FROM NEONATAL TO ADULT BRAIN MR IMAGE SEGMENTATION IN A FEW SECONDS USING 3D-LIKE FULLY CONVOLUTIONAL NETWORK AND TRANSFER LEARNING

Yongchao Xu [1,2], Thierry Géraud [1]   *Isabelle Bloch [2]*

[1] EPITA Research and Development Laboratory
14-16, rue Voltaire
FR-94270 Le Kremlin-Bicêtre, France

[2] LTCI, Télécom ParisTech
Université Paris-Saclay
FR-75013, Paris, France

ABSTRACT

Brain magnetic resonance imaging (MRI) is widely used to assess brain development in neonates and to diagnose a wide range of neurological diseases in adults. Such studies are usually based on quantitative analysis of different brain tissues, so it is essential to be able to classify them accurately. In this paper, we propose a fast automatic method that segments 3D brain MR images into different tissues using fully convolutional network (FCN) and transfer learning. As compared to existing deep learning-based approaches that rely either on 2D patches or on fully 3D FCN, our method is way much faster: it only takes a few seconds, and only a single modality (T1 or T2) is required. In order to take the 3D information into account, all 3 successive 2D slices are stacked to form a set of 2D "color" images, which serve as input for the FCN pre-trained on ImageNet for natural image classification. To the best of our knowledge, this is the first method that applies transfer learning to segment both neonatal and adult brain 3D MR images. Our experiments on two public datasets show that our method achieves state-of-the-art results.

Index Terms— Brain MRI, Neonatal/Adult brain segmentation, Deep learning, Fully convolutional network.

1. INTRODUCTION

Accurate automatic segmentation of brain magnetic resonance (MR) images into different tissues – cortical gray matter (CoGM), basal ganglia and thalami (BGT), (unmyelinated) white matter ((u)WM), brainstem (BS), cerebellum (CB), ventricles (Vent), and cerebrospinal fluid in the extra-cerebral space (CSF) – is of significant value for assessing neonatal brain development and diagnosing many neurological diseases. This problem remains challenging due to the size and the complexity of 3D brain MR volumes. In the case of neonates, it is even more challenging because there is a significant partial volume effect, and because of the poor contrast between white and gray matters. This makes particularly difficult the development of a universal method for both neonates and adults.

In this paper, we propose a fast versatile method based on fully convolutional network (FCN) [1], transfer learning,
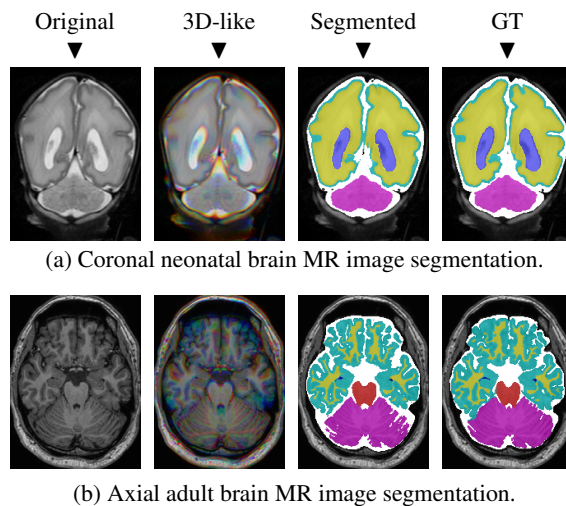

(a) Coronal neonatal brain MR image segmentation.

(b) Axial adult brain MR image segmentation.

Fig. 1: Some examples of segmentation results (obtained for a whole 3D MRI volume in a few seconds).

and a novel formulation of *3D-like* images from 3D volumes. Simply put, we stack successive 2D slices of a 3D volume to form a set of 2D "color" images (see the 2nd column of Fig. 1); these 2D images constitute the input of a FCN based on VGG network [2], pre-trained on the ImageNet dataset. We discard the fully connected layers, and we add specialized convolutional layers at the end of each of the five convolutional stages in VGG network. A linear combination of these specialized layers (i.e. fine to coarse feature maps) results in the final segmentation. Some examples are depicted in Fig. 1.

As compared to some related works, the main contributions of this paper are: **1.** A fast automatic method that segments a whole 3D MRI volume in a few seconds (2 to 6s), thanks to the 3D-like approach; **2.** A versatile framework that works for both neonates and adults; **3.** The use of transfer learning to segment whole 3D MRI volumes, so only a very few training images are required (e.g., only one training image was used to get the result depicted in Fig. 1(a)).

