Magnetic Resonance Imaging 63 (2019) 70–79

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

A multi-view pyramid network for skull stripping on neonatal T1-weighted MRI

Yan Gao[a], Jie Li[a,*], Haojun Xu[a], Miaomiao Wang[b], Congcong Liu[b], Yannan Cheng[b], Mengxuan Li[b], Jian Yang[b], Xianjun Li[b,*]

[a] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[b] Department of Radiology, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

ARTICLE INFO

Keywords:
Skull stripping
Brain extraction
Neonatal brain
Magnetic resonance image
Pyramid neural networks

ABSTRACT

Skull stripping or brain extraction on magnetic resonance imaging is a crucial step for structure analyses. In spite of good performances of conventional methods on adult brains, the skull stripping for T1-weighted imaging (T1WI) images on the neonatal brain remains a challenge because of the low image contrast. Therefore, this paper proposes a multi-view pyramid skull stripping network (PSSNet) for neonatal T1WI. To achieve superior skull stripping performance, the conventional pyramid scene parsing network was modified through (1) adding the spatial information of raw feature maps by squeezing the channel information during the feature extraction; (2) increasing the receptive field and adding boundary repair block instead of direct up-sampling; (3) obtaining the final mask through a fusion module on multi-view 2D slices. The 3D skull stripping problem was decomposed into multi-view 2D segmentation tasks to improve the efficiency. We enrolled T1WI images of 70 neonates from the local hospital and 7 infants from the publicly available dataset NeuroBrainS12 (MICCAI 2012). Images of 51 and 26 subjects were used for model training and validation. We compared the proposed method with 7 commonly used methods by using the Dice ratio, sensitivity, specificity, and efficiency. The proposed multi-view PSSNet with the highest Dice ratio (95.44–97.33%) was superior to other methods. Meanwhile, the sensitivity (93.19–97.02%), specificity (97.52–99.68%), and efficiency (8.59–9.30 s per subject) of the proposed method were comparable with the state-of-the-art method. In conclusion, the proposed skull stripping network was robust on neonatal T1WI datasets and feasible in clinical applications.

1. Introduction

Magnetic resonance imaging (MRI) is widely used in neonatal examinations and researches on the brain development due to its advantages of safety and relatively high spatial resolution [1,2]. Many neurodevelopment and psychiatric disorders are associated with the abnormal brain development in neonates [3]. Evaluation of neonatal brain structures by T1-weighted imaging (T1WI) can quantitatively characterize the brain development, which holds the potential to predict outcomes [4,5]. Skull stripping or brain extraction of T1WI images is a critical step in structure analyses [1,6]. The accuracy of skull stripping is one of the factors potentially influencing the subsequent morphological analyses [1,7]. Therefore, it is of great significance to perform accurately skull stripping on the neonatal MRI.

Manual segmentation by radiologists is always considered as the "gold standard" [8]. However, manual labeling each voxel of 3D T1WI images is time-consuming. Apart from the manual segmentation, various automatic skull stripping methods have been proposed. Some of them have been widely used, such as the brain extraction tool (BET) [9], 3dSkullStrip [10], and brain surface extractor (BSE) [11]. BET uses a deformable spherical mesh model to fit the brain surface through the local adaptive evolution [9]. 3dSkullStrip [10] is a modified version of BET that addresses the shortcomings of errors around ventricles and eyes. BSE uses anisotropic diffusion filters, edge detection and morphological operations to complete the segmentation [8]. However, these methods do not work well on neonatal T1WI images due to the

Abbreviations: MRI, magnetic resonance imaging; T1WI, T1-weighted imaging; BET, brain extraction tool; BSE, brain surface extractor; iBEAT, infant brain extraction and analysis toolbox; CNN, convolutional neural networks; PSSNet, pyramid skull stripping network; GAP, global average pooling; PSPNet, pyramid scene parsing network; LC, convolution layer; GCN, global convolutional network; BR, boundary repair; FP, false positive; NeoWA, neonates without abnormalities; NeoWML, neonates with white matter lesions; $\mathcal{SNR}$, signal-to-noise ratio; ResNet, residual network; lr, learning rate; *TP*, true positive; *TN*, true negative

* Corresponding authors.
E-mail addresses: leejie@mail.xidian.edu.cn (J. Li), xianj.li@mail.xjtu.edu.cn (X. Li).

https://doi.org/10.1016/j.mri.2019.08.025
Received 29 April 2019; Received in revised form 12 July 2019; Accepted 15 August 2019

lower contrast than adults. To improve the performance of the skull stripping on pediatric subjects, brain extraction meta-algorithm is developed [12] and has been integrated into the infant brain extraction and analysis toolbox (iBEAT) [6]. This algorithm combines BET and BSE to optimize the parameters, and uses a level set based label fusion algorithm to get the final segmentation results. In addition, another strategy to repair the BET results has been proposed based on the two-level Markov-Gibbs random field and the geometric model of brain isosurfaces [13]. The hybrid method attempts to combine features of multiple methods to get better segmentation results. The classifier with combined multiple atlases [14] holds obvious advantages than conventional methods. However, these methods require registration of atlas as a preprocessing step [15]. The target position alignment, parameter setting and initialization would seriously affect the segmentation quality.

Recently, deep learning enables the convolutional neural networks (CNN) to automatically learn the distribution of data [8], which provides suitable strategy for the neonatal skull stripping. Among different networks, U-net holds fewer parameters than fully convolution networks, which is suitable for the segmentation task of medical images. For parsing the context information of 3D images, 3D U-Net [16] uses 3D convolution to replace the 2D convolution, enabling end-to-end voxel segmentation. By using the 3D convolutional layers, a brain extraction algorithm has been performed on adults [8]. Furthermore, Auto-Net uses the previous segmentation result obtained by the network as the auxiliary context information to improve the segmentation accuracy [17]. However, these deep learning-based methods, especially the 3D models, require a mass of computing resources. Moreover, these methods also face the problem of low efficiency during training.

In this paper, we tried to propose a multi-view fusion network method to solve the above problems. The task of neonatal skull stripping, with simple and clear semantic information, requires the separation of brain parenchyma from 3D T1WI images. Therefore, we fully considered the spatial dependence between the skull and the brain parenchyma. This study would use the brain attention mechanism to focus on the position of brain tissue. We extract features of the brain region with the multi-scale pooling component to reduce the interference caused by brain injuries. Then the bounding repairs were made to refine and aggregate edges of the brain tissue. In order to improve the efficiency of the algorithm, this study would decompose the 3D T1WI image into a 2D sequence and carry out the skull stripping layer by layer. The network conducted the simple binary classification and integrated predictions of three views to obtain the final segmentation results. We named this automatic method as the multi-view pyramid skull stripping network (multi-view PSSNet).

This study compared the proposed method with 7 automatic skull stripping methods on the local and public datasets. Three out of 7 methods are based on deep learning: 3D U-Net [16], 2D U-net [18], and Auto-Net [17]. The others have been integrated into publicly available software: iBEAT [6], BET [9], 3dSkullStrip [10], and BSE [11]. We hypothesized that the proposed multi-view PSSNet would be effective for neonatal T1WI images. Meanwhile, it would be robust on neonates with brain lesions.

2. Materials and methods

2.1. Theory

This study designed a 2.5D hybrid network system to address the neonatal skull stripping task, named multi-view PSSNet (as shown in Fig. 1). In this section, we would introduce components of our method and analyze their roles in the neonatal skull stripping task. Then, other design details of the method would be described.

