Biomedical Signal Processing and Control 54 (2019) 101613

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation

Toan Duc Bui, Jitae Shin*, Taesup Moon

Department of Electrical and Computer Engineering, Sungkyunkwan University, 2066 Seobu-ro, Suwon 16419, Republic of Korea

ARTICLE INFO

Article history:
Received 5 September 2018
Received in revised form 11 June 2019
Accepted 11 July 2019
Available online 9 August 2019

Keywords:
Infant brain segmentation
Fully convolutional neural networks
DenseNet
skip-connection

ABSTRACT

Automatic 6-month infant brain tissue segmentation of magnetic resonance imaging (MRI) is still less accurate owing to the low intensity contrast among tissues. To tackle the problem, we introduce an accurate segmentation method for volumetric infant brain MRI built upon a densely connected network that achieves state-of-the-art accuracy. Specifically, we carefully design a fully convolutional densely connected network with skip connections such that the information from different levels of dense blocks can be directly combined to achieve highly accurate segmentation results. The proposed network, called 3D-SkipDenseSeg, exploits the advantage of the recently DenseNet for classification task and extends this to segment the 6-month infant brain tissue segmentation of magnetic resonance imaging (MRI). Experimental results demonstrate a competitive performance with regard to both segmentation accuracy and parameter efficiency of the proposed method over the existing methods; namely, the proposed 3D-SkipDenseSeg achieved the best dice similarity coefficient (DSC) of $90.37 \pm 1.38\%$ (WM), $92.27 \pm 0.81\%$ (GM), and $95.79 \pm 0.54\%$ (CSF) among the 21 participating teams in the 6-month infant brain dataset (iSeg-2017) and required only 10–30% of the parameters compared to similar deep learning-based methods.

1. Introduction

Volumetric infant brain segmentation has a critical role in studying of early brain development. It aims to separate the brain tissues into non-overlapping regions. The manual segmentation of a volumetric infant brain is generally performed by clinical experts and is referred to as the ground-truth; however, it is time-consuming, frequently subject to intra or inter observation variabilities, and requires skilled experts. Therefore, automatic segmentation algorithms are being considered as alternatives for addressing the aforementioned issues of manual segmentation because they can provide consistent results and are scalable. However, the low contrast between tissues, increased noise, and on-going white matter myelination [1] in the infant brain MRI frequently cause tissues to be misclassified and reduce the accuracy of the segmentation algorithms. For example, Fig. 1 displays samples of a slice in the T1 and T2 scans and the intensity distributions of brain tissue in the 6-month infant brain MRI. Fig. 1(d) and (e) indicate that the distributions between gray matter (GM) and white matter have large overlaps, which makes it difficult for the algorithms to accurately classify the tissues with the intensities in the overlapping regions. Therefore, developing accurate automatic infant brain-image segmentation algorithms remains an active area of research.

There are three major categories of automatic brain-image segmentation approaches: statistical-based, atlas-based, and learning-based. Namely, the statistical-based approaches assume that the intensity of each tissue belongs to a certain probabilistic distribution, e.g., Gaussian [2] or a mixture of Gaussians [3,4], and classify each tissue with respect to the assumed distribution. However, this approach fails in the case of heavy intensity distribution overlaps as in Fig. 1(e), e.g., for the images acquired from the isointense stage. As an alternative, atlas-based approaches also have been widely used [1,5–7]. They address the segmentation problem as registering the template images and their manual segmentation onto the target image. Because the accuracy of the registration step can negatively influence the final segmentation performance, the approach needs a large number of template images to capture the wide variability in the brain anatomy, which makes the registration step computationally prohibitive. Learning-based approaches, conversely, do not require the registration step and are considered as a promising direction for neonatal brain-image segmentation. The approaches use extracted features from the training images to learn a supervised segmentation model. For example, van Opbroek et al. [8] utilized both intensity and spatial features to learn a brain

* Corresponding author.
E-mail address: jtshin@skku.edu (J. Shin).

https://doi.org/10.1016/j.bspc.2019.101613
1746-8094/

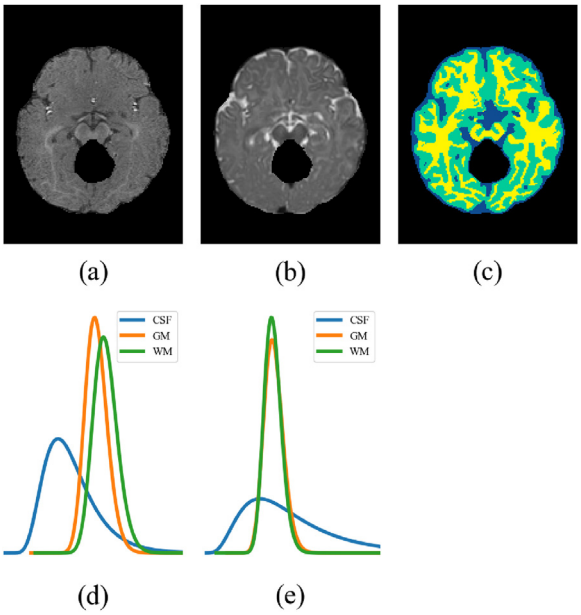

Fig. 1. Example of brain structure from multimodality images in iSeg challenge (a) T1 images, (b) T2 images, and (c) manual segmentation, and intensity distributions for (d) T1 images and (e) T2 images.

segmentation model using support vector machine. Wang et al. [9] introduced learning-based multisource images using 3D Haar-like features, and Pereira et al. [10] employed intensity, intensity-derived features, and spatial features.

Whereas the above-mentioned learning-based methods are effective to some extent, they also have limitations in mainly using handcrafted features. Recently, deep learning models were extremely successful in jointly learning features and prediction models [11], thus, overcoming the limitations of using the handcrafted features. Deep convolutional neural networks (DCNNs) [12–14] has been widely used in deep learning models and have reported significant success in diverse applications including brain image segmentation. For examples, Zhang et al. [15] introduced a DCNN architecture to segment brain tissue on isointense stage using multimodal MRI. Ronnenberger et al. [16] devised a 2D U-Net network that demonstrated superb performance for medical image segmentation. Moeskops et al. [17] introduced a DCNN using multiscale features for both neonatal and adult images at different ages. These deep learning-based methods provided significant improvement the accurate segmentation in comparison with the other learning-based methods employing handcrafted features; however, those studies primarily focused on processing two-dimensional (2D) segmentation tasks, where each slice was processed independently. Hence, for three-dimensional (3D) volumetric data such as brain MRI, it is desirable to design a DCNN to perform 3D volumetric segmentation, where the network is allowed to explore the spatial information among the adjacent image slices.

