation performance in Dice Coefficient (DC) obtained on the $MRBrains$ dataset. The best performance for each tissue class is highlighted in bold.

Model	Dice Coefficient (DC) Accuracy		
	CSF	GM	WM
U-net	86.2%	80.1%	81.1%
Standard-GAN	87.5%	89.2%	82.4%
One-stage	91.3%	93.8%	90.3%
Multi-stage	**95%**	**94%**	**92%**

TABLE IX. Average execution time (in minutes) and standard deviation (SD) in the `MRBrains` dataset

Model	Time (SD)
U-net	36.54(0.12)
Standard-GAN	30.52(0.31)
One-stage	25.30(0.16)
Multi-stage	**22.61(0.21)**

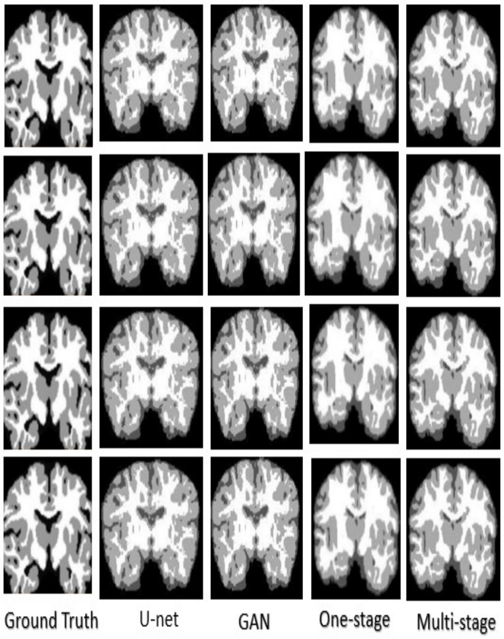

Fig. 7. Visualization results on MRBrains dataset

VI. Threats to Validity

Threats to external validity are related to the generalizability of our results. In our work, we used the two datasets that belong to two organizers. The total number of the subjects in the two datasets are 43 subjects. One could argue that the datasets do not have enough samples. We mitigate such threat by using two datasets that (a) contain both infant and adult brain data and (b) were previously used by prior studies.

Our model obtains higher performance than prior models. We believe that our model similar with the process of human learning in clinical practice. Moreover, we have only targeted three tissues in our work. However, our proposed model can be easily extended to more tissues segmentation as it does not require more labeled data.

The intuition behind this model is that multi-stage compares the perform in a few-shot learning case where a few labeled data are available.

Threats to internal validity are related to experimental errors and bias. Our model is constructed using data extracted from medical images in which contracts might be low. We use the small-size kernels, deconvolution layer (to upsample the input), PReLU, dropout and normalization methods to reduce the risk of overfitting. Hence, any potential deficiency in the data should deficient all the implemented models. Nevertheless, our model obtains higher performance than previous models.

VII. Conclusion

In this study, we propose multi-stage generative adversarial networks(GAN) model for brain segmentation supported by (i) generate a coarse outline for (i) background and (ii) brain tissues. Then, generate outline for (i) white matter (WM), (ii) gray matter (GM) and (iii) cerebrospinal fluid (CSF).

Our results are evaluated by using the infant and adult data sets and found to be fairly close to the manual reference. In addition, we compare our model with three baseline state-of-the-art models and observe that our model achieves an improvement of up to 4%. In particular, we obtain dice coefficients that range between 88% and 95%. Such results indicate that the adoption of the multi-stage generative adversarial networks has significantly improved segmentation results. We argue that our model is more efficient and accurate in practice for both infant and adult brain segmentation.

Despite the promising results obtained from our proposed model, we believe that further improvements can be achieved in the future. We aim in the future to include more datasets in our study. Furthermore, we intend to expand our multi-stage model to investigate more number of brain tissues. Lastly, we would like to point out that, we will expand our model and evaluate how multi-stage model can segment brain tissues.

Acknowledgment

This work is supported in part by the National Science Foundation of China (NSFC). Awards 61872411 and 61472150.

References

[1] R. C. Xiaohao Cai and T. Zeng, “Two-stage image segmentation method using a convex variant of themumford–shah model and thresholding,” *Society for Industrial and Applied Mathematics*, vol. 6, pp. 368–390, 2013.

