



Co-optimization Learning Network for MRI Segmentation of Ischemic Penumbra Tissues

Liangliang Liu¹, Jing Zhang², Jin-xiang Wang³, Shufeng Xiong¹ and Hui Zhang^{1*}

¹ College of Information and Management Science, Henan Agricultural University, Zhengzhou, China, ² Department of Computer Science, Henan Quality Engineering Vocational College, Pingdingshan, China, ³ Department of Computer Science, University of Melbourne, Parkville, VIC, Australia

OPEN ACCESS

Edited by:

Shuai Wang,
Shandong University, China

Reviewed by:

Wenxing Hu,
University of Pittsburgh, United States

Fa Zhang,

Institute of Computing Technology,
Chinese Academy of Sciences (CAS),
China

Jin Liu,

Central South University, China

*Correspondence:

Hui Zhang
zhhnau@163.com

Received: 24 September 2021

Accepted: 25 November 2021

Published: 16 December 2021

Citation:

Liu L, Zhang J, Wang J-x, Xiong S and
Zhang H (2021) Co-optimization
Learning Network for MRI
Segmentation of Ischemic Penumbra
Tissues.

Front. Neuroinform. 15:782262.
doi: 10.3389/fninf.2021.782262

Convolutional neural networks (CNNs) have brought hope for the medical image auxiliary diagnosis. However, the shortfall of labeled medical image data is the bottleneck that limits the performance improvement of supervised CNN methods. In addition, annotating a large number of labeled medical image data is often expensive and time-consuming. In this study, we propose a co-optimization learning network (COL-Net) for Magnetic Resonance Imaging (MRI) segmentation of ischemic penumbra tissues. COL-Net base on the limited labeled samples and consists of an unsupervised reconstruction network (R), a supervised segmentation network (S), and a transfer block (T). The reconstruction network extracts the robust features from reconstructing pseudo unlabeled samples, which is the auxiliary branch of the segmentation network. The segmentation network is used to segment the target lesions under the limited labeled samples and the auxiliary of the reconstruction network. The transfer block is used to co-optimization the feature maps between the bottlenecks of the reconstruction network and segmentation network. We propose a mix loss function to optimize COL-Net. COL-Net is verified on the public ischemic penumbra segmentation challenge (SPES) with two dozen labeled samples. Results demonstrate that COL-Net has high predictive accuracy and generalization with the Dice coefficient of 0.79. The extended experiment also shows COL-Net outperforms most supervised segmentation methods. COL-Net is a meaningful attempt to alleviate the limited labeled sample problem in medical image segmentation.

Keywords: co-optimization learning network, ischemic penumbra tissues, segmentation network, reconstruction network, transfer block

1. INTRODUCTION

1.1. Clinical Motivation and Challenges

Labeling target tissue from medical images is of great significance for disease diagnosis and treatment. Recently, the application of computer technology in brain image tissue segmentation has become a hot field of medical imaging analysis (Lutnick et al., 2019; Huseyn, 2020; Zhao et al., 2020). Brain penumbra is a common affliction of ischemic stroke diseases in men. Ischemic penumbra segmentation on magnetic resonance image (MRI) is important for stroke diagnosis and pre-operative planning (Dora et al., 2017; Maier et al., 2017; Liu et al., 2020b). In medical physiology, the penumbra tissue is introduced to designate regions of brain tissue with “almost ischemia” (Lassen et al., 1991). The infarct region of stroke has been necrotic, while the penumbra

tissues can be saved. Accurate annotation of penumbra tissues will help clinicians to locate the penumbra tissues and make an effective individualized therapy. As shown in **Figure 1A**, the dark area is the infarct tissue of stroke, it is surrounded by penumbra and normal tissues. The treatment of stroke disease is mainly based on vascular recanalization in an effective time window. In other words, it is to save the penumbra tissues and restore the penumbra tissues to normal tissue through intravenous thrombolysis and intravascular treatment. As shown in **Figure 1B**, the penumbra tissues are usually distributed around the infarct tissues. Accurate annotation of penumbra tissues plays a key role in taking active measures to save penumbra tissues or transform them into normal tissues. Unfortunately, consistent and correct labeling of penumbra tissues based on MRI is challenging for experienced doctors, due to 1) unclear tissue boundaries, 2) differences in subjective experience 3) boring and time-consuming work.

Computer technology (image processing, machine learning, deep learning, etc.) has been applied to brain image analysis. Especially the supervised segmentation convolutional neural networks (CNNs) can provide the possibility for medical image lesion segmentation. Some supervised deep learning methods have shown advantages in many segmentation tasks. However, a fine segmentation model still faces some challenges: 1) The lack of the massive labeled sample is difficult to meet the demand of a supervised deep learning method to train a fine model. Supervised segmentation models are hard to achieve the feature representation on the limited labeled data set. 2) Rough and non-uniform manual labels impair the ability of the prediction model to mine the key hidden features of the target tissues. 3) The accuracy of semi-supervised and unsupervised segmentation methods is not satisfactory. In the case of limited labeled samples, how to develop a segmentation method to improve the efficiency of delineation of unlabeled samples is our concern.

1.2. Related Work

Deep learning is the most representative technology in computer vision (Hu et al., 2018, 2019; Zhang et al., 2021; Zhou et al., 2021). Many deep learning-based studies have been applied in medical image segmentation tasks in past decades (Dora et al., 2017; Liu et al., 2020b; Hu et al., 2021). These studies are mainly divided into two types: the methods based on supervised learning strategy and the methods based on semi-supervised learning strategy.

The supervised learning strategy trains models automatically to improve the learning performance by using labeled samples (Knudsen, 1994). The supervised CNNs are representative of this kind of strategy. In the training process, CNNs learn the highly representative hidden features under the guidance of the labeled samples. To a certain extent, these methods promote the development of medical image segmentation (Moeskops et al., 2018; Liu et al., 2019, 2020a; Nakarmi et al., 2020). For example, the Ischemic Stroke Lesion Segmentation (ISLES) 2015 challenge showed that the supervised segmentation methods could obtain good segmentation results on stroke MRI images (Maier et al., 2017). Kamnitsas et al. (2016) proposed a dual pathway 3D supervised CNN for brain tumor segmentation in BraTS 2015 challenge. In 2017, Esteva et al.

developed a supervised deep neural network for analyzing skin cancer, which greatly inspired the research in skin cancer analysis (Andre et al., 2019). In addition, supervised learning strategies are used to improve the performance of image analysis models, including: spine cancer (Esteva et al., 2017), optic disc (Fu et al., 2018), liver and prostate (Litjens et al., 2014; Li et al., 2018), etc. The supervised learning strategy improves the accuracy of the model, however, the limited sample in the medical image analysis task is an unavoidable challenge to the supervised segmentation models. These models can not utilize the extra information from unlabeled samples.

The semi-supervised learning strategy trains models automatically to improve the learning performance by using labeled samples and unlabeled samples (Van Engelen and Hoos, 2020). Semi-supervised learning aims to improve the performance of the prediction model by learning from a limited number of labeled data and an arbitrary amount of unlabeled data. This strategy is more suitable for medical image analysis with limited samples and assists clinicians to delineate unlabeled samples (Weston et al., 2012; He et al., 2020; Li et al., 2020a). For example, Li et al. proposed a semi-supervised model for medical image segmentation (Li et al., 2020b), they optimized the model by using two loss functions: public supervision loss from labeled samples and regularization loss from unlabeled samples. They introduced a transformation-consistent strategy to enhance the regularization effect for pixel-level predictions in the semi-supervised segmentation tasks. The proposed model showed superior performance on the 3 challenging 2D/3D medical image data sets. Liu et al. provided a relation-driven semi-supervised framework for medical image classification (Liu et al., 2020e). The framework is a consistency-based method. Firstly, this method used unlabeled samples by encouraging the prediction consistency of a given input under disturbance and then used self-scrambling model to generate high-quality consistency targets for unlabeled samples. This method outperformed many state-of-the-arts semi-supervised learning methods on ISIC 2018 challenge and thorax disease classification with Chest X-ray images. Kumar et al. integrated the idea of semi-supervised learning into the model for retinopathy and cancer anomaly detection (Kumar and Awate, 2020). This model achieved high-quality outlier labeling by a small amount of expert calibration data. Kumar et al. adopted the maximum semi-supervised robust hybrid model strategy to improve the detection performance of the model. In addition, as a special case of semi-supervised learning, few-shooting learning is proposed to solve the problem of the limited labeled samples (Fei-Fei et al., 2006; Vinyals et al., 2016). Shaban et al. (2017) proposed a conditioning branch and segmentation branch network to solve the semantic segmentation problem on limited samples. The whole network could predict the type of a test image at the pixel level. Rakelly et al. proposed a train of guided and conditional networks with few labeled samples and sparse annotations (Rakelly et al., 2018). This approach was evaluated on the PASCAL VOC computer vision benchmark. These methods usually used unlabeled samples to assist models to make a prediction. However, these methods are easily affected by pixel values with large differences in the medical image. Especially

