Title: 3D-MASNet: 3D Mixed-scale Asymmetric Convolutional Segmentation Network for 6-month-old Infant Brain MR Images

Author affiliations: Zilong Zeng[1,2,3], Tengda Zhao[1,2,3*], Lianglong Sun[1,2,3], Yihe Zhang[1,2,3], Mingrui Xia[1,2,3], Xuhong Liao[4], Jiaying Zhang[1,2,3], Dinggang Shen[5,6], Li Wang[7], Yong He[1,2,3,8*]

[1] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

[2] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

[3] IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

[4] School of Systems Science, Beijing Normal University, Beijing 100875, China

[5] School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China

[6] Department of Research and Development, Shanghai Unitied Imaging Intelligence Co., Ltd., Shanghai 200030, China

[7] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA

[8] Chinese Institute for Brain Research, Beijing 102206, China

* Correspondence: yong.he@bnu.edu.cn (Y.H.) or tengdazhao@bnu.edu.cn (T.Z)

Manuscript information: 20 text pages, 6 figures, 6 tables.

Abstract

Precise segmentation of infant brain MR images into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) are essential for studying neuroanatomical hallmarks of early brain development. However, for 6-month-old infants, the extremely low-intensity contrast caused by inherent myelination hinders accurate tissue segmentation. Existing convolutional neural networks (CNNs) based segmentation models for this task generally employ single-scale symmetric convolutions, which are inefficient for encoding the isointense tissue boundaries in baby brain images. Here, we propose a 3D mixed-scale asymmetric convolutional segmentation network (3D-MASNet) framework for brain MR images of 6-month-old infants. We replaced the traditional convolutional layer of an existing to-be-trained network with a 3D mixed-scale convolution block consisting of asymmetric kernels (MixACB) during the training phase and then equivalently converted it into the original network. Five canonical CNN segmentation models were evaluated using both T1- and T2-weighted images of 23 6-month-old infants from iSeg-2019 datasets, which contained manual labels as ground truth. MixACB significantly enhanced the average accuracy of all five models and obtained the most considerable improvement in the fully convolutional network model (CC-3D-FCN) and the highest performance in the Dense U-Net model. This approach further obtained Dice coefficient accuracies of 0.931, 0.912, and 0.961 in GM, WM, and CSF, respectively, ranking first among 30 teams on the validation dataset of the iSeg-2019 Grand Challenge. Thus, the proposed 3D-MASNet can improve the accuracy of existing CNNs-based segmentation models as a plug-and-play solution that offers a promising technique for future infant brain MRI studies.

Keywords: infant brain segmentation, MRI, convolutional neural networks, mixed-scale convolution

1. Introduction

The accurate tissue segmentation of infant brain magnetic resonance (MR) images into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) are essential for researchers to chart the normal and abnormal early brain development of cortical regions, white matter connections, and wiring topologies (Cao et al., 2017; Hazlett et al., 2017; Wang et al., 2019a; Wen et al., 2019; Xu et al., 2019; Zhao et al., 2019). Notably, the tissue segmentation of 6-month-old infants is the biggest challenge in baby brain segmentation tasks due to the isointense phase in which the intensity distributions of GM and WM voxels become dramatically overlapped in the cortical regions (Fig. 1). The effective manual annotation, which is guided by longitudinal tracking of brain images with high tissue contrast in the latter children period (Wang et al., 2019b), is limited by the extremely high labor costs, the requirement of specialized expert knowledge (almost one week per image for an experienced neuroradiologist) and high inter- and intra-rater variations (Makropoulos et al., 2018). Developing fast, automatic, and accurate brain segmentation approaches is a crucial and ongoing goal for MR images of infants at 6 months of age (Sun et al., 2021; Wang et al., 2019b).

1.1. Convolutional neural networks (CNNs) based methods become the mainstream

In the past years, many efforts have been made for the segmentation task of 6-month-old infant brain MR images. Generally, emerging CNNs-based segmentation methods that exhibiting faster computational speed and higher accuracy than conventional atlas-based (Wang et al., 2014; Wang et al., 2012) or machine learning based methods (Sanroma et al., 2018; Wang et al., 2015; Wang et al., 2018a; Wang et al., 2014) become the mainstream. A typical example is that seven of the eight top teams in iSeg-2017 challenge has utilized CNNs to segment infant brain tissues.

Current CNNs-based approaches for infant brain segmentation are usually variants of canonical FCN (Long et al., 2015) and U-Net (Ronneberger et al., 2015) architecture. By adjusting or adding specific connectional pathways within or across neural layers on classical CNNs models, these approaches enhance the extraction and fusion of the semantic information in multimodal features to counteract the noisy and isointense tissues boundaries in 6-month-old infant brain images (Bui et al., 2019; Dolz et al., 2020; Dolz et al., 2019; Nie et al., 2019; Nie et al., 2016; Wang et al., 2018b; Wang et al., 2020; Zeng and Zheng, 2018; Zhang et al., 2015). Specifically, Bui et al. improved densely connected network (DenseNet) (Huang et al., 2017) by

concatenating fine and coarse feature maps from multiple densely connected blocks and won the iSeg-2017 competition (Bui et al., 2019). Dolz et al. proposed a semi-dense network by directly connecting all of the convolutional layers to the end of the network (Dolz et al., 2020) and further extended it into a HyperDenseNet by adding dense connections between multimodal network paths (Dolz et al., 2019). Similarly, Zeng et al. modified the classical U-Net network by constructing multi-encoder paths for each modality to effectively extract targeted high-level information (Zeng and Zheng, 2018). Wang et al. designed a global aggregation block in the U-Net model to consider global information in the decoder path of feature maps (Wang et al., 2020). Interestingly, inspired by the superiority of DenseNet and U-Net, the densely connected U-Net (DU-Net) model with a combination of these two types of networks was proposed for both tissue segmentation and autism diagnosis (Wang et al., 2018b).

1.2. Improvements from fine-grained convolution kernel designs are underestimated

Although great efforts have been made, the above CNN-based segmentation models have several limitations. First, the image appearance of 6-month-old infant brain MR images is quite noisy (Li et al., 2019; Mostapha and Styner, 2019) which makes the effective feature extraction difficult for the traditional convolution kernel design in previous works. Adopting enhanced convolution kernel designs (Ding et al., 2019; Li et al., 2020; Zhang et al., 2022) that emphasizes key features in the skeleton center of kernels may facilitate feature extractions throughout the network. Second, the voxel-wise fuzzy tissue boundaries in infant brain images are constrained by the anatomical morphology of gyrus at large spatial scales (Wang et al., 2018a). Although previous infant segmentation approaches try to fuse multi-scale features by skip-connections in variants of FCN and U-Net, they overlook capturing rich multi-scale features in kernel space, which contains more stable and homogeneous semantic information than features between layers (Fan et al., 2019). Third, all these studies focused on modifications of network layouts which need seasoned expertise experience, time-consuming hyperparameter tuning, and may also bring excessive graphics processing unit (GPU) burdens (Dolz et al., 2019; Wang et al., 2020) and architecture incompatibility. Recent CNN studies move eyes on building architecture-independent designs such as SE blocks (Hu et al., 2018), or automatically configuring models such as nnU-net (Isensee et al., 2021), which requires neither rare expert knowledge nor expensive manual interventions.

1.3. Our contribution

Our goal is to obtain a CNN-based building block for 6-month-old infant brain image segmentation which is 1) with fine-grained kernel designs to enhance the representation and abundance of features; 2) transplantable in up-to-date segmentation models in a plug-and-play way; 3) without much additional hyperparameter tuning or computational burden. To this end, we construct a 3D mixed-scale asymmetric segmentation network (3D-MASNet) framework by embedding a well-designed 3D mixed-scale asymmetric convolution block (MixACB) into existing segmentation CNNs for 6-month-old infant brain MR images (Fig. 2). The MixACB design is comprised by 1) four parallel 3D convolutional layers including a symmetric kernel ($d \times d \times d$) and three asymmetric 2D kernels ($1 \times d \times d$, $d \times 1 \times d$, $d \times d \times 1$) (Fig. 3A), respectively; 2) multiple groups on input feature maps with different kernel sizes (Fig. 3B) independently ; 3) parameter fusion for each MixACB after the training process to lower inference-time computations compare to the original network. We first evaluated the effectiveness of the MixACB on five canonical CNN networks using the iSeg-2019 training dataset. We next compared the performance of our method with that of top-4 approaches proposed in the MICCAI iSeg-2019 Grand Challenge on the iSeg-2019 validation dataset. The experimental results revealed that the MixACB significantly improved the segmentation accuracy of various CNNs, among which DU-Net (Wang et al., 2018b) with MixACB achieved the best-enhanced average performance and obtained the highest Dice coefficients of 0.931 in GM, 0.912 in WM, and 0.961 in CSF, ranking first in the iSeg-2019 Grand Challenge. All codes are publicly available at https://github.com/RicardoZiTseng/3D-MASNet.

