Journal of Information and Computing Science
J. Info. Comput. Sci., Vol. 19, No. 2, pp. 131-154

DOI: 10.4208/JICS-2024-008
December 2024

A Dual-Task Semi-supervised Neural Network Based on Skew Normal Mixture Model for Brain MR Image Segmentation

Shengyang Liao[1,*] and Yunjie Chen[1]

[1] *School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Abstract. Accurate segmentation of brain magnetic resonance (MR) images is critical in brain disease research and treatment. While deep learning methods have advanced image segmentation by extracting hierarchical features, they typically require large labeled datasets for precise results. Acquiring annotated medical data remains challenging due to the need for specialized expertise and privacy restrictions. To address this, we propose a semi-supervised model combining dual tasks: segmentation and boundary feature regression. For class imbalance in segmentation, the network employs focal loss to extract common features from annotated data. To handle asymmetric data distributions, a skew Normal Mixture-based Level set loss guides the network to learn individual image characteristics, enhancing class distribution fitting. This dual-feature integration enables strong performance on limited datasets. In regression, Level set signed distance functions focus the model on boundary information, mitigating partial volume effects on focal loss. Experiments on IBSR and MRBrainS18 datasets demonstrate our method's advantages over current state-of-the-art approaches.

AMS subject classifications: 62E99, 68T10

Key words: Brain MR image segmentation, Semi-supervised Learning, Skew Normal Mixture model, Level set functional, Dual task.

1 Introduction

The phenomenon of aging population has become one of the important issues in contemporary society, and more and more diseases are beginning to harm the health of the elderly. Among them, brain diseases have a high incidence rate and mortality, and are more

*Corresponding author. *Email addresses:* 20211215002@nuist.edu.cn (S. Liao), priestcyj@nuist.edu.cn (Y. Chen)

frequent in the elderly. From a medical perspective, effective prevention, accurate diagnosis, and rational treatment of brain diseases are beneficial for promoting healthy aging. Medical imaging technology is widely used in the diagnosis of brain diseases. This technology can effectively obtain tissue information and pathological localization. Among numerous medical imaging technologies, magnetic resonance imaging has a higher resolution of soft tissues and is non-invasive and radiation free for the human body. In addition to its advantages, magnetic resonance (MR) images often exhibit noise, offset fields and partial volume effects, and there are also differences between different individuals. Meanwhile, brain tissue segmentation requires doctors to manually implement, which can consume a significant amount of time. Numerous studies are focusing on developing advanced automatic segmentation techniques to enhance generalization capability.

Currently, various techniques [1] are employed for brain MR image segmentation including unsupervised models [2], supervised models [3,4], semi-supervised models [5], and weakly supervised models [6]. Unsupervised models rely solely on the information within the image itself for analysis, often without utilizing any prior knowledge. This techniques can be further grouped into statistical model-based methods [7] and PDE-based methods [8, 9], etc. Statistical model-based methods typically involve fuzzy C-means clustering models [10], finite Mixture models [11, 12], and more. On the other hand, the level set model [13], a classical curve/surface evolution model based on PDE theory, is widely used in brain MR image segmentation as it can effectively segment multiple objects in an image. Despite its strengths, unsupervised methods face challenges in achieving optimal segmentation results due to factors such as bias field, noise, and weak boundaries.

Over the past few years, deep learning technology [14] has become increasingly prominent in the field of brain MR image segmentation. Deep learning techniques, such as the fully convolutional network (FCN) [15] and U-Net [16], have emerged as powerful tools in this domain, demonstrating remarkable performance. These models have the ability to extract pertinent features from bottom to top, enabling more accurate and effective segmentation of brain MR images. It is regrettable that deep learning techniques are heavily dependent on rich labeled data [17]. Brain MR image segmentation poses unique challenges due to the requirement of extensive clinical knowledge and time-consuming efforts in acquiring annotated data. Furthermore, ethical and privacy concerns regarding patient information restrict the availability of such data, making its acquisition a costly and difficult task. Additionally, different imaging devices from various manufacturers often produce images with distinct distributions, compounding the challenge of training accurate segmentation models.

Semi-supervised methods [18] have been widely employed in the field of medical image segmentation to address the challenge posed by limited sample sizes and have demonstrated promising results. For example, Guo et al. [19] expanded the limited annotated data by applying operations such as translation, rotation, and linear combinations to the training data to alleviate overfitting. However, it should be noted that the generated augmented data still falls within the distribution of the annotated data, which

may lead to inaccurate derivation of decision boundaries. Shi et al. [20] presented a novel framework that enhances unlabeled data using two types of augmentations: strong and weak. These augmented data are then fed into a mutually dual-consistent learning module guided by uncertainty.

The current semi-supervised techniques still face the following main problems: firstly, due to the small scale of annotated data, its distribution cannot fully characterize the overall sample distribution. Therefore, there is a need to explore how to utilize unlabeled data to improve the model's ability to capture the overall sample characteristics. Secondly, individual variations in human bodies may result in certain individual features present in the images. It is important to leverage both the overall sample characteristics and individual features to enhance the final segmentation accuracy. Lastly, the presence of weak boundaries caused by partial volume effects [21] can lead to low confidence in the analysis results of that region, resulting in insufficient boundary preservation capability in the final output.

In this work, we introduce Skew Normal Semi-Supervised U-Net (SNSS-UNet), a novel model which aims at segmenting brain MR images. Our network focuses on achieving accurate segmentation results even when only limited labeled data is available. To accomplish this, our network consists of two modules: an image segmentation branch and a boundary feature regression branch. In the image segmentation branch, we introduce a Level Set loss based on the Skew Normal Mixture model. This loss function aids in steering the network toward grasping the typical features of the input data. Furthermore, the Focal loss function is applied to facilitate the network's understanding of the general characteristics of the labeled data. In boundary feature regression branch, we employ mean square errors of level set signed function between input and output images as the loss function. This helps mitigate the influence from low contrast regions, which often result in insufficient confidence in the network's segmentation. By focusing more on the target boundary information, the network can improve its segmentation accuracy. In order to ensure consistency between two different modules, this paper also introduces a residual connection module to couple the segmentation module and the regression module. Meanwhile, the two modules both share the same encoding structure to extract image features.

This paper details the following contributions. (1) For a few of annotated data and asymmetric data, we design a novel semi-supervised image segmentation network based on Skew Normal Mixture Model combined with level set energy functional loss function, coupled with fully supervised focal loss, which can improve the accuracy of deep neural models in small sample situations. The C-V level set energy functional uses the Euclidean distance between pixels and center points as a similarity measure, and the advantage of this loss function is that it combines with Skew Normal mixture models to consider the true distribution characteristics of image pixels, thus solving the problem of Euclidean distance being sensitive to noise and outliers. (2) To solve the weak boundary problem caused by partial volume effects in images, we propose a boundary feature regression module. The regression module uses the mean square error of prediction level set and

label level set as the loss function, and combines it with the level set loss based Skew Normal Mixure Model and focal loss in the segmentation module. Thus, the module can effectively extract boundary features while fully extracting regional features. (3) To ensure that the features extracted by the segmentation module and regression module are of the same dimension, we add the residual structure to couple the features of decoding structures under two different modules. The coupling of image segmentation and boundary feature regression complements each other to guide the network to achieve better convergence. The final coupled network is called as SNSS-UNet.

The content of this paper consists of five sections: Section 1 outlines the research background and significance of brain MR image segmentation, as well as various problems and improvements that exist. Section 2 focuses on introducing various related methods of brain MR image segmentation. Section 3 elaborates on our proposed Skew Normal Semi-Supervised U-Net (SNSS-UNet) structure and its loss function. Section 4 introduces the brain MR image dataset, evaluation indicators used in the experiment and corresponding experimental results. Section 5 discusses and summarizes the experimental results, quantitatively analyzing the relationship between the values of hyperparameters in the network and the final segmentation accuracy.

2 Related Works

2.1 Unsupervised image segmentation methods.

Level set model [22] is widely used in medical imaging segmentation by virtue of curve or surface evolution theory [9]. The CV model [23] is one of the classical methods based on level set function. It has boundarylessness and computability, which will lay the foundation for the proposed method in this paper.