To tackle the aforementioned issues, the paper introduces a fully convolutional densely connected network with skip connections for volumetric infant brain tissue segmentation. This typically requires limited supervised training data and demonstrated superb performance on the 6-month infant brain MRI dataset (iSeg-2017). To devise the proposed network architecture, we first summarize the recent research results and challenges on devising DCNN-based volumetric segmentation algorithms and outline the main contributions of the paper in Section 2. Then, Section 3 briefly reviews the key concepts of the related network architectures and presents the proposed method in the details. We shows the experimental results in Section 4. Finally, we discuss our method in Section 5 and summarize our conclusions in Section 6.

2. Related work

2.1. Benefit of 3D features for volumetric segmentation

Effectively exploiting the correlation of adjacent frames or images has been demonstrated to improve the accuracy of a target task that involves 3D volumetric data; for example, Tran et al. [18] demonstrated that learning from spatiotemporal (i.e., 3D) features achieves a significant improvement compared to using 2D features for object detection tasks. Inspired by this advantage of learning 3D features, many solutions have been proposed to use 3D features for biomedical volumetric data, e.g., [19–25]. Çiçek et al. [21] presented 3D U-Net architecture by concatenating skip layers and learns the upsampling part. Kamnitsas et al. [24] built a dual path convolutional neural networks (CNN) architecture for the brain-lesion segmentation task. Although the above work demonstrated the effectiveness of 3D features compared to 2D features in volumetric brain-image segmentation, they employed relatively shallow architectures that could limit representation capability. A deeper network may improve the representational power; however, can cause two main challenges, described in the following subsection.

2.2. Challenges in designing deep network for volumetric segmentation

Two major, generic challenges that arise when designing a deep network are the following: (1) the training becomes more difficult as the network increases the depth and (2) the computational and memory complexities increase significantly as the depth increases.

Regarding the first challenge, it has been demonstrated that the performance of a deep network architecture tends to become saturated when the depth of the network is increased by simply stacking additional layers [14,26]. Residual network (ResNet) [14], which adds skip connections to every other convolution layer, is an attempt to address this challenge and has been shown to be extremely effective in 2D image classification tasks. Furthermore, the architecture has been extended to 3D volumetric segmentation [27–29]. To further improve ResNet, Huang et al. [30] proposed DenseNet, which consists of dense blocks that connect each layer to subsequent layers through direct connections, and achieved consistent improvement over ResNet in 2D image classification tasks. Following this success, Jégou [31] proposed a 2D fully convolutional DenseNet for semantic segmentation; Yu et al. [32] recently extended 2D fully convolutional DenseNet to DenseVoxNet for 3D volumetric cardiac image segmentation. DenseVoxNet includes two dense blocks followed by pooling layers to increase receptive field, then restores the original resolution using stacks of learnable deconvolution layers. Although the method achieved reasonable segmentation performance, it may not be able to appropriately capture multiscale contextual information useful for accurate segmentation as there are no direct connections between the dense blocks and the final prediction layer.

The second challenge mentioned above arises because of the increasing number of learning parameters as the depth of a network increases leads to optimization difficulty. For example, in DenseVoxNet, the stacked deconvolution layers generate a considerable number of parameters subject to learning, which requires significant memory and computational time during training. One method to address this challenge is to use a bottleneck architecture that can maintain the performance and reduce the parameters [14,30,31,33]. Such an architecture, however, has not been investigated for DenseNet-based 3D volumetric segmentation schemes.

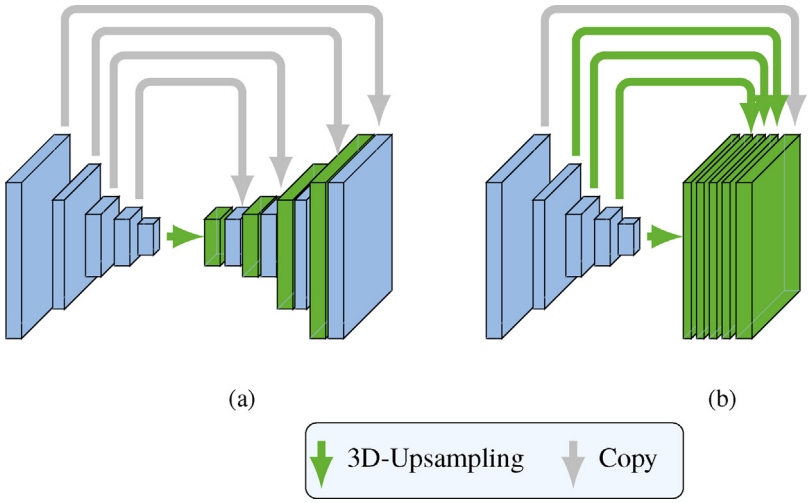

Fig. 2. A comparison between (a) 3D U-Net architecture and (b) proposed 3D-SkipDenseSeg architecture for infant brain segmentation.

2.3. Our contributions

Motivated by the aforementioned challenges, the paper introduces a novel fully convolutional DenseNet for 3D volumetric infant brain MRI segmentation, named 3D-SkipDenseSeg. Preliminary results of the proposed algorithm have appeared in [33] as a part of the recent infant brain MRI segmentation challenge, iSeg-2017 [34]. Following are our main contributions in terms of designing the model architecture:

- The skip-connections have been shown to be successful in object detection to exploit the multiple contextual information [35]. We utilize the skip-connections for the fully convolutional DenseNet architecture to concatenate information from lower to higher dense blocks. The skip-connections enable the proposed network to exploit multiple contextual information from the deep convolutional network and yields accurate and detailed segmentation results. The dense connection that provides superior information and gradient flows during training. Unlike the others semantic segmentation [36,37] that used multiple scale features on the top of the ResNet architecture [14], we utilize the multiple scale features in the intermediate dense blocks and concatenate them together via skip connections. Moreover, unlike the stacked deconvolution in 3D-Unet [21], DenseVoxNet [32], the skip connections do not increase the parameters significantly, which is a desirable property for applications with limited memory resource.
- We report the efficiency of the bottleneck with compression (BC) and $2 \times 2 \times 2$ convolution layers and show they play critical roles in achieving competitive performance for segmentation accuracy of the infant brain MRI.

Fig. 2 illustrates the difference between U-Net architecture [21] and proposed architecture. The U-Net architecture utilizes the stacked deconvolution where the low resolution features from deeper layers are gradually upsampled and concatenated with high resolution features in the shallow layers, thus the U-Net increases the number of training parameters. Unlike U-Net architecture, we directly upsample the low resolution features to input resolution and concatenated together that provides smaller training parameters than U-Net architecture. We believe that this is the first attempt of concatenating multilevel feature maps generated by dense blocks via skip-connections for volumetric 6-month infant brain tissue segmentation. In the same year, Dolz et al. introduced HyperDenseNet [38] a hyper-dense connection for infant brain segmentation. The HyperDenseNet extends the dense connection between multi-modal images, which can be viewed as a hyper fusion method. In contrast to the HyperDenseNet, the proposed 3D-SkipDenseSeg can be viewed as an early fusion where the weights that act on multi-modal images were learned using the first convolutional layer. The performance of early and hyper fusions will be reported in Section 4. Furthermore, as detailed in Section 4, the proposed network architecture uses considerably more layers than the existing state-of-the-art; however, it uses significantly less parameters per layer. The performance on the recent iSeg-2017 dataset demonstrates the significant advantage of the proposed 3D-SkipDenseSeg architecture over the existing architectures such as VoxResNet [27], DenseVoxNet [32], and 3D U-Net [21] in terms of both accurate segmentation and parameter efficiency.