2. Methods and Implementations

2.1.1. Mathematical formulation of basic 3D convolution

Consider a feature map $I \in \mathbb{R}^{U \times V \times S \times C}$ with a spatial resolution of $U \times V \times S$ as input and a feature map $O \in \mathbb{R}^{R \times T \times Q \times K}$ with a spatial resolution of $R \times T \times Q$ as output of a convolutional layer with a kernel size of $H \times W \times D$ and K filters. Then, each filter's kernel is denoted as $F \in \mathbb{R}^{H \times W \times D \times C}$, and the operation of the convolutional layer with a batch normalization (BN) layer can be formulated as follows:

$$O_{:,:,:,j} = \left(\sum_{k=1}^{C} I_{:,:,:,k} * F^{(j)}_{:,:,:,k} - \mu_j \right) \cdot \frac{\gamma_j}{\sigma_j} + \beta_j = \left(\sum_{k=1}^{C} I_{:,:,:,k} * \frac{\gamma_j}{\sigma_j} F^{(j)}_{:,:,:,k} \right) - \frac{\mu_j \gamma_j}{\sigma_j} + \beta_j \quad (1)$$

where $*$ is the 3D convolution operator, $I_{:,:,:,k}$ is the k^{th} channel of the input feature map I, $F^{(j)}_{:,:,:,k}$ is the k^{th} channel of the j^{th} filter's kernel, μ_j and σ_j are the channel-wise mean value and standard deviation value, respectively, γ_j and β_j are the scaling factor and bias term to restore the representation ability of the network, respectively.

2.1.2. Design of 3D asymmetric convolutions (3D-AC) during training and inference phases

3D-AC was designed behaving differently during training and inference phases (Fig. 3A). Concretely, for each kernel of each layer in the network during the training phase, a 3D-AC contains 4 parallel convolutional branches, namely one standard 3D convolution layer and three orthogonal 2D asymmetric convolutional layers ($1 \times d \times d$, $d \times 1 \times d$, $d \times d \times 1$) at kernel center for the enhancement of features along axial, sagittal and coronal directions, respectively. The input feature maps are fed into these 4 branches, and the outputs of these branches are summed to fuse the knowledge learned by these 4 independent branches. During the inference phase, 3D-AC contains one standard convolutional layer with equivalently fused kernel of the training-time 3D-AC (described in section 2.1.4), thus the input feature maps only need feed into this single branch which brining low inference computations.

2.1.3. Constructing MixACB by multiple 3D-ACs with varying kernel scales

To process the input feature map at different scales of detail, we propose the MixACB by mixing multiple 3D-ACs with different kernel sizes, as illustrated in Fig. 3B. Notably, since we used the 3D-AC to strength the core skeleton part of the convolutional kernel, thus the kernel size of 3D-AC must be odd, such as 3, 5, 7. Since directly adopting multiple 3D-ACs on all feature maps then concatenating outputs will dramatically increase the models' parameters and computations, we leverage the grouped convolution approach by splitting original input feature maps into groups and apply 3D-AC independently in each input feature map's group. Assume that we split the input feature maps into g groups of tensors such that their total number of channels is equal

to the original feature maps' channels: $C_1 + C_2 + ... + C_g = C$ with $C_1 > C_2 > ... > C_g$; similarly, the output feature maps also have g groups: $K_1 + K_2 + ... + K_g = K$ with $K_1 > K_2 > ... > K_g$. We denote $I^{<i>} \in \mathbb{R}^{U \times V \times S \times C_i}$ as the i^{th} group of input, $\widehat{O}^{<i>} \in \mathbb{R}^{R \times T \times Q \times K_i}$ as the MixACB's i^{th} group output, and $F'^{<i>} \in \mathbb{R}^{H_i \times W_i \times D_i \times C_i}$ as the equivalent kernel of the i^{th} group of the 3D-AC whose equivalent kernel size is $H_i \times W_i \times D_i$. Thus, we have following equations:

$$\widehat{O}^{<i>}_{:,:,:,j} = \left(\sum_{q=1}^{C_i} I^{<i>}_{:,:,:,q} * F_s'^{<i>(j)} \right) + b'^{<i>}_j \qquad (2)$$
$$s.t. \quad 1 \le i \le g, \quad 1 \le j \le K_i$$

The final output of MixACB is the concatenation of all groups' outputs:

$$\widehat{O} = concat\left(\widehat{O}^{<1>}, \widehat{O}^{<2>}, ..., \widehat{O}^{<g>} \right) \qquad (3)$$

In this paper, we only split the input and output feature maps into 2 groups, and define the mix ratio as the ratio between C_1 and C_2. For simplicity, the ratio between K_1 and K_2 is set to be equal to the mix ratio. We set a kernel size of 3 for the 1st group of 3D-AC and a kernel size of 5 for the 2nd group of 3D-AC, with the mix ratio set to 3:1. In this situation, the number of FLOPs (floating point operations) required for inference-time MixACB is $3^3 \times \frac{3}{4}C \times \frac{3}{4}K + 5^3 \times \frac{1}{4}C \times \frac{1}{4}K = 23CK$, which is smaller than that ($3^3 \times C \times K = 27CK$) for standard convolutions with kernel size of 3.

2.1.4. Equivalently fusing kernel of each 3D-AC inside MixACB

Once the training process of 3D-MASNet is completed, we equivalently fused the kernels of each 3D-AC inside the MixACB to retain the same output results as the original network. Due to the additivity of convolutional kernels, the kernels of 3D-AC's four branches can be fused to obtain an equivalent kernel in a 3D convolutional layer to produce the same output, which can be formulated as the following equation:

$$I * F + I * \widetilde{F} + I * \widehat{F} + I * \overline{F} = I * \left(F \oplus \widetilde{F} \oplus \widehat{F} \oplus \overline{F} \right) = I * F' \qquad (4)$$

where I is an input feature map, F, $\tilde{F}$, $\hat{F}$ and $\overline{F}$ are the 4 branches' kernels of 3D-AC. $\oplus$ is an elementwise operator that performs parameter addition on the corresponding positions, and F' is the equivalent fused kernel of the 4 branches' kernels.

Here, we took a kernel size of 3 as an example. We first fused the BN parameters into the convolutional kernel term and bias term following Eq. (1). Then, we further fused the four parallel kernels by adding the asymmetric kernels onto the skeletons of the cubic kernel. Formally, we denote $F'^{(j)}$ as the j^{th} filter at the $1\times3\times3$, $3\times1\times3$ and $3\times3\times1$ layer, respectively. Hence, we obtain the following formulas:

$$F'^{(j)} = \frac{\gamma_j}{\sigma_j} F^{(j)} \oplus \frac{\tilde{\gamma}_j}{\tilde{\sigma}_j} \tilde{F}^{(j)} \oplus \frac{\hat{\gamma}_j}{\hat{\sigma}_j} \hat{F}^{(j)} \oplus \frac{\overline{\gamma}_j}{\overline{\sigma}_j} \overline{F}^{(j)} \quad (5)$$

$$b'_j = -\frac{\mu_j \gamma_j}{\sigma_j} - \frac{\tilde{\mu}_j \tilde{\gamma}_j}{\tilde{\sigma}_j} - \frac{\hat{\mu}_j \hat{\gamma}_j}{\hat{\sigma}_j} - \frac{\overline{\mu}_j \overline{\gamma}_j}{\overline{\sigma}_j} + \beta_j + \tilde{\beta}_j + \hat{\beta}_j + \overline{\beta}_j \quad (6)$$

Then, we can write any output of j^{th} filter as:

$$O_{:,:,:,j} + \tilde{O}_{:,:,:,j} + \hat{O}_{:,:,:,j} + \overline{O}_{:,:,:,j} = \sum_{k=1}^{C} I_{:,:,:,k} * F'^{(j)}_{:,:,:,k} + b'_j \quad (7)$$

where $O_{:,:,:,j}$, $\tilde{O}_{:,:,:,j}$, $\hat{O}_{:,:,:,j}$ and $\overline{O}_{:,:,:,j}$ are the outputs of the original $3\times3\times3$, $1\times3\times3$, $3\times1\times3$ and $3\times3\times1$ branch, respectively.