As we advocate reproducible research, all information to reproduce our results are publicly available from [3], along
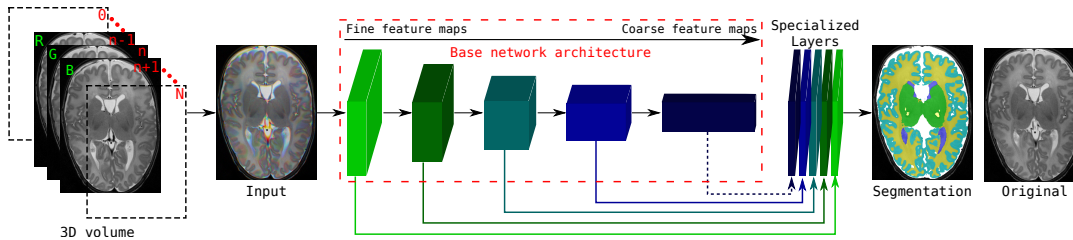
Fig. 2: Architecture of the proposed network. We fine tune it and combine linearly fine to coarse feature maps of the pre-trained VGG network [2]. The coarsest feature maps are discarded for the adult images.

with additional result images and comparative tables.

2. RELATED WORKS

Brain MRI segmentation methods can be roughly divided into three categories: 1) Atlas-based methods, which are very popular in the neonatal case [4, 5, 6]; unfortunately, they usually involve a computationally expensive registration process. 2) Classical image processing and pattern recognition methods [7, 8, 9, 10]; they are based on hand-crafted features, which may result in less accurate segmentation. 3) Methods based on deep learning, more particularly convolutional neural networks (CNN) [11]. Some examples are: [12] for segmenting WM, GM, and CSF; [13] for hippocampus segmentation; [14] for whole brain segmentation. These methods are based either on 2D patches, on a 2.5D technique, or on 3D patches. In [15], the authors use three individual FCNs respectively on three input modalities, and merge their results. The work in [16] relies on a fully 3D deep residual network [17]. All these methods are not so fast due to the use of patches, of multi-networks, or of a 3D network; in addition, they are trained from scratch, which usually requires a large training set. More references for neonatal and adult brain MR image segmentation can be found in [18, 19].

Deep learning methods have shown promising results in various medical imaging problems [20]. Many works rely on 2.5D patches, i.e., 2D orthogonal patches in the sagittal, coronal and axial planes. Two examples are [21] for headneck region image segmentation, and [22] for knee cartilage segmentation, where three individual CNNs are applied on the tri-planar patches. The work in [23], focusing on lymph node detection in CT images, extends the idea of [22] by forming color patches from the tri-planar patches, and training a CNN on them.

The usefulness of a network pre-trained on ImageNet dataset for medical imaging problems has also been demonstrated. In [24], for horaco-abdominal lymph node detection and interstitial lung disease classification, the authors extract 2.5D patches, form color images, and resize them in order to accommodate a network pre-trained on ImageNet. In [25], the authors use the two connected layers of DeCAF [26] as descriptors for chest pathology detection. Besides, in [27], the authors have also demonstrated that using transfer learning that fine tunes a network pre-trained on ImageNet dataset

greatly improves some segmentation problems in medical imaging. They have used a FCN and combined fine to coarse feature maps to analyze color retinal images.

3. PROPOSED METHOD

Pre-processing. Our segmentation method requires only one modality: a T2 (resp. T1) sequence of brain MRI for neonates (resp. for adults). We begin with an histogram equalization of the whole MR volume; this very ordinary image normalization can be achieved instantly.

Deep FCN for brain MR image segmentation. Efficient natural image segmentation can be achieved thanks to deep fully convolutional network (FCN) and transfer learning [1]. In this paper, we propose to rely on this same method to segment 3D brain MR images, although those images are very different from natural images. We rely on the 16 layers VGG network [2] pre-trained on millions of natural images in ImageNet for image classification. For our application, we discard the fully connected layers at the end of VGG network, and keep the 5 stages of convolutional parts called "*base network*". This base network is mainly composed of convolutional layers: $z_i = w_i \times x + b_i$, Rectified Linear Unit (ReLU) layers for non linear activation function: $f(z_i) = \max(0, z_i)$, and max pooling layers between two successive stages, where x is the input of each convolutional layer, w_i is the convolution parameter, and b_i is the bias term. The four max pooling layers divide the base network into five stages of fine to coarse feature maps. Inspired by the work in [1, 27], we add specialized convolutional layers (with a 3×3 kernel size) with K (e.g. $K = 16$) feature maps after the convolutional layers at the end of each stage. We resize all the specialized layers to the original image size, and concatenate them together. A last convolutional layer with kernel size 1×1 is appended at the end that combines linearly the fine to coarse feature maps in the concatenated specialized layers, to produce the final segmentation result. The proposed network architecture is depicted in Fig. 2. For the neonatal images having large BGT regions, we use all the specialized layers. For the adult images having relatively small BGT regions, we discard the coarsest feature maps from the last stage.