Finite Mixture Model (FMM) [11] is also a statistical clustering segmentation algorithm widely used in unsupervised image segmentation. The Gaussian Mixture Model (GMM) [24] based on symmetric normal distribution is one of the most widely used finite mixture models, and its parameter estimation is often completed through the EM algorithm. However, the model is prone to divergence and the data needs to follow a symmetric distribution, resulting in low segmentation quality for complex data.

The Gaussian Mixture Model has excellent fitting performance for data with an approximate normal distribution. In our task, the role of GMM is to divide the pixels in the image into background and three target classes. The probability density function of a single normal distribution is defined as

$$\Phi(\mathbf{x}_i|\theta_j)=\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma_j|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\left(\mathbf{x}_i-\mu_j\right)^T\Sigma_j^{-1}\left(\mathbf{x}_i-\mu_j\right)\right\}$$

Among them, μ and Σ represent the mean vector and covariance matrix, respectively. And, D represents the dimension of the observed data. According to the probability

density function, the logarithmic likelihood function of the Gaussian mixture model is as follows:

$$L\left(x_i|\pi_j,\left(\mu_j,\Sigma_j\right)\right)=\sum_{i=1}^{N}\log\sum_{j=1}^{K}\pi_j p\left(\mathbf{x}_i|\left(\mu_j,\Sigma_j\right)\right)$$

Among them, π_j is the prior probability that the pixel belongs to the jth normal distribution, and p represents the set of different normal distributions.

2.2 Fully supervised image segmentation methods based on deep learning.

The Fully Convolutional Neural Network [15] is the first to use 1*1 convolution for feature fusion, opening up a new chapter in semantic segmentation and guiding the development direction [25] in the future. The FCN extracts semantic features of images through continuous convolution and pooling, and combines shallow and deep features through skip connections for feature fusion to capture lost information. On the basis of FCN, Ronneberger et al. designed the UNet network [16]. This network is similar to FCN, consisting of encoding structure, decoding structure, and skip connection structure. The encoding structure adopts four sets of convolution and pooling to extract spatial features of the image, increase the number of channels in the feature map, and reduce the size of the feature map. The decoding structure can be thought of as a reverse process of the coding structure, which in turn recovers images. The fusion of encoding shallow features and decoding deep features is achieved through skip connections, and dimension concatenation is performed based on channels. The receptive field of UNet network is larger than that of FCN, and the robustness of the model is also higher. Considering semantic gap that comes from different functional layers, Zhou et al. improved the skip connection structure and introduced the UNet++ network [26]. A high number of parameters in UNet++ can slow down network convergence, and the main consideration is single scale feature fusion. Huang et al. further designed the UNet3+ network [27] based on UNet++ to address the problem of multi-scale feature fusion. However, these networks often pay less attention to the correlation between target regions and different channel feature maps in a single image, and perform poorly in effectively mining useful features. In order to implicitly learn to suppress irrelevant regions in the input image while highlighting salient features that are helpful for specific tasks, Oktay et al. proposed Attention U-Net [28]. Thus, the attention mechanism can improve the sensitivity of the model to foreground pixels. While these methods perform well with enough training data, they often struggle to achieve satisfactory results when the sample size is limited. To extract individual features from samples, it is often necessary to rely on unsupervised models.

2.3 Semi-supervised image segmentation methods based on deep learning.

The Mean Teacher model [29] was first proposed by Tarvainen and is one of the classic representatives of semi-supervised segmentation models. It is a method of sliding average of model weights rather than label prediction. The Mean Teacher model deeply

implements the core principles of π-model [30] and Temporary Ensembling [31], and is further improved and supplemented. Currently, most model training is accomplished by perturbing data or models. However, complex network designs and diverse perturbation methods may affect the accuracy of segmentation results. In order to clearly construct task-oriented regularization rather than implicitly constructing perturbations, an innovative dual-task consistent semi-supervised network (DTC) was developed [32]. In order to better obtain boundary features, Kristiadi et al. proposed ACM_CNN model [33], combining deep convolutional neural networks with level set active contour models. The advantage of this design is that it can simultaneously predict category boundaries in both labeled and unlabeled data. In addition to the aforementioned models, Chen et al. also proposed a semi-supervised semantic segmentation method [34] that enhances prediction confidence, which is crucial for improving the reliability of segmentation outcomes in scenarios with limited labeled data. Similarly, Hu et al. introduced an approach based on adaptive equalization learning [35], which aims to balance the learning process and enhance the model's generalization capability. Chen et al. further explored the use of cross pseudo supervision [36], leveraging pseudo-labels generated from the model itself to implement self-training. Wang et al. addressed the challenge of using unreliable pseudo-labels by developing a robust framework [37] that can effectively handle noisy labels, thus enhancing the performance of semi-supervised semantic segmentation.

3 Proposed method

Considering scarcity of labeled samples and asymmetric data in brain MR images, we introduce Skew Normal Semi-Supervised U-Net (SNSS-UNet), a novel network for segmenting brain MR images. This network merges the power of deep CNNs with a dual-task framework, which has an image segmentation module and a boundary feature regression module.

The SNSS-UNet network simultaneously undertakes image segmentation task and boundary feature regression task. To ensure consistency between two different tasks, we directly use residual connections to extract the correlation between features in the two tasks. Figure 1 showcases the network structure of the SNSS-UNet, providing a visual representation of its architecture and the flow of information within the model. The SNSS-UNet architecture consists of two main components: an image segmentation module and a boundary feature regression module. The two modules share the same encoding structure, which is consistent with the UNet++ network.

The segmentation module utilizes the UNet++ [26] architecture as the backbone network. The UNet++ network extends the UNet network by incorporating nested and skip connections, allowing for more comprehensive feature extraction and better segmentation performance.

On the other hand, the regression module employs the standard UNet [16] as its backbone network. The UNet architecture is well-known for its effectiveness in image-to-

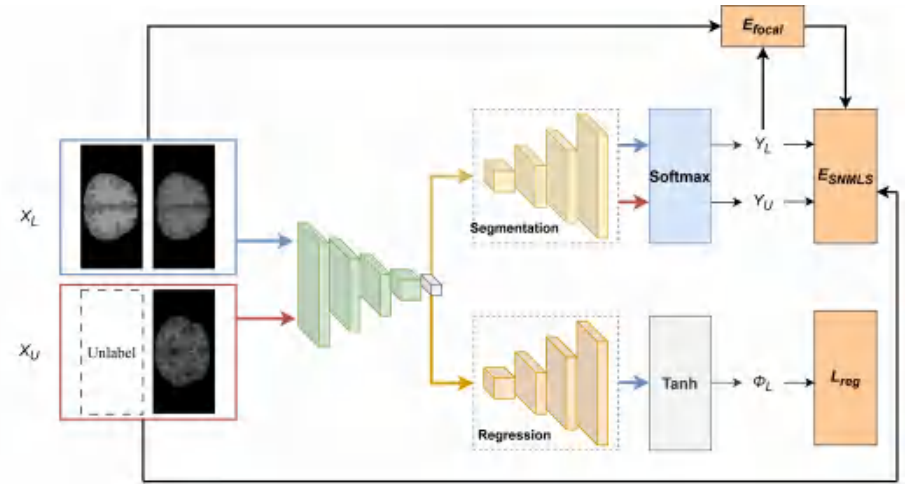

Figure 1: Structure of the proposed SNSS-UNet

image regression tasks and serves as a suitable choice for the boundary feature regression task in this work.

To establish coupling between the segmentation and regression modules, a residual module is employed, as depicted in Figure 1 [38]. This module aids in preserving and transferring relevant information between the two modules, facilitating their collaboration in the overall network.

All functional layers in the SNSS-UNet architecture inherit the configuration of the backbone network. For upsampling, a 3x3 transposed convolution operation is utilized.

Regarding the choice of stimulus response functions, the segmentation module employs the softmax function, which allows for probabilistic outputs representing different class probabilities. Meanwhile, the regression module uses the hyperbolic tangent function, enabling the regression model to predict continuous values.

By utilizing these architectural choices and stimulus response functions, the SNSS-UNet aims to achieve accurate segmentation and boundary feature regression for brain MR image analysis.