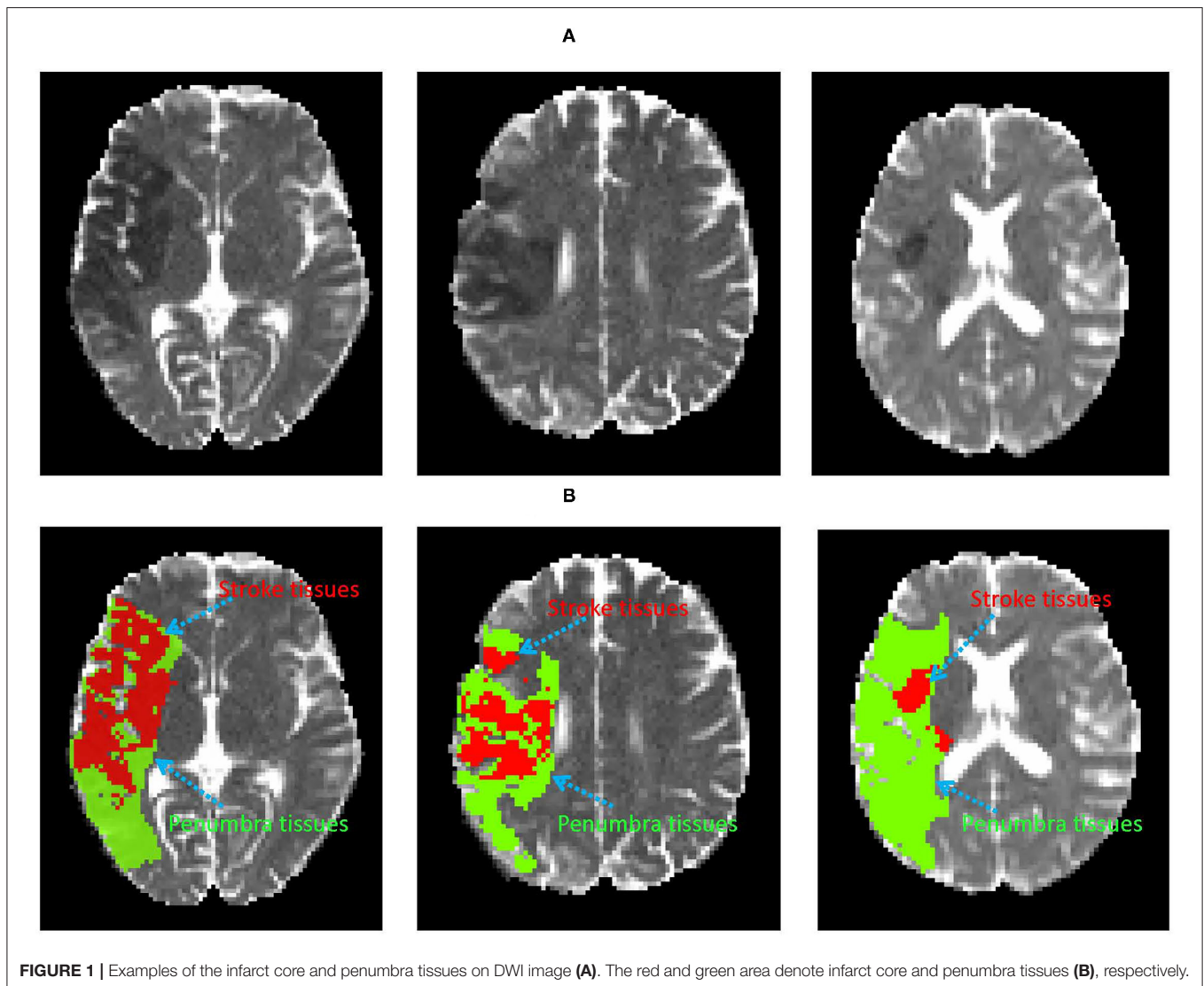


FIGURE 1 | Examples of the infarct core and penumbra tissues on DWI image (A). The red and green area denote infarct core and penumbra tissues (B), respectively.

in low contrast brain images, semi-supervised based methods usually mistakenly regard the infarct area as the optimization target in stroke imaging segmentation task, which will result in the interference of the training process and the degradation of model performance.

1.3. Contributions

In this study, we present a semi-supervised learning network (COL-Net) based on co-optimization learning for brain image segmentation. COL-Net can be trained by jointly using the limited labeled and unlabeled MRI images. The whole network is co-optimized by combining the supervised and unsupervised sub-networks (loses). To leverage the unlabeled samples, we assume each labeled sample also has a pseudo unlabeled state. We take the advantage of co-optimization learning works into delineating penumbra tissues. Based on the limited number of labeled and pseudo unlabeled samples, we leverage

pseudo unlabeled samples as additional information for the segmentation model and optimize the model. Our contributions include:

(1) We propose a co-optimization learning network for ischemic penumbra tissues segmentation which is an attempt to assist clinicians to delineate penumbra tissues under the limited labeled samples.

(2) COL-Net consists of 3 parts: segmentation network (S), reconstruction network (R), and transfer block (T). R is used as the auxiliary branch of S with the pseudo unlabeled samples, S is used to predict the results with the limited labeled samples and the supervise of R, and T is used to co-optimize the feature maps between the bottlenecks of R and S.

(3) We use a mix loss function (segmentation loss and reconstruction loss) to co-optimize COL-Net.

(4) Extensive experiments and ablation study demonstrates that COL-Net substantially outperforms well-tuned state-of-the-art methods.

Algorithm 1 COL-Net Algorithm pseudocode.

Input: $x_i \in L, y_i \in L, x'_i \in UL$
Output: \tilde{X}_k

- 1: $R(x_i | x'_i) \leftarrow$ unsupervised reconstruction network
- 2: $S(x_i) \leftarrow$ supervised segmentation network
- 3: $feature_R \leftarrow$ middle layer in $R(x_i, x'_i)$
- 4: $feature_S \leftarrow$ bottleneck layer in $S(x_i)$
- 5: $T(R, S) \leftarrow add(feature_R, feature_S)$
- 6: **for** $i = 1$ to numepochs **do**
- 7: randomly update $R(x_i | x'_i), S(x_i)$
- 8: $loss_r(x_i | x'_i) \leftarrow$ update loss $(R(x_i | x'_i))$
- 9: $loss_s(x_i) \leftarrow$ update loss $(S(x_i))$
- 10: $loss_{mix} \leftarrow min(loss_r + \lambda loss_s)$
- 11: **end for**
- 12: **return** $loss_{mix}$

The rest of this paper is organized as follows. The methodology of the proposed network is introduced in section 2. Section 3 introduces the materials and provides the evaluation metrics of our segmentation network. We introduce the experimental setup and analyze the experimental results in detail. Section 5 conducts and discusses the contributions of individual components in COL-Net. Finally, we make the conclusions of this study in section 6.

2. METHODOLOGY

2.1. The Architecture of COL-Net

The architecture of COL-Net is shown in **Figure 2** and Algorithm 1 summarizes the pseudocode. COL-Net mainly consists of 3 parts: an unsupervised reconstruction network (R), a supervised segmentation network (S), and a transfer block (T).

2.2. Network Architecture

2.2.1. The Encoder-Decoder Framework

The encoder-decoder architecture is a common framework in deep learning (Long et al., 2015; Ronneberger et al., 2015; Liu et al., 2020b,d). It can be used to analyze any text, voice, image, and video data. The encoder-decoder architecture is an end-to-end learning algorithm and usually is used as the backbone of CNN, RNN, BiRNN, LSTM, GRU, and so on. The encoder converts the input sequence into a fixed-length vector. The decoder converts the previously generated fixed vector into the output sequence. In medical image analysis tasks, residual or dense mechanisms, and skip connection operation are usually embedded into the encoder-decoder framework to improve the model performance (Ronneberger et al., 2015; He et al., 2016; Zhang et al., 2018).

In our study, the unsupervised network (R) and supervised network (S) are the main parts of COL-Net, both of them adopt the encoder-decoder architecture as the backbone. COL-Net is trained in a co-optimization manner: 1) We train R by using the pseudo unlabeled samples to obtain representative features. 2) We joint the feature maps of R and S as the input of

up-sampling by transfer block and further use the labeled samples to fine-tune the S.

2.2.2. Reconstruction Network (R)

The reconstruction network (R) is trained in the unsupervised strategy. R is illustrated in **Figure 2**, the encoder part of R comprises followed that in ResNet network (He et al., 2016). The decoder part of R adopts 4 continuous Conv-BN-Relu (Conv: convolution, BN: batch normalization, and ReLU: rectified linear unit) blocks to reconstruct the feature map. Finally, a convolution layer and *Tanh* function are used to reconstruct the image. We also use a discriminator network (D) to distinguish the input image from the generated image by the reconstruction network. The architecture of D is illustrated in **Figure 3**, which contains 4 convolutional layers with a 3×3 kernel, a fully connected (FC) layer with 512 neurons, and an FC layer followed by the *softmax* activation function.