2.2. Candidate CNNs for the evaluation of the MixACB on 6-month-old infant brain image segmentation

We choose five representative networks to evaluate the effectiveness of the 3D-MixACB in improving the segmentation performance, including BuiNet (Bui et al., 2019), 3D-UNet (Çiçek et al., 2016), convolution and concatenate 3D fully convolutional network (CC-3D-FCN) (Nie et al., 2019), non-local U-Net (NLU-Net) (Wang et al., 2020) and DU-Net (Wang et al., 2018b). Notably, these five networks are either variants of the U-type architecture (3D U-Net, NLU-Net, and DU-Net) or the FCN-type architecture (BuiNet and CC-3D-FCN) and encompass major CNN frameworks in infant brain segmentation. After replacing their original convolution layers with the 3D-MixACB design, we followed the training configurations set in the candidate CNN's release codes (Table 1) and adopted the Adam optimizer to update these models' parameters. Except for the CC-3D-FCN, which used the Xavier algorithm (Glorot and Bengio, 2010) to

initialize network weights, all other networks adopted the He initialization method (He et al., 2015). The configuration parameters are as follows:

(1) BuiNet adopted four dense blocks consisting of four $3\times3\times3$ convolutional layers for feature extraction. Transition blocks were applied between every two dense blocks to reduce the feature map resolutions. 3D up-sampling operations were used after each dense block for feature map recovery, and these upsampled features were concatenated together. (2) 3D-UNet has 4 levels of resolution, and each level adopts one $3\times3\times3$ convolution, which is followed by BN and a rectified linear unit (ReLU). The $2\times2\times2$ max pooling and the $2\times2\times2$ transposed convolution, each with a stride of 2, are employed for resolution reduction and recovery. Feature maps of the same level of both paths were summed. (3) CC-3D-FCN used 6 groups of $3\times3\times3$ convolutional layers for feature extraction, in which the $2\times2\times2$ max pooling with a stride of 2 was adopted between two groups of layers. The $1\times1\times1$ convolution with a stride of 1 was added between two groups with the same resolution for feature fusion. (4) DU-Net used 7 dense blocks to construct the encoder-decoder structure with 4 levels of resolution and leveraged transition down blocks and transition up blocks for down-sampling and up-sampling, respectively. Unlike the implementations in (Wang et al., 2018b), the bottleneck layer is introduced into the dense block to constrain the rapidly increasing number of feature maps, and the transition down block consisted of two $3\times3\times3$ convolutions, each followed by BN and ReLU. In addition, we used the $1\times1\times1$ convolution followed by a softmax activation function in the last layer. (5) NLU-Net leveraged five different kinds of residual blocks to form the U-type architecture with 3 levels of resolution. BN with the ReLU6 activation function was adopted before each $3\times3\times3$ convolution. The global aggregation block replaced the two convolutional layers of the input residual block to form the bottom residual block for the integration of global information.

We fed the same multimodal images into these five networks and employed the same inference strategy. We extracted overlapping patches of the same size as that used during the training phase. The overlapping step size had to be smaller than or equal to the patch length size to form the whole volume. Following the common practice in (Bui et al., 2019; Nie et al., 2019; Wang et al., 2018b; Wang et al., 2020), we set the step size to 8. Since the effect of the overlapping step size in the proposed framework remains unknown, we further evaluated it in section 3.3. Voxels inside the overlapping regions were averaged.

3. Experiments and Results

3.1. iSeg-2019 dataset and image preprocessing

Twenty-three isointense phase infant brain MRIs, including T1w and T2w images, were offered by the iSeg-2019 (http://iseg2019.web.unc.edu/) organizers from the pilot study of the Baby Connectome Project (BCP) (Howell et al., 2019). All the infants were term-born (40±1 weeks of gestational age) with an average scan age of 6.0±0.5 months. All experimental procedures were approved by the University of North Carolina at Chapel Hill and the University of Minnesota Institutional Review Boards. Detailed imaging parameters and preprocessing steps that were implemented are listed in (Sun et al., 2021). Before cropping the MR images into patches, we normalized the T1w and T2w images by subtracting the mean value and dividing by the standard deviation value.

The iSeg-2019 organizers offered the ground truth labels, which were obtained by a combination of initial automatic segmentation using the infant brain extraction and analysis toolbox (iBEAT) (Dai et al., 2013) on follow-up 24-month scans of the same baby and manual editing using ITK-SNAP (Yushkevich et al., 2006) under the guidance of an experienced neuroradiologist. The MR images of 10 infants with manual labels were provided for model training and validation. The images of 13 infants without labels were provided for model testing. The testing results were submitted to the iSeg-2019 organizers for quantitative measurements.

3.2. Evaluation metrics

We employed the Dice coefficient (DICE), modified Hausdorff distance (MHD) and average surface distance (ASD) to evaluate the model performance on segmenting 6-month-old infant brain MR images.

3.2.1. Dice coefficient

Let A and B be the manual labels and predictive labels, respectively. The DICE can be defined as:

$$DICE(A,B)=\frac{2|A\cap B|}{|A|+|B|} \quad (8)$$

where $|\cdot|$ denotes the number of elements of a point set. A higher DICE indicates a larger overlap between the manual and predictive segmentation areas.

3.2.2. Modified Hausdorff distance

Let C and D be the sets of voxels within the manual and predictive segmentation boundary, respectively. MHD can be defined as:

$$MHD(C,D) = \max\left\{h(C,D), h(D,C)\right\} \quad (9)$$

where $h(C,D) = \frac{1}{N_c}\sum_{c\in C} d(c,D)$, and $d(c,D) = \min_{d\in D}\|c-d\|$ with $\|\cdot\|$ representing the Euclidean distance. We follow the calculation described in (Wang et al., 2020) by computing the average MHD based on the three different vectorization directions to obtain a direction-independent evaluation metric. A smaller MHD coefficient indicates greater similarity between manual and predictive segmentation contours.

3.2.3. Average surface distance

The ASD is defined as:

$$ASD(C,D) = \frac{1}{2}\cdot\left(\frac{\sum_{v_i\in S_C}\min_{v_j\in S_D}\|v_i - v_j\|}{\sum_{v_i\in S_C} 1} + \frac{\sum_{v_j\in S_D}\min_{v_i\in S_C}\|v_j - v_i\|}{\sum_{v_j\in S_D} 1}\right) \quad (10)$$

where S_C and S_D represent the surface meshes of C and D, respectively. A smaller ASD coefficient indicates greater similarity between cortical surfaces reconstructed from manual and predictive segmentation maps.

3.3. Exploring the effectiveness of the MixACB

We performed several experiments to evaluate the effectiveness of the MixACB, including 1) ablation tests on five representative segmentation networks (section 2.2); 2) comparisons with state-of-the-art approaches in iSeg-2019; 3) component analysis of MixACB; and 4) validation of the impact of the overlapping step size.