3. Methods

We first briefly review the key concepts of DenseNet [30] to introduce notations and backgrounds for the proposed scheme. Then, we describe the proposed architecture for 6-month infant brain tissue segmentation in detail.

3.1. DenseNet: densely connected convolutional network

The output x_ℓ of the ℓth layer of a traditional convolution feed-forward networks is computed by

$$x_\ell = H_\ell(x_{\ell-1}) \tag{1}$$

where H_ℓ is a nonlinear transformation of the ℓth layer that consists of layers such as convolution, Batch Normalization (BN) [39].

Recently, Huang [30] introduced a dense connection that improves gradient flows in the network. We can then write the output x_ℓ of the ℓth layer with the dense connections as:

$$x_\ell = H_\ell([x_0, x_1, \ldots, x_{\ell-1}]) \tag{2}$$

where $[\ldots]$ denotes concatenating operation.

The DenseNet architecture provides superior information and gradient flows during training. If the input has k_0 channels and each function $H_\ell(.)$ generates k feature maps, then the input feature maps number at the ℓth layer can be computed as $k_0 + (\ell - 1) \times k$. The hyperparameter k refers to the growth rate of the dense network. A transition layer which includes a 1×1 convolutional layer followed by a 2×2 max-pooling layer [12], is applied to increase receptive field. With m input feature maps, the transition layer produces $m \times \theta$ output feature maps, where $\theta \in [0, 1]$ refers as the

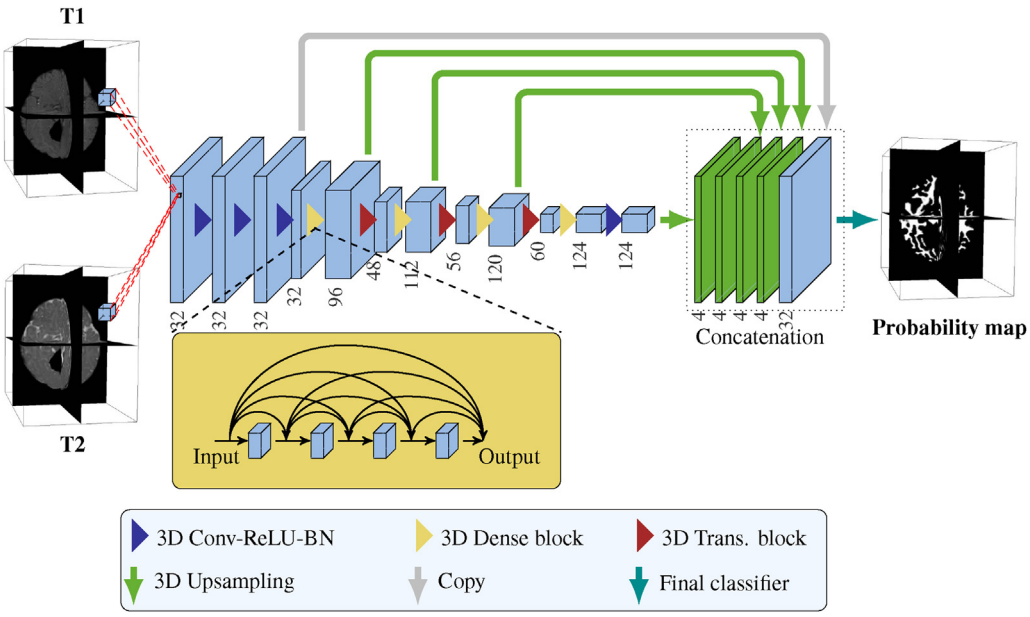

Fig. 3. Our proposed architecture, 3D-SkipDenseSeg, for 6-month infant brain tissue segmentation.

Table 1
3D-SkipDenseSeg configuration for volumetric segmentation.

Layer	Kernel size	Output
3D Conv-BN-ReLU	$\begin{bmatrix} 3\times3\times3, s=1, p=1 \end{bmatrix} \times 3$	$32\times64\times64\times64$
3D Conv	$2\times2\times2, p=0, s=2$	$32\times32\times32\times32$
3D Dense block 2 ($k=16$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 3\times3\times3, s=1, p=1 \end{bmatrix} \times 4$	$96\times32\times32\times32$
3D Trans. block 2 ($\theta=0.5$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 2\times2\times2, p=0, s=2 \end{bmatrix}$	$48\times16\times16\times16$
3D Dense block 3 ($k=16$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 3\times3\times3, s=1, p=1 \end{bmatrix} \times 4$	$112\times16\times16\times16$
3D Trans. block 3 ($\theta=0.5$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 2\times2\times2, p=0, s=2 \end{bmatrix}$	$56\times8\times8\times8$
3D Dense block 4 ($k=16$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 3\times3\times3, s=1, p=1 \end{bmatrix} \times 4$	$120\times8\times8\times8$
3D Trans. block 4($\theta=0.5$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 2\times2\times2, p=0, s=2 \end{bmatrix}$	$60\times4\times4\times4$
3D Dense block 5 ($k=16$)	$\begin{bmatrix} 1\times1\times1, s=1, p=1 \\ 3\times3\times3, s=1, p=1 \end{bmatrix} \times 4$	$124\times4\times4\times4$
3D Upsampling 2	$4\times4\times4, s=2, p=1, g=4$	$4\times64\times64\times64$
3D Upsampling 3	$6\times6\times6, s=4, p=1, g=4$	$4\times64\times64\times64$
3D Upsampling 4	$10\times10\times10, s=8, p=1, g=4$	$4\times64\times64\times64$
3D Upsampling 5	$18\times18\times18, s=16, p=1, g=4$	$4\times64\times64\times64$
Concat-BN-ReLU	–	$48\times64\times64\times64$
Classification	$1\times1\times1, s=1, p=0$	$4\times64\times64\times64$

Note: The output shape arranged as (channel × depth × height × width), s, p, g, k denote stride, padding, group and growth rate resepectively.

compression factor. Compared to the ResNet architecture [14], it has been demonstrated in [30] that the DenseNet yields a more compact, efficient, and accurate model.