2.1.1. Multi-scale attention maps for brain focus

2.1.1.1. Brain focus module. To tackle the issue of neonatal T1WI images, we firstly consider the semantic information choice in features. The convolution head is more suitable for brain tissue and

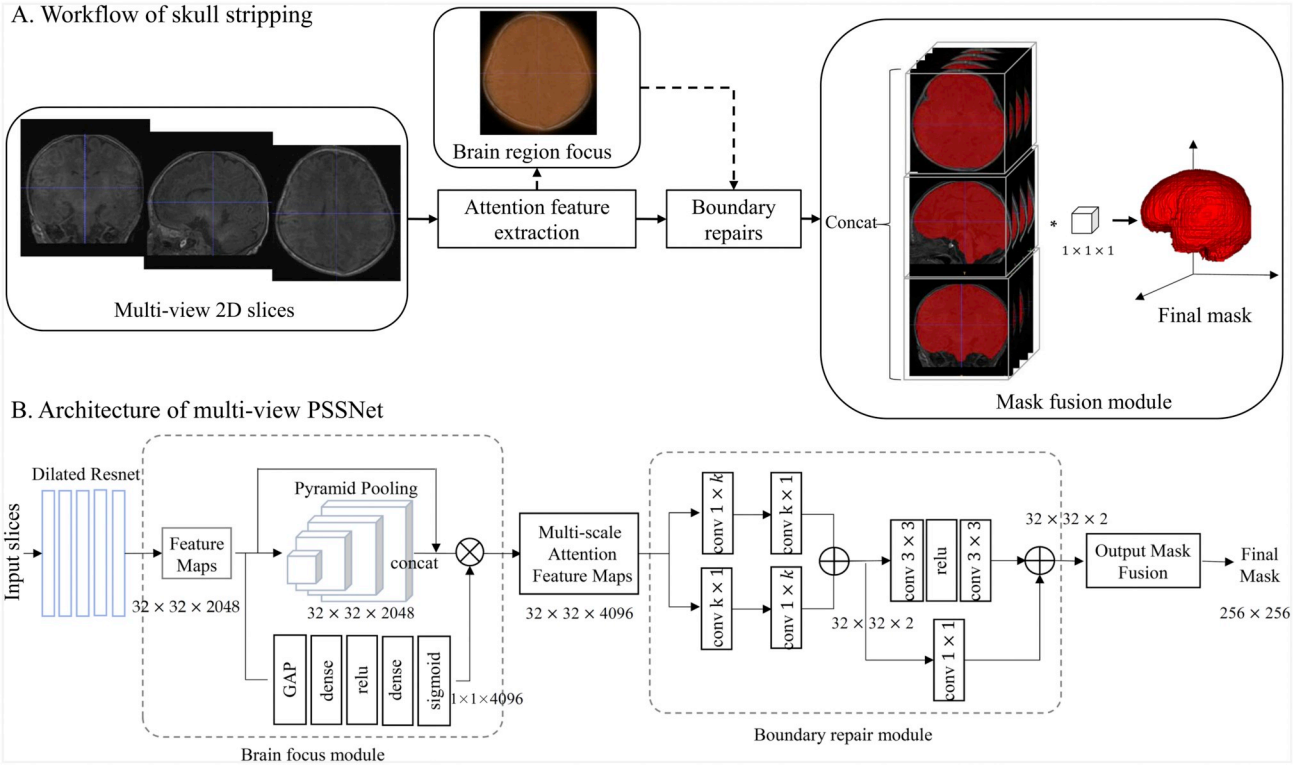

Fig. 1. The workflow of the neonatal skull stripping and the architecture of the multi-view pyramid skull stripping network (PSSNet). GAP stands for the global average pooling; dense stands for the full connected layer; and conv stands for the convolutional layer.

skull localization, considering that lower-level features contain more location information and are able to capture the context for the entire header. However, the low contrast of neonatal images and variations in tissue intensity cause that the localization needs more global features. In addition, channel relationships obtained from the convolution is essentially local. Therefore, we try to access the entire spatial information (i.e. channel information) prior to separating the brain parenchyma. Considering that the fully connected head is spatially sensitive, the structure of the brain focus module is designed as shown in Fig. 1. By constructing the dependency between channel information and multi-scale features, we can obtain the brain attentional features containing the context and spatial information.

2.1.1.2. Spatial information. To capture the spatial information of raw feature maps $\mathscr{F} \in \mathbb{R}^{C\times H\times W}$ (C, H, W are the channel number, height, width of feature maps), this work would squeeze and excitated the channel information of $\mathscr{F}$ firstly [19]. Squeeze uses a global average pooling (GAP) layer to generate channel-wise statistics $z \in \mathbb{R}^{1\times 1\times C}$ with the cth element:

$$z_c = \frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W} p_c(i,j) \tag{1}$$

where, H, W are the height, width of feature maps. c represents the cth element in z.

Excitation$\hat{z}$ is would feed z into two fully connected layers (dense) for the dimensionality-reduction with a non-linearity activation relu δ in the middle:

$$\hat{z} = W_2\delta(W_1 z) \tag{2}$$

where, $W_1 \in \mathbb{R}^{\frac{C}{2}\times C}$ and $W_2 \in \mathbb{R}^{C\times\frac{C}{2}}$ are weights of two fully connected layers. In other word, the excitation operation encodes the channel dependencies. Finally, a sigmoid activation layer transfers $\hat{z}$ to interval [0, 1]. The extraction of the spatial information depends on the dynamic neonatal MRI input, which is helpful to improve the brain tissue discrimination.

2.1.1.3. Multi-scale global information. Unlike the channel attention mechanism [19], this study would enhance the multi-scale context feature maps rather than the raw feature maps $\mathscr{F}$ to cater to neonatal T1WI images. According to the pyramid scene parsing network (PSPNet) [20], the pyramid pooling structure is used to generate multi-scale feature maps. Specifically, this work defined 4 global average pooling layers with different kernel sizes of 1×1, 2×2, 3×3, and 4×4. The differences in the receptive field size of the multi-level pooling control can make the feature map include multi-scale changes between different regions [21]. We denote the feature map through the multi-level pooling as $\mathscr{P}$. Finally, we use the bilinear interpolation to directly upsample the $\mathscr{P}$ to the $\mathscr{F}$ size. In addition to properly representing medical images for localization, the pyramid pooling structure can also reduce the impact of highlighted lesions on segmentation. This is because that expanding the scale of pooling will weaken the lesion information and enhance the neighborhood information. The brain focus module combines $\mathscr{F}$ and $\mathscr{P}$ as multi-scale global feature maps $\mathscr{M}$. Then, we would multiply the output of the channel attention branch by $\mathscr{M}$ to obtain the multi-scale attention feature map. The multi-scale attention feature map is the input of the boundary repair module.

2.1.2. Boundary repairs for brain score maps

2.1.2.1. Semantic score map. The structure of the boundary repair module is shown in Fig. 1. The per-pixel classification requires the effective receptive field of feature map, large enough to cover the entire object [22]. And the convolution with a large kernel can directly increase the size of the receptive field. Therefore, we apply a large separable convolution layer (LC) with 2 output channels [23,24] to generate the final semantic score map based on the multi-scale attention feature map.

This layer is the sum of $1\times n + n\times 1$ and $n\times 1 + 1\times n$ convolution. It can instead of $n\times n$ convolution, which would reduce parameters about $\frac{n}{4}$ times. The brain tissue is usually at the center of the image. Global convolutional network (GCN) [24] has proved that the large separable convolution layer could learn the relationship between pixels within the segmentation object. We assume that it has boundary sensitivity, which would be verified in this study.