3.3.1. Performance improvement on five representative CNN architectures

For a given network architecture without the MixACB design, we regarded it as the baseline model and further transformed it into a 3D-MASNet design. All pairs of the baseline models and their corresponding 3D-MASNet followed the training strategies described in Table 1. To

balance the training and testing sample sizes, we adopt a 2-fold cross-validation (one fold with five random selected participants for training and the left for testing) for model evaluation on the iSeg-2019 training dataset. Table 2 and Table 3 and Fig. 4 show that the performance of all the models was significantly improved across almost all tissue types in terms of the DICE and MHD, which demonstrates the effectiveness of the MixACB on a wide range of CNN layouts. Specifically, DU-Net with the MixACB achieved the highest average DICE of 0.928 and the lowest average MHD value of 0.436; CC-3D-FCN with the MixACB gained the most considerable DICE improvement and reached a higher average DICE than that attained by BuiNet, which was a champion solution in the MICCAI iSeg-2017 grand challenge, indicating that a simple network could reach excellent performance by advanced convolution designs. Fig. 5 further provides a visual segmentation comparison between networks with and without the MixACB. The MixACB could effectively correct misclassified voxels which are indicated by red squares.

3.3.2. Comparison with state-of-the-art methods on iSeg-2019

Since DU-Net, which was combined with MixACB, has achieved the highest accuracy among all candidate models, we compared it with methods developed by the 29 remaining teams that participated in the iSeg-2019 challenge. We employed a majority-voting strategy on 10 trained networks' outputs to improve the model generalization.

Table 4 reports the segmentation results achieved by our proposed method and those of other teams' methods that ranked in the top 4 on the validation dataset of the iSeg-2019. The mean DICE, MHD value and ASD value are presented for CSF, GM, and WM, representatively. Compared with other teams, our method yielded the highest DICE and lowest ASD value for the three brain tissues in the validation test of iSeg-2019, with comparable MHD values. The superior average value of the 3 types of brain tissues also indicates that our method has the best overall performance.

3.3.3. Component analysis of MixACB

We analyzed the effect of the mix ratio on model segmentation performance. Table 5 shows that segmentation accuracy reaches the highest value when the mix ratio is set to 3:1. Then, we further performed an ablation test to verify the effectiveness of each part of the proposed

MixACB, as shown in Table 6. The segmentation accuracy was improved with large variations when using different 3D-ACs alone. Moreover, when these 3D-ACs were mixed in scales for a MixACB design, the model was able to achieve the best performance in both DICE and MHD metrics.

3.3.4. Impact of overlapping step sizes

We further performed experiments to evaluate the effectiveness of the MixACB on overlapping step sizes, which controls the trade-off between accuracy and inference time. Based on 2-fold cross-validation, which has been done previously, we tested the overlapping impact when the step size is set to 4, 8, 16, and 32 on the DU-Net in the proposed 3D-MASNet framework. Fig. 6A and Fig. 6B present the changes in the segmentation performance in terms of DICE and MHD, respectively, for different overlapping step sizes. Fig. 6C presents the changes in the average number of inference patches for different overlapping step sizes. We found that a step size of 8 is a reasonable choice for achieving fast and accurate results.

4. Discussion

Instead of designing a new network architecture to segment the brain images of 6-month-old infants, we proposed a 3D-MASNet framework by replacing the standard convolutional layer with MixACB on an existing mature network and reduced model parameters and computations by equivalently performing fusion during the inference phase. The experimental results revealed that the MixACB significantly improved the performance of several CNNs by a considerable margin, in which DU-Net with MixACB showed the best average segmentation accuracy. The proposed framework obtained the highest average DICE of 0.935 and lowest ASD of 0.244, which ranked first among all 30 teams on the validation dataset of the iSeg-2019 Grand Challenge. In addition, the CC-3D-FCN model showed the largest improvement, which indicates that a simple model could achieve relatively better performance by implementing our convolution design.

4.1. Effectiveness of the MixACB on improving segmentation accuracy

The wide improvement in the segmentation accuracy of different models by the MixACB is derived from several aspects. First, the mixed-scale design of MixACB enables the network to collect multiscale details of local features with different receptive fields, facilitating the

integration of coarse-to-fine information inside the input patches at low-to-high semantic levels. Second, the isointense intensity distribution and heterogeneous tissue contrasts hamper effective feature extraction in baby brain images. The mix rate at 3:1 of feature maps for multi-scale kernel size enriches small receptive fields with enough detail features while enabling large receptive fields for capturing coarse global features. We also employed the 3D-AC inside the MixACB by adding multiple orthogonal 3D asymmetric convolutional layers to emphasize informative feature patterns in the central place (Fig. 3). Meanwhile, the asymmetric design is also shown robustness to image rotational distortion (Ding et al., 2019), which may help the network cope with the residual head motion of infants, even though these images have been linearly aligned to standard space. Third, the significant improvable performance of MixACB on various segmentation networks (Table 3) indicates that inter-layer architecture design may not be sufficient for multi-scale information fusion. Notably, besides providing better performance than the previous networks, 3D-MASNet is also more efficient than the baseline models, requiring fewer model parameters once its parameters were fused in the inference phase. For example, the baseline DU-Net's number of parameters is 2,492,795, while the corresponding 3D-MASNet's number of parameters is reduced to 2,341,141 during the inference phase.

4.2. Well-designed convolution operations

In recent years, researchers have begun to shift their interests from macro network layout to micro neuron units by studying specific convolution operators rather than touching the overall network. Previous works have proposed several advanced convolution operators by combining well-designed filters, such as pyramidal convolution (PyConv), dynamic group convolution (DGC), and asymmetric convolution block (ACB). PyConv employs multiple kernels in a pyramidal way to capture different levels of image details (Duta et al., 2020); DGC equips a feature selector for each group convolution conditioned on the input images to adaptively select input features (Su et al., 2020); ACB introduces asymmetry into 2D convolution to power up the representational power of the skeleton part of the kernel (Ding et al., 2019). These operators implanted into existing mature networks have achieved better performance on image classification or semantic segmentation tasks than in original networks. Due to the "easy-to-plug-in" property, this type of design could be conveniently adopted in various advanced CNNs and avoided high cost of network re-designing. However, these studies mainly concentrate on natural

image tasks, few were applied to infant brain segmentation tasks. Here, we design a novel convolution block by combining three basic characteristics including 3D spatial convolution, group convolution containing mixed-scale of kernel sizes, and asymmetry convolution. Due to blurred image appearance, large individual variation of brain morphology, and limited labeled sample sizes, we emphasize that effective and robust feature extraction, especially in a plug-and-play form, is essential for the infant brain segmentation task. Nevertheless, exhausting the combination of various convolution designs is beyond the scope of the article.

4.3. Limitations and future directions

The current study has several limitations. First, the patching approach may cause spatial consistency loss near boundaries. Although we adopted a small overlapping step size to relieve this issue, it is necessary to consider further integrating guidance from global information. Second, the small sample sizes of infant-specific datasets limit the generalizability of our method for babies across MRI scanners and acquisition protocols. Further validation on large samples is needed. Third, image indexes, such as the fractional anisotropy derived from diffusion MRI, contain rich white matter information (Liu et al., 2007), which could be beneficial for insufficient tissue contrast (Nie et al., 2019; Zhang et al., 2015). Importantly, determining how to leverage mixed-scale asymmetric convolution to enhance specific model features needs to be further explored. Fourth, we only explored the effectiveness of MixACB when input feature maps are split into 2 groups. Further combination configurations of convolutional kernel sizes and mix ratios are warranted.

5. Conclusion

In this paper, we proposed a 3D-MASNet framework for brain MR image segmentation of 6-month-old infants, which ranked first in the iSeg-2019 Grand Challenge. We demonstrated that the designed MixACB could easily migrate to various network architectures and enable performance improvement without extra inference-time computations. This work shows great adaptation potential for further improvement in future studies on brain segmentation.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (Nos. 31830034, 82021004 and 81801783), Changjiang Scholar Professorship Award (T2015027), and the China Postdoctoral Science Foundation (2020TQ0050 and 2022M710433).

Conflicting Interests

The authors have declared that no conflicting interests exist.

Data and Code Availability Statement

The 6-months-old infant brain MRI data was publicly offered by the iSeg-2019 (http://iseg2019.web.unc.edu/) organizers.

Codes developed for the proposed segmentation algorithm are released at https://github.com/RicardoZiTseng/3D-MASNet.

References

Bui, T.D., Shin, J., Moon, T., 2019. Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. Biomedical Signal Processing and Control 54, 101613.

