

An Infant Brain MR Image Segmentation Method with Attention Mechanism, Multi-Resolution Deep Supervision and Class Weight Tuning Strategy

Mohamad Reza Akbarzadeh Lari (PhD Candidate)^{1*},
Habibollah Danyali (PhD)^{1*}, Kamran Kazemi (PhD)¹

ABSTRACT

Background: Infant brain tissue segmentation is critical for understanding early brain development and detecting neurodevelopmental abnormalities. However, accurate segmentation remains challenging due to partial-volume effects, imaging noise, and low contrast between tissues in isointense regions.

Objective: This study aimed to automatically segment infant brain Magnetic Resonance Imaging (MRI) images into white matter, gray matter, and cerebrospinal fluid.

Material and Methods: In this analytical study, we addressed the challenges of infant brain tissue segmentation. We proposed a deep, lightweight dense backbone architecture incorporating transition blocks. The backbone was enhanced with a spatial-channel attention mechanism to capture global dependencies and emphasize task-relevant features, as well as multi-resolution deep supervision to mitigate vanishing gradients and reduce overfitting. Moreover, we used class weighting strategy to deal with class imbalance and low contrast between brain tissues.

Results: Our method attains Dice similarity scores of 91.80%, 92.07% and 94.58% for white matter, gray matter, and cerebrospinal fluid, respectively. Quantitative and qualitative evaluations show that the proposed approach accurately segments infant brain MR images and reduces errors in isointense regions.

Conclusion: The results demonstrate that the proposed method precisely delineates infant brain tissues and effectively overcomes common segmentation challenges.

Keywords

Infant Brain Segmentation; Magnetic Resonance Imaging; Deep Learning; Convolutional Neural Networks; Attention Mechanism

Introduction

Structural Magnetic Resonance Imaging (MRI) is a noninvasive imaging technique widely used to examine brain structures and tissues. Extensive longitudinal and cross-sectional studies have used MRI to track brain changes across the lifespan, monitor treatment effects, and identify biomarkers that distinguish healthy controls from patients [1]. To this end, brain tissue segmentation to the white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) is a main primary step to analysis brain structures. However, manual segmentation is time-consuming and susceptible to inter- and intra-rater errors,

¹Department of Electrical Engineering, Shiraz University of Technology, Shiraz, Iran

*Corresponding author:
Habibollah Danyali
Department of Electrical Engineering, Shiraz University of Technology, Shiraz, Iran
E-mail: danyali@sutech.ac.ir

Received: 13 December 2025
Accepted: 30 May 2026

and differences in segmentation protocols can introduce additional inconsistencies. Consequently, accurate automatic segmentation remains a critical and continuous goal in brain MRI research.

The first year of life is the most dynamic period of postnatal brain development [2], and accurate infant brain tissue segmentation is essential for understanding early growth. Reliable brain tissue segmentation enables the extraction of clinically meaningful biomarkers, including cortical thickness, inner and outer cortical surfaces, and gyrification patterns, which are important for the early detection of neurodevelopmental disorders such as schizophrenia and autism [3-6]. Consequently, accurate infant brain segmentation is critical, as even minor segmentation errors can substantially affect downstream analyses. For example, in a $1 \times 1 \times 1$ mm³ MRI scan, a single voxel misclassification can result in a 1 mm error in cortical thickness estimation, which is particularly significant in applications such as autism spectrum disorder prediction [6].

Infant brain MRI segmentation is particularly challenging during the isointense phase because of the low contrast between tissue types. In early infancy, GM typically appears brighter than WM on T1-weighted images. During the first year, WM signal increases as myelination progresses, and by around 12 months the contrast resembles the adult pattern. Between roughly 6 and 9 months (a.k.a. isointense phase), WM and GM exhibit similar intensities, producing low-contrast images that make automated tissue discrimination especially difficult [7]. Besides reduced tissue contrast, severe noise and partial volume make the accurate tissue segmentation challenging.

Among prior works on infant brain segmentation, atlas-based and deformable-model methods were historically the most popular. Atlas-based approaches register a labeled atlas to a target image and propagate manual labels to that target. While effective when registration is accurate, these methods are highly

sensitive to registration errors [8]. Deformable-model approaches, by contrast, employ shape priors and iteratively deform a template until it fits the target anatomy. However, they typically struggle with low-contrast regions and can fail when intensity differences between tissues are small [8]. In recent years, deep learning methods have shown promising results and become the prevailing approach. Among these methods, 3D-SkipDenseSeg employed densely connected network with skip connections for infant tissue segmentation [9]. They replaced the successive decoder blocks by upsampling operators and reduced the network size while preserving multi-scale feature fusion. In a similar manner, Dolz et al. introduced ensemble of semi-dense networks for infant brain parcellation [8]. They transferred all intermediate feature map of the encoder network to the fully connected layers, which acts as a decoder. ADU-Net designed two cascade anatomy-guided image segmentation networks [6]. The first network is designed to segment (WM_GM) vs, CSF. The output of this network passed through a level-set function and then in companion with MR images, input to the second network, which is designed specially to delineate WM from GM. In this setting the second network concentrate the low-contrast GM/WM with the anatomy-guided additional input. However, they designed two separate networks and used hand-crafted level-set construction, which increase the complexity and time. In a similar manner, Hashemi et al. designed a mutually exclusive dense architecture, which the network concentrate to segment WM and CSF [10]. Alternatively, in this work, we adopt simple but effective strategy to specifically address class imbalance and to improve discrimination of low-contrast WM/GM regions.