2.1.2.2. Boundary repair. For the boundary pixels of the target, the performance is mainly affected by the network locating capability [24]. We hypothesized that the edge of the target would be better fitted by combining semantic features of each pixel and the neighborhoods. The boundary repair (BR) block is composed of two 3×3 convolution layers and an relu activation layer in the middle. Results in the ablation experiment section demonstrated the improvement by adding the BR block.

2.1.3. Mask fusion module

The proposed framework combines two dimensional networks (as shown in Fig. 1). Considering the consumption of GPU memory, the proposed method uses multi-view 2D slices. And the discrete 2D prediction probability map in three views is obtained by the boundary repair module. However, it may reduce the segmentation accuracy since the 2D network ignores the context of adjacent slices. In order to further improve the precision of the 2D prediction, the output probability map is stacked into a 3D entirety, and sent to a $1\times 1\times 1$ 3D convolution layer (Fig. 1A).

2.1.4. Inference

The training of the proposed multi-view PSSNet is divided into two phases. Firstly, probability maps for each 2D slice are predicted and then stacked into a 3D volume. Secondly, the 3D volume is saved as an input to the mask fusion module. The softmax cross entropy loss $\mathscr{S}$ is used to train the final classifier. The overall loss function of the multi-view PSSNet during training is:

$$\mathscr{L}_\theta = \mathscr{S}_{pss} + \alpha\mathscr{S}_{aux} + \mathscr{S}_{mkf} \tag{3}$$

where θ is the parameter of the multi-view PSSNet; $\mathscr{S}_{pss}$ is the softmax loss of the whole network; $\mathscr{S}_{mkf}$ is the softmax loss of 3D mask fusion module; $\mathscr{S}_{aux}$ is the auxiliary loss. We set $\alpha = 0.2$ to scale the gradient according to the results in the ablation experiment, which would speed up the network convergence.

2.2. Datasets

2.2.1. Neonates from the local hospital

The in-house dataset included 3D T1WI images of 70 neonates from the local hospital. The median postnatal age at MRI was 1.29 weeks with range from 0.14 to 3.71 weeks. The median gestational age was 37.29 with range from 30.71 to 41.57 weeks. The median post-menstrual age (gestational age + postnatal age at MRI) was 38.64 weeks with range from 32.14 to 44.14 weeks. This study was approved by the local institutional review board. Informed written consents were obtained from parents of the neonates. The three-dimensional fast spoiled gradient-recalled echo T1WI was performed on a 3 T MRI scanner (Signa HDxt; GE Healthcare, USA) with an 8-channel head coil. Repetition time = 10.468 ms; Echo time = 4.764 ms; Inversion time: 400 ms; Field of view = $240\times 240\,mm^2$; Acquisition matrix = 240×240; Reconstruction matrix = 256×256; Slice thickness = 1 mm.

The masks of brains were manually drawn by radiologists as the “ground truth”. In this work, we included the subarachnoid cerebrospinal fluid, gray matter, white matter, ventricles and cerebellum. We divided the data into two kinds, the neonates without abnormalities

(NeoWA) dataset and the neonates with white matter lesions (NeoWML) dataset. The lesions were identified as hyperintensity on T1WI [25,26]. Among the NeoWML data, some neonates were accompanied by other lesions, such as the parenchymal and/or extracerebral haemorrhage.

2.2.2. NeoBrainS12

The public dataset was the preterm infant brain segmentation competition dataset, NeoBrainS12, including 7 subjects (2012 MICCAI: http://neobrains12.isi.uu.nl/). The MRI images were obtained at the term equivalent age (postmenstrual age = 40 weeks). The median gestational age was 27.57 with range from 25.57 to 30.86 weeks. The median postnatal age at MRI was 12.43 weeks with range from 9.14 to 14.43 weeks. Considering that the postmenstrual age = 40 weeks, contrasts of these T1WI images were the neonatal pattern.

2.2.3. Datasets grouping

2.2.3.1. Image quality evaluation. We calculated the signal-to-noise ratio ($\mathscr{SNR}$) and tissue contrast (Contrast) to evaluate the image quality of datasets. The $\mathscr{SNR}$ is defined as: $\mathscr{SNR} = \frac{\mathscr{R}}{\mathscr{B}}$, where, $\mathscr{R}$ is signal in the brain region of MRI; $\mathscr{B}$ is the signal in the background region. Also, assuming that the averaged value of the pixel intensity in the brain region is $\mathscr{L}_b$, and the averaged value of the pixel intensity in the skull region is $\mathscr{L}_s$, the Contrast is defined as: $Contrast = \frac{\mathscr{L}_b}{\mathscr{L}_s}$. As shown in Fig. 2 and data in Table 1, datasets of NeoWA, NeoWML, and NeoBrainS12 have different mean and standard deviation of $\mathscr{SNR}$ and image contrast. These differences are related to scanning protocols and/or brain lesions. Variations of the image quality across different datasets may bring challenges to the skull stripping. We will validate the generalization capabilities of algorithms for different datasets.

2.2.3.2. Training and validation datasets. In total, this study used images of 77 subjects (70 local and 7 public subjects) for model training and validation. These samples were divided into two groups: training samples (n = 51) and validation samples (n = 26). For the NeoWA dataset, the training dataset included 24 subjects, while the validation dataset included 10 neonates. In NeoWML dataset, the training dataset included 25 neonates, while the validation dataset included 11 neonates. For the NeoBrainS12 dataset, 2 and 5 subjects were considered as the training and validation data, respectively.

Table 1
Single-to-noise ratio (SNR) and tissue contrast (Contrast) of three datasets.

Dataset	SNR	Contrast
NeoWA (n = 34)	20.31 (± 5.82)	0.84 (± 0.17)
NeoWML (n = 36)	15.01 (± 4.36)	0.88 (± 0.89)
NeoBrainS12 (n = 7)	13.52 (± 10.37)	1.14 (± 0.05)

Note: NeoWA: neonates without abnormalities; NeoWML: neonates with white matter lesions; NeoBrainS12: 2012 MICCAI infant brain segmentation competition dataset.

All the 7 methods were compared with the proposed multi-view PSSNet on the same validation datasets.

2.3. Data processing and analysis

2.3.1. Preprocessing

To prevent overfitting of the model, this study used strategies of data enhancement [22]. Firstly, random flip and crop were used to convert the data [27]. Then, the same transformation was applied on their masks. In addition, for NeoBrainS12, we found that one case of the training set had an offset between the data and the mask. Thus, we reposition the mask to the brain by using the alignment method [28] prior to the model training. And for the local datasets, the image was cropped to retain only the head area, so that the effects of noise and artifacts on the training would be reduced.