3.2. 3D-SkipDenseSeg: a skip-connected 3D fully convolutional DenseNet

Fig. 3 presents the proposed network architecture that includes a contracting path and an expanding path. The proposed network architecture is presented in Table 1. Specifically, the contracting path aims to increase the receptive fields of feature maps. To extract the image features, we use four dense blocks. We use four $3\times3\times3$ Conv-BN-ReLU in each dense block. To overcome the overfitting, the dropout layer [40] with a dropout rate of 0.2 is used after each dense block. The transition block that comprises a convolutional layer with the kernel size of $1\times1\times1$ and the compression rate of half (i.e., $\theta=0.5$) followed by a convolution layer with a stride of two are employed to increase the receptive filed while preserving the spatial information. The transition blocks provide deep supervisions on the lower layers directly from the target labels to allow processing of a limited dataset. For simplicity, we ignore the tuning weight balance between the auxiliary and main losses for the transition layers [30,32]. The two input MRI modalities, T1 and T2 images, are concatenated as different channels and are used directly as input to the network. Moreover, before the first dense block, three ordinary $3\times3\times3$ convolution layers that generate 32 output feature maps are used.

In the expanding path, the input resolution is recovered by 3D-Upsampling operators. Specifically, to capture multiple levels of contextual information, we concatenate the upsampled feature maps from dense blocks. As indicated in Fig. 3, we employ skip-connections such that the feature maps from the shallow layers (with finer resolutions that capture detailed shapes) and deeper layers (with the coarse resolution that capture overall structures) can all be used for classifying the segmentation label of each pixel. This architecture, without applying stacked deconvolutions [21,32], is considered effective in 6-month infant brain segmentation. A $1\times1\times1$ convolution is used in the last layer to classify each pixel to one of the four class outputs using the concatenated feature maps.

Finally, the final concatenated feature map of the proposed network can be expressed as follows:

$$x^c = [x^1, x^2_{21}, \ldots, x^i_{2^{i-1}}, \ldots, x^n_{2^{n-1}}] \quad (3)$$

where $x^i_{2^{i-1}}$ is the output of the i'th dense block by upsampling with factor of 2^{i-1}, $(i=1,\ldots,n)$, x^1 is the output from the first three ordinary $3\times3\times3$ convolution layers. The final concatenated feature map x^c contains the multiple scale feature maps information and yields accurate and detailed segmentation results.

4. Experiments

4.1. Dataset

We used the iSeg-2017 dataset [34] to evaluate the robustness of the proposed 3D-SkipDenseSeg and existing architectures. The iSeg-2017 dataset consists of 10 subjects with ground-truth

Table 2
Performance on the validation set of the proposed 3D-SkipDenseSeg method and three recent deep learning-based methods: 3D U-Net [21], VoxResNet [27], and DenseVoxNet [32] in term of DSC: %, MHD: mm, and ASD: mm.

Method	Depth	Params	WM			GM			CSF		
			DSC	MHD	ASD	DSC	MHD	ASD	DSC	MHD	ASD
3D-SkipDenseSeg (ours)	47	1.55M	**91.02**	5.92	**0.39**	**91.64**	5.75	**0.34**	**94.88**	13.64	**0.13**
DenseVoxNet (2017) [32]	32	4.34M	85.46	7.07	0.64	88.51	**5.48**	0.50	91.26	**11.58**	0.23
3D-Unet (2015) [21]	18	19M	89.58	5.39	0.44	90.70	4.90	0.38	94.39	13.86	0.15
VoxResNet (2017) [27]	25	1.33M	89.87	**5.20**	0.44	90.64	5.48	0.38	94.28	13.93	0.15

Note: Bold values indicate best performance.

Table 3
Leaderboard of top five teams in the iSeg-2017 test set (DSC: %, MHD: mm, ASD: mm).

Method	WM			GM			CSF		
	DSC	MHD	ASD	DSC	MHD	ASD	DSC	MHD	ASD
3D-SkipDenseSeg (ours)	**90.37 (1.38)**	6.62 (1.02)	**0.38 (0.05)**	**92.27 (0.81)**	6.00 (1.26)	**0.32 (0.04)**	**95.79 (0.54)**	9.11 (0.94)	**0.12 (0.01)**
LIVIA (HyperDense-Net)	89.71 (1.47)	6.98 (1.14)	0.38 (0.08)	91.86 (0.88)	6.42 (1.22)	0.34 (0.04)	95.70 (0.69)	**9.03 (1.09)**	0.14 (0.10)
Bern_IPMI	89.60 (1.24)	6.78 (1.16)	0.40 (0.05)	91.63 (0.72)	6.46 (1.24)	0.34 (0.04)	95.35 (0.68)	9.62 (1.35)	0.13 (0.02)
LRDE	86.12 (1.33)	**6.61 (1.06)**	0.52 (0.04)	88.69 (0.74)	**5.85 (0.69)**	0.46 (0.03)	92.82 (0.62)	9.88 (0.97)	0.20 (0.01)
nic_vicorob	88.47 (1.54)	7.15 (0.62)	0.43 (0.04)	90.96 (0.82)	7.65 (0.81)	0.37 (0.04)	95.06 (0.50)	9.18 (1.26)	0.14 (0.01)

Note: Bold values indicate best performance.

labels for training and 13 subjects without ground-truth labels for testing. We used the training dataset for fine-tuning the segmentation models. Each subject includes T1 and T2 images with size of $144 \times 192 \times 256$, and image resolution of $1 \times 1 \times 1$ mm^3. The challenge aims to segment the infant brain into four non-overlapping regions: CSF, WM, GM, and background; the ground-truth labels were provided by an experienced neuroradiologist. The performance of each participant team was quantitatively compared with the manual segmentation using the following three performance metrics, which were calculated with the public evaluation tool [41].

4.2. Evaluation metrics

4.2.1. Dice similarity coefficient

To quantify the overlap between the segmentation result A and the ground-truth G, the Dice similarity coefficient defines as follows:

$$DSC = \frac{2\left|A \cap G\right|}{|A||G|}, \tag{4}$$

where the $|\cdot|$ notation represents the size of a set and $|A \cap G|$ represents the size of the intersection of A and G. A higher DSC value indicates a superior segmentation accuracy.

4.2.2. Modified Hausdorff distance

To calculate the distance between segmentation and ground-truth boundaries, a modified Hausdorff distance (MHD) is defines as follows:

$$MHD(A, G) = \max\left(h_{95}(A, G), h_{95}(G, A)\right) \tag{5}$$

where $h_{95}(A, G) = {}^{95}K_{a \in A} \min_{g \in G} \| g - a \|$ is the 95th-percentile of the Hausdorff distance [42] to ignore the sensitive to outliers. A smaller MHD value represents a more accurate segmentation.