While effective, these methods still have limitations in modeling global dependencies. To handle the problem of local operation of convolutions, Wang et al. [11] used non-local Unet to exploit long-range dependencies

among feature pixels. They proposed a global aggregation block based on the self-attention operator, enabling to aggregate global information without a deep encoder. Similarly, Luan et al. [12] used a rotation transformer at the end of the encoder to calculate the similarity between two patches on the basis of rotation, resulting accurate tissue segmentation. However, these methods are complex and need high computational resource and rely on matrix multiplication for computing self-attention and transformer coefficients which may cause instability of the attention weight [13]. In this study, inspired by recent attention mechanisms [14], we use global pooling and bottlenecks to capture global dependencies, avoiding high computational and unstable matrix multiplications.

Recently, several 2D deep learning networks [15-17] have been proposed to reduce computational complexity. RubikNet [15] rotates the axes of an input slices across the axial, sagittal, and coronal plane, enabling 2-D convolution kernels to extract features from each plane simultaneously. Karimi et al. [16] concatenated three successive 2D slices to construct a 2D color image and then fed it to a 2D Unet for infant tissue delineation. Dénes-Fazakas et al. [17] proposed a (2+1)D convolutional approach, in which a 3D $n \times n \times n$ convolution operation is decoupled into two successive $n \times n \times 1$ and $1 \times n \times n$ operations. Although 2D methods reduce computational complexity, they may not fully exploit the rich contextual information inherent in 3D medical images. To address this limitation, we propose a fully 3D lightweight densely connected convolutional network. Despite employing 3D operations, the network incorporates bottleneck layers to reduce the number of feature-map channels and lightweight attention modules, resulting in a compact architecture with fewer than one million trainable parameters. Moreover, we exploit multi-resolution deep supervision to ease the training process and alleviate the vanishing gradients, and channel-spatial attention

mechanism to extract global dependencies and more meaningful feature representations. We also use a weight tuning cross entropy loss function to simultaneously handle the class imbalance and low-contrast between infant brain tissues.

Material and Methods

Our analytical study provides a novel method to deal with infant brain segmentation challenges. This section describes the dataset, the preprocessing procedures used to prepare the input data, and the proposed segmentation architecture.

Data description

In this study, the dataset is the infant segmentation iSeg challenge [7]. The MR images were randomly selected from Baby Connectome Project (BCP, <http://babyconnectomeproject.org>) and acquired on a Siemens head-only 3T scanners with a circular polarized head coil. Infants were scanned while asleep and unsedated, with ear protection in place and their heads secured using a vacuum-fixation device. The average age of infants is 6.0 ± 0.5 month. T1-weighted and T2-weighted MR images were obtained with 144 sagittal slices and 64 axial slices, respectively. Acquisition parameters are presented in Table 1.

Preprocessing performed by the organizers include linearly aligning T2-w onto their

Table 1: Parameters of infant brain MRI segmentation (iSeg) dataset.

MR image	Parameter	Value
T1-weighted	TR/TE	1900/4.38 ms
	Angle	7°
	Resolution	1×1×1 mm ³
T2-weighted	TR/TE	7380/119 ms
	Angle	150°
	Resolution	1.25×1.25×1.95 mm ³

MRI: Magnetic Resonance Imaging, TR: Repetition Time, TE: Echo Time

corresponding T1-w, skull stripping, intensity inhomogeneity correction, and cerebellum and brain stem removing. All images were resampled into an isotropic $1 \times 1 \times 1 \text{ mm}^3$ resolution with $144 \times 192 \times 256$ pixels³. The ground-truth from 10 infants (5 male and 5 female) were generated by the organizers and considered as a training dataset. The initial segmentation was performed by using follow-up 2-month scans with iBEAT tool. Then the manual editing, including filling the holes or cutting the handles was performed by the radiologist to correct the errors by using ITK-SNAP tool. Figure 1 displays a slice in the T1 and T2 scans and the corresponding intensity distributions from T1. It is evident that GM and WM regions demonstrate similar intensity distributions.

Proposed method

Our proposed network architecture is depicted in Figure 2. It consists of backbone, multi-resolution deep supervision and attention mechanism.