2.3.2. Evaluation of the performance

In order to evaluate the performance of the proposed method, the Dice ratio, sensitivity and specificity would be used:

$$Dice = \frac{2 \times TP}{FP + 2 \times TP + FN} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

where, *TP* is the true positive rate, *FP* is the false positive rate, *TN* is the true negative rate, and *FN* is the false negative rate. The higher the Dice

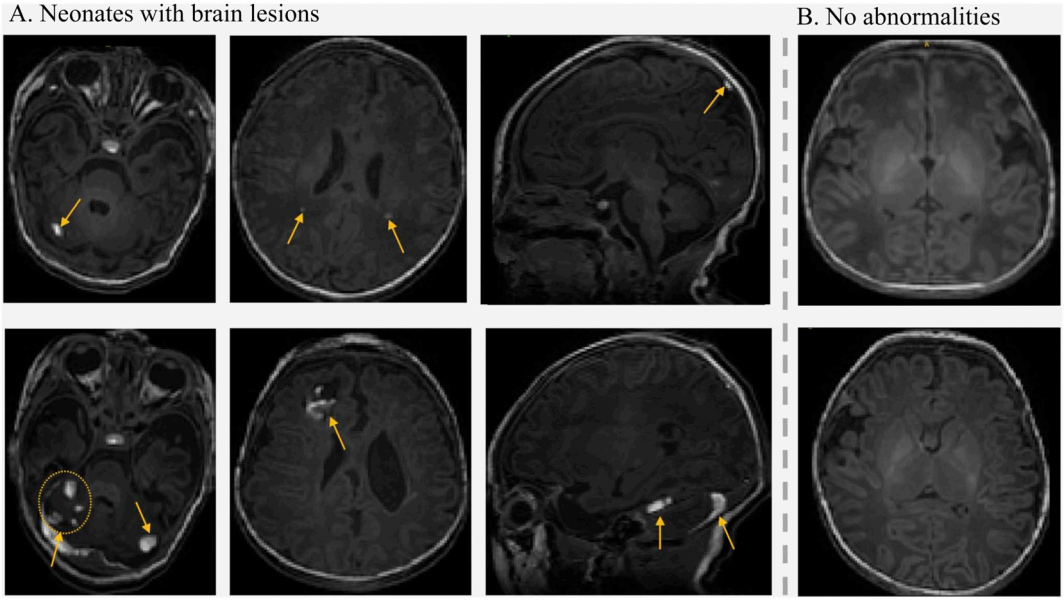

Fig. 2. Instances of T1WI images for neonates with brain lesions and those without MRI abnormalities.

coefficient obtained in the segmentation result, the better the skull stripping. Sensitivity describes how sensitive the algorithm is to brain tissues. It shows that probability of brain tissues being correctly identified. Specificity describes another aspect, probability of non-brain tissue being correctly identified. We averaged the *FP*, *FN*, and absolute errors across subjects to generate mean maps of these metrics.

2.3.3. Implementation and the ablation experiment

2.3.3.1. Implementation. iBEAT, BET, 3dSkullStrip, and BSE are publicly accessible software. And the source codes for 3D U-Net, U-net, and Auto-Net are also available. Therefore, we implemented these 7 methods by using default parameters. The proposed and the other deep learning methods were implemented by using the same hyperparameters and training strategies. We used the default values for parameters of the deep learning algorithm optimizer. The initialization of network parameters was performed by using the method of Xavier [29,30], where the magnitude = 2. According to the reference [20], the "poly" learning rate policy was used to dynamically update the learning rate for model convergence, where the learning rate = 0.001, *momentum* = 0.9, *weight decay* = 0.0001. The mini-batch stochastic gradient descent [29] was used to optimize the network. All experiments were implemented on a personal computer of Intel core i7-7700k CPU and NVIDIA Geforce GTX 1080ti GPU.

In order to determine effects of the multi-view PSSNet architecture and parameters on the performance of the skull tripping, this study conducted the ablation experiment. In this section, we would discuss the block selection by the module analysis (including the selection of the proportion α of the auxiliary loss), the kernel size selection in the LC, and the kernel size selection in the 3D convolution layer.

2.3.3.2. Module analysis. The proposed method was based on the PSPNet [20]. The baseline head was the dilated residual network (ResNet) [31]. Three depths of 50, 101, and 152 were tested. Table 2 showed that the Dice ratio of PSSNet increased from 95.38% to 96.30% when the depth of ResNet increased from 50 to 101. However, when the depth increased to 152, the performance of the network decreased. Therefore, we chose the ResNet101 with the highest Dice ratio as the pre-trained model. In Table 3, the Dice ratio, sensitivity and specificity of PSP ResNet101 were 95.86%, 95.73% and 97.76%, respectively. Then, we added Sinfo, LC, and BR block, respectively. As shown in Table 3 and Fig. 3, the LC is more beneficial to the pixel classification for the object center, while the BR can improve the classification accuracy of the boundary. Additionally, we found that the symmetric surface distance decreased from 1.62 ± 0.36 to 1.20 ± 0.18 after adding the LC and BR. Note that the complete multi-view PSSNet architecture held the highest Dice ratio of 96.33% in this ablation experiment. It achieved a better trade-off between sensitivity and specificity. To determine the proportion α of auxiliary loss, Dice ratios were calculated for different selections. Dice ratios for α of 0, 0.2, 0.4, 0.6, 0.8, 1.0 were 96.30% ± 0.67%, 96.33% ± 0.25%, 96.22% ± 0.31%, 96.32% ± 0.30%, 96.18% ± 0.57%, 96.22% ± 0.97%, respectively. Therefore,we set the α = 0.2 to scale the gradient.

Table 2
Performances (Dice ratio, sensitivity, and specificity) of the pre-trained model with different depths (%).

Method (depth)	Dice	Sensitivity	Specificity
ResNet(50)	95.38 (± 0.66)	97.79 (± 1.86)	95.92 (± 1.22)
ResNet(101)	96.30 (± 0.62)	97.20 (± 2.10)	97.37 (± 0.97)
ResNet(152)	96.19 (± 0.76)	96.22 (± 1.95)	97.83 (± 0.80)

Note: ResNet: residual network.

2.3.3.3. Kernel size selection in the LC. Dice ratios for the kernel size of 3 × 3, 7 × 7, 11 × 11, 15 × 15, 19 × 19, 23 × 23, 27 × 27, and 31 × 31 were 97.20% ± 0.76%, 85.81% ± 6.56%, 97.21% ± 0.85%, 87.40% ± 1.61%, 97.21% ± 0.73%, 97.00% ± 0.92%, 97.30% ± 0.30%, and 97.36% ± 0.57%, respectively. Therefore, the kernel size of 31 × 31 held the highest Dice ratio and was selected in this work.

2.3.3.4. Kernel size selection in the 3D convolution layer. The Dice ratios for the kernel size of 1 × 1 × 1, 3 × 3 × 3, 5 × 5 × 5, and 7 × 7 × 7 were 96.33% ± 0.69%, 96.21% ± 0.55%, 96.20% ± 0.52%, and 96.27% ± 0.53%, respectively. Therefore, the kernel size of 1 × 1 × 1 held the highest Dice ratio and was selected in this work.

2.3.4. Statistical analysis

The Wilcoxon signed rank test in MATLAB (7.11, MathWorks, Natick, Massachusetts, USA) was used to investigate differences in the Dice ratio, sensitivity, and specificity between the proposed method and other methods. *P* values < 0.05 were considered as the statistical significance.

3. Results

3.1. Visual evaluation

Figs. 4, 5, and 6 showed the representative skull stripping results of different methods on three datasets. All algorithms were compared on the same slice of the same neonate.

3.1.1. NeoWA dataset

Results in Fig. 4 demonstrated that iBEAT, 2D U-net, and 3D U-Net had slightly over-segmentation problems on neonates without abnormalities. There were partial masks covering non-brain tissue. The under-segmentation phenomenon of BET was obvious. The masks only covered part of the brain tissue. The results of BSE and 3dSkullStrip were a complete coverage of the entire brain including the skull. According to the visual evaluation, results of the proposed method were closest to the ground truth.

3.1.2. NeoWML Dataset

For the NeoWML dataset, the key is whether it can successfully segment the areas with brain lesions. Size of white matter lesions varies

Table 3
Performances (Dice ratio, sensitivity, and specificity) of models with different structures in the ablation experiment (%).