The architecture described above is very similar with the one used in [27] for retinal image analysis, where the retinal images are already 2D color images. For our application, the

Dataset	Image set	# voxels	Size (mm³)	Experiment (# images)	Code	Results		Timing
NeoBrain12	Axial / 40 weeks	512×512×50	0.35×0.35×2.0	2 training / 5 test T2	EXP1	Fig. 3(a)	Table 2 (top)	3.5 s
	Coronal / 30 weeks	384×384×50	0.34×0.34×2.0	2 training / 5 test T2	EXP2	Fig. 3(b)	Table 2 (mid.)	2.2 s
	Coronal / 40 weeks	512×512×110	0.35×0.35×1.2	2 + 2 training / 5 test T2	EXP3	Fig. 3(c)	Table 2 (bot.)	6.5 s
MRBrainS13	Axial / 70 years	240×240×48	0.96×0.96×3.0	5 training / 15 test T1	EXP4	Fig. 3(d)	Table 3	1.7 s

Table 1: A summary of the training + test datasets and of the different conducted experiments.

question amounts to how to prepare appropriate inputs given that a brain MR image is a 3D volume. To get RGB input images, we propose to stack successive 2D slices. Precisely, to form an input artificial color image for the pre-trained network to segment the nth slice, we use the slices $n - 1$, n and $n + 1$ as respectively the red, green and blue channels. If the slice $n - 1$ or $n + 1$ is out of the MR volume, an image with constant 0 is used instead. This process is depicted in Fig. 2 (left). Each 2D color image thus forms a *3D-like* representation of a part (3 slices) of the MR volume. This representation enables us to incorporate some 3D information, while avoiding the expensive computational and memory requirements of fully 3D FCN. This novel idea is a main contribution of the paper.
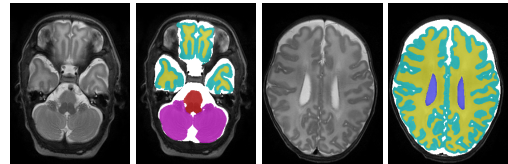
For the training phase, we use the multinomial logistic loss function for a one-of-many classification task, passing real-valued predictions through a softmax to get a probability distribution over classes. During training, we use the classical data augmentation strategy by scaling and rotating, and also subtract the mean value for each channel in the training images. We fine tune the entire network for 100k iterations using a very small learning rate ($lr = 10^{-10}$ for neonates, and $lr = 10^{-8}$ for adults). We rely on stochastic gradient descent with momentum to minimize the loss function with $momentum = 0.99, weight\_decay = 0.005$. The loss function is averaged over 20 images.

At test time, after having pre-processed the 3D volume (histogram equalization), we prepare the set of 3D-like color images. Then we subtract the mean values for each channel, and pass every image through the network.
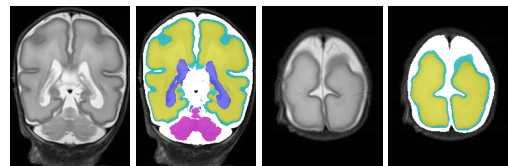
Post-processing. During inference, a region R of each slice that is small or does not superpose importantly with the brain mask M (e.g. $|R| < 5$ mm^2 or $|R \cap M|/|R| < 0.8$) is automatically considered as background. For each brain MR image, the brain mask M is roughly given by the largest region in the inferred result for the middle slice of the MR volume.
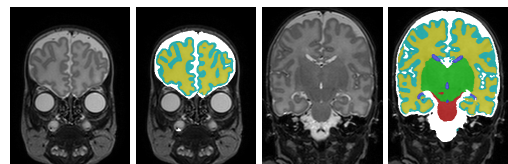
4. EXPERIMENTAL RESULTS