We aim at improving the segmentation performance for the COL-Net, reconstruction network plays as an auxiliary and correction network of the segmentation network in COL-Net. Reconstruction network is trained under the pseudo unlabeled samples in an unsupervised way. Given the i -th input sample $X_i, i \in [1, I]$, the loss function of R is defined as follow:

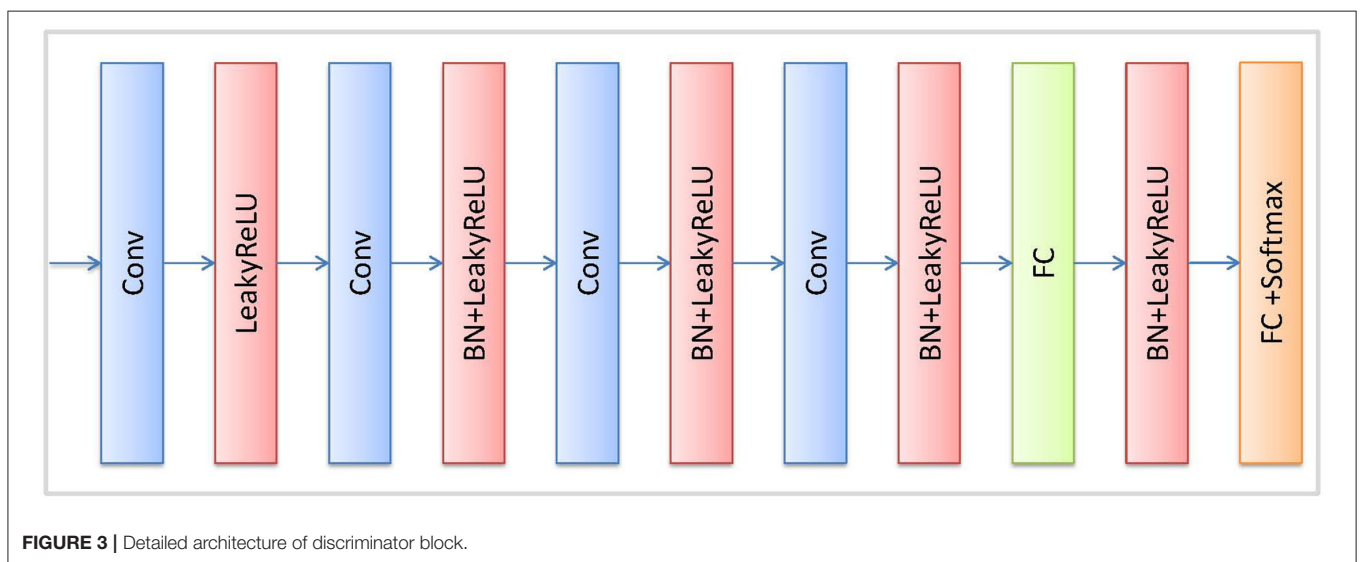
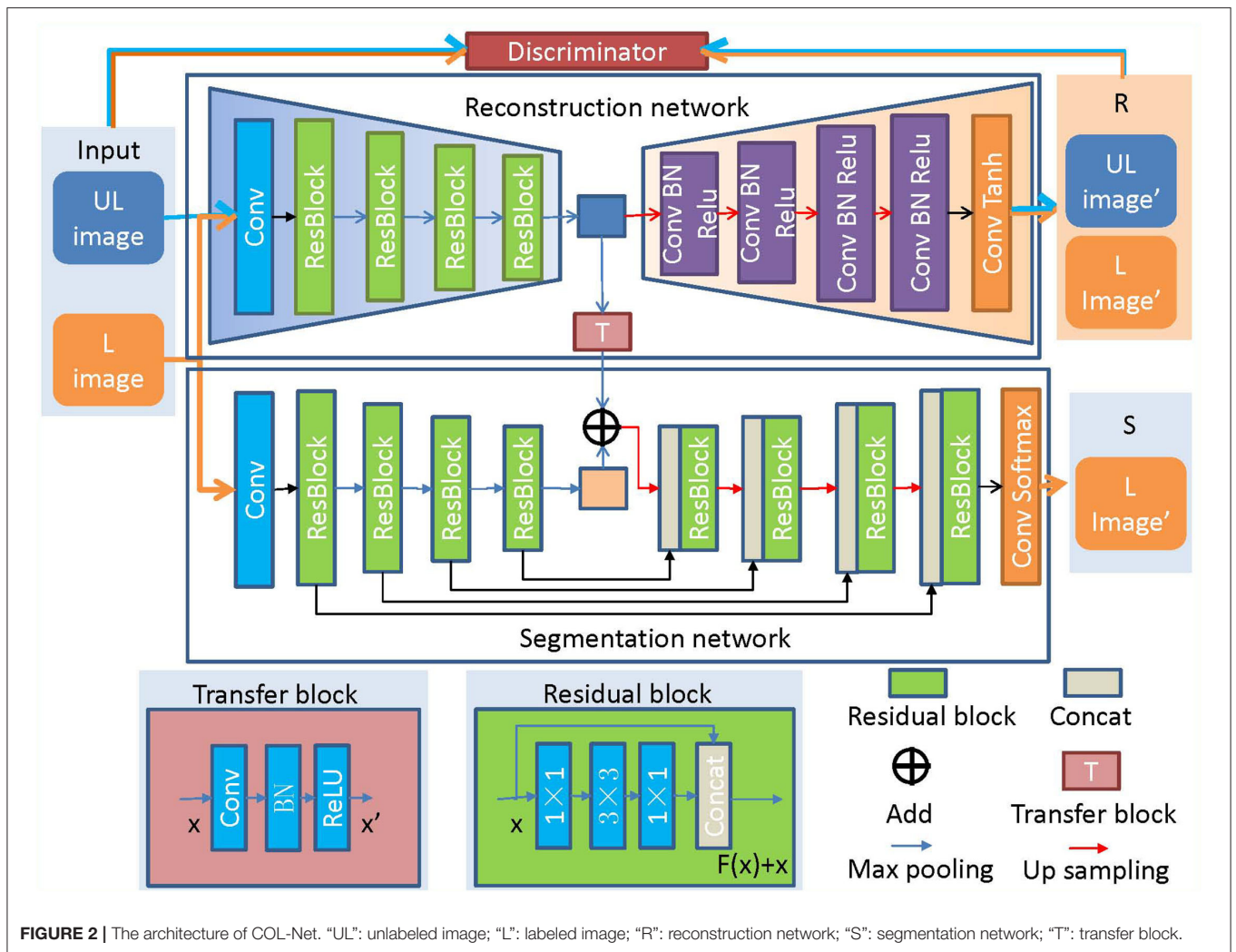
$$loss_r(X_i) = loss_{mse}(R(X_i), X_i) - [loss_{ce}(D(X_i), 0) + loss_{ce}(D(X_i), 1)], \quad (1)$$

where $loss_{mse}()$ is the mean square error (MSE) loss (Masci et al., 2011), which measures the discrepancy between the original image and the reconstructed image. The $loss_{ce}()$ is the cross-entropy adversarial loss for D, which measures the discrepancy between the distributions of an input image and the reconstructed one. We use the Adam optimizer to fine optimize the $loss_r()$ loss function.

2.2.3. Segmentation Network (S)

As illustrated in **Figure 2**, the segmentation network is inspired by the encoder-decoder architecture and is trained under the supervised strategy. The encoder-decoder architecture has been proved to perform well in medical image segmentation tasks (Long et al., 2015; Ronneberger et al., 2015). We construct the backbone of the segmentation network (S) based on ResNet (He et al., 2016). The encoder in S is the same as that in R. In addition, the skip connections make the feature mapping and fuse the feature mappings between the encoder layer and decoder layer, and the continuous up-samplings are used to recognize feature and gradually recover the location of each pixel. Finally, we use the *softmax* activation function to predict the segmentation result. The detail of the residual block is shown in the subgraph of **Figure 2**, it consists of 3 convolution layers ($1 \times 1, 3 \times 3, 1 \times 1$) and a fusion layer (*Concat()*).

We use the labeled samples to train the segmentation network. The loss function ($loss_s$) of the segmentation network bases on the Dice coefficient (DC) (Milletari et al., 2016). In the medical image analysis task, DC is usually based on the pixel level, it is calculated the similar proportion between the prediction result and the label image. The larger the DC value, the better the segmentation



result. Given the j -th labeled sample X_j and corresponding label Y_j , where $j \in [1, J]$, the $loss_s$ comes from DC, which is defined as follows:

$$loss_s = 1 - DC(X_j, Y_j) = 1 - \frac{2|Y_j \cap X_j|}{|Y_j| + |X_j|}, \quad (2)$$

where DC is a measure of similarity between image X_j and label Y_j .

In our study, we hope the reconstruction network can supervise the segmentation network to achieve the aim of using the features of the unlabeled sample to assist segmentation. In our model, we use a mixed loss function to train and co-optimize COL-Net. The mixed loss function is based on the segmentation network and the reconstruction network. The proposed COL-Net is co-optimized during the training process, the optimizing strategy compose of two parts: $loss_r$ and $loss_s$. The $loss_r$ is trained on the pseudo unlabeled samples in an unsupervised manner. The $loss_s$ is trained on the labeled samples in a supervised manner. We use a mix loss function to train COL-Net by reducing to the minimum of the following function:

$$\begin{aligned} loss_{mix} &= \min(l_s + \lambda l_r) \\ &= \min\left(1 - \frac{2|Y_j \cap X_j|}{|Y_j| + |X_j|}\right) \\ &\quad - \lambda l_{mse}(R(X_i), X_i) - [l_{ce}(D(X_i), 0) + l_{ce}(D(X_i), 1)] \end{aligned} \quad (3)$$

where λ is employed to weight the importance of the reconstruction network. It should be noted that when l_s is set to 0, the parameter λ should be set to 0 to ensure the training process of the model.

