
VoxResNet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation

Hao Chen, Qi Dou, Lequan Yu and Pheng-Ann Heng

Department of Computer Science and Engineering

The Chinese University of Hong Kong

Hong Kong, China

hchen@cse.cuhk.edu.hk

Abstract

Recently deep residual learning with residual units for training very deep neural networks advanced the state-of-the-art performance on 2D image recognition tasks, e.g., object detection and segmentation. However, how to fully leverage contextual representations for recognition tasks from volumetric data has not been well studied, especially in the field of medical image computing, where a majority of image modalities are in volumetric format. In this paper we explore the deep residual learning on the task of volumetric brain segmentation. There are at least two main contributions in our work. First, we propose a deep voxelwise residual network, referred as *VoxResNet*, which borrows the spirit of deep residual learning in 2D image recognition tasks, and is extended into a 3D variant for handling volumetric data. Second, an auto-context version of *VoxResNet* is proposed by seamlessly integrating the low-level image appearance features, implicit shape information and high-level context together for further improving the volumetric segmentation performance. Extensive experiments on the challenging benchmark of brain segmentation from magnetic resonance (MR) images corroborated the efficacy of our proposed method in dealing with volumetric data. We believe this work unravels the potential of 3D deep learning to advance the recognition performance on volumetric image segmentation.

1 Introduction

Over the last few years, deep learning especially deep convolutional neural networks (CNNs) have emerged as one of the most prominent approaches for image recognition problems in various domains including computer vision [15, 31, 18, 33] and medical image computing [25, 26, 3, 30]. Most of the studies focused on the 2D object detection and segmentation tasks, which have shown a compelling accuracy compared to previous methods employing hand-crafted features. However, in the field of medical image computing, volumetric data accounts for a large portion of medical image modalities, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), etc. Furthermore, the volumetric diagnosis from temporal series usually requires data analysis even in a higher dimension. In clinical practice, the task of volumetric image segmentation plays a significant role in computer aided diagnosis (CADx), which provides quantitative measurements and aids surgical treatments. Nevertheless, this task is quite challenging due to the high-dimensionality and complexity along with volumetric data. To the best of our knowledge, there are two types of CNNs developed for handling volumetric data. The first type employed modified variants of 2D CNN by taking aggregated adjacent slices [5] or orthogonal planes (i.e., axial, coronal and sagittal) [25, 27] as input to make up complementary spatial information. However, these methods cannot make full use of the contextual information sufficiently, hence it is not able to segment objects from volumetric data accurately. Recently, other methods based on 3D CNN have been developed to detect or

segment objects from volumetric data and demonstrated compelling performance [8, 14, 17]. Nevertheless, these methods may suffer from limited representation capability using a relatively shallow depth or may cause optimization degradation problem by simply increasing the depth of network. Recently, deep residual learning with substantially enlarged depth further advanced the state-of-the-art performance on 2D image recognition tasks [10, 11]. Instead of simply stacking layers, it alleviated the optimization degradation issue by approximating the objective with residual functions.

Brain segmentation for quantifying the brain structure volumes can be of significant value on diagnosis, progression assessment and treatment in a wide range of neurologic diseases such as Alzheimer’s disease [9]. Therefore, lots of automatic methods have been developed to achieve accurate segmentation performance in the literature. Broadly speaking, they can be categorized into three classes: 1) Machine learning methods with hand-crafted features. These methods usually employed different classifiers with various hand-crafted features, such as support vector machine (SVM) with spatial and intensity features [37, 22], random forest with 3D Haar like features [38] or appearance and spatial features [24]. These methods suffer from limited representation capability for accurate recognition. 2) Deep learning methods with automatically learned features. These methods learn the features in a data-driven way, such as 3D convolutional neural network [6], parallelized long short-term memory (LSTM) [32], and 2D fully convolutional networks [23]. These methods can achieve more accurate results while eliminating the need for designing sophisticated input features. Nevertheless, more elegant architectures are required to further advance the performance. 3) Multi-atlas registration based methods [1, 2, 28]. However, these methods are usually computationally expensive, hence limits its capability in applications requiring fast speed.

To overcome aforementioned challenges and further unleash the capability of deep neural networks, we propose a deep voxelwise residual network, referred as *VoxResNet*, which borrows the spirit of deep residual learning to tackle the task of object segmentation from volumetric data. Extensive experiments on the challenging benchmark of brain segmentation from volumetric MR images demonstrated that our method can achieve superior performance, outperforming other state-of-the-art methods by a significant margin.