Cao, M., Huang, H., He, Y., 2017. Developmental connectomics from infancy through early childhood. Trends in neurosciences 40, 494-506.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International conference on medical image computing and computer-assisted intervention. Springer, pp. 424-432.

Dai, Y., Shi, F., Wang, L., Wu, G., Shen, D., 2013. iBEAT: A Toolbox for Infant Brain Magnetic Resonance Image Processing. Neuroinformatics 11, 211-225.

Ding, X., Guo, Y., Ding, G., Han, J., 2019. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. Proceedings of the IEEE International Conference on Computer Vision, pp. 1911-1920.

Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., Ayed, I.B., 2020. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. Computerized Medical Imaging and Graphics 79, 101660.

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., Ayed, I.B., 2019. HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. IEEE Transactions on Medical Imaging 38, 1116-1126.

Duta, I.C., Liu, L., Zhu, F., Shao, L., 2020. Pyramidal convolution: Rethinking convolutional neural networks for visual recognition. arXiv preprint arXiv:2006.11538.

Fan, J., Cao, X., Yap, P.-T., Shen, D., 2019. BIRNet: Brain image registration using dual-supervised fully convolutional networks. Medical Image Analysis 54, 193-206.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, pp. 249-256.

Hazlett, H.C., Gu, H., Munsell, B.C., Kim, S.H., Styner, M., Wolff, J.J., Elison, J.T., Swanson, M.R., Zhu, H., Botteron, K.N., 2017. Early brain development in infants at high risk for autism spectrum disorder. Nature 542, 348-351.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV), 1026-1034.

Howell, B.R., Styner, M.A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., 2019. The UNC/UMN baby connectome project (BCP): an overview of the study design and protocol development. NeuroImage 185, 891-905.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods 18, 203-211.

Li, G., Wang, L., Yap, P.T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, J., Shi, F., Rekik, I., Lin, W., Shen, D., 2019. Computational neuroanatomy of baby brains: A review. Neuroimage 185, 906-925.

Li, R., Duan, C., Zheng, S., 2020. MACU-Net Semantic Segmentation from High-Resolution Remote Sensing Images. arXiv preprint arXiv:2007.13083.

Liu, T., Li, H., Wong, K., Tarokh, A., Guo, L., Wong, S.T., 2007. Brain tissue segmentation based on DTI data. Neuroimage 38, 114-123.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. IEEE transactions on pattern analysis and machine intelligence 39, 640-651.

Makropoulos, A., Counsell, S.J., Rueckert, D., 2018. A review on automatic fetal and neonatal brain MRI segmentation. Neuroimage 170, 231-248.

Mostapha, M., Styner, M., 2019. Role of deep learning in infant brain MRI analysis. Magn Reson Imaging 64, 171-189.

Nie, D., Wang, L., Adeli, E., Lao, C., Lin, W., Shen, D., 2019. 3-D Fully Convolutional Networks for Multimodal Isointense Infant Brain Image Segmentation. IEEE Trans Cybern 49, 1123-1136.

Nie, D., Wang, L., Gao, Y., Shen, D., 2016. Fully Convolutional Networks for Multi-Modality Isointense Infant Brain Image Segmentation. Proc IEEE Int Symp Biomed Imaging 2016, 1342-1345.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234-241.

Sanroma, G., Benkarim, O.M., Piella, G., Lekadir, K., Hahner, N., Eixarch, E., Ballester, M.A.G., 2018. Learning to combine complementary segmentation methods for fetal and 6-month infant brain MRI segmentation. Computerized Medical Imaging and Graphics 69, 52-59.

Su, Z., Fang, L., Kang, W., Hu, D., Pietikäinen, M., Liu, L., 2020. Dynamic group convolution for accelerating convolutional neural networks. European Conference on Computer Vision. Springer, pp. 138-155.

Sun, Y., Gao, K., Wu, Z., Li, G., Zong, X., Lei, Z., Wei, Y., Ma, J., Yang, X., Feng, X., 2021. Multi-site infant brain segmentation algorithms: The iSeg-2019 Challenge. IEEE Transactions on Medical Imaging 40, 1363-1376.

Wang, F., Lian, C., Wu, Z., Zhang, H., Li, T., Meng, Y., Wang, L., Lin, W., Shen, D., Li, G., 2019a. Developmental topography of cortical thickness during infancy. Proceedings of the National Academy of Sciences 116, 15855-15860.

Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J.H., Lin, W., Shen, D., 2015. LINKS: Learning-based multi-source IntegratioN frameworK for Segmentation of infant brain images. NeuroImage 108, 160-172.

Wang, L., Li, G., Adeli, E., Liu, M., Wu, Z., Meng, Y., Lin, W., Shen, D., 2018a. Anatomy-guided joint tissue segmentation and topological correction for 6-month infant brain MRI with risk of autism. Hum Brain Mapp 39, 2609-2623.

Wang, L., Li, G., Shi, F., Cao, X., Lian, C., Nie, D., Liu, M., Zhang, H., Li, G., Wu, Z., Lin, W., Shen, D., 2018b. Volume-Based Analysis of 6-Month-Old Infant Brain MRI for Autism Biomarker Identification and Early Diagnosis. Med Image Comput Comput Assist Interv 11072, 411-419.

Wang, L., Nie, D., Li, G., Puybareau, E., Dolz, J., Zhang, Q., Wang, F., Xia, J., Wu, Z., Chen, J., Thung, K.H., Bui, T.D., Shin, J., Zeng, G., Zheng, G., Fonov, V.S., Doyle, A., Xu, Y., Moeskops, P., Pluim, J.P.W., Desrosiers, C., Ayed, I.B., Sanroma, G., Benkarim, O.M.,

Casamitjana, A., Vilaplana, V., Lin, W., Li, G., Shen, D., 2019b. Benchmark on Automatic 6-month-old Infant Brain Segmentation Algorithms: The iSeg-2017 Challenge. IEEE Trans Med Imaging.

Wang, L., Shi, F., Gao, Y., Li, G., Gilmore, J.H., Lin, W., Shen, D., 2014. Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation. NeuroImage 89, 152-164.

Wang, L., Shi, F., Yap, P.-T., Gilmore, J.H., Lin, W., Shen, D., 2012. 4D multi-modality tissue segmentation of serial infant images. PLoS One 7, e44596.

Wang, Z., Zou, N., Shen, D., Ji, S., 2020. Non-Local U-Nets for Biomedical Image Segmentation. Proceedings of the AAAI Conference on Artificial Intelligence 34, 6315-6322.

Wen, X., Zhang, H., Li, G., Liu, M., Yin, W., Lin, W., Zhang, J., Shen, D., 2019. First-year development of modules and hubs in infant brain functional networks. NeuroImage 185, 222-235.

Xu, Y., Cao, M., Liao, X., Xia, M., Wang, X., Jeon, T., Ouyang, M., Chalak, L., Rollins, N., Huang, H., 2019. Development and emergence of individual variability in the functional connectivity architecture of the preterm human brain. Cerebral Cortex 29, 4208-4222.

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31, 1116-1128.

Zeng, G., Zheng, G., 2018. Multi-stream 3D FCN with multi-scale deep supervision for multi-modality isointense infant brain MR image segmentation. international symposium on biomedical imaging, pp. 136-140.

Zhang, J., Jiang, Z., Liu, D., Sun, Q., Hou, Y., Liu, B., 2022. 3D asymmetric expectation-maximization attention network for brain tumor segmentation. NMR in Biomedicine 35, e4657.

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D., 2015. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage 108, 214-224.

Zhao, T., Xu, Y., He, Y., 2019. Graph theoretical modeling of baby brain networks. NeuroImage 185, 711-727.

