ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 17, No. 2, 2022, *pp*.138-153

Gaussian Mixture Model Loss Functional for Brain MRI Segmentation With Deep Learning

Lianhong Ma[1]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
(Received October 11, 2022, accepted November 29, 2022)

Abstract: Because of the excellent performance and fast speed of deep neural network, U-Net has become the most popular network framework for medical image segmentation. For various specific image segmentation tasks, researchers have proposed a series of U-Net related methods. However, on the one hand, due to the inherent limitations of convolutional neural networks, the variants of U-Net still cannot model long-range information well while maintaining detailed texture information. On the other hand, since medical images are difficult to obtain a large number of high-quality semantic pixel-level annotations, it is difficult to use supervised deep learning networks. To address these issues above, we proposed a modified U-Net structure and a Gaussian mixture model (GMM) based loss function. This modified U-Net can be well applied to brain MR image segmentation, which can not only restore the detailed information well, but also take into account the relatively large-scale local information. The proposed GMM loss can be used for unsupervised training of neural networks. It effectively alleviates the shortcomings of difficult access to medical image annotation data and improves the performance of deep neural networks. In the experiments in this paper, the GMM loss function can also be used as a regular term to assist supervised learning to achieve better results. Experimental results on brain MR images demonstrate the superior performance of the proposed model.

Keyword: Semi-supervised learning; Unsupervised Deep learning; Modified U-Net; Gaussian mixture model

1. Introduction

Image segmentation is the division of an image into a set of non-overlapping regions. Accurate medical image segmentation technology can help doctors make accurate clinical diagnosis. Magnetic resonance imaging (MRI) has high resolution and high contrast, so it is often used as a theoretical study of medical images. MR images of the brain are segmented into three non-overlapping parts, including white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), which is important for physicians studying brain disorders. However, due to machine imaging, MRI also has some problems such as uneven gray scale, blurred border, complex texture and noise, which brings severe resistance to the task of brain MRI segmentation.

In recent years, a large number of deep learning-based methods [1] have emerged for brain MR image segmentation, in order to improve the accuracy of image segmentation results. Most of these methods are based on the supervised convolutional neural network (CNN), which has become popular in recent years, which requires high data set size [1,2]. But the cost of acquiring massive labeled datasets is relatively high and labor-intensive for medical images,.

In the case of limited amount of labeled data, it is difficult to make it difficult to reach the correct decision boundary by training the deep learning model., as shown in Fig. 1(a). To overcome this problem, many

[1] Corresponding Author *E-mail*: mlh19491001@163.com.

Journal of Information and Computing Science, Vol. 17(2022) No. 2, pp 138-153

researchers use data augmentation [3,4] to solve the overfitting problem and obtain more precise decision boundaries, which can be found in Fig.1(b). However, these studies still fail to reach a better decision boundary [5]. In order to solve the problem of insufficient labeled datasets, a large number of weak/semi-supervised image segmentation methods have been proposed [6,7,8]. These image segmentation methods require very little high-quality labeled data, which makes the application of deep learning methods in various image segmentation tasks gradually expand, such as brain MR image segmentation.

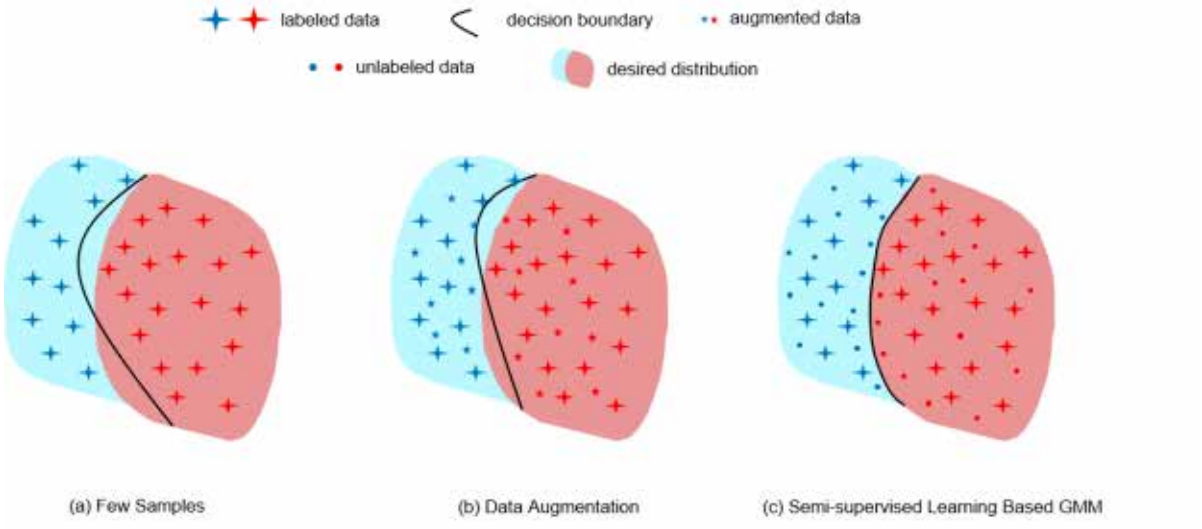

Fig. 1 Comparison of feature visualization.

Based on the observation of traditional image segmentation methods, the classic finite mixture model method for image segmentation is unsupervised [9,10,11,12,13,14]. However, these classical algorithms are computationally complex, and it is very easy to obtain local optimal solutions. In addition, they only consider the individual characteristics of a single individual rather than the common characteristics of image sets, so they are relatively limited in semantic image segmentation.

Through observations from previous studies, we found that classical finite mixture models can help improve the performance of CNN-based deep learning image segmentation models, especially when training data is small. Therefore, we propose a loss function based on the Gaussian mixture likelihood function and couple it to any deep learning network. The proposed loss function is matched with the deep learning network model, which not only utilizes the rich semantic information of the network, but also avoids the shortcomings of insufficient spatial information of the limited mixture model. If learning in a supervised way, our loss function will act as a regularization term, because the loss function considers the feature of pixel similarity, so it can improve the accuracy of the entire image segmentation model. Moreover, because our proposed loss function is a self-supervised loss, we can easily graft it to other deep learning networks for semi-supervised or unsupervised learning. Our method can obtain the best boundary, as shown in Fig.1(c).