Method	Dice	Sensitivity	Specificity
PSP ResNet101	95.86 (± 0.46)	95.73 (± 2.14)	97.76 (± 1.03)
PSP ResNet101 + Sinfo	96.10 (± 0.65)	94.82 (± 2.46)	98.04 (± 0.84)
PSP ResNet101 + Sinfo + LC	96.30 (± 0.65)	95.81 (± 2.46)	98.23 (± 0.84)
PSP ResNet101 + Sinfo + BR	96.00 (± 0.68)	95.32 (± 2.61)	98.16 (± 0.91)
PSP ResNet101 + Sinfo + LC + BR	96.30 (± 0.67)	97.20 (± 2.28)	97.37 (± 1.05)
Multi-view PSSNet	96.33 (± 0.69)	97.02 (± 2.18)	97.52 (± 0.96)

Note: PSSNet: pyramid skull stripping network; PSP: pyramid scene parsing; ResNet: residual network; LC: convolution layer; BR: boundary repair.

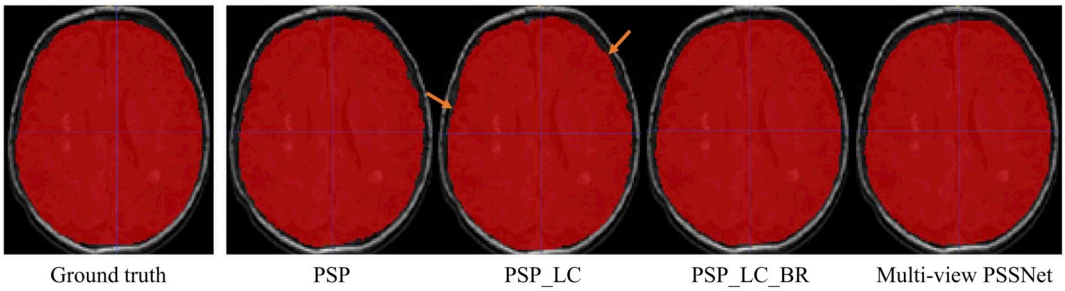

Fig. 3. Representative slices of the skull stripping in the ablation experiment.

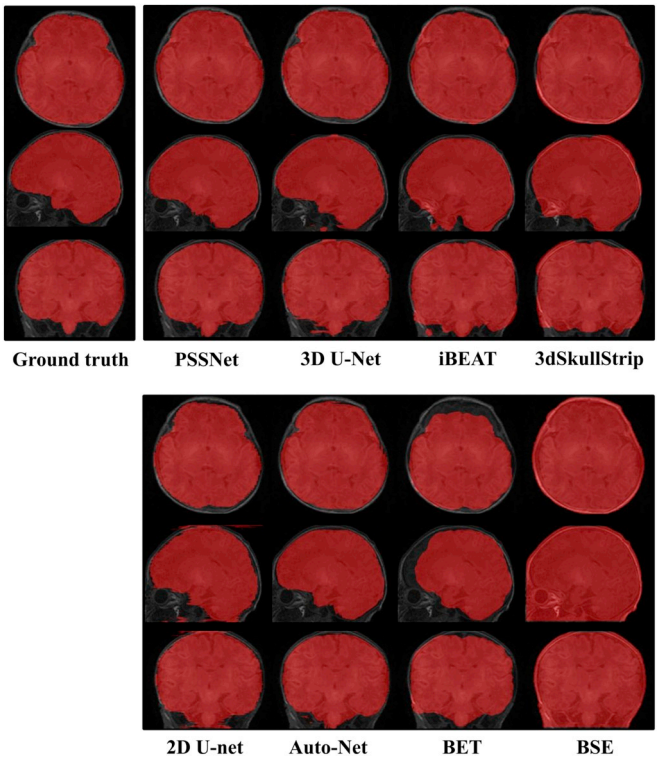

Fig. 4. Representative slices of the skull stripping by using different methods on neonates without abnormalities (NeoWA).

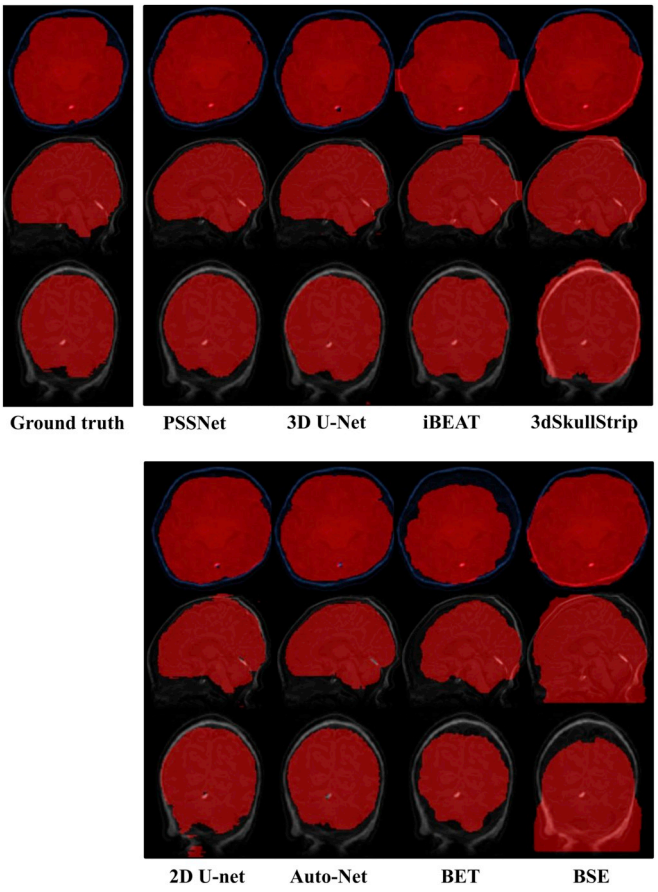

Fig. 5. Representative slices of the skull stripping by using different methods on neonates with white matter lesions (NeoWML).

across regions and/or individuals. The larger the lesion, the greater the impact on the accuracy of segmentation. Especially for the learning-based methods, they automatically learn the texture, intensity and other characteristics of images. The similarity in intensity between the skull region and the lesion region could affect the segmentation accuracy. The proposed multi-view PSSNet performed better on NeoWML data than other CNN methods (as shown in Fig. 5). The multi-view PSSNet only had false positives at locations with subtle gaps in ground truth. This suggests that the effective spatial aware representation is important for the segmentation of regions with lesions. The deformable model-based approach was robust to brain lesions. However, it was difficult to achieve appropriate parameters for neonatal MRI, which led to over-segmentation and/or under-segmentation.

3.1.3. NeoBrainS12 dataset

Consistent with results on local datasets, the proposed multi-view PSSNet performed well on the NeoBrainS12 dataset. Note that the 3D U-net also held similar results. However, jagged edges of the 3D U-net had disadvantage of missing some brain tissues. Though the 2D U-net had comparatively smooth edges, there were obvious false negative detections on the brain tissue. The edges of masks by iBEAT and BET were wavy. This may be associated with the complexity of the cerebral surface. The robustness of Auto-Net and BSE was poor. Two extremes could be found on Fig. 6.