2 Method

2.1 Deep Residual Learning

Deep residual networks with residual units have shown compelling accuracy and nice convergence behaviors on several large-scale image recognition tasks, such as ImageNet [10, 11] and MS COCO [7] competitions. By using identity mappings as the skip connections and after-addition activation, residual units can allow signals to be directly propagated from one block to other blocks. Generally, the residual unit can be expressed as following

$$x_{l+1} = x_l + \mathcal{F}(x_l, W_l) \quad (1)$$

here the \mathcal{F} denotes the residual function, x_l is the input feature to the l -th residual unit and W_l is a set of weights correspondingly associated with the residual unit. The key idea of deep residual learning is to learn the additive residual function \mathcal{F} with respect to the input feature x_l . Hence by unfolding above equation recursively, the $x_L (L > l \geq 1)$ can be derived as

$$x_L = x_l + \sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i) \quad (2)$$

therefore, the feature x_L of any deeper layers can be represented as the feature x_l of shallow unit l plus summarized residual functions $\sum_{i=l}^{L-1} \mathcal{F}(x_i, W_i)$. The derivations imply that residual units can make information propagate through the network smoothly.

2.2 VoxResNet for Volumetric Image Segmentation

Although 2D deep residual networks have been extensively studied in the domain of computer vision [10, 11, 29, 39], to the best of our knowledge, seldom studies have been explored in the field of medical image computing, where the majority of dataset is volumetric images. In order to leverage

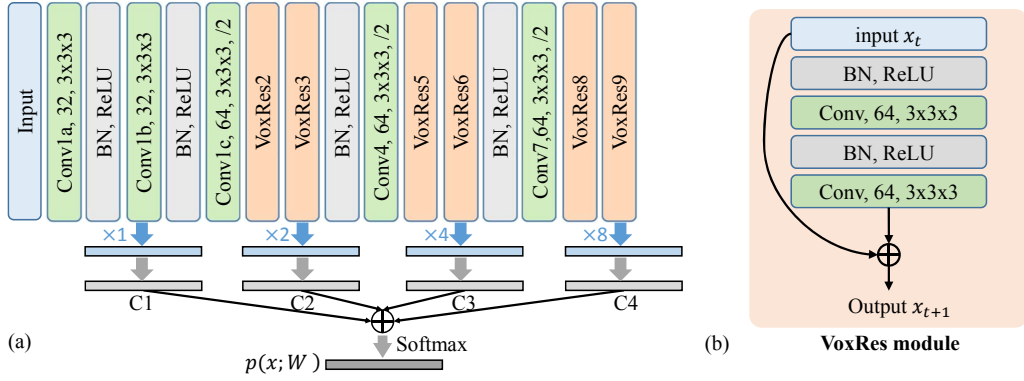


Figure 1: (a) The architecture of proposed VoxResNet for volumetric image segmentation; (b) The illustration of VoxRes module.

the powerful capability of deep residual learning and tackle the object segmentation tasks from high-dimensional volumetric images efficiently and effectively, we extend the 2D deep residual networks into a 3D variant and design the architecture following the spirit from [11] with full pre-activation, i.e., using asymmetric after-addition activation, as shown in Figure 1(b).

The architecture of our proposed VoxResNet for volumetric image segmentation is shown in Figure 1(a). Specifically, it consists of stacked residual modules (i.e., VoxRes module) with a total of 25 volumetric convolutional layers and 4 deconvolutional layers [18], which is the deepest 3D convolutional architecture so far. In each VoxRes module, the input feature x_l and transformed feature $\mathcal{F}(x_l, W_l)$ are added together with skip connection as shown in Figure 1(b), hence the information can be directly propagated in the forward and backward passes. Note that all the operations are implemented in a 3D way to strengthen the volumetric feature representation learning. Following the principle from VGG network [31] and deep residual networks [11], we employ small convolutional kernels (i.e., $3 \times 3 \times 3$) in the convolutional layers, which have demonstrated the state-of-the-art performance on image recognition tasks. Three convolutional layers are along with a stride 2, which reduced the resolution size of input volume by a factor of 8. This enables the network to have a large receptive field size, hence enclose more contextual information for improving the discrimination capability. Batch normalization layers are inserted into the architecture intermediately for reducing internal covariate shift [12], hence accelerate the training process and improve the performance. In our network, the rectified linear units, i.e., $f(x) = \max(0, x)$, are utilized as the activation function for non-linear transform [15].