3.1 Segmentation branch

3.1.1 Backbones

Inspired by the UNet++ network concept, our proposed image segmentation network also satisfies the encoding and decoding structures, and connects them through convolutional blocks at different levels. In this way, SNSS-UNet is capable of thoroughly extracting features of brain MR images and alleviating the semantic gap phenomenon to a certain extent.

We chose the UNet++ framework because it has demonstrated high segmentation accuracy in various medical image segmentation tasks. To promote feature reuse, we

have employed the densenet [38] architecture as encoder of our suggested segmentation branch. The decoder of our network is enabled through transposed convolution. To combine the feature maps from the decoding and encoding stages of the same layer, we utilize dimensional concatenation.

3.1.2 Loss Function

Deep learning models typically use cross-entropy to guide network training, thereby learning common features from annotated data. However, when there is a lack of annotated data, overfitting may occur. In this case, combining individual features of unlabeled samples can effectively strengthen the network's potential in understanding the overall characteristics of samples. The final network output of the segmentation branch depends on the softmax function activation result of the feature fusion layer, which indicates the likelihood of each pixel associated with a particular category. In brain MR images, the heat maps of the signed distance functions is shown in Figure 2.

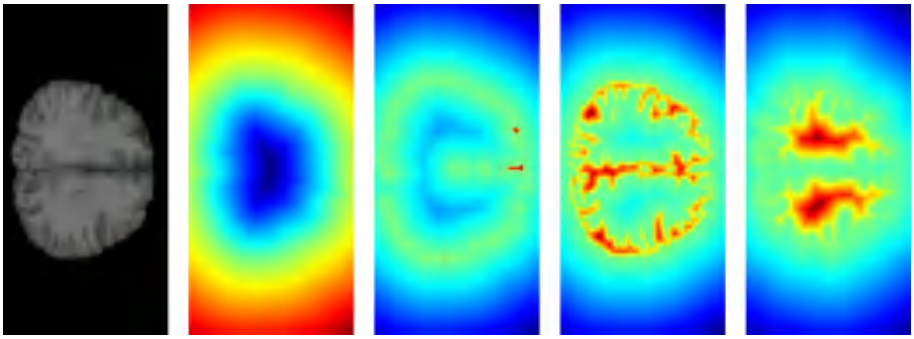

Figure 2: The heat maps of the signed distance functions. Starting with the first figure, the sequence is signed distance function heat maps of the original image, background, cerebrospinal fluid, gray matter, and white matter. The red portion of each map represents the distribution area of the target class in the heat map. The three heat maps on the right provide a perfect overview of the level set structures of cerebrospinal fluid, gray matter, and white matter. In the heat maps, the imbalance in the proportions of the three target classes is also evident.

Similarly, $M(\phi_1,\phi_2)$ in the C-V level set energy function can also approximate the likelihood of each pixel associated with each category. However, the energy function in the C-V level set model only considers the mean center of the image and does not calculate the corresponding variance, resulting in poor robustness to noise. Therefore, using a loss function based on Euclidean distance may not produce the best segmentation results. To mitigate the perturbation of segmentation outcomes by noise and offset field elements, we combine FMMs with the C-V model. Thus, the improved energy function can simultaneously calculate the mean and variance of various categories. We once again combine this energy function as a loss function with the cross-entropy in deep learning, thereby promoting the network to capture the common features of the all samples and the individual features of the each independent sample.

For example, the GMM assumes that all pixels in a brain MR image satisfy a mixture of standard normal distributions. Then a GMM based on C-V level set energy function

can be defined as

$$\begin{aligned} E=&\int_{\Omega}\sum_{k=1}^{4}-log\left(p(u(x,y),\theta_j)M_k(\phi_1,\phi_2)dxdy\right.\\ &+v\int_{\Omega}|\nabla H(\phi_1)|dxdy+v\int_{\Omega}|\nabla H(\phi_2)|dxdy \end{aligned} \tag{3.1}$$

Based on the observation in Kim et al. [39], M_k can be replaced by using the softmax results $s(x,y)$ of the preceding layer. Where, $s(x,y)$ is defined as

$$s_j(x,y)=\frac{e^{z_j(x,y)}}{\sum_{k=1}^{K}e^{z_k(x,y)}},j=1,2,...,K, \tag{3.2}$$

$z_j(x,y)$ represents the the preceding layer. Then Eq. (3.1) can be written as

$$E_{GMLS}=\int_{\Omega}\sum_{k=1}^{4}log\left(p(u(x,y),\theta_k)s_k(x,y)dxdy+v\int_{\Omega}|\nabla s_k(x,y)|dxdy.\right. \tag{3.3}$$

According to the EM algorithm, the parameters of the model can be calculated as

$$\mu_k=\frac{\sum_{i=1}^{N}u_i s_{i,k}}{\sum_{i=1}^{N}s_{i,k}} \quad \text{and} \quad \Sigma_k=\frac{\sum_{i=1}^{N}(u_i-\mu_k)^T(u_i-\mu_k)s_{i,k}}{\sum_{i=1}^{N}s_{i,k}},$$

where N is the number of the pixels in the image. Since the Gaussian mixture distribution is employed to characterize the image distribution, it provides a better representation of sample personality features. In order to obtain the overall characteristics of the samples, we combine cross-entropy loss as the final loss function for the segmentation model

$$L=\alpha E_{CE}+\beta E_{GMLS},$$

where

$$E_{CE}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}y_{ij}log s_{i,j}.$$

When u is an annotated sample, $\beta=1$, otherwise, $\beta=0$. The segmentation module can not only capture the feature information of annotated samples, but also obtain the feature information of unannotated samples, thus reducing the risk of model overfitting.

For some symmetric data, the Gaussian mixture model (GMM) is able to fit the distributions of different categories. However, the distribution of pixels in brain MR images is usually asymmetric. For data with asymmetric normal distribution, the skew normal mixture model (SNMM) can more accurately fit the distribution of various categories in brain MR images compared to the traditional Gaussian mixture model, especially for some categories with high skewness, which helps to better extract the individual features of sample images. A visualization of the SNMM probability density distribution is shown in the Figure 3.

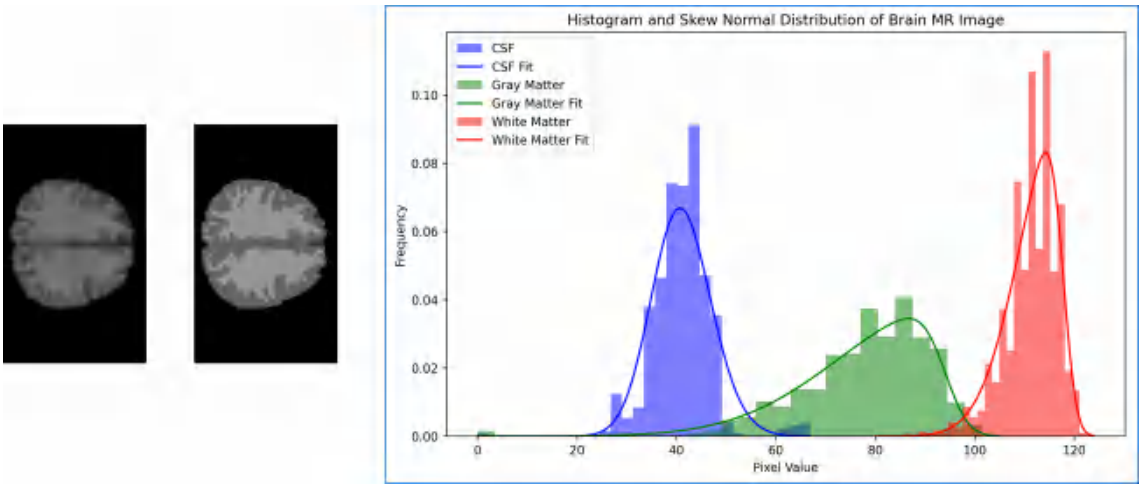

Figure 3: Visualization of skew normal mixed distribution of brain MR images. The first subgraph is the original image, the second subgraph is the ground truth, and the third subgraph is the histogram and skew normal distribution curves of cerebrospinal fluid, gray matter, and white matter. Taking this whole figure as an example, it can be seen that the skew normal mixture model is very accurate for the fitting of gray matter and white matter.