1. Backbone

Each layer at the encoder and decoder consists of dense block [18, 19] to extract semantic rich features. A dense block is defined as $y_k = H_k[y_0, y_1, \dots, y_{d-1}]$, where y_i is the output of the i^{th} layer, H_k is the k^{th} transition layer, and $[y_0, y_1, \dots, y_{d-1}]$ denotes the concatenation of the output featuremaps of all preceding layers. We set $d=4$, resulting in four convolutional

layers and one transition layer within each dense block. Consequently, each dense block comprises a total of five convolutional layers. At the encoder path, the transition down block consists of bottleneck block, and max-pooling operation is employed. We design the bottleneck to keep the channel dimension to a fix number at each layer. Unlike conventional U-Net architectures, which typically double the number of feature channels after each downsampling stage, our network maintains a fixed number of channels throughout the architecture, leading to a compact model with fewer than one million trainable parameters. The role of max-pool is to reduce the feature map resolution to attain deeper semantic representation. At the decoder path, the transition up block recovers the spatial resolution with up-sampling operations. The transition up block is identical to transition down block, except that max-pooling is replaced by deconvolution operations. The skip connections between encoder and corresponding decoder layers are employed to facilitate information flow from the contracting path to the expanding path.

2. Attention mechanism

Inspired by Liu et al. [14] work, we design a spatial channel attention module to attain richer semantic representation. Different from conventional spatial-channel attention modules, our attention module consists three

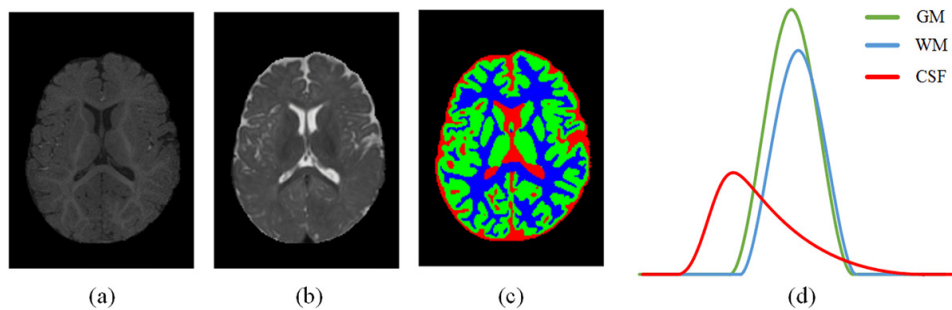


Figure 1: Example of infant brain slice from multimodality images in infant brain MRI segmentation (iSeg) dataset (a) T1 images, (b) T2 images, and (c) manual segmentation and (d) intensity distributions from T1. (MRI: Magnetic Resonance Imaging, GM: Gray Matter, WM: White Matter, CSF: Cerebrospinal Fluid)

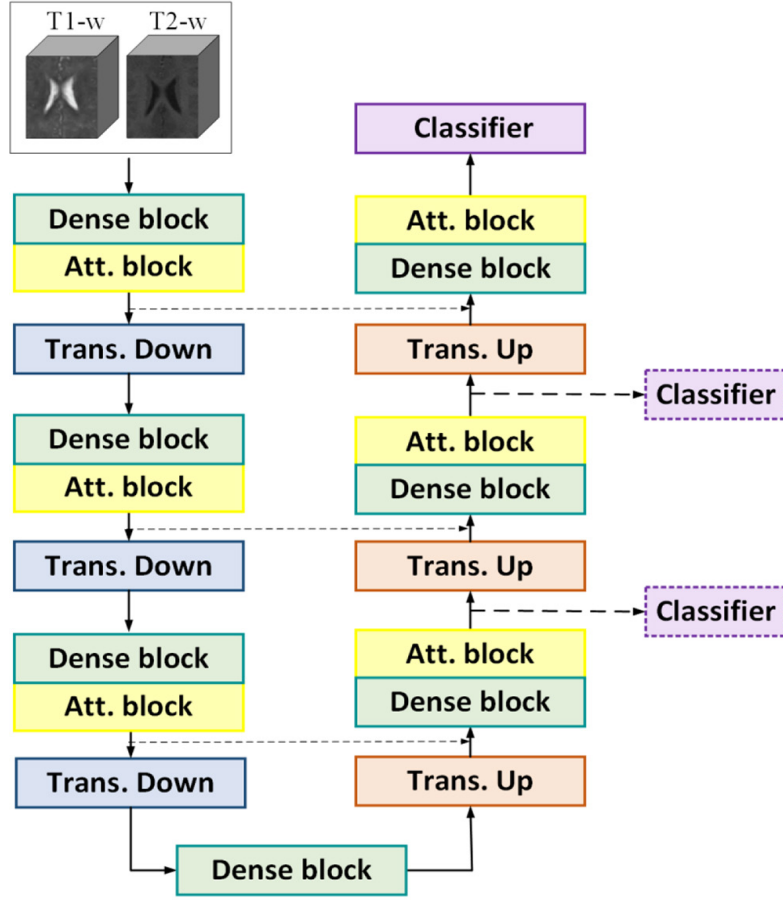


Figure 2: The proposed network architecture. (Att. block: Attention block, Trans. Down: Transition down block, Trans. Up: Transition up block, T1-w: T1-weighted MRI, T2w: T2-weighted, MRI: Magnetic Resonance Imaging)