Figures and Tables

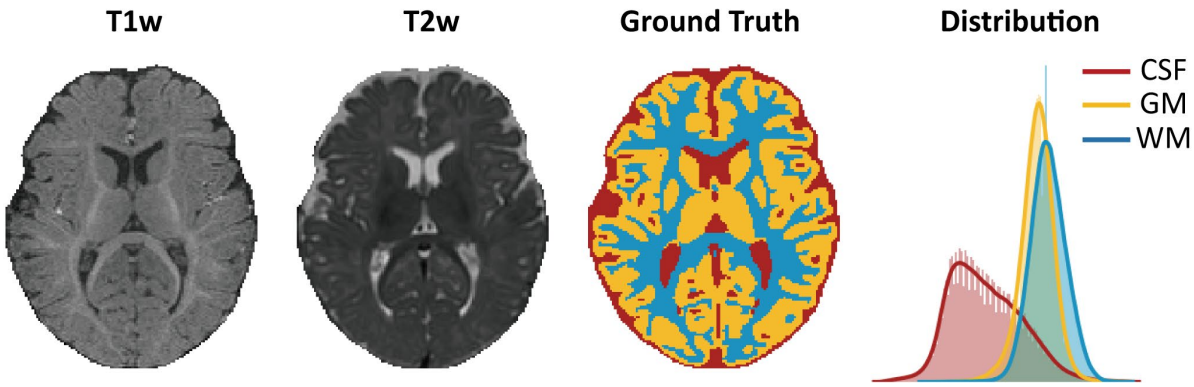

Figure 1. Data of a 6-month-old infant from the training set in iSeg-2019. The isointense brain appearance of an axial slice in T1-weighted (T1w) and T2-weighted (T2w) images. An axial view of the manual segmentation label (ground truth) and the corresponding brain tissue intensity distribution of the T1w image (distribution).

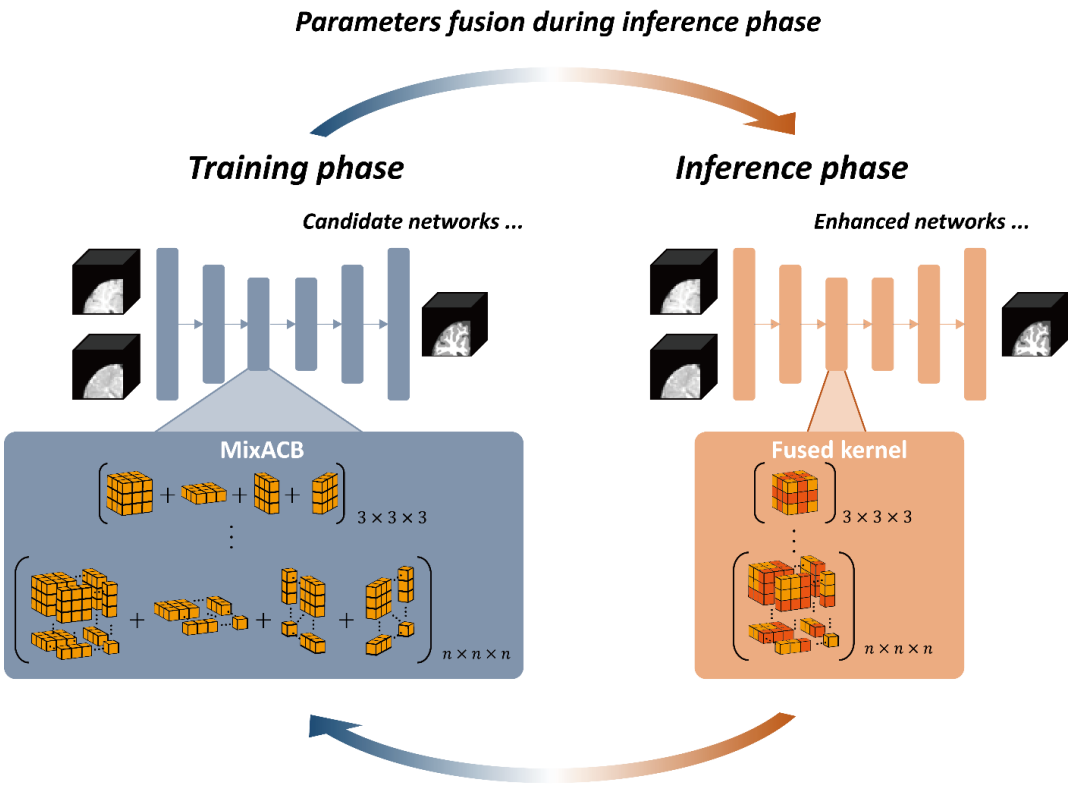

Figure 2. Overview of the 3D-MASNet framework. For a candidate network, we replace its traditional convolutional layers with MixACB during the training phase. Once the training process is complete, we fuse the parameters of MixACB to obtain an enhanced model containing fewer parameters after equivalent fusion.

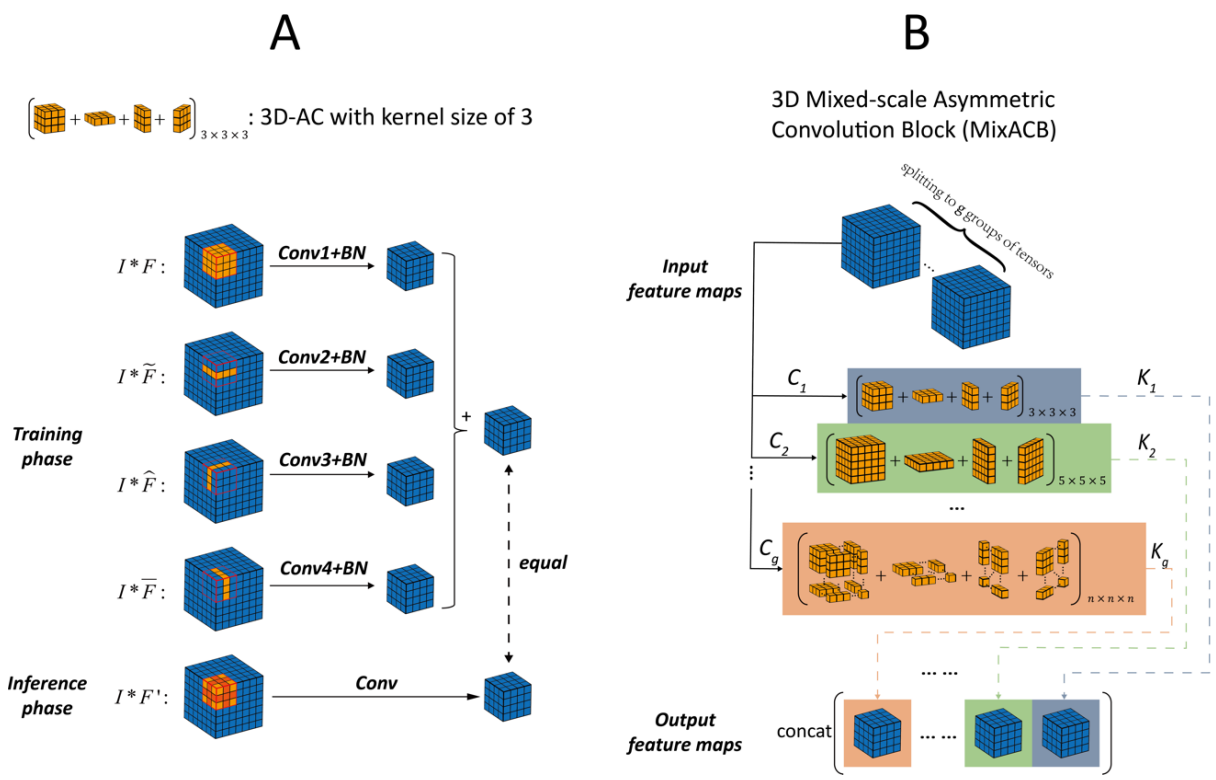

Figure 3. (A) Diagram of 3D-AC (taking a kernel size of 3 as an example), which has four convolutional layers during the training phase and one convolutional layer once kernel parameters have been fused during the inference phase. (B) Diagram of MixACB, which is composed of multiple 3D-ACs with different kernel sizes. MixACB splits input feature maps into several groups, applies asymmetric convolution on each group of feature maps, and then concatenates each group's output as the output feature maps.