3.2. Quantitative analysis

Besides the visual evaluation, this study calculated the Dice ratio, sensitivity, and specificity for different methods. The results in Fig. 7 and Table 4 revealed that the proposed method had good performance, with the highest Dice ratios on all datasets ($P < 0.05$). The Dice ratio for 3D U-Net was also prominent. In addition, the Dice ratios of 2D U-net, iBEAT, and BET on the NeoWA dataset were higher than 90%. However, Dice ratios of these three methods on NeoBrainS12 were 87.13%, 84.28%, and 87.03%, respectively. The performance of iBEAT suggests that the combination of BET and BSE parameters by the meta-method could improve the reliability of segmentation. But the accuracy

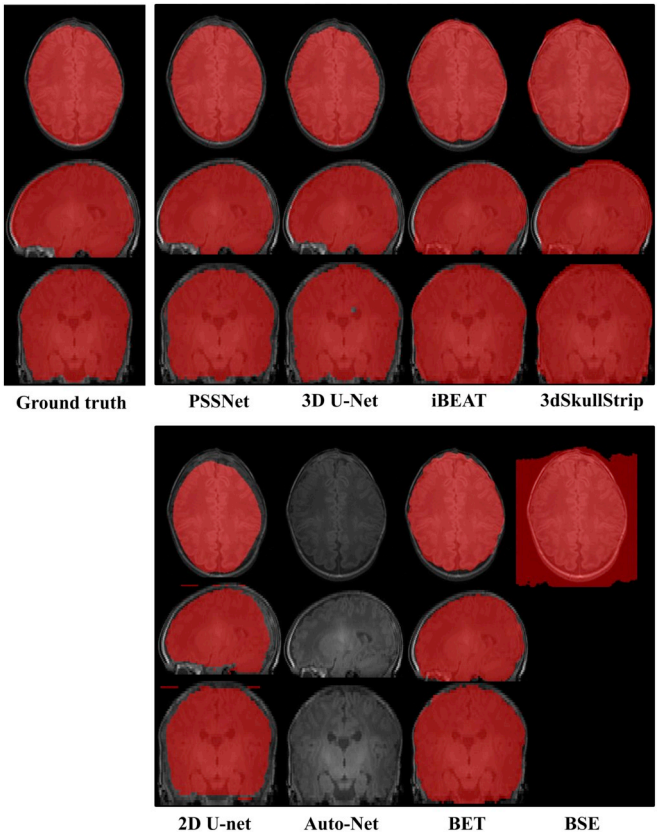

Fig. 6. Representative slices of the skull stripping by using different methods on the public dataset (NeoBrainS12).

of iBEAT is limited by the ability of the combined algorithms themselves. The Dice ratio of Auto-Net was obviously low, compared to other methods. The performances of Auto-Net varied across slices, which led to the low Dice ratio.

As shown in Fig. 7 and Table 4, the multi-view PSSNet, 2D and 3D U-Net demonstrated relative good balance between sensitivity and specificity. iBEAT, 3dSkullStrip, and BSE achieved high sensitivity at the expense of specificity. Their segmentation mask could cover most of the neonatal brain tissues, while it is likely to cause skull stripping errors (as shown in Fig. 4). Obvious low specificity of BSE was revealed. As a comparison, Auto-Net was specific but not sensitive. Similarly, the specificity of BET was above 95% on all datasets, but sensitivity was low. The specificity and sensitivity of U-net varied dependent on the image quality.

The absolute error, FP, and FN were shown in Fig. 8, where absolute error maps were the sum of FP and FN. It was normalized with the range between 0 and 1. Higher brightness indicates more severe errors. We displayed it from top to bottom based on the overall performance of different methods. The multi-view PSSNet and 3D U-Net were top two methods with few errors. It was difficult to suppress noise and artifacts by 2D U-net. It could be seen in Fig. 8 that the 2D U-net held obvious errors at the top and bottom of the head. Note that locations of errors on all the maps were spatial symmetrical. This suggests that skull stripping does not distinguish between the left and right hemispheres. And the methods always fail at the similar structural position, except for the Auto-Net.

3.3. Efficiency evaluation

The computation time was averaged across subjects for each method (Table 5). Results demonstrated that the computation time of the proposed method (8.59–9.30 s per subject) was acceptable during applications. It should be noted that the time for all methods was measured on the computer with Intel core i7-7700K CPU and NVIDIA Geforce GTX 1080ti GPU devices. The 2D U-net architecture was fast on all datasets. In addition, the splitting time of 3D U-Net was related to the amount of data. All algorithms that required the registration took a long time. Both deformable and parametric models were efficient.

4. Discussion

This study proposed a clinically feasible skull stripping method based on the pyramidal network for the neonatal T1WI. The main contribution of this current work is to adapt the PSPNet to serve for the skull striping on neonates. The adaptations of the conventional PSPNet were as follows: (1) adding the spatial information of raw feature maps by squeezing the channel information during the feature extraction; (2) increasing the receptive field and adding the boundary repair block instead of direct up-sampling; (3) obtaining the final mask through a fusion module on multi-view 2D slices. Compared with conventional methods, the proposed multi-view PSSNet achieved great overall performances.

4.1. Model-based vs. deep learning-based methods

Conventional skull stripping methods are always designed for adults, such as BSE [11], BET [9], and 3dSkullStrip [10]. They are not

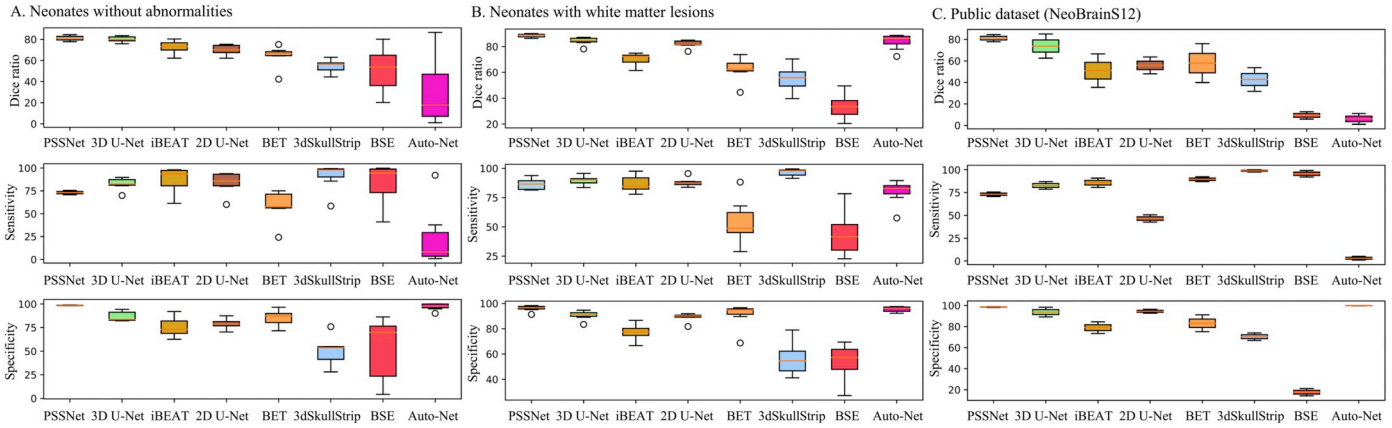

Fig. 7. Dice ratio, sensitivity, and specificity of different methods on different datasets. The values of metrics were transformed by an exponential function with the base of 100.

Table 4
Performances (Dice ratio, sensitivity, and specificity) of different skull stripping methods on three datasets (%). *P* values reveal the significances for the comparisons between the proposed method with other methods.