paths: the spatial attention path, channel attention path, and connection path, which connect the input feature map to the output. The conventional attention mechanism consists of two paths spatial attention and channel attention whose outputs are summed to produce the output feature map: $f_{out} = SA(f_{in}) + CA(f_{in})$, where SA and CA denote spatial and channel attention, respectively. In our method, we set $f_{out} = SA(f_{in}) + CA(f_{in}) + CP(f_{in})$. The extra part CP is calculated as: $CP = BT(cov3D(BT(f_{in})))$, where BT is the bottleneck operation, and $cov3D$ represents three-dimensional convolution operation. The connection path enabling the attention block to capture more relevant and denser feature maps.

After each dense block, we employ our

channel-spatial attention block to exploits global representation and emphasis to relevant features at both channel and spatial dimensions.

3. Loss function

There are two main problems in infant brain MR images, as follows: 1) the low contrast between white matter and gray matter and 2) the class imbalance between them, where GM voxels outnumber white matters. As a result, the predictions may lean towards the GM class in low-contrast isointense regions. To handle the problem, some methods introduced similarity loss functions [10, 20]. Alternatively, we try to handle the issue by designing a loss function composed of cross entropy (CE) loss and weighing strategy. Similarity losses helps

to handle the problem of class unbalance [10]. However, they may not properly handle the low contrast issue. We use cross entropy loss function and assign different weights for each class voxels as follows:

$$loss_{WCE} = \frac{-1}{SV} \sum_{s=1}^S \sum_{v=1}^V \sum_{i=1}^C \log(\alpha_v^s p(c_v^s | \theta)) \delta(c_v^s = c_i) \quad (1)$$

here, s , v , and i denote the batch index, voxel index within a batch, and class index, respectively, while S , V , and C represent the total numbers of batches, voxels per batch, and classes. δ denotes the Kronecker delta function, which equals 1 when its argument is zero and 0 otherwise. c_v^s and $p(c_v^s | \theta)$ correspond to the ground-truth label and the network output for voxel v in batch s , respectively. α_v^s is a class-weighting factor used in the loss function. Background voxels are assigned a weight of zero, whereas tissue voxels receive positive weights. To encourage the network to focus on isointense regions, higher weights are assigned to GM and WM voxels. Furthermore, WM is given a larger weight than GM to address class imbalance and improve WM recall. The specific weight values used in the experiments are provided in the Results section.

4. Multi-resolution Deep supervision

Besides from main classifier, as depicted in Figure 2, we also employed two axillary side classifiers at the decoder part of the network, i.e. the down-sampled versions of ground-truth is supervised at decoder layers. As a result, the multi-resolution segmentation is performed at multiple output layers. Thus, the total loss function can be expressed as follows:

$$loss_T = \sum_{m=0}^M L_{WCE}^m = \frac{-1}{S} \sum_{m=0}^M \sum_{s=1}^S \sum_{v \in V_m} \sum_{i=1}^C \frac{1}{V_m} \log(\alpha_v^s p(c_{vm}^s | \theta)) \delta(c_{vm}^s = c_i) \quad (2)$$

here, C_{v0}^s is the label at voxel $v \in V_0$ in the s^{th} patch sampled from full resolution (original) ground truth, and $p(c_{v0}^s | \theta)$ is the corresponding output of the main classifier. $c_{vm}^s | m \neq 0$ belong to the batch of ground-truth labels while its patches are downscaled

version of original ground-truth patches by factor of 2^m and are the label of voxel $v \in V_m$ in the s^{th} patch, and $p(c_{vm}^s | \theta)$ is the corresponding output of the m^{th} side-classifier. γ_m weights classifiers in the loss function.

In this setting, the network takes the advantages of multi-scale supervision, and the final loss can be efficiently back propagated to shallow layers, alleviating the overfitting and vanishing gradient problems.

Preprocessing

Besides primary preprocessing conducted by the iSeg organizers, we apply further preprocessing steps to prepare the MR images for input the network. We first remove the background pixels around the brain tissues. In this setting, the computational burden is reduced, and the network concentrates more on relevant brain tissues. Second, we normalize both T1-w and T2-w images to zero mean and unit variance.

Training procedure and implementation details

We train the network with 20000 iterations and use AdamW optimizer with initial learning rate equal to 0.002. The learning rate is halved if the training loss keeps constant in successive iterations. For weight initialization, we use the He-Normal initialization. During training, the loss and Dice coefficient are continuously monitored, and training is terminated once both metrics reached a plateau.

For data augmentation, input images are randomly cropped to patches of size $32 \times 32 \times 32$ voxels. The T1-w and T2-w are concatenated at the input. As a result, the input patch is a tensor with size $2 \times 32 \times 32 \times 32$ and the output is $4 \times 32 \times 32 \times 32$, where each output channel corresponds to the one of the four class (WM, GM, CSF, and background) labels. The training batch size is 2. The model was implemented in Keras API and trained on a Tesla T4 GPU.