In recent years, deep learning-based methods, especially U-Net based methods [2], have been successfully applied to brain MR image segmentation and achieve outstanding performance. However, these methods have deep semantic gaps between the encoders and decoders [15]. Furthermore, the skip connections used in the U-Net can not retrieve the spatial information lost by pooling operations [2]. In order to deal with these problems, many improved U-Net methods have been proposed [15,16,17,18]. Zhou et al. [16] proposed nested and dense skip connections to capture fine-grained details of the objects. Chen et al. [19] used the transformers to improve the U-Net and can preserve more detail information; however, this method has high computational complexity [21]. In this paper, we proposed a modified U-Net, which can effectively makes up for the defect of U-Net modeling long-distance information. Compared with transformer-based models, our method contains fewer parameters and is easier to converge.

2. Backgrounds

2.1 Gaussian Mixture model

Let $x_i \in R^D$, $i=(1,2,...,N)$, denotes an observation at the i th pixel of an image. ∂_i is its neighborhood. In order to divide an image into K regions $(\Omega_i,\Omega_2,...,\Omega_k)$, GMM assumes that x_i is independent and its probability density function (pdf) is:

$$f(\mathbf{x}_i \mid \mathbf{\Pi},\mathbf{\Theta}) = \sum_{k=1}^{K} \pi_{ik} p(\mathbf{x}_i \mid \mathbf{\Theta}_k) \tag{2.1}$$

where $\mathbf{\Pi}$ and $\mathbf{\Theta}$ are model parameters. $\mathbf{\Pi}=\{\pi_1,...,\pi_K\}$ is the prior probabilities and satisfies: $0 \le \pi_{ik} \le 1$ and $\sum_{k=1}^{K} \pi_{ik} = 1$. $p(\mathbf{x}_i \mid \mathbf{\Theta}_k)$ is a normal distribution and defined as:

$$p(\mathbf{x}_i \mid \mathbf{\Theta}_k) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\mathbf{\Sigma}_k|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}_i-\mu_k)^{\top} \mathbf{\Sigma}_k^{-1}(\mathbf{x}_i-\mu_k)\right) \tag{2.2}$$

where μ_k and $\mathbf{\Sigma}_k$ are mean and covariance matrices, respectively.

Just like the other independent assumptions in statistical analysis, we suppose that $\mathbf{x}_i$ is independent. Therefore, the maximum likelihood function of GMM is written as;

$$L(\mathbf{\Pi},\mathbf{\Theta}) = \ln p(X \mid \mathbf{\Pi},\mathbf{\Theta}) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_{ik} p(x_i \mid \mathbf{\Theta}_k) \tag{2.3}$$

Eq.(2.3) simply solves the model parameters iteratively using the expectation maximization (EM) algorithm [22] as follows:

E step: Define a Q function of Eq.(2.3), with respect to the conditional distribution of the latent variables Z:

$$\begin{aligned} &Q(\mathbf{\Theta} \mid \mathbf{\Theta}^{(t)}) = E_{\mathbf{Z}|\mathbf{X},\mathbf{\Theta}^{(t)}}\left[L(\mathbf{\Theta};\mathbf{X},\mathbf{Z})\right] \\ &= \sum_{k=1}^{K}\{\sum_{i=1}^{N}(E_{z_{ik}})\log\pi_{ik} + \sum_{i=1}^{N} E_{z_{ik}}\left[\log\frac{1}{(2\pi)^{\frac{D}{2}}} - \frac{1}{2}\log|\mathbf{\Sigma}_k| - \frac{1}{2}(\mathbf{x}_i-\mu_k)^{\top}\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i-\mu_k)\right]\} \end{aligned} \tag{2.4}$$

where $E_{z_{ik}} = \hat{z}_{ik} = P(Z_i = k \mid \mathbf{x}_i;\theta_k^{(t)})$ and can be calculated as:

$$\hat{z}_{ik} = \frac{\pi_{ik}^{(t)} p(\mathbf{x}_i \mid \theta_k^{(t)})}{\sum_{k=1}^{K} \pi_{ik}^{(t)} p(\mathbf{x}_i \mid \theta_k^{(t)})} \tag{2.5}$$

M step: find the parameters that maximize this quantity:

Journal of Information and Computing Science, Vol. 17(2022) No. 2, pp 138-153

$$\theta^{(t+1)} = \arg\max Q(\theta \mid \theta^{(t+1)}) \tag{2.6}$$

Then the mean, covariance matrix and component weights can be calculate as:

$$\begin{aligned} \mu_k &= \frac{\sum_{i=1}^{N} \hat{z}_{ik} \mathbf{x}_i}{\sum_{i=1}^{N} \hat{z}_{ik}}, \\ \mathbf{\Sigma}_k &= \frac{\sum_{i=1}^{N} \hat{z}_{ik} \left(\mathbf{x}_i - \mu_k\right)^T \left(\mathbf{x}_i - \mu_k\right)}{\sum_{i=1}^{N} \hat{z}_{ik}}, \\ \hat{\pi}_{ik} &= \frac{1}{N} \sum_{i=1}^{N} \hat{z}_{ik} \end{aligned} \tag{2.7}$$

Although GMM can obtain satisfactory results for natural images; however, it is sensitive to noise, intensity inhomogeneity and weak boundary [10,11,12,13,14].

2.2 Bias field formulation

To reduce the effect of intensity inhomogeneity (also named as bias field) in the MR image, the observed image is usually modeled as :

$$I = BJ + N \tag{2.8}$$

Where I is an observed image, J is the clean image, B is the unknown bias field and N is additive noise. Each tissue in brain MR images should have a specific value of the measured physical properties. Therefore, it can be assumed that the real image is piecewise approximately constant. The bias field is smooth and slowly changing in image domain [24,25,26], so the models [25,26] use a low-pass filter to ensure the smoothness of the bias field, which also causes the models to easily lose boundary information. To improve robustness, [27,28] use a linear combination of a set of orthogonal polynomial basis functions to fit the bias field.

2.3 CNN-Based Deep Learning methods

1) *Supervised methods:* When there are enough semantic pixel level labels, the deep neural network method becomes the main tool of modern segmentation technology because of its high performance and fast running complexity. Because the output maps generated by full convolutional network (FCN [1] have the same input size, it is widely used in medical image segmentation. U-Net [2] is a U-shaped framework composed of encoder and decoder inspired by FCN, which can achieve feature extraction and precise positioning, thus obtaining more accurate results than FCN.