Level set methods drives curve evolution by minimizing an energy functional, while the Skew-Normal Mixture Model (SNMM) provides statistical modeling of data distributions. The mathematical integration of these methodologies requires the definition of a unified energy functional comprising two critical components: a data fidelity term derived from SNMM likelihood estimation and a geometric regularization term inherited from level set theory. Specifically, the Euclidean distance metric in the conventional Chan-Vese (C-V) energy functional is reformulated as the SNMM-based likelihood function, enhancing noise robustness by probabilistically weighting pixel intensities according to their skew-normal distribution characteristics. This formulation enables simultaneous optimization of regional statistical characteristics and boundary smoothness constraints. The integration process employs an alternating optimization strategy that iteratively updates segmentation boundaries and hyperparameters of SNMM, ensuring that the energy function decreases monotonically during optimization. This dual-update mechanism guarantees convergence while maintaining computational stability, thereby facilitating the convergence of our proposed SNSS-UNet architecture. Therefore, the C-V level set energy function combined with the skew normal mixing model is

$$\begin{aligned} L=&\int_{\Omega}\sum_{j=1}^{K}\log\left(P(u(x,y)|\Theta_j\right)^{Z_{i,j}}M_j(\phi_1,\phi_2)dxdy \\ &+v\int_{\Omega}|\nabla H(\phi_1)|dxdy+v\int_{\Omega}|\nabla H(\phi_2)|dxdy. \end{aligned} \tag{3.4}$$

The ϕ_1 and ϕ_2 in the above equation(3.4) are biphasic level sets, which can divide the pixels in the image into four categories. There are K categories of samples. Ω means the

entire image area; $u(x,y)$ represents each pixel in a two-dimensional image, and (x,y) can obtain its position in the image; Θ_j represents the set of parameters under different categories;$\left(P(x_i|\Theta_j\right)$ is a skew normal distribution probability density function, defined as

$$P\left(x_i|\Theta_j\right)=\frac{1}{2\pi^{\frac{1}{2}}}\frac{1}{|\Gamma_j|^{\frac{1}{2}}}exp\left(-\frac{1}{2}\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)^T\Sigma_j^{-1}\Gamma_j^{-1}\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)-\frac{1}{2}\tau_i^2\right), \quad (3.5)$$

where τ_i is a latent variable, which is derived by maximizing the joint probability density or conditional posterior probability. Specifically:

We assume the joint distribution of the observed variable x_i and latent variable τ_i is

$$P(x_i,\tau_i|\Theta_j)\propto\exp\left(-\frac{1}{2}\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)^T\Sigma_j^{-1}\Gamma_j^{-1}\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)-\frac{1}{2}\tau_i^2\right), \quad (3.6)$$

where the optimal estimate of τ_i is obtained by maximizing the log-likelihood of the joint distribution

$$\mathcal{L}(\tau_i)=\frac{1}{2}\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)^T A\left(x_i-b_i\mu_j-\Delta_j\tau_i\right)+\frac{1}{2}\tau_i^2, \quad (3.7)$$

where $A=\Sigma_j^{-1}\Gamma_j^{-1}$. And then, we take the derivative with respect to τ_i and setting it to zero to get

$$\frac{\partial\mathcal{L}}{\partial\tau_i}=-\Delta_j^T A(x_i-b_i\mu_j)+\Delta_j^T A\Delta_j\tau_i+\tau_i=0. \quad (3.8)$$

Finally, we solve for τ_i yields

$$\tau_i=\left(\Delta_j^T A\Delta_j+I\right)^{-1}\Delta_j^T A(x_i-b_i\mu_j), \quad (3.9)$$

where I is the identity matrix. τ_i quantifies the skewness characteristics of the observed data x_i. When $\Delta_j=0$, the model reduces to a symmetric normal distribution with $\tau_i=0$. In the E-step of the EM algorithm, τ_i is computed to update model parameters Θ_j, enabling precise modeling of asymmetric data.

Subsequently, we present a comprehensive derivation of closed-form parameter expressions through the Expectation-Maximization (EM) algorithm framework, with particular emphasis on elucidating the functional role of parameter τ_i in governing the mathematical tractability of the optimization process. The analytical derivation systematically addresses the interdependencies between latent variables and observable data, where τ_i serves as a critical mediator in maintaining probabilistic consistency during the iterative maximization of the complete-data likelihood. This formulation explicitly reveals how structural constraints imposed by τ_i enable stable convergence while preserving the statistical interpretability of skewness coefficient within the SNMM.

In conclusion, the parameters in the above probability density function $P\left(x_i|\Theta_j\right)$ are given by the following mathematical expressions:

$$\Delta_j=\Sigma_j^{\frac{1}{2}}\delta_j,$$

$$\Gamma_j = \Sigma_j - \Delta_j \Delta_j^T,$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^{N} \hat{z}_{i,j} b_i \left(\mathbf{x}_i - \Delta_j \hat{s}_{1,i,j}\right)}{\sum_{i=1}^{N} \hat{z}_{i,j} b_i^2},$$

$$\hat{\pi}_j = \frac{1}{N} \sum_{i=1}^{N} \hat{z}_{i,j},$$

$$\widehat{\Gamma}_j = \frac{\sum_{i=1}^{N} \hat{z}_{i,j} \left[\left(\mathbf{x}_i - b_i \mu_j - \Delta_j \hat{s}_{1,i,j}\right) \left(\mathbf{x}_i - b_i \mu_j - \Delta_j \hat{s}_{1,i,j}\right)^T + \left(\hat{s}_{2,i,j} - \hat{s}_{1,i,j}^2\right) \Delta_j \Delta_j^T \right]}{\sum_{i=1}^{N} \hat{z}_{i,j}},$$

$$\widehat{\Delta}_j = \frac{\sum_{i=1}^{N} \hat{z}_{i,j} \hat{s}_{1,i,j} \left(\mathbf{x}_i - b_i \mu_j\right)}{\sum_{i=1}^{N} \hat{z}_{i,j} \hat{s}_{2,i,j}},$$

$$\omega = U^{-1} V. \tag{3.10}$$

According to the estimation formula of the above parameters, the covariance matrix and skewness coefficient can be expressed as

$$\widehat{\Sigma}_j = \widehat{\Gamma}_j + \widehat{\Delta}_j \widehat{\Delta}_j^T,$$

$$\hat{\lambda}_j = \frac{\widehat{\Sigma}_j^{-1/2} \widehat{\Delta}_j}{\left(1 - \widehat{\Delta}_j^T \widehat{\Sigma}_j^{-1} \widehat{\Delta}_j\right)^{1/2}}.$$

The selection of skewness coefficient in the Skew Normal Mixture Model (SNMM) critically influences the segmentation accuracy of brain MR images. When $\lambda = 0$, the skew normal distribution will degenerate to a standard normal distribution, equivalent to the Gaussian Mixture Model (GMM). As λ increases, the distribution exhibits enhanced skewness to better capture tail characteristics in intensity histograms, thereby improving the model's capacity to fit asymmetric intensity patterns prevalent in neuroimaging data. However, excessive λ value may erroneously interpret noise components as distribution tails, resulting in irregular segmentation boundaries and potential model overfitting. To address this trade-off, our methodology integrates an expectation-maximization (EM) algorithm framework to adaptively estimate tissue-specific skewness coefficient λ for cerebrospinal fluid, gray matter, and white matter. This optimization strategy achieves equilibrium between model complexity and fitting efficacy, ensuring robust segmentation performance while maintaining plausibility in tissue boundary delineation.