[2] Y. H. X. D. H. G. L. Sun, J. Wang and J. Paisley, "An adversarial learning approach to medicalimage synthesis for lesion detection," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 24, 2020.

[3] E. W. X. Yia and P. Babyn, "Generative adversarial network in medical imaging: A review", medical image analysis," *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS*, vol. 58, 2019.

[4] M. B. H. A. Helaly and A. Y. Haikal, "Toward deep mri segmentation for alzheimer's disease detection," *Neural Computing and Applications*, 2021.

[5] A. K. B. V. G. N. N. S. A. A. M. S. Kazeminia, C. Baur, "Gans for medical image analysis," *Neural Computing and Applications*, vol. 109, 2020.

[6] W. T. A. Khaled, C. Own and T. A. Ghaleb, "Improved brain segmentation using pixel separation and additional segmentation features," *APWeb-WAIM 2020: The 4th APWeb-WAIM International Joint Conference on Web and Big Data*, pp. 85–100, 2020.

[7] X. S. T. Chen and C. Wang, "Preserving-texture generative adversarial networks for fast multi-weighted mri," *IEEE Access*, vol. 6, pp. 71 048–71 059, 2018.

[8] H. A. T. Iqbal, "Generative adversarial network for medical images(mi-gan)," *journal of Medical Systems*, 2018.

[9] S. V. A. A. Adegun and R. O. Ogundokun, "Deep learning approach for medical image analysis," *Computational Intelligence and Neuroscience*, 2021.

[10] N. R. E. G. R. K. S. Dara, P. Tumma, "Feature extraction in medical images by using deep learning approach," *International Journal of Pure and Applied Mathematics*, vol. 120, pp. 305–312, 2018.

[11] C. Liang and S. Xin, "Research status and prospects of deep learning in medical images," *International Conference on Communications, Information System and Computer Engineering (CISCE)*, pp. 380–382, 2020.

[12] Y. P. Y. Gu and H. Li, "Aids brain mris synthesis via generative adversarial networks based on attention-encoder," *IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 629–633, 2020.

[13] Y. Z. D. W. M. Chen, X. Shi and M. Guizani, "Deep feature learning for medical image analysis with convolutional autoencoder neural network," *IEEE Transactions on Big Data*, vol. 4, pp. 750–758, 2021.

[14] N. Yamanakkanavar and B. Lee, "Using a patch-wise m-net convolutional neural network for tissue segmentation in brain mri images," *IEEE Access*, vol. 8, pp. 120 946–120 958, 2020.

[15] D. Y. T. S. S. H. B. T. B. J. W. H. R. A. M. D. X. Z. X. L. Zhang, X. Wang, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IIEEE Transactions on Medical Imaging*, vol. 39, pp. 2531–2540, 2020.

[16] M. Özbey and T. Çukur, "T1-weighted contrast-enhanced synthesis for multi-contrast mri segmentation," *28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, 2020.

[17] J. M. W. E. M. P. P. A. M. S. R. A. P. T. T. L. M. A. V. I. I. J. M. H. Noothout, B. D. Vos, "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 4011–4022, 2020.

[18] L. W. S. L. E. K. Y. Bar, I. Diamant and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 294–297, 2015.

[19] J. Y. L. Z. L. J. Z. J. B. Q. H. L. H. K. J. Y. K. Y. H. S. H. H. Z. Y. Q. R. Y. L. M. Geng, X. Meng, "ontent-noise complementary learning for medical image denoising," *IEEE Transactions on Medical Imaging*, 2021.

[20] X. Y. L. S. X. Chen, C. Xu and D. Tao, "Gated-gan: Adversarial gated networks for multi-collection style transfer," *IEEE Transactions on Image Processing*, vol. 28, pp. 546–560, 2018.

[21] A. S. S. U. Niyaz and Devanand, "Advances in deep learning techniques for medical image analysis," *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 271–277, 2018.

[22] S. Singh and N. Singh, "Object classification to analyze medical imaging data using deep learning," *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–4, 2017.