2) *Weakly/Semi-Supervised methods:* In the training set with a large number of pixel level annotations, the supervised learning method can obtain high-performance semantic segmentation results, but for medical image analysis, it is difficult to obtain a large number of tag images with pixel level annotations. To solve this problem, a common semi-supervised semantic segmentation method is to generate training samples by using Generative Adversarial Networks(GANs) [29]. Hung et al. [30] proposed a discriminator based on GAN to distinguish the confidence map from the prediction of labeled and unlabeled data.

On the other hand, consistent regularization is widely studied for semi-supervised segmentation, which enhances the consistency of prediction/intermediate features and various disturbances [31]. In addition, the bounding box annotation [32], bounding box annotation [33], image level annotation [34,35] and scribbled annotations [36] have been used for semantic segmentation.

3. Method

As mentioned earlier, in deep learning-based medical image segmentation methods, the lack of high-quality labeled data and the limitation of the receptive field of CNN itself are the two main problems that limit the segmentation accuracy of the entire model. In order to solve the defects brought by the dataset and CNN, we invented a loss function based on Gaussian mixture likelihood function to improve the accuracy of the model. At the same time, we propose an improved U-Net to enhance its ability of remote information modeling.

3.1 Key observation

An important research result of this paper is that the softmax layer in the deep learning network can be approximated as a probability output, so it is feasible to directly maximize the likelihood function of the Gaussian mixture model. It uses the EM algorithm to obtain the posterior probabilities, which is different from the classic Gaussian mixture model approach. Specifically, the k th channel softmax output is written as:

$$s_{i,k}(i)=\frac{e^{z_k(i)}}{\sum_{i=1}^{K}e^{z_k(i)}},\qquad k=1,2,...,K,i\in\Omega \tag{3.1}$$

where $z_k(i)$ is the output of the network before the softmax. When i belongs to the k th class, $s_{i,k}$ is close to 1. Furthermore, like $\sum_{k=1}^{K}s_{ik}=1$, $\hat{z}_{ik}$ in Eq.(2.5) also meets this condition. This similarity illustrates that the softmax output can be used as an approximation of the posterior probability of each pixel. Therefore, we define the GMM based loss function as:

$$L_{GMM}=-\sum_{k=1}^{K}\sum_{i=1}^{N}s_{ik}[\log\pi_{ik}-\frac{1}{2}\log|\boldsymbol{\Sigma}_k|-\frac{1}{2}(\mathbf{x}_i-\mu_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i-\mu_k)]]+C \tag{3.2}$$

Where C is a constant independent of variables, $\mathbf{x}_i$ is the input and s_{ik} is the output of softmax layer and

$$\begin{aligned}\mu_k&=\frac{\sum_{i=1}^{N}s_{ik}(\mathbf{x}_i)\mathbf{x}_i}{\sum_{i=1}^{N}s_{ik}(\mathbf{x}_i)},\qquad k=1,2,\ldots,K\\ \boldsymbol{\Sigma}_k&=\frac{\sum_{i=1}^{N}s_{ik}(\mathbf{x}_i)(\mathbf{x}_i-\mu_k)^T(\mathbf{x}_i-\mu_k)}{\sum_{i=1}^{N}s_{ik}(\mathbf{x}_i)},\qquad k=1,2,\ldots,K\\ \pi_{ik}&=\frac{1}{N}\sum_{i=1}^{N}s_{ik}(\mathbf{x}_i),\qquad k=1,2,\ldots,K\end{aligned} \tag{3.3}$$

It is worth noting that Eq.(3.2) is differentiable and can be derived for each parameter. Therefore, it can find the minimum value by gradient descent method. The loss function is a self supervised loss function that can be used for unsupervised segmentation. In addition, the loss function takes into account the statistical distribution

information of image pixel values, so adding statistical distribution information to the network can enhance robustness.

3.2 Application of GMM based Loss Function

In order to make the edge of the segmented image smoother, we introduce a TV regularization term to combine with the Gaussian mixture loss to form the following loss function:

$$L_{GMM\_TV} = L_{GMM} + \gamma \sum_{k=1}^{K} \sum_{i=1}^{N} \nabla \left| s_k(\mathbf{x}_i) \right|. \tag{3.4}$$

where γ is a non-negative constant.

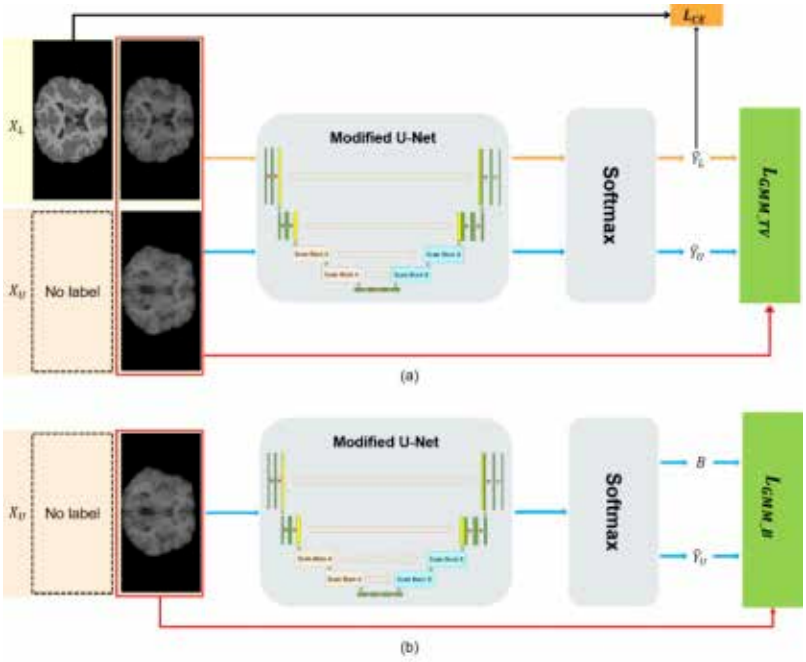

Fig. 2 The segmentation flow of the proposed method. (a) Semi-supervised segmentation: depth learning network with pixel level annotation image X_S and unlabeled image X_U. Use L_{CE} and L_{GMM_TV} training the network. (b) Unsupervised segmentation with bias field estimation.

Fig.2(a) shows the use of our loss in semi-supervised segmentation tasks. Additionally, GMM-based losses can be used with any deep neural network, even with little or no high-quality labeled data.