4.2.3. Average surface distance

The third measurement metric is the average surface distance (ASD) error, defined as

$$ASD = \frac{1}{2}\left(\frac{1}{n_A}\sum_{a \in surf(A)} d(a, G) + \frac{1}{n_G}\sum_{g \in surf(G)} d(g, A)\right) \tag{6}$$

where $surf(A)$, $surf(G)$ denote the surface of the segmentation result A, ground-truth G, respectively. The n_A, n_G are the total number points in the segmentation and ground-truth surfaces, and $d(a, G)$ measures the closest the Euclidean distance from a point a on the boundary of surface A to the surface G. A smaller value of ASD represents a superior segmentation accuracy.

4.3. Training

The proposed network is trained with a 12GB Titan X GPU and implemented with Caffe framework [43,21]. The intensity of the T1 and T2 images was normalized to mean of zero and variance of one. Owing to the limited memory of the Titan X, we randomly crop $64 \times 64 \times 64$ sub-volume samples for input to the network. We used the approach that introduced by He et al. [44] for weights initialization. Then, the weights were optimized by Adam [45] optimizer with a batch size of four. We used the initialized learning rate as 0.0002 and was decreased ten times every 50,000 iterations. We regularized the proposed model with a weight decay of 0.0005. The total number of iteration was 200,000. The growth rate $k = 16$ of dense block were chosen to be a small integer for parameter efficiency small for parameter efficiency. To obtain the above hyperparameters, we first trained the network with nine subjects with the ground-truth label and used the one subject as a validation. After obtaining the appropriate hyperparameters, the final model was trained using all ten subjects. The final prediction is obtained by the majority voting strategy on the results of overlapping with a stride of 8. The training process required approximately two days and the inference process takes five minutes for each subject. The implementation of proposed 3D-SkipDenseSeg is available at the website[1] for further analysis.

4.4. Performance evaluation

4.4.1. Quantitative evaluation

Table 2 reports the accuracy metrics on a validation image, depths of the networks, and the number of learnable parameters in the compared models. All models were trained with the nine subjects with ground-truth labels. From the table, it is clear that the proposed architecture, 3D-SkipDenseSeg, outperformed the existing architectures for six out of nine metrics. Averaging the DSC accuracy metrics over the brain tissue classes, we note that 3D-Unet achieved 91.58%, VoxResNet achieved 91.59%, and DenseVoxNet

[1] https://github.com/tbuikr/3D_DenseSeg

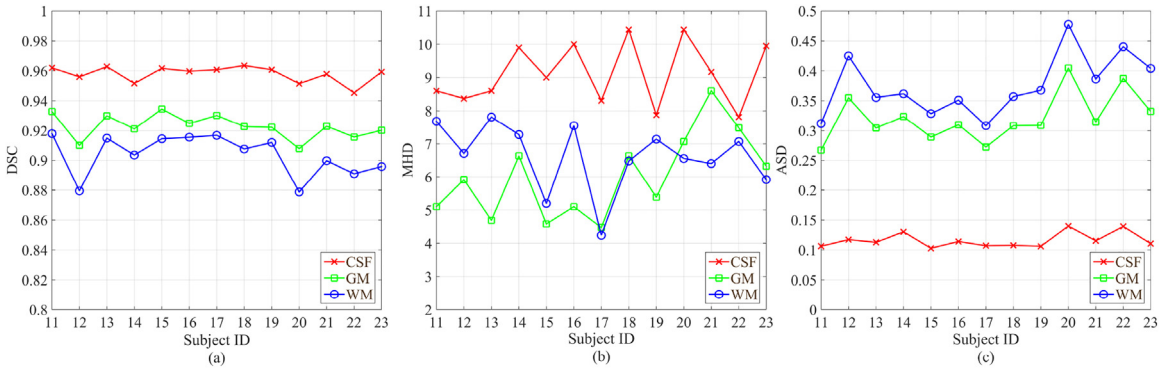

Fig. 4. Performance of 3D-SkipDenseSeg on different subjects of test set. (a) DSC, (b) MHD, (c) AHD.

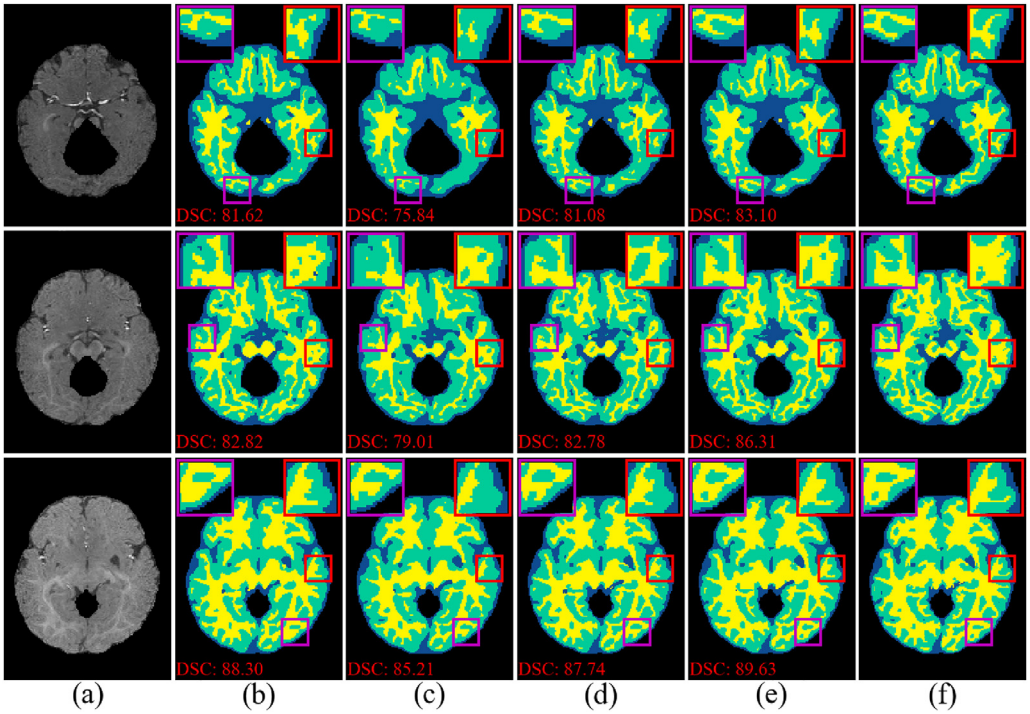

Fig. 5. Segmentation results on different slices of validation sample (a) T1 image, (b) VoxResNet, (c) DenseVoxNet, (d) 3D-Unet, (e) 3D-SkipDenseSeg, and (f) manual segmentation.

achieved 89.23%, whereas the proposed 3D-SkipDenseSeg achieved 92.51%, outperforming the others. Furthermore, the proposed 3D-SkipDenseSeg had a significantly deeper architecture (47 layers) compared to the others with only 1.55 million parameters to learn. For comparison, a similar DenseNet-based method, DenseVoxNet, had 32 layers and 4.34 million parameters, which is approximately three times more parameters than the proposed approach with less depth. Thus, the proposed method increased the depth with efficient usage of the parameters and achieved highly accurate segmentation results particularly for data-sparse applications (i.e. infant brain tissue segmentation).