2.2.4. Transfer Block (T)

The transfer block is used as a bridge between R and S, which transfers the features maps by the convolution, BN, and ReLU operations. We transfer the output of the middle layer in R and fuse it with the output of the bottleneck layer in S. The transfer block is illustrated in **Figure 4**, the T block is a 1×1 convolutional layer with a stride of 1 followed by BN and ReLU activation.

We use $add()$ method to fuse two feature maps that come from two bottleneck layers (the reconstruction network and the segmentation network). $add()$ superimposes pixel values. It is a pixel information overlay method, which increases the amount of information describing the features of an image, the channel information in the image does not increase. Each feature map is fused according to channels. Let $*$ be the convolution operation and f^i be the i -th feature map. f^i consists of channels $A_1^i, A_2^i, \dots, A_c^i$. There has n feature maps. The equation of the output feature map with $add()$ can be expressed as:

$$\begin{aligned} O_{add} &= \sum_{j=1}^c (A_j^1 * K_j^1 + \dots + A_j^n * K_j^n) \\ &= \sum_{j=1}^c A_j^1 * K_j^1 + \dots + \sum_{j=1}^c A_j^n * K_j^n, \end{aligned} \quad (4)$$

where K is the convolution kernel, c is the number of channels.

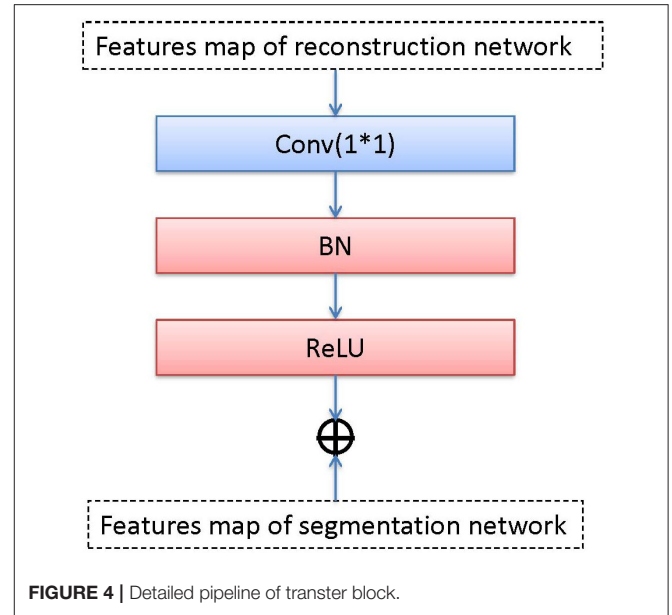


FIGURE 4 | Detailed pipeline of transfer block.

3. EVALUATION DATASET

3.1. SPES Image Data

The Ischemic Stroke Lesion Segmentation (ISLES) launched a challenge to estimate the regions of the penumbra in acute stroke in 2015 (SPES). SPES challenge provides 50 acute ischemic stroke participants (30 for training and 20 for testing). All participants had been diagnosed and collected by the University Hospital of Bern between 2005 and 2013. All participants are older than 18. There are 7 MRI modalities for each participant: T1 contrast-enhanced (T1c), T2, DWI, cerebral blood flow (CBF), cerebral blood volume (CBV), time-to-peak (TTP), and time-to-max (Tmax). All MRIs were performed on either a 1.5T Siemens Magnetom Avanto or a 3T MRI system Siemens Magnetom Trio. All MRI modalities were skull-stripped, re-sampled to an isotropic spacing of $2^3 mm$, and co-registered to the T1w contrast modality.

The ground truth of training samples are manually labeled in DWI images by a medicine doctor who obeyed the following steps: First, medicine doctor obtained the fusion restriction label region through a semi-manual segmentation method based on the Tmax modality. Then, he obtained the diffusion restriction label region through the currently accepted ADC threshold ($600 \times 10^{-6} mm^2/sec$). Finally, the penumbra region was obtained by the mismatch of these two labels which were defined as follows: $(perfusion - restrictionlabel) - (diffusion - restrictionlabel)$. The ground truth of training samples are provided on the challenge web page¹. The ground truths of testing samples are not available on the challenge web page.

¹www.isles-challenge.org

3.2. Evaluation Metrics

To demonstrate the advantage of COL-Net, we inherit the evaluation metrics of the SPES challenge and compare it with the supervised methods in the challenge and other supervised \ semi-supervised methods. SPES challenge offers 3 evaluation metrics: DC, Hausdorff distance (HD) (Huttenlocher et al., 1993), and Average Symmetric Surface Distance (ASSD). The DC is used to assess the similar proportion of overlap between the predicted image and the ground truth, DC is defined in Equation (3). ($0 \leq DC \leq 1$, the higher the DC value, the better the segmentation result). The HD is used to quantify the maximum distance between two surface pixels between the predicted image and ground truth. It is defined as follows:

$$HD(X, Y) = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(y, x) \right\}, \quad (5)$$

where X and Y are predicting image and ground truth, respectively. x and y are the pixels in X and Y , and $d(\cdot)$ is the Euclidean distance between points x and y . The HD values are expressed in millimeters (mm).

The ASSD is used to evaluate the average distance between predict image and ground truth on pixel level, which is defined as:

$$ASSD(X, Y) = \frac{1}{2} \left(\frac{\sum_{x \in X} \min_{y \in Y} d(x, y)}{|X|} + \frac{\sum_{y \in Y} \min_{x \in X} d(y, x)}{|Y|} \right), \quad (6)$$

where the ASSD values are expressed in millimeters (mm). For HD and ASSD evaluation measures, the smaller the value is, the better is the segmentation results.

4. EXPERIMENTS AND RESULTS

4.1. Experimental Setup

The SPES dataset provides 30 labeled training samples and 20 unlabeled testing samples. The reconstruction network is trained in an unsupervised manner and the segmentation network is trained in a supervised manner. In the training process, according to two different scenarios, we assume that each training sample has two label states: labeled and pseudo unlabeled samples. The labeled samples are used to train the segmentation network, the pseudo unlabeled samples are used to train the reconstruction network, the testing samples are just used to test the model. In the details, we use all pseudo unlabeled samples to train the reconstruction network and use 5 different percentages of labeled samples to train the segmentation network: 10% labeled samples (3 labeled training samples), 20% labeled samples (6 labeled training samples), 50% labeled samples (15 labeled training samples), 80% labeled samples (24 labeled training samples), 100% labeled samples (30 labeled training samples).

Under the unsupervised manner, the reconstruction network is optimized by the Adam optimizer, the mini-batch is 3, the epoch is 300, the learning rate is 10^{-5} , the drop-out rate is 10^{-2} , and the random weight initialization. We use the early stop strategy to train COL-Net and save the best weight of the model.

TABLE 1 | The results of different strategies on the SPES dataset.

Method	DC	ASSD	HD
COL-Net1 (10%) (Semi)	37.00 (± 12.00)	9.86 (± 7.53)	76.57 (± 43.17)
COL-Net2 (20%) (Semi)	53.00 (± 9.00)	7.23 (± 5.40)	62.61 (± 32.40)
COL-Net3 (50%) (Semi)	68.00 (± 12.00)	4.00 (± 4.01)	59.30 (± 25.91)
COL-Net4 (80%) (Semi)	79.00 (± 8.00)	2.41 (± 2.19)	34.20 (± 24.05)
COL-Net5 (100%) (Supe)	79.00 (± 9.00)	1.83 (± 0.52)	39.20 (± 25.38)

DC:%, ASSD:mm, HD:mm. Semi, Semi-supervised training strategy; Supe, Supervised training strategy. The bold values mean the best results.

In the segmentation network, the labeled training samples are randomly split into two parts: training part and validation part. In our study, all experiments are work on the sample level. We use the validation split method to divide the training and validation samples. The value of split parameter is set of 0.2. The training samples are used to train the models, the validation samples are used to fine-tune the trained models. Although segmentation and reconstruction networks have the same architecture, the initialization parameters of segmentation network are different from that of the reconstruction network. The segmentation network uses Adam function as an optimizer, the mini-batch size is 3, the epoch is 80, the learning rate is 10^{-4} , the drop-out rate is 10^{-2} , and the random weight initialization. Both networks work on the same computer server with an NVIDIA GeForce Titan X Pascal CUDA GPU with 16 GB memory. We save the best weight parameters of the segmentation model by maximizing the performance of the training process. And then we use the best model weight parameters to predict the segmentation results on the 20 testing samples.

4.2. Result and Analysis

In order to obtain the best semi-supervised segmentation model, we evaluate 5 training strategies respectively. In these 5 models, the last model (COL-Net5: using 100% labeled samples in the training process) is the only one under the supervised training strategy, others are under the semi-supervised training strategy. For each model, we test the best weight parameters on the 20 testing samples and submit the predicted results to the SPES challenge web page. These 5 predicated scores of testing samples are shown in **Table 1**. From COL-Net1 to COL-Net5, the results of the proposed COL-Net achieve relatively consistent improvements on 3 metrics. When the proportion of labeled samples used for training is less than 50%, the segmentation result is not satisfactory. When the number of labeled samples for training reaches 80% (COL-Net4), the DC value of COL-Net4 can achieve 79% which is the same as the DC value of using 100% labeled samples model (COL-Net5), which demonstrates that the reconstruction network provides the useful information for the segmentation network. It denotes that our proposed the co-optimization learning strategy has the effectiveness on limited labeled samples.