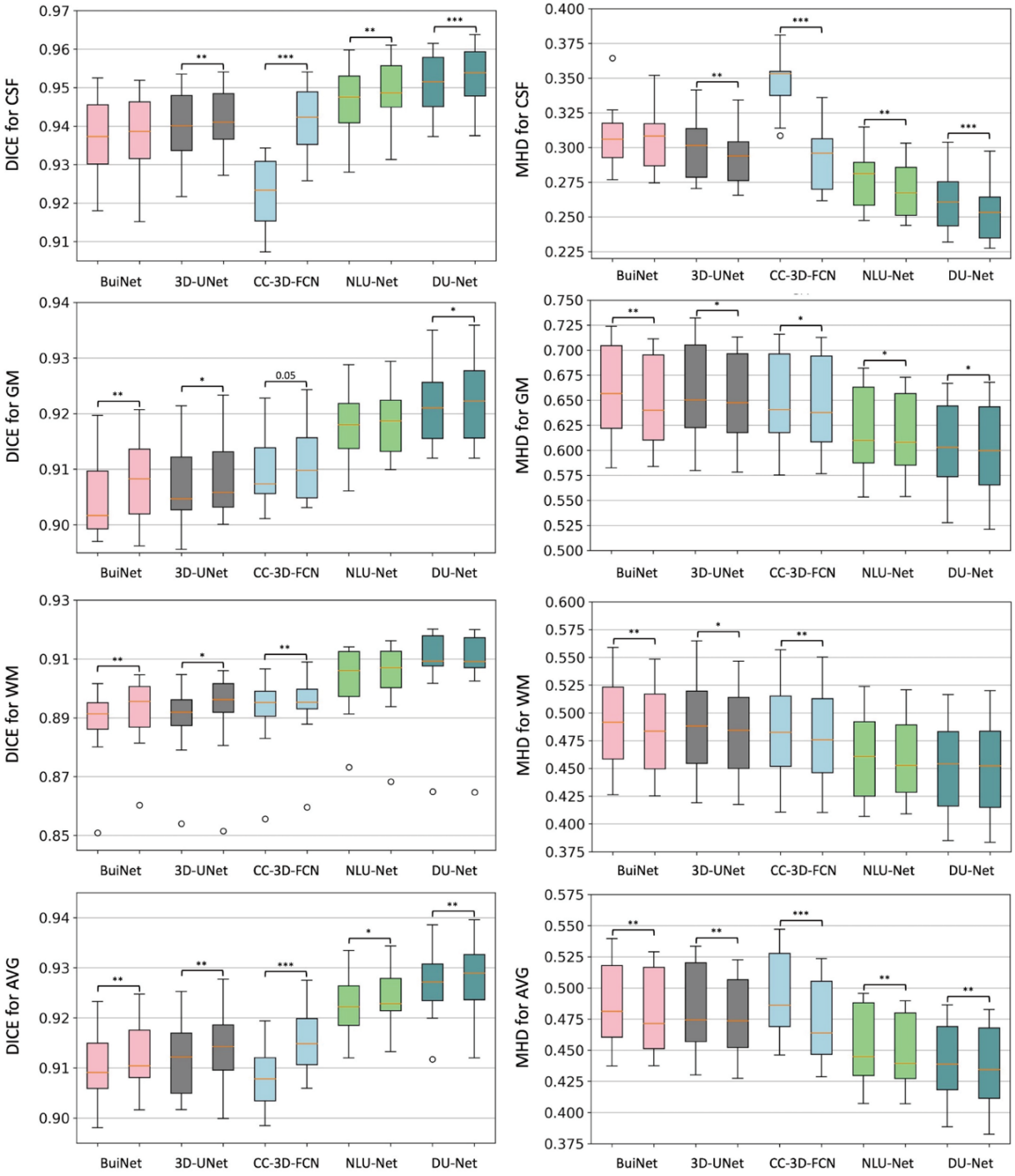

Figure 4. Box plot of the segmentation performance improvement on five candidate CNN architectures in the 3D-MASNet framework. The first column shows the measurement of DICE to represent the segmentation accuracy for each tissue type. The second column shows the results of MHD. In each subgraph, we use two neighbor box plots to represent a candidate model (first bar) and its corresponding 3D-MASNet (second bar). The significance of model comparison is evaluated by 2-fold cross-validation. "*" denotes that $0.01 \leq p < 0.05$, "**" denotes that $0.001 \leq p < 0.01$, and "***" denotes that $p < 0.001$.

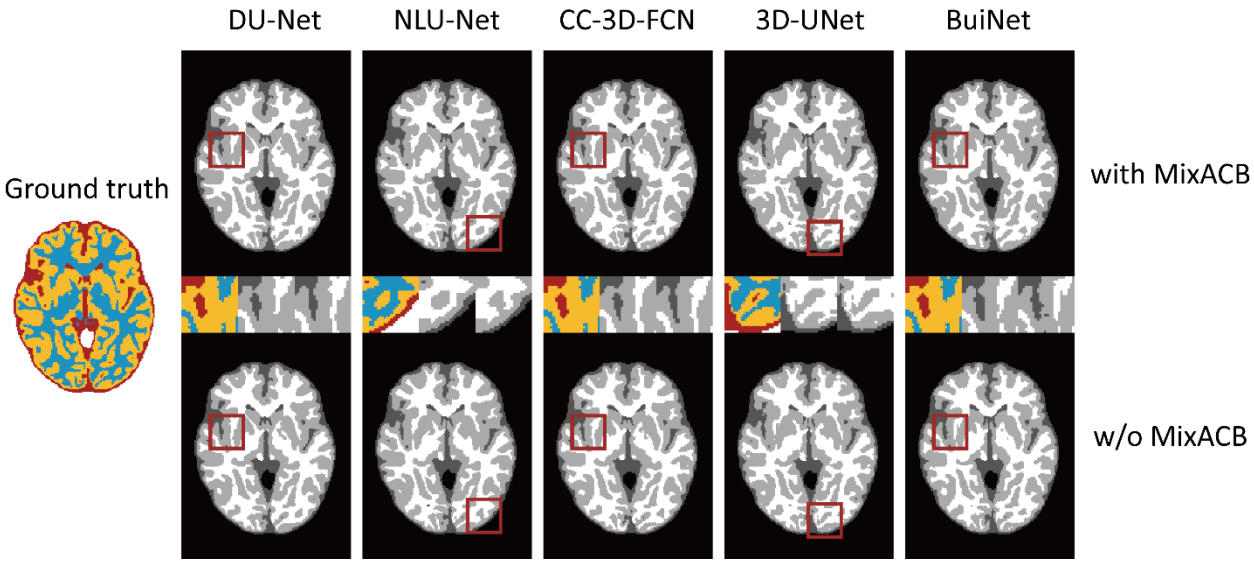

Figure 5. Visualization of the segmentation results on different models with (w) and without (w/o) the MixACB. The ground truth map is shown in color, and CNNs-based segmentation maps are shown in the gray scale. The regions in the red square are magnified in the middle row following an order from with MixACB to without MixACB.

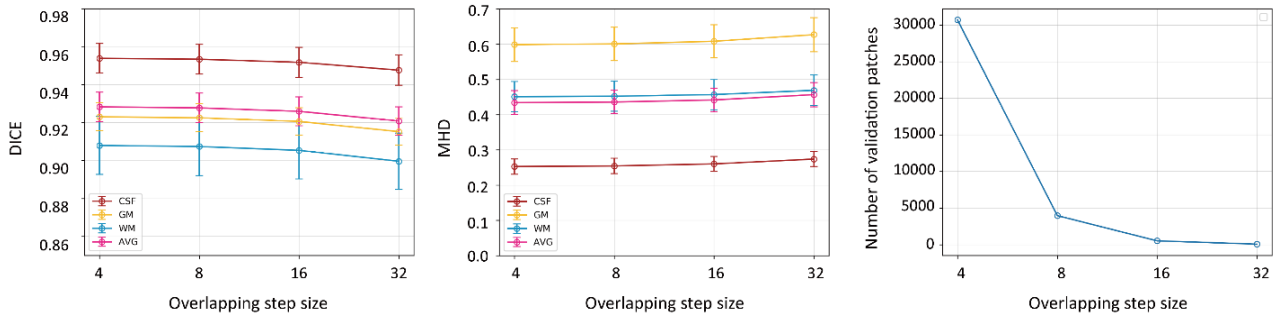

Figure 6. Changes in segmentation performance in terms of DICE (A) and MHD (B) with respect to different overlapping step sizes on 10 subjects during inference, where 2-fold cross-validation is used. (C) Changes of the average number of the 10 subjects' patches with respect to different overlapping step sizes during inference.