We have evaluated the proposed method on two types of brain MR images: the dataset of the MICCAI challenge of Neonatal Brain Segmentation 2012 (NeoBrainS12) [28], and the dataset of the MICCAI challenge of MR Brain Image Segmentation (MRBrainS13) [29]. A summary of the two datasets (along with the different experiments we have conducted) are depicted in Table 1; more details are available from [28, 29], and from the dedicated web page [3] support-
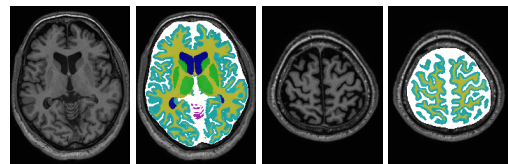

(a) On an axial neonatal image acquired at 40 weeks.

(b) On a coronal neonatal image acquired at 30 weeks.

(c) On a coronal neonatal image acquired at 40 weeks.

(d) On an adult image in MRBrainS13 dataset.

| CoGM | BGT | UWM | MWM/WML | BS | CB | Vent | CSF |

Fig. 3: Some qualitative segmentation results.

ing this present paper. These datasets are interesting because they are considered as references, and they serve as an ongoing benchmark for evaluating algorithms. Given an automatic segmentation S and a reference segmentation G of a corresponding tissue, the quantitative evaluation is based on the Dice coefficient, defined as $DC = 2 \times |S \cap G| / (|S| + |G|)$, a robust modified Hausdorff Distance (HD) [30, 29], and the mean surface distance (MSD) between the boundaries of S and G.

On the NeoBrainS12 dataset, we have conducted three types of experiments as depicted in Table 1. We have used all the two given training images and evaluated on the five test images for the axial images acquired at 40 weeks and respectively coronal images at 30 weeks. Some segmentation results are depicted in Fig. 3(a) and Fig. 3(b). The

Code	Method	CoGM DC	CoGM MSD	BGT DC	BGT MSD	UWM DC	UWM MSD	BS DC	BS MSD	CB DC	CB MSD	Vent DC	Vent MSD	CSF DC	CSF MSD
EXP1	*Our*	**0.87**	**0.11**	0.91	0.51	**0.93**	**0.11**	**0.85**	0.49	**0.94**	0.33	**0.87**	0.24	**0.83**	**0.20**
	UPF_SIMBioSys [6]	0.85	0.15	**0.93**	**0.29**	0.91	0.17	**0.85**	**0.15**	**0.94**	**0.28**	0.83	0.44	0.79	0.29
	UNC-IDEA [9]	0.86	**0.11**	0.92	0.33	0.92	0.13	0.83	0.27	0.92	0.45	0.79	0.25	0.79	0.25
	5 next (median)	0.84	0.18	0.88	0.62	0.88	0.25	0.79	0.69	0.91	0.53	0.81	0.32	0.73	0.54
EXP2	*Our*	**0.79**	**0.14**	0.89	0.42	**0.95**	**0.14**	0.84	0.37	0.91	**0.30**	0.87	0.33	**0.89**	**0.13**
	UPF_SIMBioSys [6]	0.75	0.16	**0.90**	**0.38**	0.93	0.22	**0.86**	**0.32**	**0.92**	0.31	**0.88**	0.25	0.85	0.17
	CIMAT_Team	0.69	0.26	0.89	0.41	0.93	0.28	-	-	-	-	0.82	**0.22**	0.82	0.22
	5 next (median)	0.60	0.38	0.82	0.77	0.87	0.46	0.71	0.88	0.87	0.40	0.86	0.41	0.74	0.46
EXP3	*Our*	**0.79**	**0.21**	0.86	0.98	**0.91**	**0.18**	0.68	1.13	0.89	0.65	0.82	0.41	**0.82**	**0.30**
	MorphoSeg [7]	0.77	**0.21**	0.86	0.96	0.89	0.24	0.72	0.95	**0.91**	**0.55**	0.78	0.39	0.78	0.39
	UPF_SIMBioSys [6]	0.73	0.27	**0.89**	**0.52**	0.87	0.30	**0.76**	**0.53**	**0.91**	0.59	**0.85**	**0.34**	0.72	0.55
	5 next (median)	0.72	0.28	0.87	0.87	0.85	0.35	0.73	0.84	0.91	0.62	0.81	0.48	0.71	0.59

Table 2: Quantitative evaluation of the presented method on the NeoBrainS12 dataset.