There is a huge variation of 3D anatomical structure shape, which demands different suitable receptive field sizes for better recognition performance. In order to handle the large variation of shape sizes, multi-level contextual information (i.e., 4 auxiliary classifiers C1-C4 in Figure 1(a)) with deep supervision [16, 4] is fused in our framework. Therefore, the whole network is trained in an end-to-end way by minimizing following objective function with standard back-propagation

$$\mathcal{L}(x, y; \theta) = \lambda \psi(\theta) - \sum_{\alpha} \sum_{x \in \mathcal{V}} \sum_c w_{\alpha} y_c^x \log p_c^{\alpha}(x; \theta) - \sum_{x \in \mathcal{V}} \sum_c y_c^x \log p_c(x; \theta) \quad (3)$$

where the first part is the regularization term (L_2 norm in our experiments) and latter one is the fidelity term consisting of auxiliary classifiers and final target classifier. The tradeoff of these terms is controlled by the hyperparameter λ . w_{α} (where α indicates the index of auxiliary classifiers) is the weights of auxiliary classifiers, which were set as 1 initially and decreased till marginal values (i.e., 10^{-3}) in our experiments. The weights of network are denoted as $\theta = \{W\}$, $p_c(x; \theta)$ or $p_c^{\alpha}(x; \theta)$ denotes the predicted probability of c th class after softmax classification layer for voxel x in volume space \mathcal{V} , and $y_c^x \in \{0, 1\}$ is the corresponding ground truth, i.e., $y_c^x = 1$ if voxel x belongs to the c th class, otherwise 0.

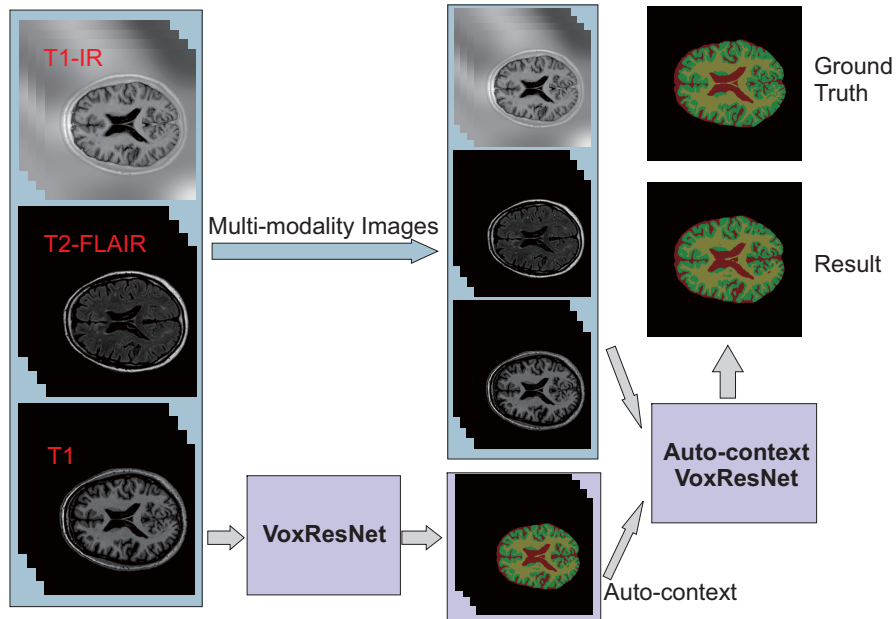


Figure 2: An overview of our proposed framework for fusing auto-context with multi-modality information.

2.3 Multi-modality and Auto-context Information Fusion

In medical image computing, the volumetric data are usually acquired with multiple imaging modalities for robustly examining different tissue structures. For example, three imaging modalities including T1, T1-weighted inversion recovery (T1-IR) and T2-FLAIR are available in brain structure segmentation task [20] and four imaging modalities are used in brain tumor (T1, T1 contrast-enhanced, T2, and T2-FLAIR MRI) [21] and lesion studies (T1-weighted, T2-weighted, DWI and FLAIR MRI) [19]. The main reason for acquiring multi-modality images is that the information from multi-modality dataset can be complementary, which provides robust diagnosis results. Thus, we concatenate these multi-modality data as input, then the complementary information is jointly fused during the training of network in an implicit way, which demonstrated consistent improvement compared to any single modality.