4.3. Ranking in the Challenge

All the top ranking methods are supervised methods in the SPES challenge web page. We use the COL-Net model which

TABLE 2 | The results of the participants on the SPES dataset.

Methods	DC	ASSD	HD
CH-Insel (Supe)	82.00 (± 8.00)	1.65 (± 1.40)	29.02 (± 12.60)
DE-UzL (Supe)	81.00 (± 9.00)	1.36 (± 0.74)	23.62 (± 12.99)
COL-Net (Semi)	79.00 (± 8.00)	2.41 (± 2.19)	34.20 (± 24.05)
BE-Kul2 (Supe)	78.00 (± 9.00)	2.77 (± 3.27)	40.27 (± 25.10)
CN-Neu (Supe)	76.00 (± 9.00)	2.29 (± 1.76)	30.65 (± 16.49)
DE-Ukf (Supe)	73.00 (± 13.00)	2.44 (± 1.93)	33.92 (± 20.88)
BE-Kul1 (Supe)	67.00 (± 24.00)	4.00 (± 3.39)	57.95 (± 28.77)
CA-USher (Supe)	54.00 (± 26.00)	5.53 (± 7.59)	59.89 (± 30.18)

DC:%, ASSD:mm, HD:mm. Semi, Semi-supervised training strategy; Supe, Supervised training strategy. The bold values mean the best results.

TABLE 3 | The results of the state-of-the-arts methods on the 2D SPES images.

Methods	DC	ASSD	HD
U-Net (Liang et al., 2017)	68.00 (± 16.00)	3.49 (± 2.77)	51.86 (± 24.93)
FCN (Long et al., 2015)	70.00 (± 12.00)	3.54 (± 2.04)	60.42 (± 28.56)
Res-FCN (Liu et al., 2018)	71.00 (± 12.00)	3.43 (± 1.94)	61.05 (± 23.75)
Res-UNet (He et al., 2016)	76.00 (± 9.00)	2.10 (± 0.57)	42.54 (± 21.14)
COL-Net	79.00 (± 8.00)	2.41 (± 2.19)	34.20 (± 24.05)
RA-UNet (Jin et al., 2018)	80.00 (± 3.00)	2.03 (± 0.86)	39.43 (± 27.07)
DRANet (Liu et al., 2020c)	80.00 (± 7.00)	1.91 (± 1.02)	30.13 (± 20.32)

DC:%, ASSD:mm, HD:mm. The bold value mean the best results.

is obtained by the third training strategy to verify the testing samples (the only 80% of the labeled training samples are used in the segmentation network). We compare COL-Net with other top 7 participate teams in the SPES challenge. The rankings of these 7 teams are frozen by the ISLES workshop². The top 7 predicated results of the test dataset on SPES are shown in **Table 2**. Compared with these supervised methods, COL-Net ranks the third. COL-Net obtains DC score is 79.00%, the ASSD is 2.41 mm, and the HD is 34.20 mm. CH-Insel obtains the best DC score (82.00%) which is higher than that of COL-Net 3.00%. DE-UzL obtains the best ASSD (1.36 mm) and HD (23.62 mm) scores which are only 1.05 mm and 10.58 mm lower (better) than that of COL-Net. Although our model does not achieve the best ranking, as a semi-supervised segmentation method, the result of COL-Net is a very competitive in all supervised methods.

4.4. Comparison With the State-of-the-Art Methods

In addition, we also compare COL-Net with the state-of-the-art methods: U-Net (Ronneberger et al., 2015), FCN (Long et al., 2015), Res-UNet (He et al., 2016), Res-FCN (Liu et al., 2018), RA-UNet (Jin et al., 2018), and DRANet (Liu et al., 2020c). These methods are all supervised methods. We either use the code released by the authors or re-implement them exactly as described by the authors. To ensure the fairness

²www.isles-challenge.org

of the comparison, we adopt multi-modalities MRI as input including DWI, CBV, Tmax, CBF, T2, and TTP as the inputs of all comparison methods. The training samples and testing samples of all supervised methods follow the SPES challenge and the evaluation scores come from the official challenge web page.

Table 3 shows the comparison results of the DC, ASSD, and HD scores of these 7 methods. Among all the methods, U-Net, FCN, Res-FCN, Res-UNet, RA-UNet, and DRANet are supervised methods. These methods use the residual mechanism or attention mechanism as the key to improve the segmentation performance of the methods. U-Net and FCN only use the end-to-end structure as the main backbone of the framework. Based on U-shape structure, the residual mechanism is added to Res-FCN and Res-UNet. RA-UNet and DRANet embed residual mechanism and attention mechanism into the encoder-decoder structure. These methods are under the supervised training strategy, they obtain the context information mainly come from the labeled samples, without the help of external auxiliary unlabeled samples. In these methods, RA-UNet and DRANet obtain the best results on 3 metrics: DC scores of 80.00–80.00%, ASSD value of 2.03–1.91 mm, and HD value of 39.43–30.13 mm. Both methods obtain more supplementary semantic information from attention mechanisms. The attention mechanisms help both methods extract the spatial features and focus on spatial semantic information. Compared with these supervised methods, COL-Net obtains a DC score is 79.00%, the ASSD is 2.41 mm, and the HD is 34.20 mm. Although the segmentation results of COL-Net are not great as that of RA-UNet and DRANet, however, under the co-optimization learning strategy and using limited samples to obtain competitive results is a meaningful attempt to assist medical image labeling.

5. DISCUSSION

The good performance of supervised deep learning mainly depends on the training of a large number of labeled data. In this study, we are committed to developing a semi-supervised co-optimization learning network for brain image segmentation, which uses unlabeled data to improve the performance of the segmentation model and reduce the labeling workload.

5.1. Reconstruction Network

In this study, we propose a co-optimization learning framework COL-Net for penumbra tissues segmentation. COL-Net consists of an encoder-decoder segmentation framework (S), reconstruction network (R), and transferblock (T). It takes the advantages of the limited labeled samples. In segmentation framework, we use the residual mechanism to extract high-quality features from the input images and skip operation to improve the processing of the features between the encoder and decoder parts of the network. We present the reconstruction network for reconstructing feature generation in a unsupervised training strategy. The transfer block is used to co-optimization lean the features maps between segmentation framework and reconstruction network. We learn the reconstruction features with that of segmentation network, we embed them into the segmentation network to optimize the model with the mix loss

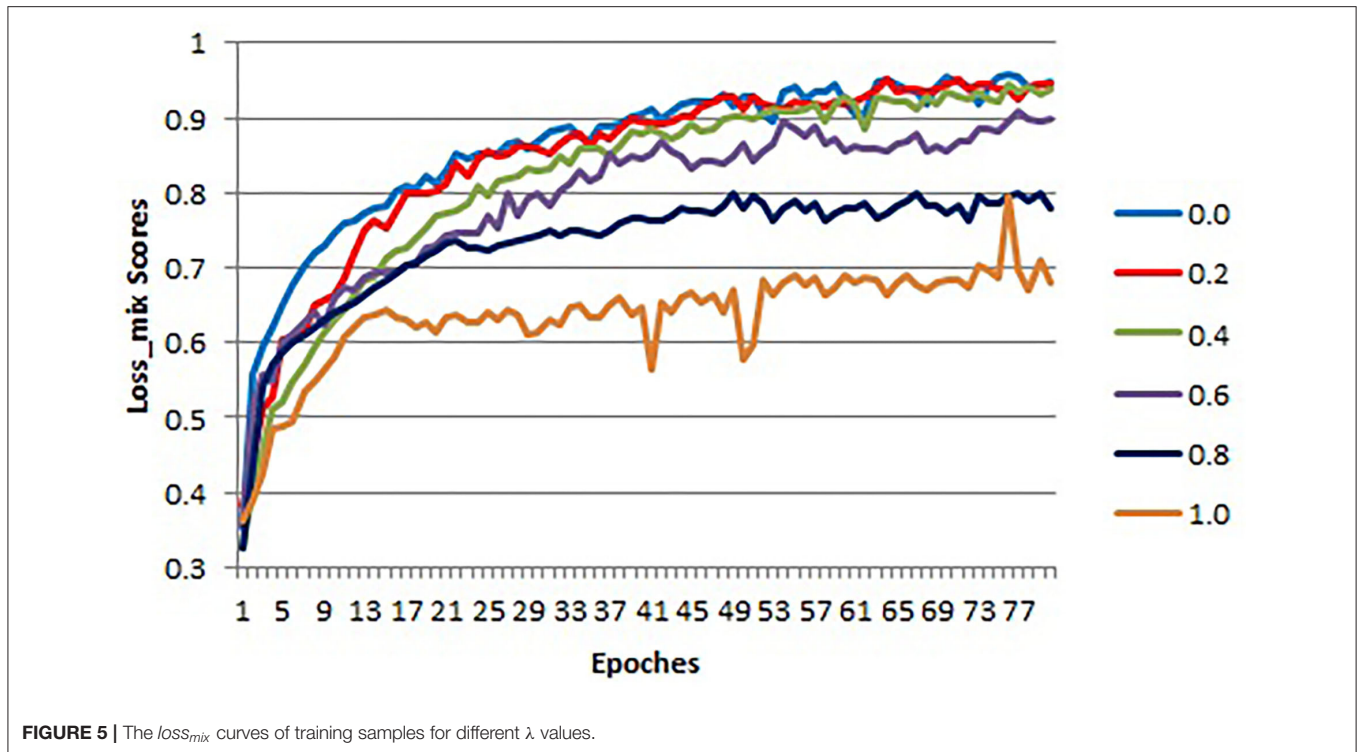


FIGURE 5 | The $loss_{mix}$ curves of training samples for different λ values.