proposed method segments accurately the neonatal brain into different tissues. Although some small segmentation errors exist (see the second image in Fig. 3(b) and [3]), this does not significantly negatively affect the result. Some quantitative results using the accompanied on-line evaluation system (see the URL in [28]) based on Dice coefficient and mean surface distance are given in Table 2 (top) and Table 2 (middle). Clearly, the proposed method achieves state-of-the-art results. Note that there are no training images for the coronal images acquired at 40 weeks; see Fig. 3(c). Alternatively, we trained on the two axial training images at 40 weeks having similar gray levels and on the two coronal training images at 30 weeks featuring similar geometrical structures; see the EXP3 part of Table 1. The segmentation results, given in Fig. 3(c) and Table 2 (bottom), are also very accurate, which demonstrates the robustness of the proposed method. The total training time for each model is about five hours for 100k iterations using one NVIDIA GeForce GTX 1080 GPU. The average runtime during test for a whole 3D MRI volume is depicted in the rightmost column of Table 2.

For the adult case, on the MRBrainS13 dataset [29], we have trained a model on the 5 training images, and evaluated the model on the 15 test images. Some qualitative results are depicted in Fig. 3(d), where one can see that the segmentation is very accurate. We have quantitatively evaluated the segmentation results on this dataset based on its on-line evaluation system (see the URL in [29]). The evaluation is performed by grouping CoGM and BGT together into gray matter (GM), UWM and WML into white matter (WM), and Vent and CSF into cerebrospinal fluid (CSF'). The top ten results (among the 38 submitted methods) are depicted in Table 3. The results in Table 3 are sorted in decreasing performance order (given by the on-line evaluation system). Our proposed method ranks the second best. The average runtime is 1.7 seconds, which contrasts with the 2 minutes required by the best method. Besides, only one sequence (T1) is required in our case. Note that the symbol ∗ in Table 3 denotes "T1, T1_IR, and FLAIR sequences", and that the symbol ∗′ denotes "T1-1mm, T1_IR, and FLAIR sequences". The manual annotation of outer border of the CSF' relies also on T1_IR sequence. This may explain why the proposed method achieves less accurate CSF' segmentation. We would expect to improve this

Method for EXP4	GM DC	GM HD	WM DC	WM HD	CSF' DC	CSF' HD	Seq	Timing
CU_DL[16]	**86.12**	**1.47**	**89.39**	1.94	83.96	2.28	*	2 min
Our	86.03	1.44	89.29	1.86	82.44	2.28	T1	**2 s**
MDGRU	85.40	1.55	88.98	2.02	**84.13**	**2.17**	*	2 min
PyraMiD	84.89	1.67	88.53	2.07	83.05	2.30	*	2 min
FBI | LMB	85.44	1.58	88.86	1.95	83.47	2.22	*'	2 min
IDSIA	84.82	1.70	88.33	2.08	83.72	**2.14**	*	2 min
STH	84.77	1.71	88.45	2.34	82.77	2.31	*	5 min
ISI-Neo	85.77	1.62	88.66	2.07	81.08	2.65	T1	**1+ hour**
UNC	84.36	1.62	88.68	2.06	82.81	2.35	*	3 min
MNAB2	84.50	1.70	88.04	2.12	82.30	2.27	*	**25 min**

Table 3: Quantitative evaluation on the MRBrainS13 dataset.

by using both T1 and T1_IR sequences. More results and details about quantitative evaluation of the corresponding methods can be found from the URL given in [29]. The training time on this dataset is about 3 hours for 120k iterations.

5. CONCLUSION

In this paper, we have presented a method to segment brain MR images, relevant from neonates to aging adults, and which runs in a very few seconds. Our major contribution is to rely on transfer learning and on *3D-like* color images obtained by stacking successive MR slices. The proposed method leverages the latest progress on deep learning, and avoids the expensive computational and memory requests of 3D or patch-based approaches. We have achieved or improved the state-of-the art results on two public datasets, while having significantly reduced the segmentation running time. Besides, thanks to transfer learning, the proposed method also obtains promising results even with only one training image. This is especially valuable for medical image segmentation problems, for which a large dataset with precise annotations can be difficult to obtain. In the future, we would like to test the proposed method on young adult brain MR images to confirm its versatility. Another major perspective is to investigate some other ways to form the color images from a 3D volume, and how to make use of multi-modalities. We also would like to apply the proposed method on some other segmentation problems in medical imaging. Last, let us recall that, for reproducible research purpose, we have made available [3] all the information to reproduce our results.