In order to better obtain spatial information of images and focus on the common features of images, we combine the skew normal mixture model based on C-V method with our segmentation branch. The softmax probability values $s_{i,j}$ outputted by the network can approximately replace the posterior probabilities $Z_{i,j}$ in the skew normal mixture

model, M_j can be replaced by using the softmax results of the previous layer in the network. According to Eq.(3.4), the unsupervised loss function is modeled as

$$E_{SNMLS} = \int_{\Omega} \sum_{j=1}^{K} -\log\left(P(u(x,y)|\Theta_j)^{s_{i,j}} s_{i,j} dxdy + v\int_{\Omega} |\nabla s_{i,j}| dxdy. \tag{3.11}$$

In the above Eq.(3.11), the various parameters $\hat{\mu}$, $\hat{\pi}$, $\widehat{\Sigma}$ and $\hat{\lambda}$ in the loss function can all be solved by $s_{i,j}$, and then the network can be trained using backpropagation algorithm to obtain numerical solutions of the parameters.

Based on observation, it can be seen that the proportion of background and target classes is relatively unbalanced, and the proportion of cerebrospinal fluid in the target class is significantly lower than that in gray matter and white matter. For imbalance categories, we replace the traditional cross entropy loss with focal loss to more accurately extract the common features of the interested categories. The focal loss function is defined as

$$E_{focal} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} \alpha_j y_{i,j}(1-p_{i,j})^{\gamma} \log p_{i,j}, \tag{3.12}$$

where $p_{i,j} = s_{i,j}$ is the probability value of the network predicting the ith pixel as the jth class; $y_{i,j}$ is the label value, with a range of values consisting of a set of 0 and 1; α_j aims to control the contribution weights of different categories in the entire loss function. Finally, a comprehensive semi-supervised loss function is constructed by combining the skew normal mixture loss of the unsupervised coupling level set with the fully supervised Focal loss as the loss function of the segmentation module for network training. The specific expression is

$$L_{Total} = E_{focal} + \beta E_{SNMLS}, \tag{3.13}$$

where the hyperparameter β serves as an important means of adjusting E_{SNMLS} and E_{focal}, which can adjust the contribution weights of the two different loss functions, and balance the difference in values between unsupervised loss function and fully supervised loss function.

Considering the presence of partial volume effects, weak boundaries may be observed, which may lead to a lack of clear differentiation between different categories. To address this issue, we introduce a boundary information regression module to facilitate the network in extracting boundary features.

The boundary feature regression module aims to overcome the challenges posed by weak boundaries in the images. By integrating this module, the network becomes more capable of distinguishing different categories and further improving overall segmentation performance.

3.2 Boundary feature Regression branch

3.2.1 Backbones

In the boundary feature regression branch, the U-Net architecture [16] is employed as the backbone, utilizing the same encoding structure as the segmentation module.

By using U-Net as the backbone, the boundary feature regression module benefits from the effective feature extraction capabilities of the U-Net architecture. This allows the module to capture and utilize important boundary-related information for improved performance in capturing boundary features.

Overall, the boundary feature regression module, with U-Net as its backbone, contributes to our SNSS-UNet to utilize essential boundary information, enhancing the segmentation model's accuracy and performance.

3.2.2 Loss Function

The level set model views curves/surfaces as zero level sets on a higher-dimensional surface, and achieves segmentation through surface evolution. Therefore, its level set function can characterize the characteristic information of curves and their neighborhoods. Exploiting this, this paper uses the signed distance function of the calibration result as the final boundary feature information and constructs a regression loss to guide the network in capturing boundary information.

The signed distance function is defined as

$$\phi(x)=\begin{cases}-\inf||x-y||_2, & x\in S_{in}\\ 0, & x\in\partial S\\ +\inf||x-y||_2, & x\in S_{out},\end{cases}$$

where $y\in\partial S$ is a point of the boundary of the object. S_{in} and S_{out} represent the inside region and outside region. Thus, the regression loss is given by using the mean square error (MSE):

$$L_{reg}=\int_{\Omega}\sum_{i=1}^{2}||\phi_i-\widehat{\phi}_i||^2dx.$$

3.3 Coupling module

To effectively integrate the feature information between the segmentation branch and the regression branch, we design a feature coupling module based on Residual (Res) blocks, which can ensure double-branch consistency.

The Res block union can be represented as

$$y_i=x_i+F(x_i,W_i),$$

$$x_{i+1}=f(y_i),$$

where x_i and x_{i+1} are the ith (input) and $(i+1)$th (output), respectively. $f(\cdot)$ is the activation function and $F(\cdot)$ is the residual function.

The purpose of this idea was to simplify the training process and ensure that information propagation does not degrade, ultimately leading to improved network performance.

In summary, the combination of the feature coupling module and the Residual blocks enables effective integration of feature information from both modules, leading to improved representation and performance in capturing complex relationships between different tasks.

Then, the entire network is ultimately trained using a combined loss function that encompasses both two branches

$$L_{total} = L_{seg} + \gamma L_{reg}, \tag{3.14}$$

where γ represents a hyperparameter that controls the weight assigned to the losses from the segmentation module and the regression module.

4 Experiments

In this section, we performed five sets of comparative experiments to showcase the effectiveness of our proposed method. Among these experiments, we compared our method against various other models, including fully supervised models such as UNet++ [26], double UNet [32], Attention u-net [28], and a semi-supervised model called ACM_CNN [39].

To analyze the effect of annotated sample quantity on results, we explored relationship between sample size and segmentation accuracy using our dual-task semi-supervised network. Our findings demonstrate that our proposed dual task UNet achieves higher segmentation accuracy compared to the other models.

4.1 Datasets

We collected data from two sources for our experiments: the MRBrainS18 dataset (which is available at https://mrbrains18.isi.uu.nl) and the Internet Brain Segmentation Repository (IBSR) dataset (which is available at http://www.cma.mgh.harvard.edu/ibsr). The sample sizes of MRBrainS18 and IBSR are 282 and 1842, respectively.

To divide the dataset for training and testing, we set the training sample size to be 80%. The test sample should be different from the training sample. This partitioning strategy ensures that the model is trained on a diverse range of images while also allowing for unbiased evaluation on unseen data.

By incorporating data from multiple sources, we can more effectively evaluate and validate the performance of our proposed method in segmenting brain images.

4.2 Training and Evaluation Metrics

We employed the Adam optimizer as the algorithm for back-propagation optimization in the network training process. We set the learning rate to 0.001, which determines the step size for gradient updates during training. The choice of Adam optimizer and the specific learning rate were based on their proven effectiveness in achieving faster convergence and better optimization results. During the training phase, we used a small batch size of 6. A smaller batch size can help in reducing memory requirements and allow for more frequent updates, leading to potentially faster convergence. We conducted a total of 120 iterations during the training process. The experiments were conducted on a hardware environment equipped with an NVIDIA GeForce GTX 3090 GPU with 24GB of memory.

In order to facilitate a statistical analysis of the segmentation accuracy, we employed the Jaccard similarity (JS) metric [40] and the misclassification rate (MCR) [41] as evaluation criteria. The Jaccard similarity measures the overlap between the predicted segmentation and the ground truth, while the misclassification rate quantifies the proportion of misclassified pixels or regions. By employing these evaluation criteria, we can effectively compare and quantify the performance of our method against other models.

4.3 Results

The first experiment primarily focuses on the IBSR dataset. To demonstrate robustness of our proposed SNSS-UNet to the variation in the annotated sample size, we conducted this experiment using two different proportions of the dataset. Specifically, we used 20% and 80% of the annotated samples as the training set, while the remaining samples were utilized as the test set.

In this study, we conducted comparative experiments. Figure 4 provides a comprehensive comparison of the performance of our proposed SNSS-UNet and the other frameworks, highlighting the advantages of our approach.

The second experiment was conducted on the MRBrainS18 dataset. Similar to the previous experiment, we used 20% and 80% of the annotated samples as the training set, while the remaining samples were used as the test set. It is worth mentioning that the MRBrainS18 dataset has a smaller sample size compared to the IBSR dataset. This smaller sample size poses a challenge to deep learning methods as they are more prone to overfitting. The segmentation results are also shown in Figure 5.

The first and second rows in Figure 4 and Figure 5 display the segmentation results obtained using 80% and 20% annotated samples, respectively. Each column from left to right represents the initial images, the ground truths, segmentation results of UNet++, Attention u-net, double UNet, ACM_CNN, and our method, respectively.