Evaluation metrics

For evaluation metrics, we use Dice similarity coefficient as given below. The Dice coefficient is the harmonic mean of recall and precision and is the main evaluation metric in the field of medical image segmentation.

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (3)$$

TP , FP , and FN represent true positives, false positives, and false negatives, respectively. We also evaluate Area Under the Precision-Recall Curve (AUPRC) to summarize the model's precision-recall performance.

Results

Similar to Bui et al. [9] and Basnet et al. [19], we trained the network with nine subjects with the ground-truth label and used the one subject (subject 9) as a test subject. Table 2 demonstrates the effectiveness of various components in the proposed segmentation method. The backbone is the architecture in Figure 2 without attention blocks and deep supervision. From the reported results, it is evident that all attention mechanism, deep supervision and class-weighting strategy have improved the network performance.

For class-weight tuning, we set the loss function hyperparameter α_v^s to 0, 0.25, 0.3, and 0.45 for background, CSF, GM, and WM voxels, respectively. In this setting, first, we set zero weights in loss function for

background voxels. Second, the WM+GM voxels' weights is 0.75, which is higher than 0.67 (equal tissue weights), thus the network more concentrates on the isointense low-contrast tissues. Third, the higher weight of WM compared to the GM compensates the imbalance class numbers and recover the recall of WM. Figure 3 shows the issue by plotting confusion matrix before and after weight tuning strategy. From Figure 3, the prominent errors prior to applying the weighting strategy arise from WM pixels that are wrongly labeled as GM tissue. Class weight tuning effectively addresses low contrast and class imbalance, leading to higher Dice score as reported in the Table 2. Empirically, when we increased the weight of CSF regions above 0.25, the network's focus on isointense regions decreased, yielding lower Dice scores. Moreover, reducing the weight for WM regions below 0.45 was insufficient to recover false negatives, while increasing it further produced more false positives; both cases led to lower Dice scores. We repeated the experiment using different test subjects (once with subject #1 and once with subject #5) and obtained similar results.

Table 3 compares the proposed segmentation method with several recent approaches [16, 17, 19, 21]. The results demonstrate that the proposed method is more effective at segmenting low-contrast and challenging GM and WM tissues than previous approaches.

Table 2: Comparison of ablation experimental results on infant brain MRI segmentation (iSeg) dataset.

Network	Dice			
	WM	GM	CSF	Mean
Backbone	90.58	91.03	94.32	91.97
Backbone + attention	91.18	91.56	94.77	92.51
Backbone + deep supervision	91.11	91.52	94.67	92.44
Backbone + weighting strategy	91.41	91.50	94.28	92.39
Proposed	91.80	92.07	94.58	92.81

MRI: Magnetic Resonance Imaging, WM: White Matter, GM: Gray Matter, CSF: Cerebrospinal Fluid

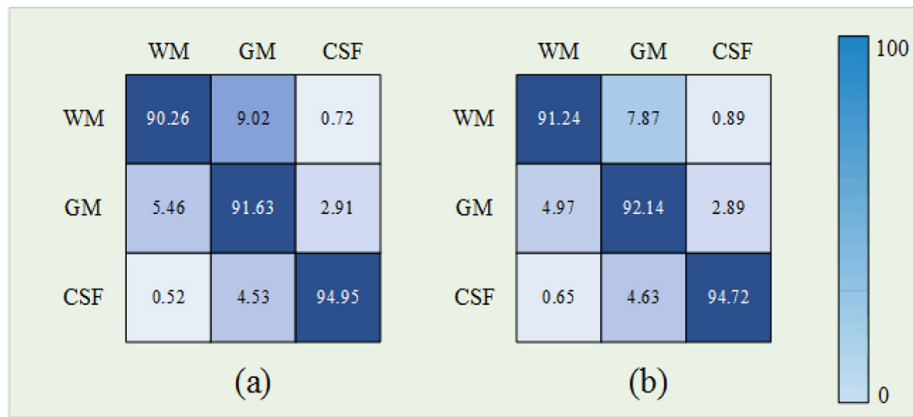


Figure 3: Confusion matrix, (a) before weight tuning and (b) after weight tuning strategy. (GM: Gray Matter, WM: White Matter, CSF: Cerebrospinal Fluid)

Table 3: Comparison with other works on infant brain MRI segmentation (iSeg) dataset.