Method	Dice	*P* value	Sensitivity	*P* value	Specificity	*P* value
Neonates without abnormalities						
PSSNet	**96.33 (± 0.69)**	NA	97.02 (± 2.18)	NA	97.52 (± 0.96)	NA
3D U-Net	95.19 (± 0.75)	3.7×10^{-5}	95.83 (± 1.82)	5.1×10^{-6}	96.91 (± 1.29)	5.1×10^{-6}
iBEAT	93.02 (± 1.87)	1.1×10^{-5}	96.66 (± 4.06)	5.4×10^{-6}	93.76 (± 2.94)	5.1×10^{-6}
2D U-net	92.25 (± 1.56)	6.8×10^{-5}	95.98 (± 3.43)	5.1×10^{-6}	94.84 (± 1.48)	5.1×10^{-6}
BET	90.10 (± 4.03)	1.7×10^{-5}	87.37 (± 8.46)	5.8×10^{-6}	96.40 (± 2.29)	5.2×10^{-6}
3dSkullStrip	86.70 (± 2.54)	3.2×10^{-5}	97.54(± 4.22)	7.6×10^{-6}	84.06 (± 6.82)	5.1×10^{-6}
BSE	83.45 (± 10.34)	1.7×10^{-5}	94.98 (± 7.29)	5.5×10^{-6}	77.95 (± 23.96)	5.1×10^{-6}
Auto-Net	57.57 (± 34.90)	1.6×10^{-5}	48.38 (± 35.12)	3.3×10^{-5}	99.34 (± 0.83)	5.2×10^{-6}
Neonates with white matter lesions						
PSSNet	**97.33 (± 0.32)**	NA	96.79 (± 1.14)	NA	99.13 (± 0.50)	NA
3D U-Net	96.28 (± 0.81)	5.0×10^{-4}	97.58 (± 0.93)	5.6×10^{-6}	97.84 (± 0.89)	5.2×10^{-6}
iBEAT	92.26 (± 1.61)	3.0×10^{-4}	96.88 (± 1.74)	5.3×10^{-6}	94.32 (± 1.82)	5.3×10^{-6}
2D U-net	95.68 (± 0.81)	1.5×10^{-2}	97.21 (± 0.91)	6.1×10^{-6}	97.47 (± 0.85)	5.1×10^{-6}
BET	89.53 (± 3.51)	2.0×10^{-5}	85.64 (± 7.66)	5.3×10^{-6}	97.74 (± 2.65)	5.3×10^{-6}
3dSkullStrip	86.68 (± 4.36)	4.0×10^{-4}	99.17 (± 0.75)	5.3×10^{-6}	86.95 (± 4.95)	5.2×10^{-6}
BSE	75.43 (± 6.65)	1.8×10^{-3}	80.38 (± 9.27)	1.9×10^{-5}	85.68 (± 7.12)	5.2×10^{-6}
Auto-Net	96.12 (± 1.69)	2.6×10^{-5}	94.83 (± 3.24)	5.8×10^{-6}	98.97 (± 0.50)	5.3×10^{-6}
Public dataset (NeoBrainS12)						
PSSNet	**95.44 (± 1.24)**	NA	93.19 (± 1.02)	NA	99.68 (± 0.12)	NA
3D U-Net	93.14 (± 4.72)	1.9×10^{-3}	95.86 (± 1.55)	1.8×10^{-2}	98.57 (± 1.51)	1.6×10^{-2}
iBEAT	84.28 (± 9.69)	1.6×10^{-3}	96.62 (± 1.82)	1.8×10^{-2}	94.83 (± 2.16)	1.6×10^{-2}
2D U-net	87.13 (± 4.35)	3.6×10^{-3}	83.31 (± 2.64)	1.6×10^{-2}	98.79 (± 0.59)	1.6×10^{-2}
BET	87.03 (± 9.88)	1.2×10^{-3}	97.61 (± 0.90)	1.6×10^{-2}	95.89 (± 2.98)	1.6×10^{-2}
3dSkullStrip	80.76 (± 8.12)	2.9×10^{-3}	99.73 (± 0.38)	1.6×10^{-2}	92.35 (± 1.57)	1.6×10^{-2}
BSE	46.79 (± 12.01)	8.0×10^{-4}	99.00 (± 1.17)	1.6×10^{-2}	61.99 (± 6.09)	1.6×10^{-2}
Auto-Net	26.13 (± 36.95)	2.5×10^{-3}	17.78 (± 25.15)	4.0×10^{-2}	99.96 (± 0.05)	1.9×10^{-2}

Note: NA: not available.
Bold indicates the largest Dice ratio.

satisfactory on the neonatal T1WI. This was revealed in the result section of this work. The BSE mask produced excessive false positives, and was not sensitive to noise. In fact, BET performs well on the adult dataset [11]. However, it often faces problems of under-segmentation and/or over-segmentation. In addition, the problem with 3dSkullStrip is similar to BSE. Although the above methods are robust to brain lesions, results in this study suggest that they are not suitable for neonatal datasets. Software iBEAT has been designed for infants. And it performs well compared with other model-based methods. However, the process procedure is complex since each image requires preprocessing and registration.

Compared with model-based methods [9–11], deep learning models could automatically learn the data distribution and image features [8,16]. In this study, performances of deep learning methods were relatively better than model-based approaches. This is consistent with the previous findings on adults [8]. The key to the successful segmentation by deep learning methods is to have an effective global information representation strategy [32]. The 3D U-Net [16] performs well, which suggests the importance of the inference combined with spatial information. There is an exception in the Auto-Net algorithm [17]. The error of Auto-Net in Fig. 8 is very prominent. The cause may be that it relies on the output mask in each epoch. During training, the segmentation of the target slice in Auto-Net was guided by the golden standard and the segmentation result from the upper slice. Therefore, the errors were easily accumulated across slices, resulting in the instability of segmentation.

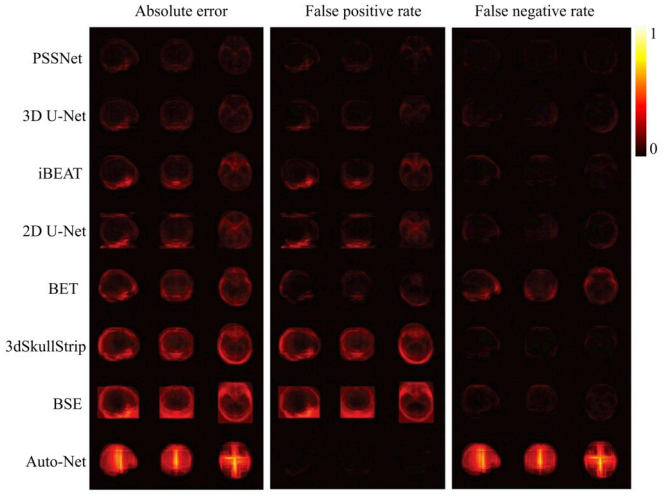

Fig. 8. Absolute error, false positive rate, false negative rate maps of the skull stripping for different methods.

4.2. Multi-view PSSNet vs. other deep learning-based methods

Global-level features are helpful to improve the skull stripping accuracy. The proposed multi-view PSSNet extracted global-level features through the pyramid pooling [21]. In this study, the 2D U-net, 3D U-Net, and Auto-Net did not perform well on the skull stripping when the dataset contained brain lesions. This may be due to the fact that the context information of the lesion area is not effectively used in the

Table 5
The averaged computation time per subject for different methods (seconds).

Method	NeoWA	NeoWML	NeoBrainS12
PSSNet	9.30	8.59	8.65
3D U-Net	2.30	2.27	0.68
iBEAT	9.45	9.03	20.24
2D U-net	2.12	2.10	2.12
BET	0.99	0.98	3.69
3dSkullStrip	41.63	40.25	70.52
BSE	2.93	2.68	20.00
Auto-Net	136.40	158.47	329.56

network. Different from the previous deep learning-based methods [10,17,18], this study fused 2D probability maps to get the final 3D segmentation. This network adaptively learned the slice-to-slice relationships in three orthorhombic directions. Through this strategy, the robustness of the algorithm to brain lesions was improved.