1) Application for Semantic segmentation: When the model has only a few pixel level semantic annotations, we use the traditional cross entropy loss and GMM loss to evaluate the model loss; When there is no annotations, we only use GMM loss. In this way, the total loss function is designed as:

$$L_{Total} = \alpha L_{CE} + \beta L_{GMM\_TV}, \tag{3.5}$$

where L_{CE} is the cross-entropy loss. When the input has labels, α is set as 1; otherwise, it is set to 0. We use L_{CE} and L_{GMM_TV} to describe semantic information and pixel similarity to improve model accuracy.

2) Application for unsupervised segmentation: Eq.(3.5) is a self-supervised loss, and when $\alpha = 0$, the model

is unsupervised. For supervised or semi-supervised models, the network can learn the characteristic information of the bias field by using the label information. In unsupervised learning, the model is affected by the bias field and can not obtain ideal results.

Let $B=\{b_i, i=1,2,\ldots,N\}$ denotes the bias field, which can be modeled by using a linear combination of a set of polynomial basis functions:

$$b_i = \sum_{j=1}^{(P+1)(P+2)/2} \omega_j g_j(i) = \omega^T G(j) \tag{3.6}$$

where $G=\{g_1, g_2, \ldots, g_M\}$ is a set of orthogonal Legendre polynomials with degree P. Then the loss can be written as:

$$L_{GMM\_B} = -\sum_{k=1}^{K}\sum_{i=1}^{N} s_{ik}[\log \pi_{ik} - \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\left(\mathbf{x}_i - b_i\mu_k\right)^T \boldsymbol{\Sigma}_k^{-1}\left(\mathbf{x}_i - b_i\mu_k\right)]] \tag{3.7}$$

where

$$\begin{aligned}
\mu_k &= \frac{\sum_{i=1}^{N} s_{ik} b_i \mathbf{x}_i}{\sum_{i=1}^{N} s_{ik} b_i^2}, \\
\boldsymbol{\Sigma}_k &= \frac{\sum_{i=1}^{N} s_{ik}\left(\mathbf{x}_i - b_i\mu_k\right)^{\top}\left(\mathbf{x}_i - b_i\mu_k\right)}{\sum_{i=1}^{N} s_{ik}}, \\
\pi_{ik} &= \frac{1}{N}\sum_{i=1}^{N} s_{ik}, \\
\omega &= U^{-1}V
\end{aligned} \tag{3.8}$$

where $U=\sum_{i=1}^{N} G(i)G(i)^T J_1$, $V=\sum_{i=1}^{N} G(i)J_2$, $J_1=\sum_{k=1}^{K} s_{ik}\mu_k^T\boldsymbol{\Sigma}_k^{-1}\mu_k$, $J_2=\sum_{k=1}^{K} s_{ik}\mu_k^T\boldsymbol{\Sigma}_k^{-1}\mathbf{x}_i$.

3.3 A modified U-Net

The original U-Net [2] is classical Encoder-Decoder architecture, which can produce almost perfect segmentation results with enough labeled training data. Although the classical u-net method has good segmentation performance, it is usually limited in explicitly modeling long-range correlation due to the inherent defects of convolution operation [19]. Therefore, we propose an improved version of U-Net.

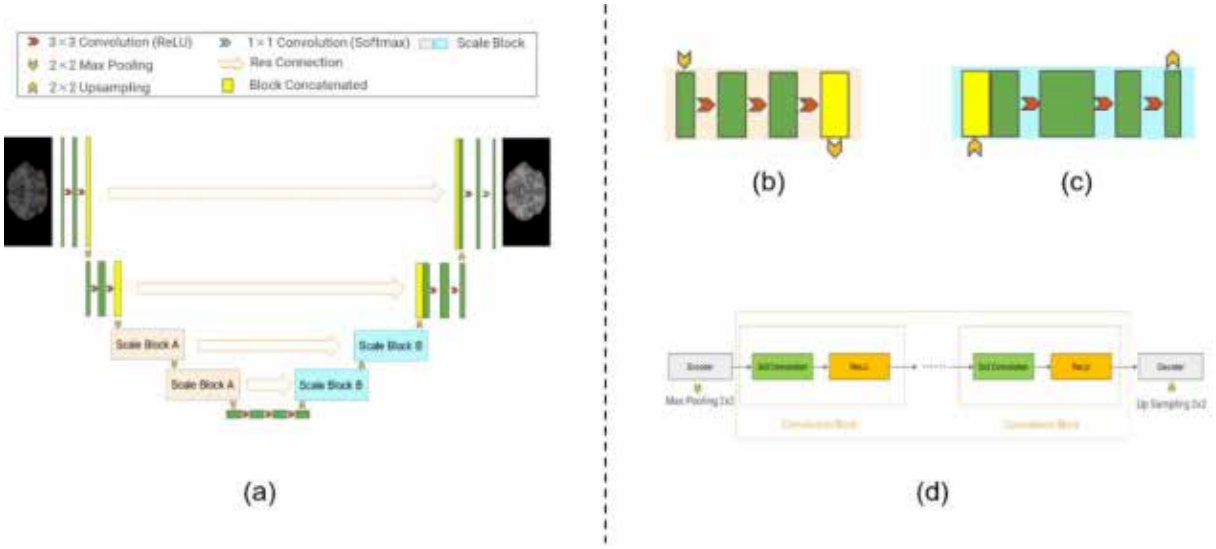

Fig. 3 The modified U-Net and its respective components. (a) is the overall framework of modified U-Net; (b) is the scaling module A ; (c) is the scaling module B; (d) is the skip connection module.

Fig.3 (a) shows the modified U-Net structure. Compared with the original u-net, the new network architecture adopts improved up scaling module and skip connection module. The network adopts the coding and decoding structure similar to that of U-net. The coding process includes two original U-net coding sequences and two scaling module(see Fig.3 (b)). To obtain a larger field of view, we add a 3×3 convolution in the scale block A. Each sequence or scale block is followed by a maximum pooling operation with a size of 2×2 and a step of 2. After each down sampling, the number of filters in the convolution layer is doubled. Similarly, the decoding process uses 2 layers of improved scale block B and 2 layers of conventional decoding sequences. In scale block B, a 3×3 convolution is added. In this paper, the bilinear interpolation is used in the up sampling process to obtain more accurate results.

The U-Net uses skip connection to fuse coded and decoded feature information, which will lead to differences between two groups of potentially incompatible features in the learning process, thus affecting the accuracy of the model [15]. In this paper, we propose an improved skip connection module, which consists of a series of convolutional modules, to reduce the difference between the encoding module and the decoding module. Each convolution module includes a 3×3 convolution and a ReLU module. The number of convolutional modules in the skip connection module decreases with the increase of the number of coding layers.