[23] M. M. B. X. D. W. S. z. A. C. Y. B. J. Goodfellow, J. Pouget, "Generative adversarial nets," *NIPS*, 2014.

[24] X. W. Y. Wang, A. K. Katsaggelos and T. B. Parrish, "A deep symmetry convnet for stroke lesion segmentation," *IEEE International Conference on Image Processing (ICIP)*, pp. 111–115, 2016.

[25] Y. Z. Z. W. C. Z. Y. Ding, F. Chen and D. Wu, "A stacked multi-connection simple reducing net for brain tumor segmentation," *IEEE Access*, vol. 7, pp. 104 011–104 024, 2019.

[26] N. N. T. Bdair and S. Albarqouni, "Roam: Random layer mixup forsemi-supervised learning in medical imaging," *Computer Science*, 2020.

[27] T. M. S. S. N. I. K. M. S. Majib, M. M. Rahman and S. K. Dey, "Vgg-scnet: A vgg net-based deep learning framework for brain tumor detection on mri images," *IEEE Access*, vol. 9, pp. 116 942–116 952, 2021.

[28] Y. F. Y. Sun, C. Zhou and X. Xue, "Parasitic gan for semi-supervised brain tumor segmentation," *EEE International Conference on Image Processing (ICIP)*, pp. 1535–1539, 2019.

[29] D. A. M. D. Cirillo1 and A. Eklund, "Vox2vox: 3d-gan for brain tumour segmentation," *IEEE Access*, vol. 9, pp. 116 920–116 930, 2021.

[30] C. C. C. D. G. D. H. M. N. B. R. F. N. P. F. R. Q. Delannoy, C. Pham, "Segsrgan: Super-resolution and segmentation using generative adversarial networks application to neonatal brain mri," *Computers in Biology and Medicine*, vol. 120, pp. 103–755, 2020.

[31] L. R. R. A. W. S. S. M. Y. F. G. M. H. N. C. Han, H. Hayashi, "Gan-based synthetic brain mr image generation," *IEEE 15th International Symposium on Biomedical Imaging*, vol. 120, pp. 734–738, 2018.

[32] R. G. P. B. R. G. A. H. D. D. M. V. H. J. W. D. R. C. Bowles, Liang Chen, "Gan augmentation: Augmenting training data using generative adversarial networks," *IEEE 15th International Symposium on Biomedical Imaging*, 2018.

[33] C. K. C. J. P. O. S. C. M. F. K.-C. B. Andrew, B. James, "High-resolution medical image synthesis using progressively grown generative adversarial networks," *arXiv:1805.03144*, 2018.

[34] J. D. A. Kumar Mondal and C. Desrosiers, "Few-shot 3d multi-modal medical image segmentation using generative adversarial learning," 2018.

[35] D. H. Y. J. B. S. J. C. Y. E. Kang, H. J. Koo, "Cycle-consistent adversarial denoising network for multiphase coronary ct angiography," *The international Journal of Medical Physics Research and Practice*, pp. 2473–4209, 2018.

[36] J. B. U. B. M.J. Chuquicusma, S. Hussein, "How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis," *IEEE 15th international symposium on biomedical imaging*, 2018.

[37] B. K. M. M. A. J. G. H. G. M. Frid-Adar, I. Diamant, "Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification," *Neurocomputing*, 2018.

[38] S. F. J. G. M. Jafari, D. Auer and X. Chen, "Dru-net: An efficient deep convolutional neural network for medical image segmentation," *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 1144–1148, 2020.

[39] S. P. G. Madhupriya, N. M. Guru and B. Nivetha, "Brain tumor segmentation with deep learning technique," *3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 758–763, 2019.

[40] J. C. X. Y. B. Li, C. Wu and G. Wang, "A deeply supervised convolutional neural network for brain tumor segmentation," *39th Chinese Control Conference (CCC)*, pp. 6262–6267, 2020.

[41] Z. Z. H. S. T. Li, F. Zhou and H. Zhu, "A label-fusion-aided convolutional neural network for isointense infant brain tissue segmentation," *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 692–695, 2018.

[42] S. C. M. Arjovsky and L. Bottou, "Wasserstein gan," *IEEE Transactions on Image Processing*, 2017.