Furthermore, in order to harness the integration of high-level context information, implicit shape information and original low-level image appearance for improving recognition performance, we formulate the learning process as an auto-context algorithm [35]. Compared with the recognition tasks in computer vision, the role of auto-context information can be more important in the medical domain as the anatomical structures are roughly positioned and constrained [36]. Different from [36] utilizing the probabilistic boosting tree as the classifier, we employ the powerful deep neural networks as the classifier. Specifically, given the training images, we first train a *VoxResNet* classifier on original training sub-volumes. Then the discriminative probability maps generated from *VoxResNet* are then used as the context information, together with the original volumes (i.e., appearance information), to train a new classifier *Auto-context VoxResNet*, which further refines the semantic segmentation and removes the outliers. Different from the original auto-context algorithm in an iterative way [36], our empirical study showed that following iterative refinements gave marginal improvements. Therefore, we chosen the output of *Auto-context VoxResNet* as the final results.

3 Experiments

3.1 Dataset and Pre-processing

We validated our method on the MICCAI MRBrainS challenge data, an ongoing benchmark for evaluating algorithms on the brain segmentation. The task of MRBrainS challenge is to segment the

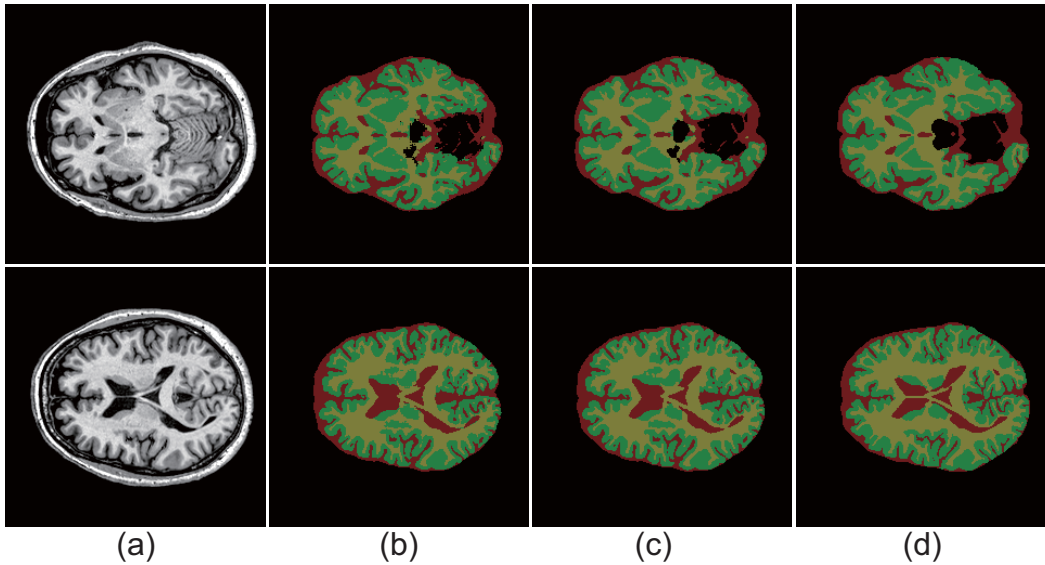


Figure 3: The example results of validation data (yellow, green, and red colors represent the WM, GM, and CSF, respectively): (a) original MR images, (b) results of VoxResNet, (c) results of Auto-context VoxResNet, (d) ground truth labels.

brain into four-class structures, i.e., background, cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM). The datasets were acquired at the UMC Utrecht of patients with diabetes and matched controls with varying degrees of atrophy and white matter lesions [20]. Multi-sequence 3T MRI brain scans, including T1-weighted, T1-IR and T2-FLAIR, are provided for each subject. The training dataset consists of five subjects with manual segmentations provided. The test data includes 15 subjects with ground truth held out by the organizers for independent evaluation.

In the pre-processing step, we subtracted Gaussian smoothed image and applied Contrast-Limited Adaptive Histogram Equalization (CLAHE) for enhancing local contrast by following [32]. Then six input volumes including original images and pre-processed ones were used as input data in our experiments. We normalized the intensities of each slice with zero mean and unit variance before inputting into the network.

3.2 Evaluation and Comparison

The evaluation metrics of MRBrainS challenge include the Dice coefficient (DC), the 95th-percentile of the Hausdorff distance (HD) and the absolute volume difference (AVD), which are calculated for each tissue type (i.e., GM, WM and CSF), respectively [20]. The details of evaluation can be found in the challenge website¹.