Table 1. Training strategy of each candidate network

Candidate Network	Training Batch Size	Training/Inference Patch Size	Learning Rate Schedule
BuiNet	4	64	Train for 20,000 iterations. The initial learning rate is set to 2e-4 and is decreased by a factor of 0.1 every 5000 iterations.
3D-UNet	10	32	Train for 80 epochs for a total of 5000 patches that are randomly extracted per epoch. The learning rate is decreased every 20 epochs and is set to 3e-4, 1e-4, 1e-5 and 1e-6. Train for 80 epochs for a total of 5000 patches that are randomly extracted per epoch. The learning rate is decreased every 20 epochs and is set to 3e-4, 1e-4, 1e-5 and 1e-6.
CC-3D-FCN	10	32	The same as 3D-UNet.
NLU-Net	5	32	Train for 80 epochs for a total of 5000 patches that are randomly extracted per epoch. The learning rate is set to 1e-3.
DU-Net	16	32	The cosine annealing strategy with a maximum learning rate of 3e-4 and a minimum learning rate of 1e-6 is adopted. The model is trained for 500 epochs and a total of 1000 patches are randomly extracted at each epoch.

Table 2. Ablation study performed by comparing the segmentation accuracy between different models and their corresponding 3D-MASNet in terms of DICE by 2-fold cross validation.

Network	CSF		GM		WM		Avg	
	Baseline	MixACB	Baseline	MixACB	Baseline	MixACB	Baseline	MixACB
BuiNet	0.938±0.010	0.938±0.011	0.905±0.007	0.908*±0.007	0.888±0.014	0.892*±0.013	0.910±0.007	0.912*±0.007
3D-UNet	0.940±0.010	0.942*±0.008	0.907±0.007	0.909*±0.007	0.889±0.014	0.892*±0.015	0.912±0.007	0.914*±0.008
CC-3D-FCN	0.923±0.010	0.942*±0.008	0.910±0.006	0.911±0.007	0.892±0.013	0.894*±0.013	0.908±0.006	0.915*±0.006
NLU-Net	0.947±0.009	0.949*±0.008	0.918±0.007	0.919±0.006	0.903±0.012	0.904±0.014	0.922±0.006	0.924*±0.006
DU-Net	0.951±0.008	**0.953***±0.008	0.922±0.007	**0.923***±0.007	0.907±0.015	**0.907**±0.015	0.927±0.007	**0.928***±0.008

Note that the best values are highlighted in bold font. “Baseline” denotes that the corresponding model adopted the standard convolutional operation; “MixACB” denotes that the corresponding model was transformed into 3D-MASNet; “*” denotes that the difference between baseline and 3D-MASNet is statistically significant ($p<0.05$).

Table 3. Ablation study performed by comparing the segmentation accuracy between different models and their corresponding 3D-MASNet in terms of MHD by 2-fold cross validation.

Network	CSF		GM		WM		Avg	
	Baseline	MixACB	Baseline	MixACB	Baseline	MixACB	Baseline	MixACB
BuiNet	0.308±0.024	0.307±0.023	0.659±0.048	0.649*±0.045	0.493±0.043	0.485*±0.042	0.487±0.035	0.480*±0.034
3D-UNet	0.299±0.022	0.293*±0.020	0.658±0.050	0.651*±0.046	0.490±0.046	0.485*±0.042	0.483±0.036	0.476*±0.033
CC-3D-FCN	0.348±0.022	0.292*±0.023	0.649±0.047	0.645*±0.048	0.485±0.046	0.480*±0.046	0.494±0.034	0.473*±0.034
NLU-Net	0.278±0.022	0.270*±0.020	0.619±0.043	0.615*±0.040	0.461±0.040	0.460±0.037	0.453±0.032	0.448*±0.030
DU-Net	0.261±0.021	**0.254***±0.022	0.605±0.046	**0.601***±0.047	0.452±0.041	**0.452**±0.043	0.439±0.032	**0.436***±0.034

Note that the best values are highlighted in bold font. “Baseline” denotes that the corresponding model adopted the standard convolutional operation; “MixACB” denotes that the corresponding model was transformed into 3D-MASNet; “*” denotes that the difference between baseline and 3D-MASNet is statistically significant ($p<0.05$).

Table 4. Comparison of segmentation performance of the proposed method and the methods of the top-4 ranked teams on the 13 validation infant MRI images of iSeg-2019.

Method (Top 5)	CSF			GM			WM			AVG		
	DICE	MHD	ASD	DICE	MHD	ASD	DICE	MHD	ASD	DICE	MHD	ASD
Brain_Tech	0.961	8.873	0.108	0.928	5.724	0.300	0.911	7.114	0.347	0.933	7.237	0.252
FightAutism	0.960	9.233	0.110	0.929	5.678	0.300	0.911	**6.678**	0.341	0.933	7.196	0.250
OxfordIBME	0.960	**8.560**	0.112	0.927	**5.495**	0.307	0.907	6.759	0.353	0.931	**6.938**	0.257
QL111111	0.959	9.484	0.114	0.926	5.601	0.307	0.908	7.028	0.353	0.931	7.371	0.258
Proposed	**0.961**	9.293	**0.107**	**0.931**	5.741	**0.292**	**0.912**	7.111	**0.332**	**0.935**	7.382	**0.244**

Note that the best values are highlighted in bold font.

Table 5. Ablation study performed by comparing the segmentation accuracy in different mix ratios by 2-fold cross validation.

Mix Ratio	CSF		GM		WM		AVG	
	DICE	MHD	DICE	MHD	DICE	MHD	DICE	MHD
1:0	0.952±0.010	0.261±0.024	0.922±0.008	0.604±0.045	0.906±0.014	0.453±0.041	0.927±0.008	0.440±0.033
1:1	0.953±0.008	0.258±0.023	0.921±0.007	0.605±0.045	0.905±0.013	0.455±0.040	0.926±0.007	0.439±0.033
3:1 (proposed)	**0.953**±0.008	**0.254**±0.022	**0.923**±0.007	**0.601**±0.047	**0.907**±0.015	**0.452**±0.043	**0.928**±0.008	**0.436**±0.034
5:1	0.953±0.009	0.257±0.025	0.922±0.008	0.601±0.047	0.907±0.015	0.452±0.042	0.926±0.008	0.437±0.034

Note that the best values are highlighted in bold font.

Table 6. Component analysis of MixACB by 2-fold cross validation.

	CSF		GM		WM		AVG	
	DICE	MHD	DICE	MHD	DICE	MHD	DICE	MHD
CONV_3	0.951±0.008	0.261±0.021	0.922±0.007	0.605±0.046	0.907±0.015	0.452±0.041	0.927±0.007	0.439±0.032
AC_3	0.952±0.010	0.261±0.024	0.922±0.008	0.604±0.045	0.906±0.014	0.453±0.041	0.927±0.008	0.440±0.033
CONV_5	0.947±0.012	0.276±0.023	0.918±0.008	0.619±0.047	0.903±0.016	0.463±0.043	0.922±0.008	0.453±0.034
AC_5	0.952±0.008	0.261±0.022	0.920±0.008	0.610±0.046	0.904±0.016	0.460±0.043	0.925±0.008	0.443±0.033
MixACB	**0.953**±0.008	**0.254**±0.022	**0.923**±0.007	**0.601**±0.047	**0.907**±0.015	**0.452**±0.043	**0.928**±0.008	**0.436**±0.034

Note that the best values are highlighted in bold font. “CONV_3” denotes that the 3D convolution with a kernel size of 3; “AC_3” denotes that the 3D-AC with a kernel size of 3; “CONV_5” denotes that the 3D convolution with a kernel size of 5; “AC_5” denotes that the 3D-AC with a kernel size of 5.