For deep learning-based methods, labels are essential for the model training. Though more or less differences in labels existed across individuals, the multi-view PSSNet achieved lowest error among the deep learning-based methods. The proposed method also complemented the semantic information needed for the segmentation. The accuracy could be maintained even when the sample size of the training data is small. Moreover, 3D convolution requires more parameters. This is the reason why it could not converge as fast as the PSSNet.

4.3. Efficiency considerations for applications

The important issue that should be considered during applications is the computation time and the required computational resources. In the conventional strategy of the skull stripping, registration is always required, which will take a long time. For deep learning-based methods, the 3D model holds better ability to acquire information between layers than the 2D model. The 3D strategy would achieve more higher accuracy than 2D strategy. However, the 3D version usually requires more computing resources (especially the computer memory) and more computation time [33]. Considering these factors, the proposed multi-view PSSNet chose a 2.5D framework to address this issue. The 2D model was used for the prediction firstly. Then the 3D model was used for the final mask fusion. This greatly reduced the memory requirement. We used CPU to read data and GPU to process data in parallel to optimize the procedure. Therefore, the efficiency of the algorithm was improved. In summary, the efficiency of the proposed method is acceptable for applications.

4.4. Limitations

Though the proposed method demonstrated great performances, there are still some limitations. Firstly, errors also existed in several cases because of noises and/or artifacts on images, although some noises had been filtered out by using a threshold method. Preprocessing by using the advanced method to reduce noises and artifacts may solve this problem [34]. Secondly, the training dataset only contained data of two sites in this work. The diversity of samples is limited. The generalization of this method may be improved in the future work based on multi-center studies. Thirdly, the accuracy of the image segmentation network depends largely on the superiority of the CNN feature extraction. Further work is needed to investigate performances of different backbone networks.

5. Conclusion

This study proposed and evaluated the multi-view PSSNet for the skull stripping, which used a hybrid framework. The proposed method was robust and efficient for the skull stripping on neonatal T1WI images. It could serve as a feasible preprocessing step during structure analyses of MRI on neonatal brains.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (81771810, 61571343, 61671339, 61432014, 61772402 and U1605252), the National High-Level Talents Special Support Program of China (CS31117200001), the National Key Research and Development Program of China (2016YFC0100300), the Fundamental Research Funds for the Central Universities (xjj2018265), the Fundamental Research Funds of the First Affiliated Hospital of Xi'an Jiaotong University (2017QN-09).

References

[1] Li G, Wang L, Yap PT, Wang F, Wu Z, Meng Y, et al. Computational neuroanatomy of baby brains: a review. Neuroimage 2019;185(1):906–25.
[2] Gilmore JH, Knickmeyer RC, Gao W. Imaging structural and functional brain development in early childhood. Nat Rev Neurosci 2018;19(3):123–37.
[3] Jeon TY, Kim JH, Yoo SY, Eo H, Kwon JY, Lee J, et al. Neurodevelopmental outcomes in preterm infants: comparison of infants with and without diffuse excessive high signal intensity on MR images at near-term-equivalent age. Radiology 2012;263(2):518–26.
[4] Zhang C, Adeli E, Wu Z, Li G, Lin W, Shen D. Infant brain development prediction with latent partial multi-view representation learning. IEEE Trans Med Imaging 2019;38(4):909–18.
[5] Jha SC, Xia K, Ahn M, Girault JB, Li G, Wang L, et al. Environmental influences on infant cortical thickness and surface area. Cereb Cortex 2019;29(3):1139–49.
[6] Dai Y, Shi F, Wang L, Wu G, Shen D. iBEAT: a toolbox for infant brain magnetic resonance image processing. Neuroinformatics 2013;11(2):211–25.
[7] Makropoulos A, Robinson EC, Schuh A, Wright R, Fitzgibbon S, Bozek J, et al. The developing human connectome project: a minimal processing pipeline for neonatal cortical surface reconstruction. Neuroimage 2018;173(1):88–112.
[8] Kleesiek J, Urban G, Hubert A, Schwarz D, Maier-Hein K, Bendszus M, et al. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. Neuroimage 2016;129:460–9.
[9] Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2010;17(3):143–55.
[10] CR W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29(3):162–73.
[11] Shattuck DW, Sandor-Leahy SR, Schaper KA. Magnetic resonance image tissue classification using a partial volume model. Neuroimage 2001;13(5):856–76.
[12] Shi F, Wang L, Dai Y, Gilmore JH, Lin W, Shen D. Label: pediatric brain extraction using learning-based meta-algorithm. Neuroimage 2012;62(3):1975–86.
[13] Alansary A, Ismail M, Soliman A, Khalifa F, Nitzken M, Elnakib A, et al. Infant brain extraction in T1-weighted MR images using BET and refinement using LCDG and MGRF models. IEEE J Biomed Health Inform 2016;20(3):925–35.
[14] Serag A, Blesa M, Moore EJ, Pataky R, Sparrow SA, Wilkinson AG, et al. Accurate Learning with Few Atlases (ALFA): an algorithm for MRI neonatal brain extraction and comparison with 11 publicly available methods. Sci Rep 2016;6(1):23470.
[15] Juan Eugenio I, Cheng-Yi L, Thompson PM, Zhuowen T. Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans Med Imaging 2011;30(9):1617–34.
[16] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention. 2016. p. 424–32.
[17] Ssm S, Erdogmus D, Gholipour A. Auto-context convolutional neural network (Auto-Net) for brain extraction in magnetic resonance imaging. IEEE Trans Med Imaging 2017;36(11):2319–30.
[18] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. 2015. [arXiv: 1505.04597].
[19] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. 2017. [arXiv: 1709.01507].
[20] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. IEEE conference on computer vision and pattern recognition. 2017. p. 6230–9.
[21] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 2015;37(9):1904–16.
[22] Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, Lecun Y. OverFeat: integrated recognition, localization and detection using convolutional networks. 2013. [arXiv: 1312.6229].
[23] Szegedy Christian, Vanhoucke Vincent, Ioffe Sergey, Shlens Jon, Wojna Z. Rethinking the inception architecture for computer vision. IEEE conference on computer vision and pattern recognition. 2016. p. 2818–26.
[24] Peng C, Zhang X, Yu G, Luo G, Sun J. Large kernel matters improve semantic segmentation by global convolutional network. IEEE conference on computer vision and pattern recognition. 2017. p. 1743–51.
[25] Li X, Gao J, Wang M, Zheng J, Li Y, Hui ES, et al. Characterization of extensive microstructural variations associated with punctate white matter lesions in preterm neonates. AJNR Am J Neuroradiol 2017;38(6):1228–34.
[26] Tortora D, Panara V, Mattei PA, Tartaro A, Salomone R, Domizio S, et al. Comparing 3T T1-weighted sequences in identifying hyperintense punctate lesions in preterm neonates. AJNR Am J Neuroradiol 2015;36(3):581–6.
[27] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. IEEE conference on computer vision and pattern recognition. 2015. p. 1–9.
[28] Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23(Suppl. 1):S208–19.

[29] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
[30] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. JMLR Proc 2010;9(1):249–56.
[31] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. arXiv:1512.03385.
[32] Nie D, Wang L, Adeli E, Lao C, Lin W, Shen D. 3-D fully convolutional networks for multimodal isointense infant brain image segmentation. IEEE Trans Cybern 2019;49(3):1123–36.
[33] Huo Y, Xu Z, Xiong Y, Aboud K, Parvathaneni P, Bao S, et al. 3D whole brain segmentation using spatially localized atlas network tiles. Neuroimage 2019;194:105–19.
[34] Jiang D, Dou W, Vosters L, Xu X, Sun Y, Tan T. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. Jpn J Radiol 2018;36(9):566–74.