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of CVPR*, 2015, pp. 3431–3440.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[3] Y. Xu, T. Géraud, and I. Bloch, "Extra materials for this paper." http://publications.lrde.epita.fr/xu.17.icip, 2017.

[4] F. Shi, Y. Fan, S. Tang, J. H. Gilmore, W. Lin, and D. Shen, "Neonatal brain image segmentation in longitudinal MRI studies," *NeuroImage*, vol. 49, no. 1, pp. 391–400, 2010.

[5] M. J. Cardoso *et al.*, "AdaPT: an adaptive preterm segmentation algorithm for neonatal brain MRI," *NeuroImage*, vol. 65, pp. 97–108, 2013.

[6] G. Sanroma *et al.*, "Building an ensemble of complementary segmentation methods by exploiting probabilistic estimates," in *Intl. Work. on Machine Learning in Medical Imaging*. Springer, 2016, pp. 27–35.

[7] L. Gui, R. Lisowski, T. Faundez, P. S. Hüppi, F. Lazeyras, and M. Kocher, "Morphology-driven automatic segmentation of MR images of the neonatal brain," *Medical Image Analysis*, vol. 16, no. 8, pp. 1565–1579, 2012.

[8] P. Moeskops *et al.*, "Evaluation of an automatic brain segmentation method developed for neonates on adult MR brain images," in *Proc. of SPIE Medical Imaging*, 2015, p. 941315.

[9] L. Wang, Y. Gao, F. Shi, G. Li, J. H. Gilmore, W. Lin, and D. Shen, "LINKS: Learning-based multi-source IntegratioN frameworK for segmentation of infant brain images," *NeuroImage*, vol. 108, pp. 160–172, 2015.

[10] B. Morel, Y. Xu, A. Virzi, T. Géraud, C. Adamsbaum, and I. Bloch, "A challenging issue: Detection of white matter hyperintensities on neonatal brain MRI," in *Proc. of EMBC*, 2016, pp. 93–96.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[13] M. Lai, "Deep learning for medical image segmentation," https://arxiv.org/pdf/1505.02000, 2015.

[14] P. Moeskops *et al.*, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.

[15] D. Nie, L. Wang, Y. Gao, and D. Sken, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. of ISBI*, 2016, pp. 1342–1345.

[16] H. Chen, Q. Dou, L. Yu, and P. Heng, "VoxResNet: Deep voxelwise residual networks for volumetric brain segmentation," https://arxiv.org/abs/1608.05895, 2016.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of CVPR*, 2016, pp. 770–778.

[18] C. N. Devi, A. Chandrasekharan, V. Sundararaman, and Z. C. Alex, "Neonatal brain MRI segmentation: A review," *Computers in Biology and Medicine*, vol. 64, pp. 163–178, 2015.

[19] I. Despotović, B. Goossens, and W. Philips, "MRI segmentation of the human brain: Challenges, methods, and applications," *Computational and Mathematical Methods in Medicine*, vol. 2015, 2015.

[20] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial—Deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. on Medical Imaging*, vol. 35, no. 5, pp. 1153–1158, 2016.

[21] K. Fritscher *et al.*, "Deep neural networks for fast segmentation of 3D medical images," in *Proc. of MICCAI*, vol. 2, 2016, pp. 158–165.

[22] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proc. of MICCAI*, 2013, pp. 246–253.

[23] H. R. Roth *et al.*, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *Proc. of MICCAI*, 2014, pp. 520–527.

[24] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[25] Y. Bar *et al.*, "Chest pathology detection using deep learning with non-medical training," in *Proc. of ISBI*, 2015, pp. 294–297.

[26] J. Donahue *et al.*, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *ICML*, vol. 32, 2014, pp. 647–655.

[27] K.-K. Maninis *et al.*, "Deep retinal image understanding," in *Proc. of MICCAI*, 2016, pp. 140–148.

[28] I. Išgum *et al.*, "Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge," *Medical Image Analysis*, vol. 20, no. 1, pp. 135–151, 2015.

[29] A. M. Mendrik *et al.*, "MRBrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans," *Computational Intelligence and Neuroscience*, vol. 2015, 2015, doi:10.1155/2015/813696.

[30] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. on PAMI*, vol. 15, no. 9, pp. 850–863, 1993.