Network	Dice				AUPRC	NP (M)	FLOPs (G)	IT (s)
	WM	GM	CSF	Mean				
3D-Unet [21]	89.80	90.27	94.12	91.36	0.871	19	28.1	3.6
Residual-denseNet [19]	90.32	91.05	94.47	91.94	0.879	0.71	20.6	1.9
2DCNN [16]	85.41	87.50	90.11	87.67	0.862	1.9	12.9	1.3
EUNet [17]	90.28	92.01	95.11	92.47	0.881	2.7	18.8	1.7
Proposed	91.80	92.07	94.58	92.81	0.885	0.9	17.3	1.5

MRI: Magnetic Resonance Imaging, WM: White Matter, GM: Gray Matter, CSF: Cerebrospinal Fluid, 3D-UNet: 3 Dimensional UNet, 2D-CNN: 2 Dimensional Deep Convolutional Neural Network, EUNet: Enhanced UNet, AUPRC: Area Under the Precision-Recall Curve, NP: Number of Parameters, FLOPs: Floating-Point Operations Per Second, IT: Inference Time

Specifically, our approach achieves the highest Dice scores for both WM and GM, as well as the highest mean Dice score across all three tissue classes. In addition, our model attains the highest AUPRC, indicating superior precision-recall performance. As shown in Table 3, the proposed network contains 0.9 million parameters, requires 17.3 GFLOPs, and performs inference in approximately 1.5 second. These computational requirements are comparable to those of the 2D baseline networks, EUNet and 2DCNN, as well as a lightweight 3D Residual-DenseNet. The low computational cost and fast inference suggests that the proposed network is lightweight and well suited for real-time segmentation

applications.

The qualitative results are shown in Figure 4, which provides a visual comparison between the proposed method and the competing methods.

To assess generalizability, we evaluated our network on the internet brain segmentation repository (IBSR18) dataset, which contains 18 T1-weighted 3D brain MR images of subjects aged 7–71 years. All volumes have a size of $256 \times 256 \times 128$ voxels. We split the data into 10 images for training and 8 for testing, using the same architecture and training procedure as for iSeg, but omitting the class-weighting strategy (which was designed to address class imbalance and low contrast in isointense infant MR

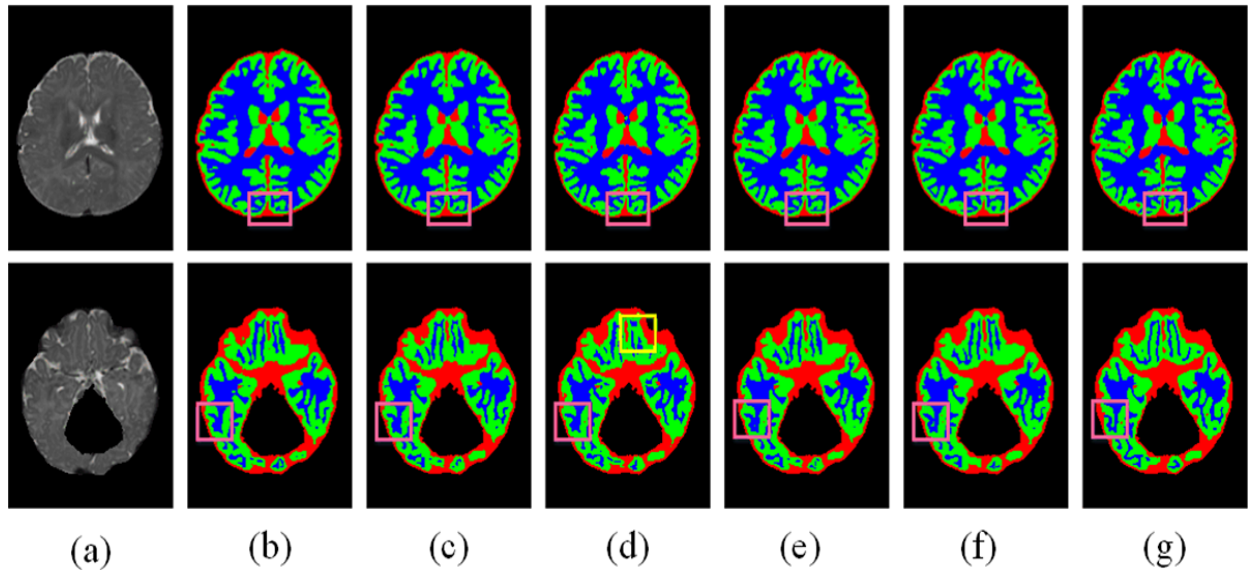


Figure 4: Segmentation results on two different slices of test sample (a) T2-w image, (b) 3D-Unet, (c) Residual-denseNet, (d) 2DCNN, (e) EUNet, (f) proposed and (g) ground truth. (3D-UNet: 3 Dimensional UNet, 2D-CNN: 2 Dimensional Deep Convolutional Neural Network, EUNet: Enhanced UNet)

images). The results are given in Table 4. For WM, GM, and CSF, we achieved Dice coefficients of 92.57, 93.33, and 89.85, respectively. We also assessed statistical significance with a two-sided paired t-test and (except for EUNet) observed significant Dice improvements relative to other baselines (P -value <0.005).

Table 4: Comparison with other works on infant brain MRI segmentation (IBSR18) dataset.