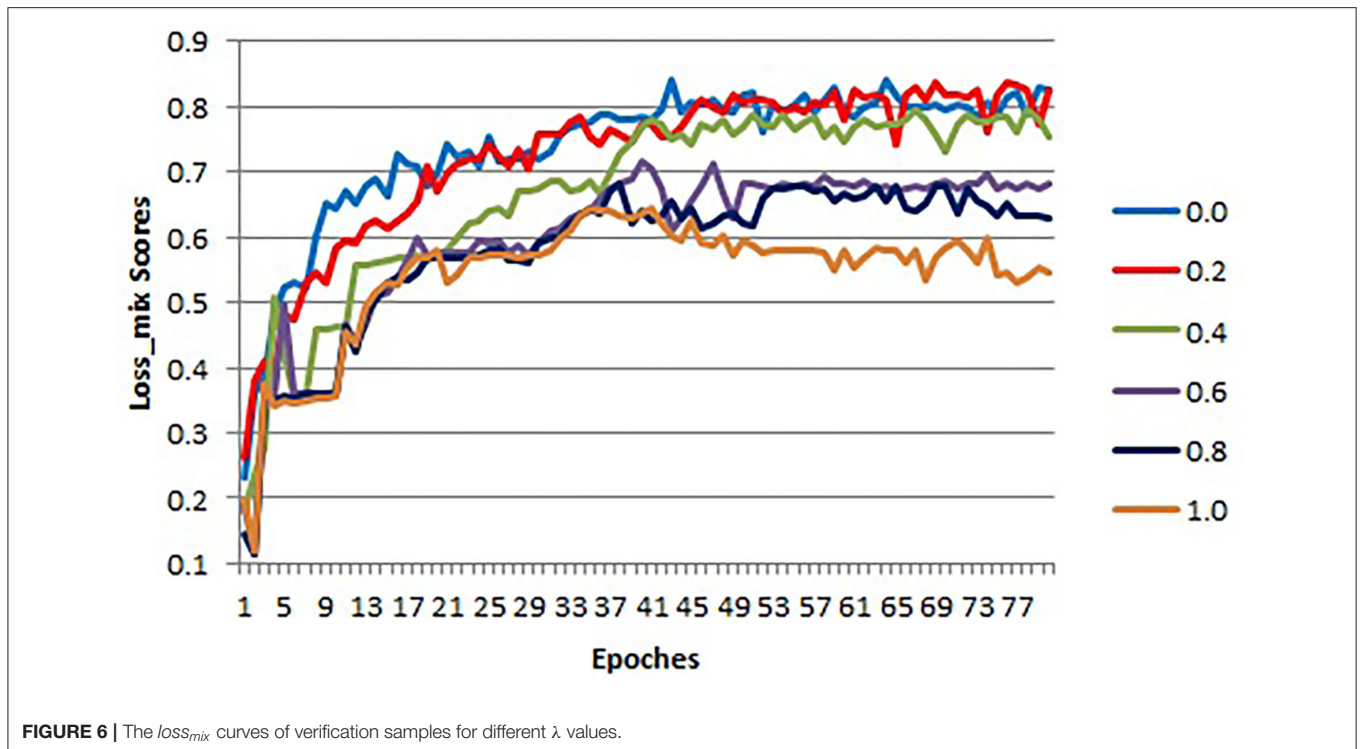


FIGURE 6 | The $loss_{mix}$ curves of verification samples for different λ values.

function. COL-Net is trained under the guidance of the mix loss function. As shown in **Table 3**, the reconstruction network demonstrates its efficacy for improving the segmentation results in the limited labeled stroke medical images. Our

study proves useful in mitigating the shortage of labeled samples of penumbra targets, with two dozens of labeled samples proving sufficient to auxiliary train a highly efficient segmentation model.

TABLE 4 | Ablation study about performance of 4 networks with DC, ASSD and HD metrics.

Methods	DC	ASSD	HD
U-net (Ronneberger et al., 2015)	75.00 (± 12.00)	3.34 (± 2.18)	64.11 (± 29.14)
Res-UNet (He et al., 2016)	76.00 (± 9.00)	2.10 (± 0.57)	42.54 (± 21.14)
COL-Net	79.00 (± 8.00)	2.41 (± 2.19)	34.20 (± 24.05)
RA-UNet (Jin et al., 2018)	80.00 (± 3.00)	2.03 (± 0.86)	39.43 (± 27.07)

DC:%, ASSD:mm, HD:mm.

5.2. Loss Functions

Loss function is used to optimize the model by estimating the inconsistency between the predicted value and the real value of the model. There have two sub-loss functions in COL-Net ($loss_r$ and $loss_s$). The $loss_r$ is used to optimize the reconstruction network, which is based on cross-entropy loss (Equation 1). The $loss_s$ is used to optimize the segmentation network, which is based on DC loss (Equation 2). In order to optimize the COL-Net model as a whole, we combine these two loss functions and propose the mix loss function ($loss_{mix}$) Equation 3. We use λ to weigh the weight of the two-loss functions. We use 5 experiments to get the optimal λ value by setting λ to 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0, respectively. These experiments use the same training and verification samples, which use 80% labeled samples in the training process. The initialization parameters of the model are in section 4.1.

In the training and verification processes, we use the $loss_{mix}$ curves corresponding to $loss_{mix}$ to measure the quality of λ values. The $loss_{mix}$ curves of training and verification samples are shown in **Figures 5, 6**, respectively. In these two coordinates, the horizontal axis corresponds to each epoch value in the training and verification processes respectively, the vertical axis represents the $loss_{mix}$ score of each epoch. As shown in **Figure 5**, when $\lambda = 0.0$ denotes the reconstruction network ($loss_r$) does not work, we only use the supervised segmentation network ($loss_s$) to training COL-Net. When $\lambda = 1.0$ means that the reconstruction network ($loss_r$) and the segmentation network ($loss_s$) have the same weight and influence on COL-Net ($loss_{mix}$). In general, with the increase of epoch value in the training process, the $loss_{mix}$ scores of all λ are gradually increasing. In the verification process, with the increase of epoch value, the $loss_{mix}$ scores of all λ are gradually increasing, except when $\lambda = 1.0$. The value of λ is opposite to that of $loss_{mix}$. When the value of λ increases from 0 to 1.0, the negative effect on $loss_{mix}$ increases gradually. In the training and verification process, the $loss_{mix}$ curves corresponding to $\lambda = 0.0$ and $\lambda = 0.2$ get the best $loss_{mix}$ score, the coincidence degree is very high, which indicates that the COL-Net method can also achieve the supervised segmentation effect where $\lambda = 0.2$. It demonstrates that our proposed method could effective work on the limited labeled samples medical image segmentation task.