Table 3 presents the accurate segmentation on the 13 subjects on the test set of the iSeg-2017 dataset. As mentioned in Section 4.1, all participating teams in the challenge were allowed to evaluate the test accuracy only twice. For brevity, Table 3 lists only the results of the top five teams.[2] From the table, it is clear that the proposed 3D-SkipDenseSeg achieved the top accuracy again for six out of nine metrics over the 21 participating teams, which is consistent with our validation results. Fig. 4 illustrates the accuracy metric values of the proposed method for the 13 subjects in the test set. We note that the DSC and AHD values for all three WM, GM, and CSF have low variance, whereas MHD has a relatively higher variance. We suspect that this result is due to the sensitivity in computing the distances to the surfaces caused by the low contrast tissues between different classes. Improving MHD could be a topic for future research.

4.4.2. *Qualitative evaluation*

Fig. 5 displays the tissue segmentation for three different slices of a validation subject image. We also display the original T1 image and segmentation results of the competing methods. The magnified views of the segmentation of the rectangular regions are also displayed at the bottom of each image. From the figures, it shows that the proposed method achieved the highest DSC for all three slides and captured the details of the ground-truth labels considerably better than the other methods. Fig. 6 displays the 3D visualization of the WM surfaces obtained using the existing ITK-SNAP tool [46]. The first row of the figure illustrates the entire volume of the WM

[2] The full table can be found in the challenge website: http://iseg2017.web.unc.edu/rules/results/ for the first round evaluation and http://iseg2017.web.unc.edu/evaluation-on-the-second-round-submission/ for the second round evaluation.

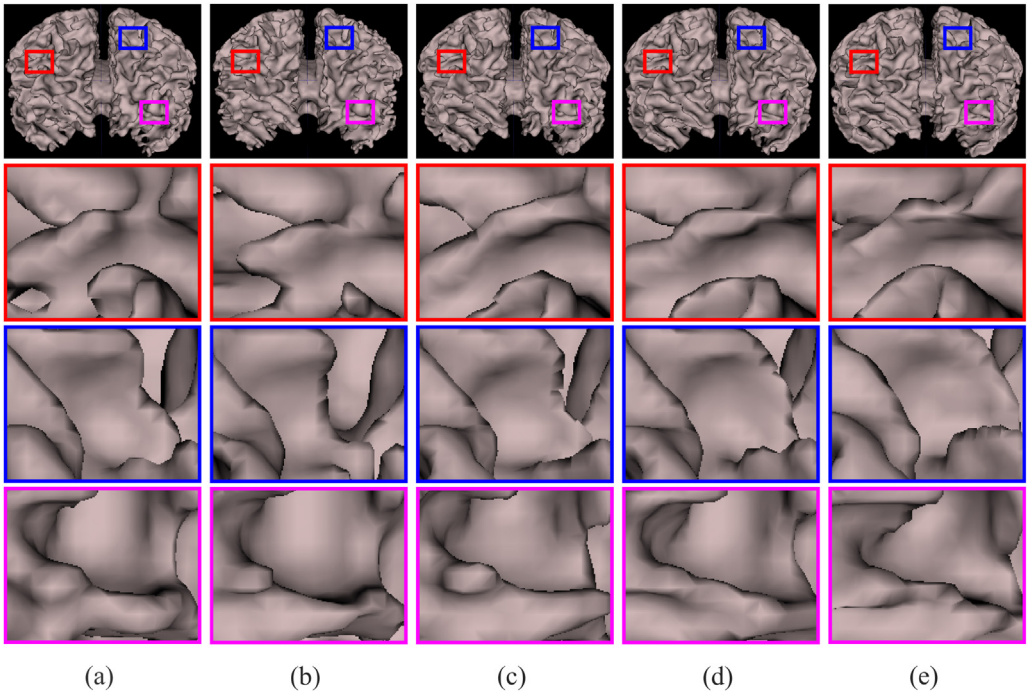

Fig. 6. 3D visualization of WM for validation sample (a) VoxResNet, (b) DenseVoxNet, (c) 3D-Unet, (d) 3D-SkipDenseSeg, (e) manual segmentation.

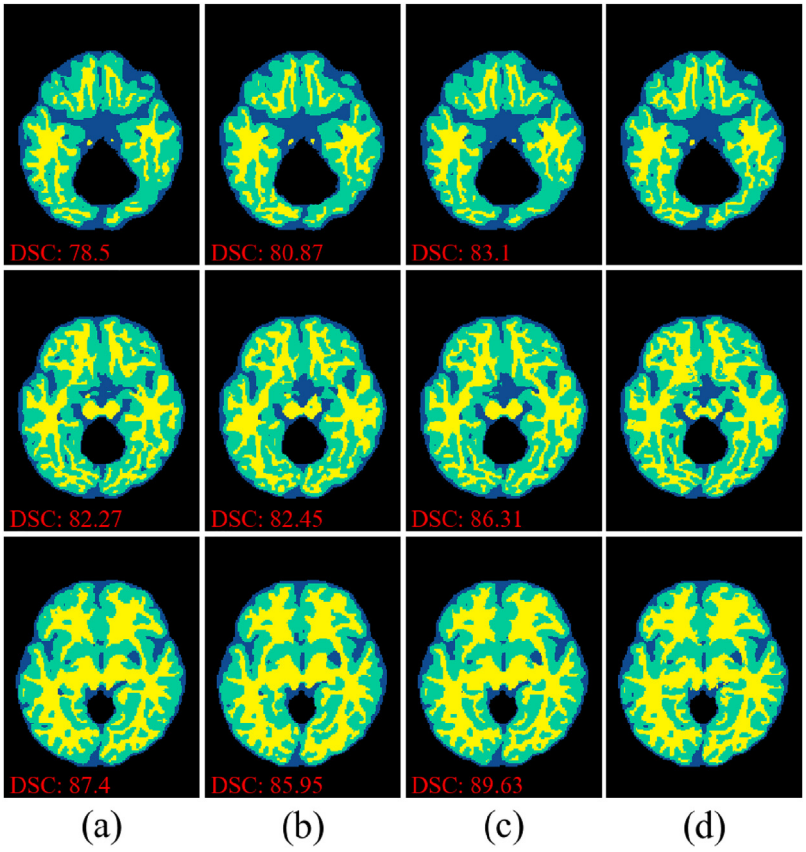

Fig. 7. Segmentation result of proposed method on different slices using (a) T1 only, (b) T2 only, (c) T1 and T2, and (d) manual segmentation.

region, whereas the second, third, and forth rows illustrate the partially enlarged views of the corresponding regions. We can again clearly observe that the proposed 3D-SkipDenseSeg captured the details of the ground-truth structures better than the baselines.