4. Experimental Results

4.1 Dataset

1) Internet Brain Segmentation Repository (IBSR):

To evaluate the semantic segmentation performance, we use data sets from IBSR (http://www.cma.mgh.harvard.edu/ibsr/), which provides 3D brain MR images and segmentation labels with a resolution of $256\times128\times256$ voxels. All these images is normalized into $[0,1]$.

2) MRBrainS18: The MRBrainS18 (https://mrbrains18.isi.uu.nl/) is a Grand Challenge on MR Brain Segmentation at MICCAI 2018, which consists 7 sets of labeled brain MR images with size $240\times240\times48$. Some data sets have bias fields.

4.2 Implementation Details

When the input image contains pixel-level semantic labels, the model trains the deep neural network using a semi-supervised method; otherwise, the network is trained in an unsupervised manner. All experiments were performed on NVIDIA GeForce GTX 3090 GPU (24GB).

1) Semi-Supervised Semantic Segmentation: To illustrate the advantages of the GMM based loss in this paper, we use use it into U-Net [2], U-Net++ [16], TRMFCN [18], ViT [20], TransUNet [19], MTUNet [21] and the improved U-Net proposed in this paper on IBSR and MRBrainS18. All the methods used in the experiments have the same initial parameters. Adam optimization algorithm is used for model training. The learning rate is 10^{-4}, the batch size is 20 and the epochs is 200.

2) Unsupervised Segmentation: Our method can transform the existing supervised or semi-supervised segmentation network into unsupervised network by using L_{GMM_B}. For this purpose, we use the unlabeled IBSR data sets to train the network.

4.3 Results

1) Semi-Supervised Semantic Segmentation: In this experiment, we couple L_{Total} into six methods and set the hyper-parameter β as 10^{-10} on randomly chosen 1/4, 1/2 and 3/4 labeled data of the IBSR and MRBrainS18. Since the labeled data can help the deep learning network to learn the characteristics of the bias field, this experiment of all semi-supervised methods do not consider any bias field estimation. We used recall to illustrate the advantages of the proposed network and L_{GMM_TV}.

Table.1 and Table.2 list the average recall values, which illustrates that L_{GMM_TV} improves the accuracy of all these methods when there is only 1/4 labeled data sets. To show the accuracy of each tissue, we use intersection over-union (IoU) to evaluate the model segmentation results and the mean values on IBSR and MRBrainS18 with 1/4 labeled data sets are shown in Table.3 and Table.4. It can be found that our method can obtain more accurate results(higher mean values). Furthermore, our method obtains a higher CSF value, which indicates that our method can retain more details, and obtains a smaller variance in most cases, which indicates that our method is more robust.

Table 1 Recall values of semi-supervised methods with different ratio of labeled data on IBSR.

		U-Net	U-Net++	TRMFCN	ViT	TransUNet	MTUNet	Our method
L_{CE}	1/4	91.95	92.79	92.67	87.76	92.27	93.13	93.62
	1/2	93.83	94.49	94.27	89.15	94.03	94.70	94.95
	3/4	94.54	95.11	95.01	89.96	94.61	95.29	95.50
$L_{CE}+L_{GMM\_TV}$	1/4	92.89	92.87	93.07	88.02	92.56	93.24	93.64
	1/2	94.17	94.61	94.43	89.37	94.11	94.76	95.01
	3/4	94.73	95.26	95.03	89.99	94.65	95.31	95.54

Table 2 Recall values of semi-supervised methods with different ratio of labeled data on MRBrainS18.

		U-Net	U-Net++	TRMFCN	ViT	TransUNet	MTUNet	Our method
L_{CE}	1/4	87.63	87.31	87.47	65.18	87.11	79.75	88.07
	1/2	88.39	88.04	88.19	65.95	87.83	80.71	88.79
	3/4	88.27	88.62	88.71	66.41	88.37	81.13	89.33
$L_{CE}+L_{GMM\_TV}$	1/4	87.91	87.58	87.72	65.17	88.04	82.09	88.09
	1/2	88.82	88.42	88.53	65.99	88.97	82.89	88.92

	3/4	89.93	88.99	88.97	66.47	89.41	83.32	89.39

Fig.4 shows four examples of the segmentation results on 1/4 labeled data. The first four rows show the segmentation results on IBSR and the other four rows show the results on MRBrainS18. Odd rows are based on L_{CE}, and even rows are based on $L_{CE}+L_{GMM\_TV}$. The detail information of the results show that the models with L_{GMM_TV} can obtain more accurate results. In addition, even in the case of a small amount of labeled data, our loss can improve the accuracy of learning methods.

2) Unsupervised Semantic Segmentation: The experiment of unsupervised learning methods directly illustrates that the combination of GMM method and deep learning network increase the accuracy of semantic segmentation methods. The bias field changes the intensity distribution of brain MRI images and greatly reduces the segmentation accuracy of unsupervised models. The bias field is coupled into the learning model to improve the segmentation performance.

Table 3 Quantitative results of semi-supervised brain MR images of IBSR val set using 1/4 labeled data.

		CSF	GM	WM
L_{CE}	U-Net	78.82±0.4390	92.13±0.0463	89.17±0.5819
	U-Net++	81.56±0.3839	92.96±0.0392	90.22±0.4993
	TRMFCN	79.91±0.4310	92.55±0.0433	89.63±0.5837
	ViT	58.77±1.0172	86.03±0.1156	79.12±1.5737
	TransUNet	77.63±0.0429	92.04±0.0429	88.91±0.5973
	MTUNet	81.19±0.3449	93.14±0.0325	90.48±0.4590
	Our	81.64±0.3422	93.20±0.0319	90.63±0.4227
$L_{CE}+L_{GMM\_TV}$	U-Net	79.25±0.4241	92.20±0.0453	89.21±0.5657
	U-Net++	81.58±0.3843	93.00±0.0397	90.28±0.5126
	TRMFCN	80.26±0.4244	92.59±0.0435	89.67±0.5887
	ViT	59.37±1.0012	86.92±0.1036	79.91±1.4331
	TransUNet	78.21±0.0413	92.54±0.0402	89.41±0.5737
	MTUNet	81.41±0.3437	93.24±0.0317	90.54±0.4523
	Our	81.74±0.3388	93.30±0.0313	90.71±0.4225

Fig.5 shows the segmentation results on images form IBSR, which have severe intensity inhomogeneity. The loss based on GMM transforms supervised/semi supervised learning model into unsupervised model with superior performance. GMM has not consider any information of bias field, which makes it hard to find satisfactory results. Coupling the bias field information improves the performance of our method. To quantify the accuracy of our method, we computed the precision, recall, F1-score and JS values as evaluation metrics, which can be found in Table.3. This images of IBSR contain a small amount of CSF, and brain MRI analysis focuses on GM and WM, so we only gives quantitative results of GM and WM. For GM and WM, we use the proposed loss function method to achieve 10% and 4% higher accuracy than GMM respectively, which shows that the depth learning network can achieve improved image segmentation without ground truth labels.