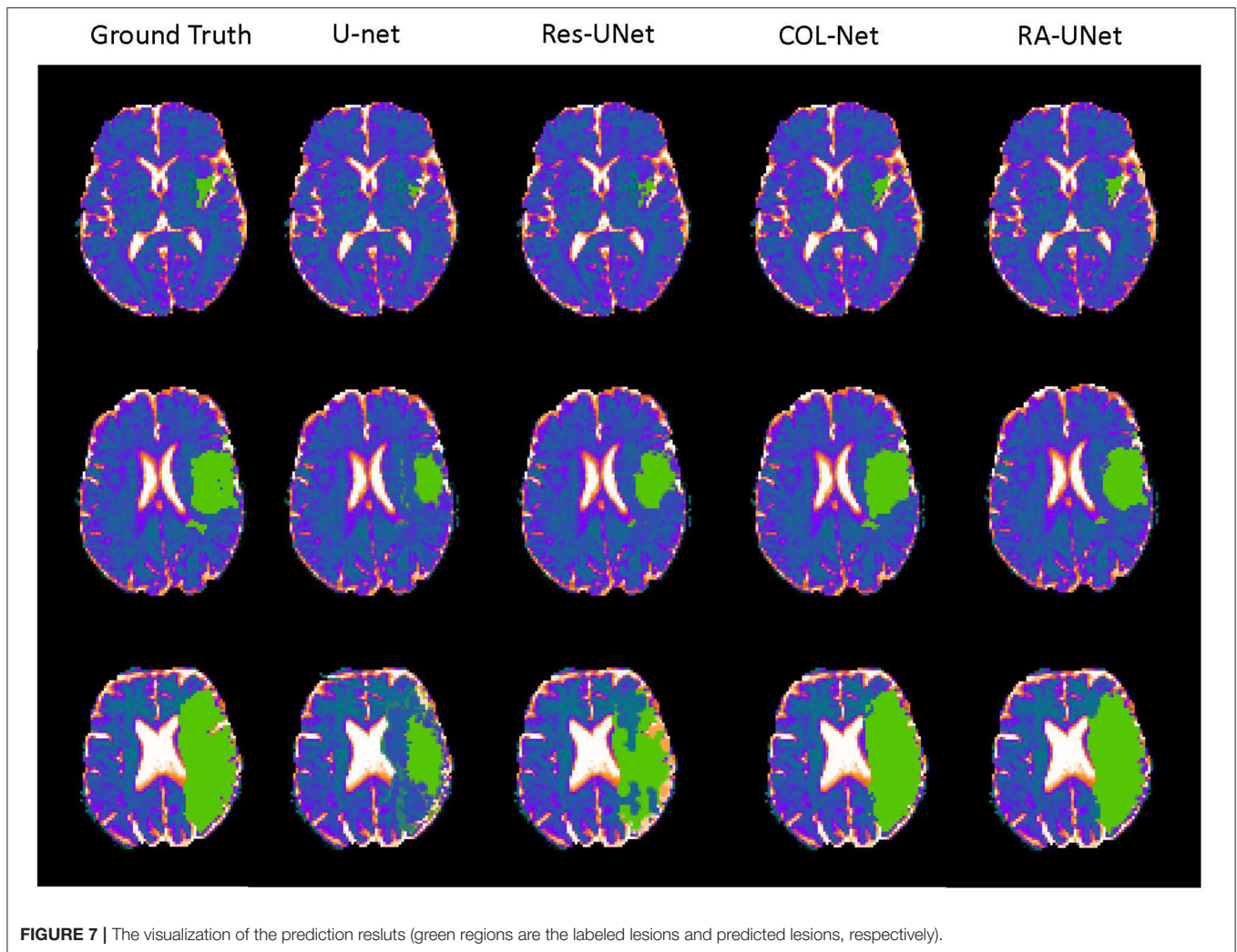
5.3. Ablation Experiments

To investigate the contributions of the reconstruction network in our proposed method, we conduct several ablation experiments

based on COL-Net. In all ablation experiments, we use the training and testing samples as designed in section 4.1. The results of ablation experiments are shown in **Table 4**. We use the representative U-Net as the baseline. We compare the improved Res-UNet based on U-Net. Res-UNet integrates the residual mechanism into the U-Net network. Consequently, the mean DC, ASSD, and HD values are improved by 1.00%, 1.24 mm, and 21.57 mm for penumbra tissues segmentation. Compared with the baseline U-Net, the Res-UNet increases the depth of the network and uses a residual mechanism to alleviates the vanishing gradient problem of the deep network. Furthermore, we compare the improved RA-UNet based on the attention mechanism. RA-UNet integrates attention mechanism into skip connection paths of Res-UNet. The main measure means the DC value of RA-UNet is improved by 4.00% than that of the Res-UNet. Compared with the Res-UNet, the attention mechanism in RA-UNet extract higher quality features from the input images, which is more conducive to improve the model's performance. Compared with other supervised models, our proposed COL-Net obtains the top-level DC and ASSD values for penumbra segmentation. The mean DC value of the COL-Net is only 1.0% lower (worse) than that of RA-UNet, the mean ASSD value of the COL-Net is only 0.38 mm higher (worse) than that of RA-UNet, the HD value of the COL-Net is 0.74 mm lower (better) than that of RA-UNet. Other supervised models use all labeled training samples to fine-tune the models, while our COL-Net model only uses 24 labeled training samples. Although the segmentation performance of the COL-Net model is not the best, it has great advantages to solve the segmentation task with the limited labeled samples.

5.4. Visualization of the Prediction Results

In order to further analyze the performance of the COL-Net, we conduct several extended experiments to visualize the predicted result of a training sample in SPES. There are 30 training samples in the SPES data set, we use 29 samples to train and adjust the models (U-net, Res-UNet, RA-UNet, and COL-Net), the last one is used as a test sample (The 1 - th sample). U-net, Res-UNet and RA-UNet are supervised segmentation methods, COL-Net is a semi-supervised segmentation method. We use the optimal weight of COL-Net trained by 80% labeled samples to test the sample. **Figure 7** shows the ground truth against the predicted results produced by these 4 models. There are 3 scenarios in **Figure 7**. The first row in **Figure 7** is an example that has small and trivial size penumbra tissues. The second row in **Figure 7** is an example that has medium size penumbra tissues. The third row in **Figure 7** is an example that has big size penumbra tissues. In the first row, the ground truth image shows the minuteness and triviality of the annotation label. The supervised U-net and Res-UNet methods are not sensitive to the small lesions, which lead to the fact that the predicted size of the segmentation lesion are much smaller than the ground truth. The prediction result of supervised RA-UNet is the closest to the ground truth. Our semi-supervised COL-Net method is not as good as RA-UNet in predicting small lesions. This is mainly because the attention mechanism in RA-UNet can capture more sensitive feature information under the guidance of a supervised training



strategy. In the second and third rows, the green lesion regions are bigger than those in the first row. The prediction results of U-net and Res-UNet methods are still significantly smaller than that in the ground truth. The prediction results of RA-UNet and COL-Net are the closest to ground truth. This shows that the proposed COL-Net model is comparable to the advanced supervised deep learning model in capturing the features of relatively large lesions.

5.5. Robustness on SISS Challenge Dataset

In order to further verify the performance of COL-Net, we verify it on SISS challenge. We retain the parameters trained on the SPES dataset, and use the reshape method to scale the size of the SISS dataset image to be consistent with the size of the SPES dataset image. The SISS challenge is a subtask of the ISLES 2015 challenge, which contains 64 samples, including 28 labeled training samples and 36 unlabeled testing samples. We perform the experiment on this data set: we use the pseudo unlabeled samples to train reconstruction network. We use the 80% labeled samples to

TABLE 5 | The results of the participants of the SISS dataset.

Methods	DC	ASSD	HD	Manner
UK-Imp2	59.00 (± 31.00)	5.96 (± 9.38)	37.88 (± 30.06)	Sup
Lian1	57.00 (± 29.00)	8.22 (± 16.25)	43.02 (± 30.48)	Sup
COL-Net	56.00 (± 28.00)	9.432 (± 17.05)	44.02 (± 45.10)	Semi
Fengc1	55.00 (± 30.00)	8.13 (± 15.15)	25.02 (± 22.02)	Sup
Martc2	50.00 (± 32.00)	14.69 (± 17.82)	80.06 (± 22.00)	Sup
Abdua1	43.00 (± 31.00)	16.85 (± 15.71)	74.66 (± 25.10)	Sup

DC:%, ASSD:mm, HD:mm. Semi, Semi-supervised training strategy; Sup, Supervised training strategy.

train segmentation network and all test unlabeled samples on the fine-trained segmentation network. Finally, we upload the prediction results of the testing samples to the challenge web page. The obtained DC, ASSD, and HD are given in **Table 5**. It shows that, in all methods, COL-Net adopts a semi-supervised training strategy and achieves the average DC,

ASSD, and HD to 56.00%, 9.432 *mm*, and 44.02 *mm*. The performance of the proposed COL-Net is second only to the best two supervised segmentation methods (UK-Imp2 and Lian1), ranking third. This not only shows that COL-Net can approach the performance of the supervised method in the case of a limited labeled samples but also proves that our model has strong generalization ability.

6. CONCLUSIONS

We propose a novel co-optimization learning network (COL-Net). The COL-Net method tries to use a co-optimization learning strategy to train the segmentation model on the basis of the limited labeled samples and the unlabeled samples. It consists of 3 parts: segmentation network (S), reconstruction network (R), and transfer block (T). The mix loss function is proposed to co-optimization COL-Net. It is verified to segment penumbra tissues from brain MRIs on the SPES challenge and an extended SISS challenge. The results demonstrate that COL-Net achieves the state-of-the-art segmentation performance. Extensive experiments show that COL-Net outperforms most supervised segmentation methods and reaches the advanced level on the limited labeled samples. However, the segmentation accuracy of this method on small lesions needs to be improved. In the face of a limited labeled samples, it is a meaningful attempt to use a co-optimization learning semi-supervised method to make segmentation and prediction. In future work, we will explore semi-supervised segmentation methods with better generalization performance.

REFERENCES

- Andre, E., Brett, K., Novoa, R. A., Justin, K., Swetter, S. M., Blau, H. M., et al. (2019). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118.
- Dora, L., Agrawal, S., Panda, R., and Abraham, A. (2017). State of the art methods for brain tissue segmentation: a review. *IEEE Rev. Biomed. Eng.* 10, 235–249. doi: 10.1109/RBME.2017.2715350
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118. doi: 10.1038/nature21056
- Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *IEEE Trans. Pattern. Anal. Mach. Intell.* 28, 594–611. doi: 10.1109/TPAMI.2006.79
- Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., and Cao, X. (2018). Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* 37, 1597–1605. doi: 10.1109/TMI.2018.2791488
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 770–778.
- He, Y., Yang, G., Yang, J., Chen, Y., and Li, S. (2020). Dense biased networks with deep priori anatomy and hard region adaptation: semi-supervised learning for fine renal artery segmentation. *Med. Image Anal.* 63:101722. doi: 10.1016/j.media.2020.101722
- Hu, W., Cai, B., Calhoun, V., and Wang, Y.-P. (2018). “Multi-modal brain connectivity study using deep collaborative learning,” in *Graphs in Biomedical*

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Ethical review and approval was not required for the animal study because this is a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI (Everyone can obtain the data from <http://www.isles-challenge.org/>).