5. Discussion

In this section, we perform ablation studies to justify our modeling choices on dropout, multimodal data fusion, growth rate k in

Table 4
Segmentation accuracy (DSC: %) of proposed method under different settings.

Method	Params	WM	GM	CSF	Avg. DSC
$k = 8$, conv. w/ $s = 2$	0.69M	90.05	90.60	94.13	91.59
$k = 16$, max-pooling	1.48M	90.54	90.92	94.16	91.87
$k = 16$, conv. w/ $s = 2$	1.55M	89.83	91.04	**94.94**	91.94
$k = 16$, conv. w/ $s = 2$, dropout	1.55M	**91.02**	**91.64**	94.88	**92.51**

Note: Bold values indicate best performance.

the dense block, usage of convolution layer in the transition layer, and level of skip-connections.

Fig. 7 presents the accurate segmentation of the proposed network architecture that evaluates on the same slices as Fig. 5; however, with different input data, namely, T1-only, T2-only, and both T1 and T2. We can observe the clear advantage of using multimodal inputs as they achieved significantly higher DSC values compared to using the unimodal image.

Table 4 presents the ablation study results on our validation subject image that justified our choice of using the growth rate $k = 16$ in each dense block and convolution layer with a stride of two rather than the max-pooling layer in each transition block. We can observe that as we increased k, the accuracy improved; however, the number of parameters also increased. We set $k = 16$ because increasing k beyond 16 did not improve the segmentation accuracy; it did, however, significantly increase the memory usage and training time. Furthermore, we observed that replacing the max-pooling layer by the convolution layer with kernel of two and a stride of two marginally improved the accuracy by preserving the spatial information in each transition layer. Moreover, as indicated in Fig. 8, using dropout further improved the accuracy.

To investigate the benefit of using the skip-connected multilevel contextual information, we performed experiments on the validation set with multiple combinations of skip-connections as indicated in Table 5. The notation of $L_1, \ldots, L_5$ is defined in Fig. 3. From the table, it is clear that skip-connecting the low-level feature maps improved the accurate segmentation. In particular, concatenating the feature maps of the first dense block, i.e., L_2, improved DSC from 91.2% to 92.66%, which is a significant improvement. Fig. 9 displays the qualitative segmentation results for different combinations of skip-connections. We note that whereas using only the higher-level feature maps resulted in coarse segmentation, combining these with lower layer feature maps successfully captured the details of the structures of the ground-truth segmen-

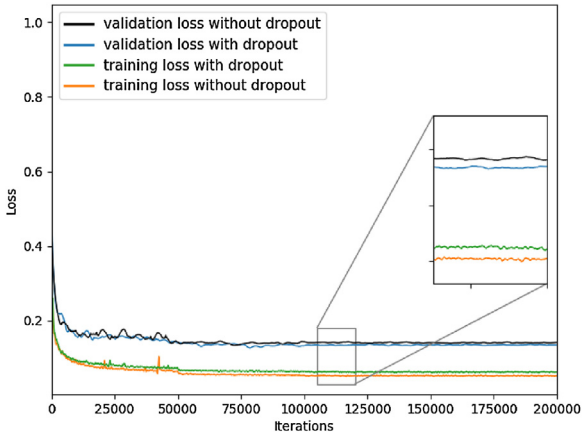

Fig. 8. Training and validating of proposed model using Adam optimization.

Table 5
Effect of concatenating multilevel feature maps using $k = 16$.

Method	Params	WM	GM	CSF	Avg. DSC
L_5	1.48M	81.67	83.05	86.5	83.74
$[L_5, L_4]$	1.53M	90.41	90.8	92.81	91.34
$[L_5, L_4, L_3]$	1.54M	89.49	90.72	93.4	91.2
$[L_5, L_4, L_3, L_2]$	1.55M	**91.25**	**91.77**	**94.95**	**92.66**
$[L_5, L_4, L_3, L_2, L_1]$	1.55M	91.02	91.64	94.88	92.51

Note: Bold values indicate best performance.

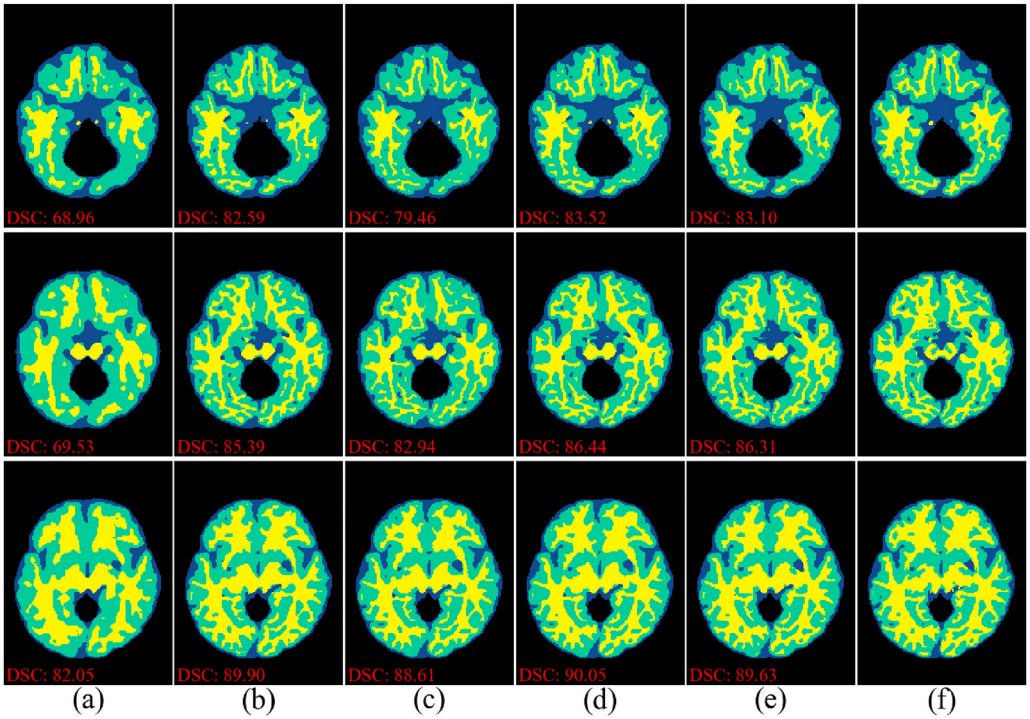

Fig. 9. Segmentation result – proposed method with concatenation of different levels of contextual information (a) L5 only, (b) L5 and L4, (c) L5, L4 and L3, (d) L5, L4, L3 and L2, (e) L5, L4, L3, L2 and L1, and (f) manual segmentation.

tations. Note also that our final model for the test set included L_1, although it resulted in reduced validation accuracy because it achieved marginally higher test accuracy.