Table 4 Quantitative results of semi-supervised brain MR images of MRBrainS18 val set using 1/4 labeled data.

		CSF	GM	WM
L_{CE}	U-Net	85.56±0.2102	75.59±0.1303	71.98±2.6008
	U-Net++	85.09±0.2316	75.35±0.1498	71.83±2.5624
	TRMFCN	85.60±0.2083	75.27±0.1452	72.15±2.3695
	ViT	47.12±0.6582	47.96±0.3079	46.01±3.7844
	TransUNet	83.84±0.2076	75.37±0.1175	71.91±2.6133
	MTUNet	82.49±0.0.2782	66.34±0.6290	59.72±2.4865
	Our	85.79±0.2000	76.91±0.1068	72.80±2.7132
$L_{CE}+L_{GMM\_TV}$	U-Net	85.95±0.1955	76.60±0.1061	73.01±2.3719
	U-Net++	85.55±0.2035	76.16±0.1437	73.25±2.5269
	TRMFCN	85.70±0.2080	76.04±0.1292	72.86±2.0678
	ViT	48.02±0.6152	48.36±0.3007	46.61±3.5724
	TransUNet	84.24±0.2011	76.11±0.1015	72.21±2.5123
	MTUNet	83.25±0.2402	67.24±0.6079	61.37±2.4215
	Our	86.09±0.1908	77.12±0.1356	73.01±2.6338

Table 5 Unsupervised segmentation accuracy based on IBSR data.

		Precision	Recall	F1-Score	JS
GMM	GM	80.90	52.60	63.42	46.90
	WM	74.79	74.86	74.17	59.92
Our method without bias field correction	GM	85.12	60.12	70.17	54.56
	WM	73.12	80.02	75.59	61.46
Our method with bias field correction	GM	89.26	68.39	77.29	63.15
	WM	77.96	84.74	80.79	68.16

Journal of Information and Computing Science, Vol. 17(2022) No. 2, pp 138-153

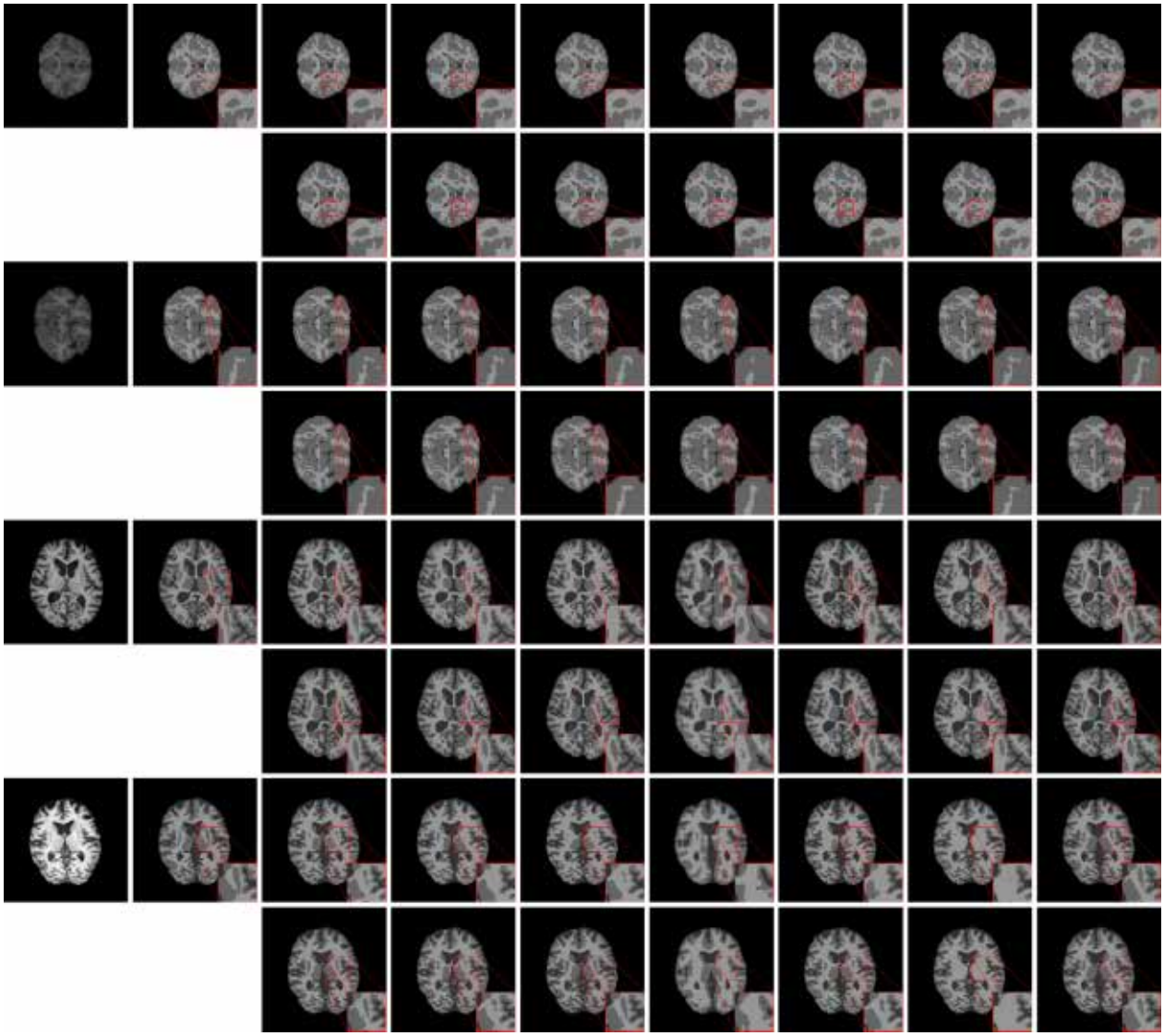

Fig. 4 Semi supervised brain segmentation results of 1/4 calibration data in IBSR and MRBrainS18. The first to last columns are the initial image, the ground truth, the segmentation results of UNet, UNet++, TRMFCN, ViT, TransUNet, MTUNet and our method, respectively. The first four rows show the segmentation results of ISBR. The other rows show the results of MRBrainS18. Odd rows are based on L_{CE}, and even rows are based on $L_{CE} + L_{GMM\_TV}$.