AUTHOR CONTRIBUTIONS

LL and HZ contributed to conception and design of the study. JZ and J-xW organized the database. LL performed the statistical analysis and wrote the first draft of the manuscript. SX wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work described in this paper was supported by the National Natural Science Foundation of China under grant nos. 61802442, 61772552, 61622213, and 61728211; the 111 project (no.B18059).

Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities (Springer), 66–73.

- Hu, W., Cai, B., Zhang, A., Calhoun, V. D., and Wang, Y.-P. (2019). Deep collaborative learning with application to the study of multimodal brain development. *IEEE Trans. Biomed. Eng.* 66, 3346–3359. doi: 10.1109/TBME.2019.2904301
- Hu, W., Meng, X., Bai, Y., Zhang, A., Qu, G., Cai, B., et al. (2021). Interpretable multimodal fusion networks reveal mechanisms of brain cognition. *IEEE Trans. Med. Imaging* 40, 1474–1483. doi: 10.1109/TMI.2021.3057635
- Huseyn, E. N. (2020). Application of deep learning in medical imaging. *Nat. Sci.* 3, 7–13. doi: 10.36719/2707-1146/04/7-13
- Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. A. (1993). Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 850–863. doi: 10.1109/34.232073
- Jin, Q., Meng, Z., Sun, C., Wei, L., and Su, R. (2018). Ra-unet: a hybrid deep attention-aware network to extract liver and tumor in ct scans. *arXiv preprint arXiv:1811.01328*.
- Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2016). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi: 10.1016/j.media.2016.10.004
- Knudsen, E. I. (1994). Supervised learning in the brain. *J. Neurosci.* 14, 3985–3997. doi: 10.1523/JNEUROSCI.14-07-03985.1994
- Kumar, N., and Awate, S. P. (2020). Semi-supervised robust mixture models in rkhfs for abnormality detection in medical images. *IEEE Trans. Image Proc.* 29, 4772–4787. doi: 10.1109/TIP.2020.2975958
- Lassen, N. A., Fieschi, C., and Lenzi, G. L. (1991). Ischemic penumbra and neuronal death: comments on the therapeutic

- window in acute stroke with particular reference to thrombolytic therapy. *Cerebrovasc. Dis.* 1(Suppl. 1):32–35. doi: 10.1159/000108878
- Li, R., Wagner, C., Chen, X., and Auer, D. (2020a). “A generic ensemble based deep convolutional neural network for semi-supervised medical image segmentation,” in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (Iowa City, IA: IEEE), 1168–1172.
- Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., and Heng, P.-A. (2018). H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Trans. Med. Imaging* 37, 2663–2674. doi: 10.1109/TMI.2018.2845918
- Li, X., Yu, L., Chen, H., Fu, C.-W., Xing, L., and Heng, P.-A. (2020b). Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Trans. Neural Netw. Learn. Syst.* 99, 1–12.
- Liang, C., Bentley, P., and Rueckert, D. (2017). Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks. *Neuroimage Clin.* 15, 633–643. doi: 10.1016/j.nicl.2017.06.016
- Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., et al. (2014). Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Med. Image Anal.* 18, 359–373. doi: 10.1016/j.media.2013.12.002
- Liu, L., Chen, S., Zhang, F., Wu, F.-X., Pan, Y., and Wang, J. (2019). Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality. *Neural Comput. Appl.* 32, 6545–6558. doi: 10.1007/s00521-019-04096-x
- Liu, L., Chen, S., Zhu, X., Zhao, X.-M., Wu, F.-X., and Wang, J. (2020a). Deep convolutional neural network for accurate segmentation and quantification of white matter hyperintensities. *Neurocomputing* 384, 231–242. doi: 10.1016/j.neucom.2019.12.050
- Liu, L., Cheng, J., Quan, Q., Wu, F. X., and Wang, J. (2020b). A survey on u-shaped networks in medical image segmentations. *Neurocomputing* 409, 244–258. doi: 10.1016/j.neucom.2020.05.070
- Liu, L., Kurgan, L., Wu, F. X., and Wang, J. (2020c). Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. *Med. Image Anal.* 65:101791. doi: 10.1016/j.media.2020.101791
- Liu, L., Wu, F. X., Wang, Y. P., and Wang, J. (2020d). Multi-receptive-field cnn for semantic segmentation of medical images. *IEEE J. Biomed. Health Inform.* 24, 3215–3225. doi: 10.1109/JBHI.2020.3016306
- Liu, Q., Yu, L., Luo, L., Dou, Q., and Heng, P. A. (2020e). Semi-supervised medical image classification with relation-driven self-ensembling model. *IEEE Trans. Med. Imaging* 39, 3429–3440. doi: 10.1109/TMI.2020.2995518
- Liu, Z., Cao, C., Ding, S., Han, T., Wu, H., and Liu, S. (2018). Towards clinical diagnosis: automated stroke lesion segmentation on multimodal mr image using convolutional neural network. *arXiv preprint arXiv:1803.05848*, 1–20. doi: 10.1109/ACCESS.2018.2872939
- Long, J., Shelhamer, E., and Darrell, T. (2015). “Fully convolutional networks for semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA: IEEE), 3431–3440.
- Lutnick, B., Ginley, B., Govind, D., McGarry, S. D., LaViolette, P. S., Yacoub, R., et al. (2019). An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nat. Mach. Intell.* 1, 112–119. doi: 10.1038/s42256-019-0018-3
- Maier, O., Menze, B. H., Von, d., G. J., Hni, L., Heinrich, M. P., Liebrand, M., et al. (2017). Isles 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Med. Image Anal.* 35, 250–269. doi: 10.1016/j.media.2016.07.009
- Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. (2011). “Stacked convolutional auto-encoders for hierarchical feature extraction,” in *International Conference on Artificial Neural Networks* (Springer), 52–59.
- Milletari, F., Navab, N., and Ahmadi, S. A. (2016). “V-net: fully convolutional neural networks for volumetric medical image segmentation,” in *International Conference on 3D Vision (3DV)* (Stanford, CA: IEEE), 565–571.
- Moeskops, P., Bresser, J. D., Kuijff, H. J., Mendrik, A. M., Biessels, G. J., Pluim, J. P. W., et al. (2018). Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in mri. *Neuroimage Clin.* 17, 251–262. doi: 10.1016/j.nicl.2017.10.007
- Nakarmi, U., Cheng, J. Y., Rios, E. P., Mardani, M., Pauly, J. M., Ying, L., et al. (2020). “Multi-scale unrolled deep learning framework for accelerated magnetic resonance imaging,” in *IEEE International Conference on Biomedical Imaging* (Iowa City, IA: IEEE), 1–4.
- Rakelly, K., Shelhamer, E., Darrell, T., Efros, A., and Levine, S. (2018). Few-shot segmentation propagation with guided networks. *arXiv preprint arXiv:1806.07373*.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Cham: Springer), 234–241.
- Shaban, A., Bansal, S., Liu, Z., Essa, I., and Boots, B. (2017). One-shot learning for semantic segmentation. *arXiv preprint arXiv:1709.03410*. doi: 10.5244/C.31.167
- Van Engelen, J. E., and Hoos, H. H. (2020). A survey on semi-supervised learning. *Mach. Learn.* 109, 373–440. doi: 10.1007/s10994-019-05855-6
- Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., and Wierstra, D. (2016). Matching networks for one shot learning. *arXiv preprint arXiv:1606.04080*.
- Weston, J., Ratle, F., Mobahi, H., and Collobert, R. (2012). “Deep learning via semi-supervised embedding,” in *Neural Networks: Tricks of the Trade* (Springer), 639–655.
- Zhang, L., Wang, L., Gao, J., Risacher, S. L., Yan, J., Li, G., et al. (2021). Deep fusion of brain structure-function in mild cognitive impairment. *Med. Image Anal.* 72:102082. doi: 10.1016/j.media.2021.102082
- Zhang, R., Zhao, L., Lou, W., Abrigo, J. M., Mok, V. C., Chu, W. C., et al. (2018). Automatic segmentation of acute ischemic stroke from dwi using 3d fully convolutional densenets. *IEEE Trans. Med. Imaging* 37, 2149–2160. doi: 10.1109/TMI.2018.2821244
- Zhao, Q., Adeli, E., and Pohl, K. M. (2020). Training confounder-free deep learning models for medical applications. *Nat. Commun.* 11:6010. doi: 10.1038/s41467-020-19784-9
- Zhou, B., Augenfeld, Z., Chapiro, J., Zhou, S. K., Liu, C., and Duncan, J. S. (2021). Anatomy-guided multimodal registration by learning segmentation without ground truth: application to intraprocedural cbct/mr liver segmentation and registration. *Med. Image Anal.* 71:102041. doi: 10.1016/j.media.2021.102041

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Liu, Zhang, Wang, Xiong and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.