To investigate the efficacy of multi-modality and auto-context information, we performed extensive ablation studies on the validation data (leave one out cross-validation). The detailed results of cross-validation are reported in Table 1. We can see that combining the multi-modality information can dramatically improve the segmentation performance than that of single image modality, especially on the metric of DC, which demonstrates the complementary characteristic of different imaging modalities. Moreover, by integrating the auto-context information, the performance of DC can be further improved. The qualitative results of brain segmentation can be seen in Figure 3 and we can see that the results of combining multi-modality and auto-context information can give more accurate results visually than only multi-modality information.

Regarding the evaluation of testing data, we compared our method with several state-of-the-art methods, including MDGRU, 3D U-net [6] and PyraMiD-LSTM [32]. The MDGRU applied a neural network with the main components being multi-dimensional gated recurrent units and achieved quite

¹MICCAI MRBrainS Challenge: <http://mrbrains13.isi.uu.nl/details.php>

Table 1: Cross-validation Results of MR Brain Segmentation using Different Modalities (DC: %, HD: mm, AVD: %).

Modality	GM			WM			CSF		
	DC	HD	AVD	DC	HD	AVD	DC	HD	AVD
T1	86.96	1.36	4.67	89.70	1.92	6.85	79.58	2.71	17.55
T1-IR	80.61	1.92	8.45	85.89	2.87	7.42	76.44	3.00	12.87
T2-FLAIR	81.13	1.92	9.15	83.21	3.00	4.99	75.34	3.03	3.77
All	86.86	1.36	7.13	90.22	1.36	5.12	81.97	2.14	9.87
All+auto-context	87.83	1.36	6.22	90.63	1.36	2.22	82.76	2.14	5.50

good performance. The 3D U-net extended previous 2D version [26] into a 3D variant and highlighted the necessity for volumetric feature representation when applied to 3D recognition tasks. The PyraMiD-LSTM parallelised the multi-dimensional recurrent neural networks in a pyramidal fashion and achieved compelling performance. The detailed results of testing data from different methods on brain segmentation from MR images can be seen in Table 2. We can see that deep learning based methods can achieve much better performance than hand-crafted feature based methods. The results of VoxResNet (see CU_DL in Table 2) by fusing multi-modality information achieved better performance than other deep learning based methods, which demonstrated the efficacy of our proposed framework. Incorporating the auto-context information (see CU_DL2 in Table 2), the performance of DC can be further improved. Overall, our methods achieved the top places in the challenge leader board out of 37 competing teams.

Table 2: Results of MICCAI MRBrainS Challenge of different methods (DC: %, HD: mm, AVD: %, only top 10 teams are shown here).

Method	GM			WM			CSF			Score*
	DC	HD	AVD	DC	HD	AVD	DC	HD	AVD	
CU_DL (ours)	86.12	1.47	6.42	89.39	1.94	5.84	83.96	2.28	7.44	39
CU_DL2 (ours)	86.15	1.45	6.60	89.46	1.94	6.05	84.25	2.19	7.69	39
MDGRU	85.40	1.55	6.09	88.98	2.02	7.69	84.13	2.17	7.44	57
PyraMiD-LSTM2	84.89	1.67	6.35	88.53	2.07	5.93	83.05	2.30	7.17	59
FBI/LMB Freiburg [6]	85.44	1.58	6.60	88.86	1.95	6.47	83.47	2.22	8.63	61
IDSIA [32]	84.82	1.70	6.77	88.33	2.08	7.05	83.72	2.14	7.09	77
STH	84.77	1.71	6.02	88.45	2.34	7.67	82.77	2.31	6.73	86
ISI-Neonatology [22]	85.77	1.62	6.62	88.66	2.07	6.96	81.08	2.65	9.77	87
UNC-IDEA [38]	84.36	1.62	7.04	88.68	2.06	6.46	82.81	2.35	10.5	90
MNAB2 [24]	84.50	1.70	7.10	88.04	2.12	7.74	82.30	2.27	8.73	109

*Score = Rank DC + Rank HD + Rank AVD; a smaller score means better performance.

3.3 Implementation Details

Our method was implemented using Matlab and C++ based on Caffe library [13, 34]. It took about one day to train the network while less than 2 minutes for processing each test volume (size $240 \times 240 \times 48$) using a standard workstation with one NVIDIA TITAN X GPU. Due to the limited capacity of GPU memory, we cropped volumetric regions (size $80 \times 80 \times 80 \times m$, m is number of image modalities and set as 6 in our experiments) for the input into the network. This was implemented in an on-the-fly way during the training, which randomly sampled the training samples from the whole input volumes. In the test phase, the probability map of whole volume was generated in an overlap-tiling strategy for stitching the sub-volume results².