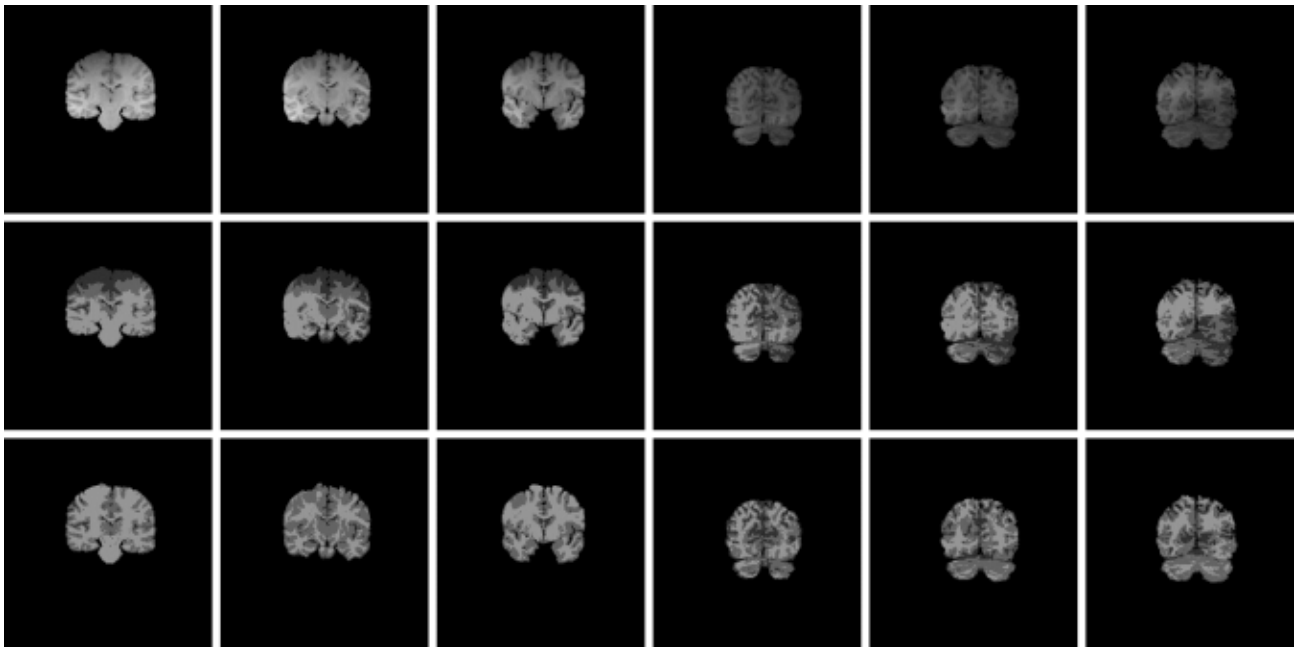

Fig. 5 Results of the unsupervised brain segmentation. The first row shows the initial images. The second row to third row shows the segmentation results of GMM, and our method, respectively.

5. Discussion

Our method can achieve good performance in both semi-supervised and unsupervised situations, but there are still some limitations. Non negative constant β in Eq.(3.5) balances the influences of L_{CE} and L_{GMM_TV}. The bigger β, the stronger the model's ability to extract sample personality characteristics. We use 200 images as a test set to verify the impact of β on the segmentation results. The average recall value is shown in Fig.6 (a). It has been shown that when β is set to about 10^{-10}, our method can obtain satisfactory results.

γ in Eq.(3.4) is a non negative constant that will affect the accuracy of the model. It should be small enough to maintain clarity, while a bigger value can reduce the effect of noise. Fig.6(b) shows the average recall value of γ and illustrates that when it is set to 80, we obtain more accurate results.

6. Conclusion

In this paper, we propose a novel GMM loss based deep learning framework, which can realize semi-supervised and unsupervised segmentation. The main motivation of the loss function proposed in this paper is that the output of the softmax layer of the deep learning model and the posterior probability in the Gaussian mixture model are surprisingly similar, so the improved GMM likelihood function can be minimized using the deep learning network. the learning of a small number of calibrated samples or data without calibrated samples can be realized. Experiments on various data sets have proved the effectiveness of the loss function.

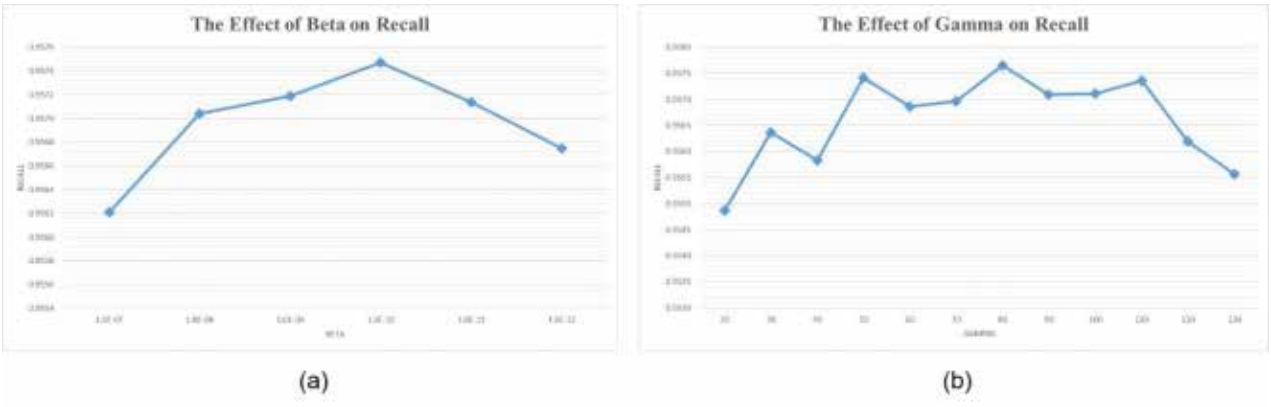

Fig. 6 The mean Recall values of β and γ on brain MR images (a) Recall values of β. (b) Recall values of γ.