All these methods can achieve comparable results when utilizing 80% annotated samples. However, when the size of annotated samples is small, supervised methods such as UNet++, double UNet, and Attention u-net struggle to produce satisfactory results. ACM_CNN is a semi-supervised method that employs the CV function as the loss, mak-

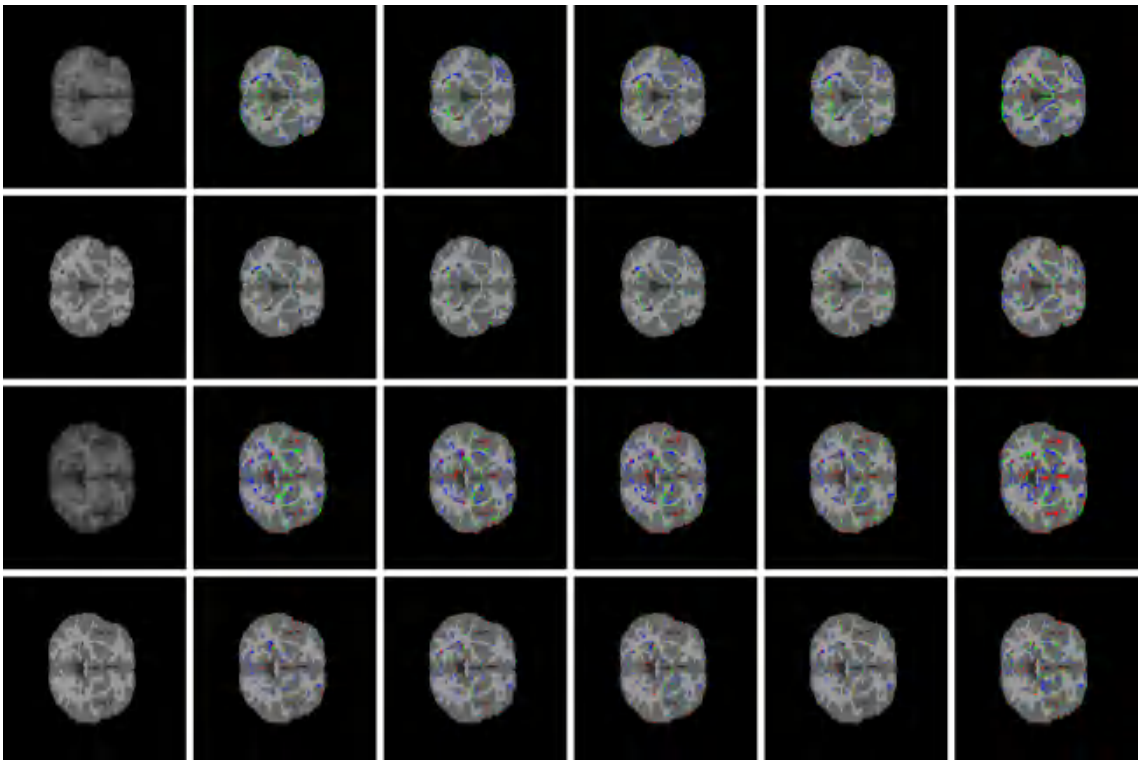

Figure 4: Segmentation results on IBSR dataset. The first row shows the results on 80% annotated samples and the second row shows the results on 20% annotated samples. From the left column to the right column show the initial images(the ground truth), the segmentation results of UNet++, Attention u-net, double UNet, ACM_CNN and our method, respectively.

ing it sensitive to weak edges. On the other hand, our method is based on a SNMM-based level set and utilizes an edge information regression model to preserve edge information.

Here we used the Jaccard similarity (JS) value and MCR values as the evaluation metric. From results of Table 1, it is evident that all the methods can achieve satisfactory segmentation results when the training set is sufficiently large. However, as the size of the labeled sample decreases, the semi-supervised methods exhibit higher accuracy compared to the fully supervised methods. This distinction is particularly noticeable in datasets like MRBrainS18, where the sample size is already small, with semi-supervised methods achieving 0.1% higher JS accuracy. Examining the segmentation accuracy of CSF, our method achieves the highest numerical value, indicating that the inclusion of boundary feature regression branches preserves detailed contours.

4.4 Ablation Analysis

To examine the impact of the skew Normal Mixture Model based level set loss and evaluate the effectiveness of dual branch networks, we performed ablation studies following

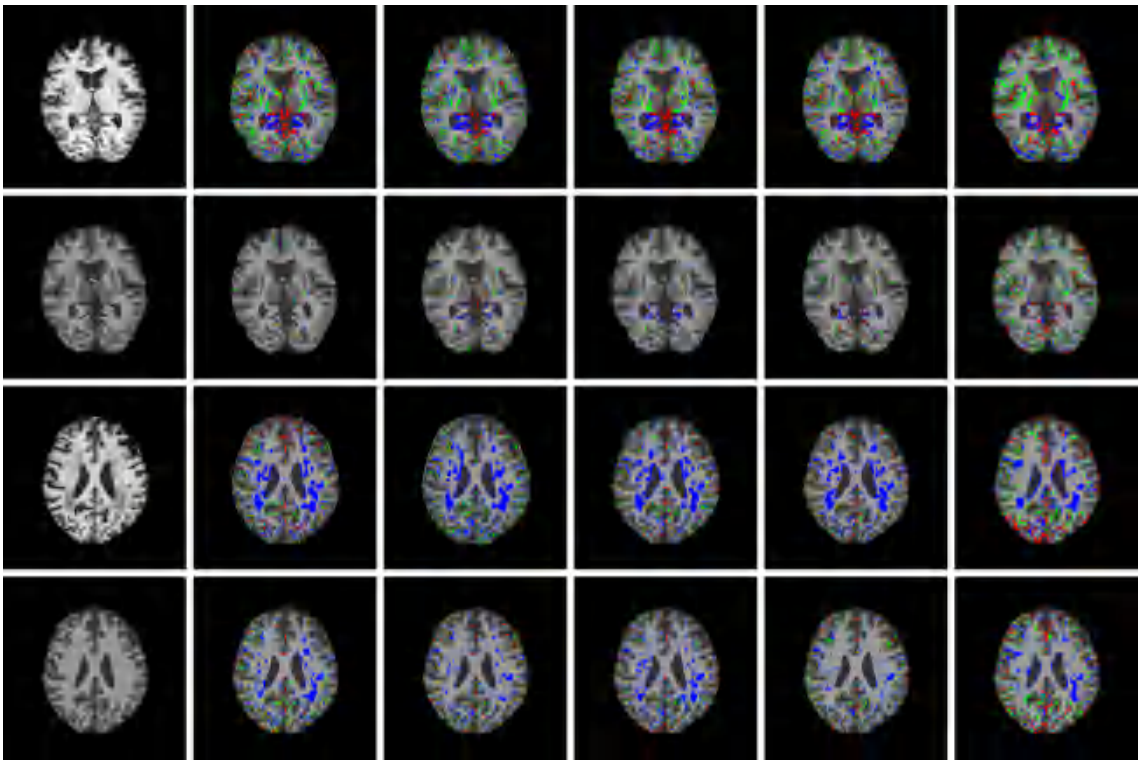

Figure 5: Segmentation results on MRBrainS18 dataset. The first row shows the results on 80% annotated samples and the second row shows the results on 20% annotated samples. From the left column to the right column show the initial images(the ground truth), the segmentation results of UNet++, Attention u-net, double UNet, ACM_CNN and our method, respectively.

the experimental setup described in Section V.B. We employ UNet++ as the backbone for our segmentation model. U-Net++ framework has garnered widespread recognition in the field. It effectively tackles the challenge of uncertain network depths by integrating U-Nets of different depths. By sharing an encoder and leveraging deep supervision, these integrated U-Nets learn collaboratively. Notably, UNet++ enhances the decoder sub-network to incorporate features with varying semantic scales, enabling a flexible and adaptable feature fusion scheme.

In order to highlight the advantages of dual branch networks, we incorporated our boundary feature regression branch into the UNet++ architecture and compared it to the original network. The Table 2 demonstrates that the inclusion of the boundary feature regression branch is effective, regardless of whether it is trained with 80% or 20% of the training set. This observation holds true for the MRBrainS18 dataset as well.