6. Conclusion

We proposed a novel 3D fully convolutional, skip-connected DenseNet architecture that addresses the challenges in volumetric infant brain MRI segmentation. The proposed 3D-SkipDenseSeg allows the combination of fine and coarse feature maps via skip connections. The proposed network architecture is deeper than the existing methods and efficiently utilizes the parameters; thus, it achieves superior predictive power for applications with limited training data. We have investigated the efficient usage of $2 \times 2 \times 2$ convolution with to preserve the spatial information, which could be necessary for improved segmentation. We also utilized the multimodal input images, which was extremely effective for segmentation accuracy. The proposed method provided consistent segmentation accuracy across different subjects, thus, it is suitable for large-scale studies. The promising results of the proposed method can provide comprehensive information for doctors making diagnoses on early brain developments. Moreover, we also extended to apply for other brain phases such as early adult-like phases [47], and achieved competitive results in comparison with existing methods.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported partly by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1D1A1B03031752) and partly by the NRF grant funded by the Korea government (MSIT) (NRF-2018R1C1B6007462). Also this research was supported partly by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2018-0-01798) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

[1] L. Wang, F. Shi, G. Li, Y. Gao, W. Lin, J.H. Gilmore, D. Shen, Segmentation of neonatal brain MR images using patch-driven level sets, NeuroImage 84 (2014) 141–158.
[2] L. Wang, F. Shi, W. Lin, J.H. Gilmore, D. Shen, Automatic segmentation of neonatal images using convex optimization and coupled level sets, NeuroImage 58 (3) (2011) 805–817.
[3] F. Shi, Y. Fan, S. Tang, J.H. Gilmore, W. Lin, D. Shen, Neonatal brain image segmentation in longitudinal MRI studies, Neuroimage 49 (1) (2010) 391–400.
[4] M.J. Cardoso, A. Melbourne, G.S. Kendall, M. Modat, N.J. Robertson, N. Marlow, S. Ourselin, Adapt: an adaptive preterm segmentation algorithm for neonatal brain MRI, NeuroImage 65 (2013) 97–108.
[5] F. Shi, P.-T. Yap, Y. Fan, J.H. Gilmore, W. Lin, D. Shen, Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation, Neuroimage 51 (2) (2010) 684–693.
[6] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S.J. Counsell, V. Doria, A. Serag, I.S. Gousias, J.P. Boardman, M.A. Rutherford, A.D. Edwards, et al., A dynamic 4D probabilistic atlas of the developing brain, NeuroImage 54 (4) (2011) 2750–2763.
[7] I.S. Gousias, A.D. Edwards, M.A. Rutherford, S.J. Counsell, J.V. Hajnal, D. Rueckert, A. Hammers, Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants, Neuroimage 62 (3) (2012) 1499–1509.
[8] A. van Opbroek, F. van der Lijn, M. de Bruijne, Automated Brain-Tissue Segmentation by Multi-Feature SVM Classification, 2013.
[9] L. Wang, Y. Gao, F. Shi, G. Li, J.H. Gilmore, W. Lin, D. Shen, Links: learning-based multi-source integration framework for segmentation of infant brain images, NeuroImage 108 (2015) 160–172.
[10] S. Pereira, A. Pinto, J. Oliveira, A.M. Mendrik, J.H. Correia, C.A. Silva, Automatic brain tissue segmentation in mr images using random forests and conditional random fields, J. Neurosci. Methods 270 (2016) 111–123.
[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[13] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems (2012) 1097–1105.
[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016) 770–778.
[15] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, D. Shen, Deep convolutional neural networks for multi-modality isointense infant brain image segmentation, NeuroImage 108 (2015) 214–224.
[16] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[17] P. Moeskops, M.A. Viergever, A.M. Mendrik, L.S. de Vries, M.J. Benders, I. Išgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE Trans. Med. Imaging 35 (5) (2016) 1252–1261.
[18] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3D convolutional networks, Proceedings of the IEEE International Conference on Computer Vision (2015) 4489–4497.
[19] M.F. Stollenga, W. Byeon, M. Liwicki, J. Schmidhuber, Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation, Advances in Neural Information Processing Systems (2015) 2998–3006.
[20] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V.C. Mok, L. Shi, P.-A. Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging 35 (5) (2016) 1182–1195.
[21] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[22] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016, pp. 565–571.
[23] S. Andermatt, S. Pezold, P. Cattin, Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data, in: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2016, pp. 142–151.
[24] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78.
[25] J. Dolz, C. Desrosiers, I.B. Ayed, 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study, NeuroImage (2017).
[26] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, Aistats, vol. 9 (2010) 249–256.
[27] H. Chen, Q. Dou, L. Yu, J. Qin, P.-A. Heng, VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images, NeuroImage (2017).
[28] L. Yu, X. Yang, H. Chen, J. Qin, P.A. Heng, Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images, Thirty-First AAAI Conference on Artificial Intelligence (2017).
[29] A. Fakhry, T. Zeng, S. Ji, Residual deconvolutional networks for brain electron microscopy image segmentation, IEEE Trans. Med. Imaging 36 (2) (2017) 447–456.
[30] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017).
[31] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, 2016, arXiv preprint arXiv:1611.09326.
[32] L. Yu, J.-Z. Cheng, Q. Dou, X. Yang, H. Chen, J. Qin, P.-A. Heng, Automatic 3D cardiovascular MR segmentation with densely-connected volumetric convnets, MICCAI, 2017 (2017).
[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016) 2818–2826.
[34] L. Wang, D. Nie, G. Li, É. Puybareau, J. Dolz, Q. Zhang, F. Wang, J. Xia, Z. Wu, J. Chen, et al., Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge, IEEE Trans. Med. Imaging (2019).
[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.
[36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, 2016, Preprint arXiv:1606.00915.
[37] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2017) 2881–2890.
[38] J. Dolz, I.B. Ayed, J. Yuan, C. Desrosiers, Hyperdense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Semantic Segmentation, 2017, Preprint arXiv:1710.05956.

[39] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015, Preprint arXiv:1502.03167.
[40] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[41] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC Med. Imaging 15 (2015) 29.
[42] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 850–863.
[43] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.
[44] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, Proceedings of the IEEE International Conference on Computer Vision (2015) 1026–1034.
[45] D. Kingma, J. Ba, Adam:, A Method for Stochastic Optimization, 2014, Preprint arXiv:1412.6980.
[46] P.A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, Neuroimage 31 (3) (2006) 1116–1128.
[47] A.M. Mendrik, K.L. Vincken, H.J. Kuijf, M. Breeuwer, W.H. Bouvy, J. De Bresser, A. Alansary, M. De Bruijne, A. Carass, A. El-Baz, et al., MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans, Comput. Intell. Neurosci. 2015 (2015) 1.