4 Conclusions

In this paper, we analyzed the capabilities of VoxResNet in the field of medical image computing and demonstrated its potential to advance the performance of biomedical volumetric image segmentation problems. The proposed method extends the deep residual learning in a 3D variant for handling

²Project page: http://www.cse.cuhk.edu.hk/~hchen/research/seg_brain.html

volumetric data. Furthermore, an auto-context version of VoxResNet is proposed to further boost the performance under an integration of low-level appearance information, implicit shape information and high-level context. Extensive experiments on the challenging segmentation benchmark corroborated the efficacy of our method when applied to volumetric data. Moreover, the proposed algorithm goes beyond the application of brain segmentation and it can be applied in other volumetric image segmentation problems. In the future, we will investigate the performance of our method on more object detection and segmentation tasks from volumetric data.

References

- [1] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.
- [2] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solórzano. Combination strategies in multi-atlas image segmentation: application to brain mr data. *IEEE transactions on medical imaging*, 28(8):1266–1277, 2009.
- [3] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *Biomedical and Health Informatics, IEEE Journal of*, 19(5):1627–1636, 2015.
- [4] H. Chen, X. J. Qi, J. Z. Cheng, and P. A. Heng. Deep contextual networks for neuronal structure segmentation. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [5] H. Chen, L. Yu, Q. Dou, L. Shi, V. C. Mok, and P. A. Heng. Automatic detection of cerebral microbleeds via deep learning based 3d feature representation. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 764–767. IEEE, 2015.
- [6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *arXiv preprint arXiv:1606.06650*, 2016.
- [7] J. Dai, K. He, and J. Sun. Instance-aware semantic segmentation via multi-task network cascades. *arXiv preprint arXiv:1512.04412*, 2015.
- [8] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. *IEEE transactions on medical imaging*, 35(5):1182–1195, 2016.
- [9] A. Giorgio and N. De Stefano. Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging*, 37(1):1–14, 2013.
- [10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. *arXiv preprint arXiv:1603.05027*, 2016.
- [12] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [14] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *arXiv preprint arXiv:1603.05959*, 2016.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [16] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. In *AISTATS*, volume 2, page 6, 2015.
- [17] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–312. Springer, 2014.
- [18] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

- [19] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, et al. Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. *Medical Image Analysis*, 35:250–269, 2017.
- [20] 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. *Computational intelligence and neuroscience*, 2015.
- [21] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.
- [22] P. Moeskops, M. A. Viergever, M. J. Benders, and I. Išgum. Evaluation of an automatic brain segmentation method developed for neonates on adult mr brain images. In *SPIE Medical Imaging*, pages 941315–941315. International Society for Optics and Photonics, 2015.
- [23] D. Nie, L. Wang, Y. Gao, and D. Sken. Fully convolutional networks for multi-modality iso-intense infant brain image segmentation. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1342–1345. IEEE, 2016.
- [24] S. Pereira, A. Pinto, J. Oliveira, A. M. Mendrik, J. H. Correia, and C. A. Silva. Automatic brain tissue segmentation in mr images using random forests and conditional random fields. *Journal of Neuroscience Methods*, 270:111–123, 2016.
- [25] A. Prason, 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 246–253. Springer, 2013.
- [26] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [27] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 520–527. Springer, 2014.
- [28] D. Sarikaya, L. Zhao, and J. J. Corso. Multi-atlas brain mri segmentation with multiway cut. In *Proceedings of the MICCAI Workshops The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS13)*, 2013.
- [29] F. Shen and G. Zeng. Weighted residuals for very deep networks. *arXiv preprint arXiv:1605.08831*, 2016.
- [30] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.
- [31] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [32] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. In *Advances in Neural Information Processing Systems*, pages 2998–3006, 2015.
- [33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [34] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Deep end2end voxel2voxel prediction. *arXiv preprint arXiv:1511.06681*, 2015.
- [35] Z. Tu. Auto-context and its application to high-level vision tasks. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [36] Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1744–1757, 2010.
- [37] A. van Opbroek, F. van der Lijn, and M. de Bruijne. Automated brain-tissue segmentation by multi-feature svm classification. 2013.

- [38] 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*, 108:160–172, 2015.
- [39] S. Zagoruyko and N. Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.