In this paper, we utilize the SNMM based Level set loss to capture the individual characteristics of the samples and use focal loss to capture the common characteristics of the sample set. This experiment is primarily conducted to test the role of the individual characteristics of the samples in network segmentation. The JS values are shown in Table

Table 1: The JS values of CSF, GM, WM by using different models for IBSR and MRBrainS18 datasets

dataset	Sample ratio	Backbones	JS_CSF(%)	JS_GM(%)	JS_WM(%)
IBSR	80%	UNet++	81.7±0.39	93.06±0.04	90.36±0.45
		Attention u-net	82.34±0.35	93.29±0.03	90.45±0.50
		double UNet	79.01±0.44	92.29±0.04	89.32±0.52
		ACM_CNN	82.27±0.38	93.39±0.04	90.65±0.46
		Ours	82.80±0.36	93.50±0.02	90.71±0.41
	20%	UNet++	65.54±0.92	85.22±0.26	79.46±1.82
		Attention u-net	65.05±1.02	85.2±0.25	79.88±1.18
		double UNet	64.85±1.03	85.05±0.17	77.87±1.47
		ACM_CNN	65.3±1.33	85.33±0.25	79.95±1.65
		Ours	66.92±1.22	85.60±0.21	80.08±1.19
MRBrainS18	80%	UNet++	82.09±0.41	73.88±0.20	74.31±1.29
		Attention u-net	82.23±0.26	74.43±0.18	75.85±0.87
		double UNet	82.12±0.24	73.85±0.18	75.37±0.96
		ACM_CNN	82.57±0.23	75.04±0.23	75.77±1.23
		Ours	83.73±0.21	76.19±0.19	76.83±0.93
	20%	UNet++	80.33±0.32	74.79±0.31	72.37±2.38
		Attention u-net	81.42±0.59	73.24±0.28	72.68±3.29
		double UNet	81.87±0.22	72.81±0.17	74.67±0.95
		ACM_CNN	81.34±0.34	75.62±0.21	75.64±1.08
		Ours	83.05±0.31	75.71±0.12	76.24±0.96

3 highlight the differences between these two loss functions. Observing the data in the table, driven by the loss function we proposed, the segmentation accuracy is enhanced by approximately 0.2% for large samples and improve the segmentation accuracy by around 1% for small samples. These findings suggest that incorporating the individual characteristics of the samples can effectively mitigate the risk of network overfitting.

Finally, in order to demonstrate that our proposed loss function and regression term have a positive effect on improving the segmentation accuracy, we conducted corresponding ablation experiments. The JS values are shown in Table 4 highlight the differences between these two methods. Combining two tables, our proposed loss function

Table 2: The JS values of CSF, GM, WM by using different models for IBSR and MRBrainS18 datasets

dataset	Sample ratio	Backbones	JS_CSF(%)	JS_GM(%)	JS_WM(%)
IBSR	80%	UNet++	81.7±0.39	93.06±0.04	90.36±0.45
		UNet++ with regression	81.94±0.36	93.27±0.04	90.65±0.43
	20%	UNet++	65.54±0.92	85.22±0.26	79.46±1.82
		UNet++ with regression	66.15±1.19	85.42±0.27	79.85±1.73
MRBrainS18	80%	UNet++	82.09±0.41	73.88±0.20	74.31±1.29
		UNet++ with regression	82.38±0.28	75.7±0.18	76.36±1.11
	20%	UNet++	80.33±0.32	74.79±0.31	72.37±2.38
		UNet++ with regression	81.72±0.30	74.29±0.25	76.23±1.23

Table 3: The JS values of CSF, GM and WM by using different loss functions for IBSR and MRBrainS18 datasets

dataset	Sample ratio	Backbones	JS_CSF(%)	JS_GM(%)	JS_WM(%)
IBSR	80%	UNet++ with CE	81.7±0.39	93.06±0.04	90.36±0.45
		UNet++ with CE+SNMLS	82.39±0.35	93.36±0.02	90.53±0.42
	20%	UNet++ with CE	65.54±0.92	85.22±0.26	79.46±1.82
		UNet++ with CE+SNMLS	66.18±0.90	85.95±0.23	79.51±1.35
MRBrainS18	80%	UNet++ with CE	82.09±0.41	73.88±0.20	74.31±1.29
		UNet++ with CE+SNMLS	83.38±0.28	75.33±0.19	76.44±1.01
	20%	UNet++ with CE	80.33±0.32	74.79±0.31	72.37±2.38
		UNet++ with CE+SNMLS	82.85±0.29	75.58±0.18	76.12±0.97

and regression term can better enhance the segmentation accuracy, especially under small samples.

5 Discussion and Conclusion

The network introduced in this article includes two hyperparameters: β and γ, whose configurations have an impact on the ultimate segmentation outcome. The parameter β in Eq.(3.13) is primarily utilized to balance the common characteristics of the sample set with the individual characteristics of the samples. A higher value of β directs the network's attention towards the individual characteristics of each sample, whereas a lower value of β emphasizes the collective common features of the sample set. Figure 6 shows the effect of β.

The parameter γ in Eq.(3.14) is primarily utilized to balance the segmentation model

Table 4: Comparison of ablation experimental results based on the addition of our loss function and regression term

dataset	Sample ratio	Backbones	JS_CSF(%)	JS_GM(%)	JS_WM(%)
IBSR	80%	UNet++ with CE	81.7±0.39	93.06±0.04	90.36±0.45
		UNet++ with CE+SNMLS and regression	82.77±0.35	93.44±0.03	90.63±0.42
	20%	UNet++ with CE	65.54±0.92	85.22±0.26	79.46±1.82
		UNet++ with CE+SNMLS and regression	66.88±1.24	85.74±0.20	80.03±1.21
MRBrainS18	80%	UNet++ with CE	82.09±0.41	73.88±0.20	74.31±1.29
		UNet++ with CE+SNMLS and regression	83.66±0.25	76.09±0.16	76.75±0.96
	20%	UNet++ with CE	80.33±0.32	74.79±0.31	72.37±2.38
		UNet++ with CE+SNMLS and regression	82.99±0.29	75.56±0.14	76.28±0.95

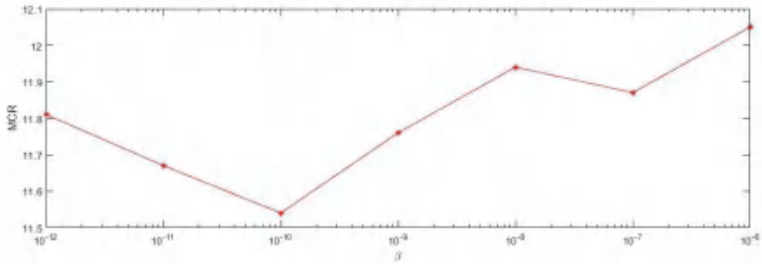

Figure 6: The MCR values of SNSS-UNet based on different β values

and the regression model. The larger γ, the more the network focuses on the boundary feature information. However, this can also lead to suboptimal segmentation results. Figure 7 shows the effect of γ.

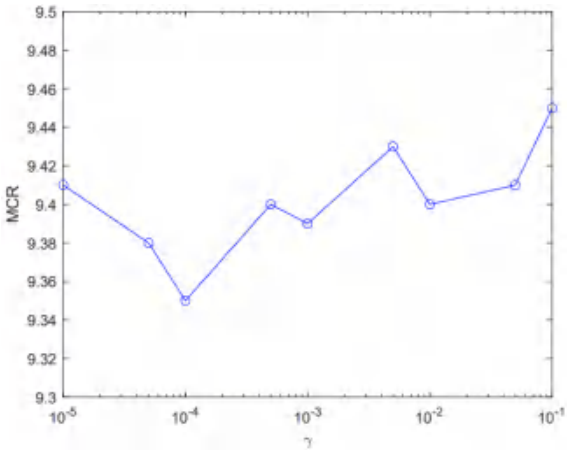

Figure 7: The MCR values of SNSS-UNet based on different γ values

Acknowledgments

This work was supported by the Open Project of Center for Applied Mathematics of Jiangsu Province (Nanjing University of Information Science and Technology).