Network	Dice			
	WM	GM	CSF	Mean
3D-Unet [21]	89.32	89.56	88.78	89.22
Residual-denseNet [19]	92.01	93.22	88.85	91.36
2DCNN [16]	87.19	88.42	84.69	86.76
EUNet [17]	91.97	93.40	89.41	91.59
Proposed	92.57	93.33	89.85	91.91

MRI: Magnetic Resonance Imaging, WM: White Matter, GM: Gray Matter, CSF: Cerebrospinal Fluid, 3D-UNet: 3 Dimensional UNet, 2D-CNN: 2 Dimensional Deep Convolutional Neural Network, EUNet: Enhanced UNet

Discussion

Infant brain tissue segmentation plays a crucial role in the assessment of early brain development and the prediction of neurodevelopmental disorders [2, 6]. Nevertheless, accurate segmentation remains challenging due to several factors, including low tissue contrast, class imbalance, and the limited availability of annotated training data. Although CSF generally exhibits clear contrast with surrounding tissues, WM and GM often present highly similar intensity distributions, making their discrimination particularly difficult. Furthermore, the imbalance between WM and GM classes frequently leads to an increased number of false-positive GM predictions, thereby degrading overall segmentation performance. As shown in Table 2, even the widely adopted 3D U-Net architecture [21], which contains more than 19 million trainable parameters, achieved only limited performance on the infant brain tissue segmentation.

To address overfitting caused by the scarcity of annotated data and to reduce computational

complexity, several recent studies have proposed 2D-based architectures for the segmentation of 3D infant brain volumes [15-17]. Although these approaches substantially reduce training difficulty and computational requirements, their performance is often constrained by their inability to fully exploit the rich spatial context and complex anatomical information inherent in 3D infant brain images.

In this study, we developed a novel 3D segmentation framework to overcome the challenges associated with low tissue contrast and class imbalance. Ablation studies were performed to assess the impact of the framework components, revealing that attention mechanisms, deep supervision, and class-weight optimization each made a positive contribution to segmentation accuracy. Despite employing 3D operations, the proposed network utilizes bottleneck layers and lightweight attention modules, resulting in a compact architecture with fewer than one million trainable parameters.

Compared with existing infant brain tissue segmentation methods, the proposed model achieved superior segmentation performance while maintaining a low computational burden. Specifically, our network achieved the highest Dice scores for WM (91.80%) and GM (92.07%), as well as the highest mean Dice score across the three tissue classes (92.81%). Furthermore, the proposed network reduced the number of trainable parameters by more than twofold and threefold compared with the 2DCNN [16] and EUNet [17] architectures, respectively, while maintaining a parameter count and FLOPs comparable to that of the lightweight 3D Residual-denseNet [19]. These findings highlight the effectiveness and efficiency of the proposed approach for infant brain tissue segmentation. Moreover, our model achieved the highest mean Dice coefficient (91.91%) on the IBSR18 dataset, suggesting strong generalizability of the proposed segmentation network to adult brain images.

Conclusion

In this paper we devised a new method for segmenting the infant brain tissues into white matter, gray matter and cerebrospinal fluid. We present a lightweight architecture by using dense blocks and transition blocks as a backbone architecture. We equipped the backbone by new channel spatial attention mechanism, multi-resolution deep supervision and class weigh tuning strategy to handle the problem of isointense low-contrast tissues and class imbalance and achieve precise segmentation results.

Authors' Contribution

MR. Akbarzadeh Lari conceived and conducted the experiments, analyzed the data, and drafted the manuscript; H. Danyali supervised the project and revised the manuscript; K. Kazemi contributed to the medical data interpretation and revised the manuscript. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

Since the data supporting this study is openly available at <https://iseg2017.web.unc.edu/download/> and <https://www.nitrc.org/projects/ibsr/>, the requirement for an additional code of ethics is not applicable.

Conflict of Interest

None

References

1. Merritt K, Luque Laguna P, Irfan A, David AS. Longitudinal Structural MRI Findings in Individuals at Genetic and Clinical High Risk for Psychosis: A Systematic Review. *Front Psychiatry*. 2021;**12**:620401. doi: 10.3389/fpsy.2021.620401. PubMed PMID: 33603688. PubMed PMCID: PMC7884337.
2. Knickmeyer RC, Gouttard S, Kang C, Evans D, Wilber K, Smith JK, et al. A structural MRI study of human brain development from birth to 2 years. *J Neurosci*. 2008;**28**(47):12176-82. doi: 10.1523/JNEUROSCI.3479-08.2008.