References:

1. Long, Jonathan, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, 3431--3440, 2015.
2. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention, 9351: 234--241, 2015.
3. S. Yun, D. Han, S. J. Oh, et al, "Cutmix: Regularization strategy to train strong classifiers with localizable features," Proceedings of the IEEE/CVF international conference on computer vision, 6023--6 032, 2019.
4. E. D. Cubuk, B. Zoph, J. Shlens, et al, "Randaugment: Practical automated data augmentation with a reduced search space," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 702--703, 2020.
5. Guo, Xiaoqing, Jie Liu, and Yixuan Yuan, "Semantic-Oriented Labeled-to-Unlabeled Distribution Translation for Image Segmentation," IEEE Transactions on Medical Imaging, 41(2): 434-445, 2021.
6. X. Li, L. Yu, H. Chen, et al, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," arXiv:1808.03887, 2018.
7. X. Li, L. Yu, H. Chen, et al, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," IEEE Transactions on Neural Networks and Learning Systems, 32(2): 523-534, 2020.
8. Li, Shuailin, Chuyu Zhang, and Xuming He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," International Conference on Medical Image Computing and Computer-Assisted Intervention. 12261: 552--561, 2020.
9. G. McLachlan and D. Peel, "Finite mixture models," John & Sons, 2000.
10. M. S. Allili, D. Ziou, N. Bouguila, et al, "Image and video segmentation by combining unsupervised generalized Gaussian mixture modeling and feature selection," IEEE Transactions on Circuits and Systems for Video Technology, 20(10): 1373-1377, 2010.
11. Adelchi Azzalini, "Skew-Normal Family of Distributions," American Cancer Society, pp: 1–8, 2015.
12. Diplaros, Aristeidis, Nikos Vlassis, and Theo Gevers, "A spatially constrained generative model and an EM algorithm for image segmentation," IEEE Transactions on Neural Networks, 18(3): 798--808, 2007.
13. Nguyen, Thanh Minh, and QM Jonathan Wu, "Fast and robust spatially constrained Gaussian mixture model for

image segmentation," IEEE transactions on circuits and systems for video technology, 23(4): 621--635, 2012.

14. Z. Ji, J. Liu, H. Yuan, et al, "A spatially constrained asymmetric gaussian mixture model for image segmentation," Image and Video Technology, 9431: 697--708, 2015.
15. Ibtehaz, Nabil, and M. Sohel Rahman. "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," Neural Networks, 121: 74-87, 2020.
16. Z. Zhou, M. M. Rahman Siddiquee, et al, "Unet++: A nested u-net architecture for medical image segmentation," Deep learning in medical image analysis and multimodal learning for clinical decision support, 11045: 3-11, 2018.
17. S. Shah, P. Ghosh, L. S. Davis, et al, "Stacked u-nets: a no-frills approach to natural image segmentation," arXiv:1804.10343, 2018.
18. Y. Chen, Y Qin, Z. Jin, et al, "A Triple Residual Multiscale Fully Convolutional Network Model for Multimodal Infant Brain MRI Segmentation," KSII Transactions on Internet and Information Systems (TIIS), 14(3): 962-975, 2020.
19. Chen, Y. Lu, Q Yu, et al, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv:2102.04306, 2021.
20. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv:2010.11929, 2020.
21. H. Wang, S. Xie, L. Lin, et al, "Mixed transformer u-net for medical image segmentation," ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022.
22. Ambroise, C, "The EM algorithm and extensions, by GM McLachlan and T. Krishnan," Journal of Classification, 15(1): 154--156, 1998.
23. Y. Chen, M. Cai, C. Ning, et al, "A Robust Spatial Information-Theoretic GMM Algorithm for Bias Field Estimation and Brain MRI Segmentation," IEEE Access, 8: 89617-89629, 2020.
24. Banerjee, Abhirup, and Pradipta Maji, "Spatially constrained Student's t-distribution based mixture model for robust image segmentation," Journal of Mathematical Imaging and Vision, 60(3): 355-381, 2018.
25. W. M. Wells, W. E. L. Grimson, R. Kikinis, et al. "Adaptive segmentation of MRI data," IEEE transactions on medical imaging}, 15(4): 429-442, 1996.
26. Banerjee, Abhirup, and Pradipta Maji, "Rough-probabilistic clustering and hidden Markov random field model for segmentation of HEp-2 cell and brain MR images," Applied Soft Computing, 46: 558-576, 2016.
27. K. Van Leemput, F. Maes, D. Vandermeulen, "Automated model-based bias field correction of MR images of the brain," IEEE transactions on medical imaging, 18(10): 885-896, 1999.
28. Li, Chunming, John C. Gore, and Christos Davatzikos, "Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation," Magnetic resonance imaging, 32(7): 913-923, 2014.
29. I. Goodfellow, J. Pouget-Abadie, M. Mirza, "Generative adversarial nets," Communications of the ACM, 63(11): 139--144, 2020.
30. W.C. Hung, Y. H. Ysai, Y. T. Liou, et al, "Adversarial learning for semi-supervised semantic segmentation," arXiv:1802.07934, 2018.
31. X. Chen, Y. Yuan, G. Zeng, et al, "Semi-supervised semantic segmentation with cross pseudo supervision," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp: 2613--2622, 2021.
32. W. Xia, C. Domokos, J. Dong, et al, "Semantic segmentation without annotating segments." Proceedings of the IEEE international conference on computer vision, pp: 2176--2183, 2013.
33. J. Zhu, J. Mao, and AL. Yuille, "Learning from weakly supervised data by the expectation loss svm (e-svm)

algorithm," Advances in neural information processing systems, 27, 2014.

34. Pinheiro, Pedro O., and Ronan Collobert, "From image-level to pixel-level labeling with convolutional networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp: 1713--1721, 2015.
35. Pathak, Deepak, Philipp Krahenbuhl, and Trevor Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," Proceedings of the IEEE international conference on computer vision, pp: 1796--1804, 2015.
36. D. Lin, J. Dai, J. Jia, et al, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp: 3159--3167, 2016.
37. T. M. Nguyen, and Q. Wu, "Fast and robust spatially constrained gaussian mixture model for image segmentation," IEEE Transactions on Circuits and Systems for Video Technology, 23(4): 621--635, 2013.
38. S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation," IEEE Transactions on Fuzzy Systems, 16(5): 1351--1361, 2008.