Conflicts of Interest

The author(s) declare no competing interests. S. Liao conducted numerical experiments and paper writing. Y. Chen provided guidance on technical matters and improved the presentation of this paper. The author(s) reviewed the manuscript. The experiments in this study have been performed in two publicly available datasets: the MRBrainS18 dataset (https://mrbrains18.isi.uu.nl) and the Internet Brain Segmentation Repository (IBSR) dataset (http://www.cma.mgh.harvard.edu/ibsr).

References

[1] T. A. Soomro, L. Zheng, A. J. Afifi, A. Ali, S. Soomro, M. Yin, and J. Gao, Image segmentation for MR brain tumor detection using machine learning: A Review, *IEEE Rev. Biomed. Eng.*, 2023, 16: 70-90.

[2] D. Selvathi, A. Arulmurgan, S. Thamarai Seivi, and S. Alagappan, MRI image segmentation using unsupervised clustering techniques, in *Proc. Sixth Int. Conf. Comput. Intell. Multimed. Appl. (ICCIMA'05)*, 2005, 105-110.

[3] B. K. Tripathy, S. Parikh, P. Ajay, and C. Magapu, Brain MRI segmentation techniques based on CNN and its variants, in *Brain Tumor MRI Image Segmentation Using Deep Learning Techniques*, Elsevier, 2022, 161–183.

[4] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, UNETR: Transformers for 3D medical image segmentation, in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2022, 574-584.

[5] N. M. Portela, G. D. C. Cavalcanti, and T. I. Ren, Semi-supervised clustering for MR brain image segmentation, *Expert Syst. Appl.*, 2014, 41(4): 1492-1497.

[6] B. Xiao, X. Cheng, Q. Li, Q. Wang, L. Zhang, D. Wei, Y. Zhan, X. S. Zhou, Z. Xue, G. Lu, and others, Weakly supervised confidence learning for brain MR image dense parcellation, in *Proc. 10th Int. Workshop Mach. Learn. Med. Imaging (MLMI 2019)*, 2019, 409-416.

[7] A. K. Jain, R. P. W. Duin, and J. Mao, Statistical pattern recognition: A review, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, 22(1): 4-37.

[8] A. Sofou and P. Maragos, PDE-based modeling of image segmentation using volumic flooding, in *Proc. 2003 Int. Conf. Image Process.*, 2003, vol. 2, II-431.

[9] A. Sofou and P. Maragos, Generalized flooding and multicue PDE-based image segmentation, *IEEE Trans. Image Process.*, 2008, 17(3): 364-376.

[10] A. A. Abdullah, A. M. Ahmed, T. Rashid, H. Veisi, Y. H. Rassul, B. Hassan, P. Fattah, S. A. Ali, and A. S. Shamsaldin, Advanced clustering techniques for speech signal enhancement: A review and metanalysis of fuzzy c-means, k-means, and kernel fuzzy c-means methods, *arXiv preprint arXiv:2409.19448*, 2024.

[11] M. A. T. Figueiredo and A. K. Jain, Unsupervised learning of finite mixture models, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, 24(3): 381-396.
[12] A. Mahdavi, N. Balakrishnan, and A. Jamalizadeh, EM algorithm for bounded generalized t mixture model with an application to image segmentation, *Comput. Appl. Math.*, 2025, 44(1): 89.
[13] B. Dong, G. Weng, Q. Bu, Z. Zhu, and J. Ni, An active contour model based on shadow image and reflection edge for image segmentation, *Expert Syst. Appl.*, 2024, 238: 122330.
[14] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, Deep learning for medical image segmentation: State-of-the-art advancements and challenges, *Inform. Med. Unlocked*, 2024, 101504.
[15] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, 3431-3440.
[16] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI 2015)*, 2015, 234-241.
[17] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, and others, A large annotated medical image dataset for the development and evaluation of segmentation algorithms, *arXiv preprint arXiv:1902.09063*, 2019.
[18] K. Han, V. S. Sheng, Y. Song, Y. Liu, C. Qiu, S. Ma, and Z. Liu, Deep semi-supervised learning for medical image segmentation: A review, *Expert Syst. Appl.*, 2024, 245: 123052.
[19] X. Guo, C. Yang, Y. Liu, and Y. Yuan, Learn to threshold: Thresholdnet with confidence-guided manifold mixup for polyp segmentation, *IEEE Trans. Med. Imaging*, 2020, 40(4): 1134-1146.
[20] J. Shi, T. Gong, C. Wang, and C. Li, Semi-supervised pixel contrastive learning framework for tissue segmentation in histopathological image, *IEEE J. Biomed. Health Inform.*, 2022, 27(1): 97-108.
[21] M. A. G. Ballester, A. P. Zisserman, and M. Brady, Estimation of the partial volume effect in MRI, *Med. Image Anal.*, 2002, 6(4): 389-405.
[22] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI, *IEEE Trans. Image Process.*, 2011, 20(7): 2007-2016.
[23] T. F. Chan and L. A. Vese, Active contours without edges, *IEEE Trans. Image Process.*, 2001, 10(2): 266-277.
[24] M. A. Balafar, Gaussian mixture model based segmentation methods for brain MRI images, *Artif. Intell. Rev.*, 2014, 41: 429-439.
[25] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.*, 2020, 162: 94-114.
[26] Z. Zhou, M. M. Siddiquee, N. Tajbakhsh, and J. Liang, Unet++: Redesigning skip connections to exploit multiscale features in image segmentation, *IEEE Trans. Med. Imaging*, 2019, 39(6): 1856-1867.
[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, Densely connected convolutional networks, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, 4700-4708.
[28] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, and others, Attention u-net: Learning where to look for the pancreas, *arXiv preprint arXiv:1804.03999*, 2018.

[29] A. Tarvainen and H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, *Adv. Neural Inf. Process. Syst.*, 2017, 1195-1204.
[30] J. Enguehard, P. O'Halloran, and A. Gholipour, Semi-supervised learning with deep embedded clustering for image classification and segmentation, *IEEE Access*, 2019, 7: 11093-11104.
[31] S. Laine and T. Aila, Temporal ensembling for semi-supervised learning, *arXiv preprint arXiv:1610.02242*, 2016.
[32] X. Luo, J. Chen, T. Song, and G. Wang, Semi-supervised medical image segmentation through dual-task consistency, in *Proc. AAAI Conf. Artif. Intell.*, 2021, 35(10): 8801-8809.
[33] A. Kristiadi and P. Pranowo, Deep Convolutional Level Set Method for Image Segmentation, *J. ICT Res. & Appl.*, 2017, 11(3): 284-298.
[34] H. Chen, Y. Jin, G. Jin, C. Zhu, and E. Chen, Semisupervised semantic segmentation by improving prediction confidence, *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, 33(9): 4991-5003.
[35] H. Hu, F. Wei, H. Hu, Q. Ye, J. Cui, and L. Wang, Semi-supervised semantic segmentation via adaptive equalization learning, *Adv. Neural Inf. Process. Syst.*, 2021, 34: 22106-22118.
[36] X. Chen, Y. Yuan, G. Zeng, and J. Wang, Semi-supervised semantic segmentation with cross pseudo supervision, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, 2613-2622.
[37] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao, and X. Le, Semi-supervised semantic segmentation using unreliable pseudo-labels, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, 4248-4257.
[38] Y. Ali, M. A. Iftikhar, Q. Abbas, and T. Wahab, Automated White Matter Segmentation in MR Images Using Residual UNet, in *Proc. 4th Int. Conf. Advancements Comput. Sci. (ICACS)*, 2023, 1-6.
[39] B. Kim and J. C. Ye, Mumford–Shah loss functional for image segmentation with deep learning, *IEEE Trans. Image Process.*, 2019, 29: 1856-1866.
[40] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, Using of Jaccard coefficient for keywords similarity, in *Proc. Int. Multiconf. Eng. Comput. Sci.*, 2013, 1(6): 380–384.
[41] T. M. Nguyen and Q. M. Jonathan Wu, Fast and robust spatially constrained Gaussian mixture model for image segmentation, *IEEE Trans. Circuits Syst. Video Technol.*, 2012, 23(4): 621–635.