- PubMed PMID: 19020011. PubMed PMCID: PMC2884385.
3. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, et al. Early brain development in infants at high risk for autism spectrum disorder. *Nature*. 2017;**542**(7641):348-51. doi: 10.1038/nature21369. PubMed PMID: 28202961. PubMed PMCID: PMC5336143.
4. Gilmore JH, Kang C, Evans DD, Wolfe HM, Smith JK, Lieberman JA, et al. Prenatal and neonatal brain structure and white matter maturation in children at high risk for schizophrenia. *Am J Psychiatry*. 2010;**167**(9):1083-91. doi: 10.1176/appi.ajp.2010.09101492. PubMed PMID: 20516153. PubMed PMCID: PMC3105376.
5. Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ben Ayed I. HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation. *IEEE Trans Med Imaging*. 2019;**38**(5):1116-26. doi: 10.1109/TMI.2018.2878669. PubMed PMID: 30387726.
6. Wang L, Li G, Shi F, Cao X, Lian C, Nie D, et al. Volume-Based Analysis of 6-Month-Old Infant Brain MRI for Autism Biomarker Identification and Early Diagnosis. *Med Image Comput Comput Assist Interv*. 2018;**11072**:411-9. doi: 10.1007/978-3-030-00931-1_47. PubMed PMID: 30430147. PubMed PMCID: PMC6231401.
7. Wang L, Nie D, Li G, Puybureau E, Dolz J, Zhang Q, et al. Benchmark on Automatic 6-month-old Infant Brain Segmentation Algorithms: The iSeg-2017 Challenge. *IEEE Trans Med Imaging*. 2019;**38**(9):2219-30. doi: 10.1109/TMI.2019.2901712. PubMed PMID: 30835215. PubMed PMCID: PMC6754324.
8. Dolz J, Desrosiers C, Wang L, Yuan J, Shen D, Ben Ayed I. Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. *Comput Med Imaging Graph*. 2020;**79**:101660. doi: 10.1016/j.compmedimag.2019.101660. PubMed PMID: 31785402.
9. Bui TD, Shin J, Moon T. Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. *Biomed Signal Process Control*. 2019;**54**:101613. doi: 10.1016/j.bspc.2019.101613.
10. Hashemi SR, Prabhu SP, Warfield SK, Gholipour A. Exclusive independent probability estimation using deep 3D fully convolutional DenseNets: Application to IsoIntense infant brain MRI segmentation. *International Conference on Medical Imaging with Deep Learning; PMLR*; 2019. p. 260-72.
11. Wang Z, Zou N, Shen D, Ji S. Non-local u-nets for biomedical image segmentation. *Proceedings of the AAAI conference on artificial intelligence*; New York, USA: AAAI Press; 2020. p. 6315-22.
12. Luan X, Wang J. Image Semantic Information Fusion for MR Brain Tissue Segmentation. *International Conference on Medical Artificial Intelligence (MedAI)*; Chongqing, China: IEEE; 2024. p. 571-9.
13. Li J, Yu H, Chen C, Ding M, Zha S. Category guided attention network for brain tumor segmentation in MRI. *Phys Med Biol*. 2022;**67**(8):085014. doi: 10.1088/1361-6560/ac628a. PubMed PMID: 35354136.
14. Liu H, Huo G, Li Q, Guan X, Tseng ML. Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation. *Expert Systems with Applications*. 2023;**214**:119166. doi: 10.1016/j.eswa.2022.119166.
15. Luan X, Zheng X, Li W, Liu L, Shu Y, Guo Y. Rubik-Net: Learning Spatial Information via Rotation-Driven Convolutions for Brain Segmentation. *IEEE J Biomed Health Inform*. 2022;**26**(1):289-300. doi: 10.1109/JBHI.2021.3095846. PubMed PMID: 34242176.
16. Karimi H, Hamghalam M. Segmentation of 3D MRI using 2D convolutional neural networks in infants' brain. *Multimed Tools Appl*. 2024;**83**(11):33511-26. doi: 10.1007/s11042-023-16790-z.
17. Dénes-Fazakas L, Kovács L, Eigner G, Szilágyi L. Enhanced U-Net for Infant Brain MRI Segmentation: A (2+1)D Convolutional Approach. *Sensors (Basel)*. 2025;**25**(5):1531. doi: 10.3390/s25051531. PubMed PMID: 40096351. PubMed PMCID: PMC11902485.
18. Saha S, Dutta S, Goswami B, Nandi D. ADU-Net: An Attention Dense U-Net based deep supervised DNN for automated lesion segmentation of COVID-19 from chest CT images. *Biomed Signal Process Control*. 2023;**85**:104974. doi: 10.1016/j.bspc.2023.104974. PubMed PMID: 37122956. PubMed PMCID: PMC10121143.

19. Basnet R, Ahmad MO, Swamy MN. A deep dense residual network with reduced parameters for volumetric brain tissue segmentation from MR images. *Biomed Signal Process Control*. 2021;**70**:103063. doi: 10.1016/j.bspc.2021.103063.
20. Liu B, Dolz J, Galdran A, Kobbi R, Ben Ayed I. Do we really need dice? The hidden region-size biases of segmentation losses. *Med Image Anal*. 2024;**91**:103015. doi: 10.1016/j.media.2023.103015. PubMed PMID: 37918314.
21. 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; Cham: Springer International Publishing; 2016. p. 424-32.