Improved Brain Lesion Segmentation with Anatomical Priors from Healthy Subjects

Chenghao Liu[1], Xiangzhu Zeng[2], Kongming Liang[3], Yizhou Yu[4], and Chuyang Ye[1(✉)]

[1] School of Information and Electronics, Beijing Institute of Technology, Beijing, China
chuyang.ye@bit.edu.cn
[2] Department of Radiology, Peking University Third Hospital, Beijing, China
[3] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
[4] Deepwise AI lab, Beijing, China

Abstract. *Convolutional neural networks* (CNNs) have greatly improved the performance of brain lesion segmentation. However, accurate segmentation of brain lesions can still be challenging when the appearance of lesions is similar to normal brain tissue. To address this problem, in this work we seek to exploit the information in scans of healthy subjects to improve brain lesion segmentation, where anatomical priors about normal brain tissue can be taken into account for better discrimination of lesions. To incorporate such prior knowledge, we propose to register a set of reference scans of healthy subjects to each scan with lesions, and the registered reference scans provide reference intensity samples of normal tissue at each voxel. In this way, the spatially adaptive prior knowledge can indicate the existence of abnormal voxels even when their intensities are similar to normal tissue, because their locations contradict with the prior knowledge about normal tissue. Specifically, with the reference scans, we compute anomaly score maps for the scan with lesions, and these maps are used as auxiliary inputs to the segmentation network to aid brain lesion segmentation. The proposed strategy was evaluated on different brain lesion segmentation tasks, and the results indicate the benefit of incorporating the anatomical priors using our approach.

Keywords: Brain lesion segmentation · Convolutional neural network · Anatomical priors

1 Introduction

Brain lesions may lead to persistent systematic changes in brain functions [4], such as motor impairments [17] and cognitive deficits [11]. Accurate segmentation of brain lesions enables quantitative measurements of the lesion size, shape, and location, and it provides important clues for intervention and prognosis [6,13,17].

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12901, pp. 186–195, 2021.
https://doi.org/10.1007/978-3-030-87193-2_18

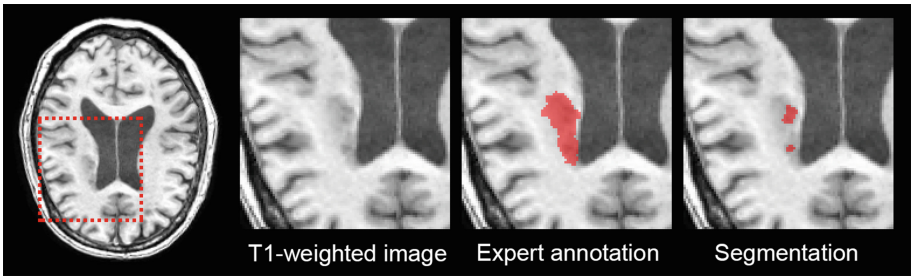

Fig. 1. An example of chronic stroke lesion segmentation on the T1-weighted image. The intensity of the lesion (indicated by the expert annotation) resembles the intensity of normal gray matter, and even a state-of-the-art segmentation pipeline nnU-Net [5] can undersegment the lesion.

The performance of automated brain lesion segmentation has been greatly advanced by *convolutional neural networks* (CNNs) [5,7]. However, accurate segmentation of brain lesions may still be challenging when the appearances of lesions are similar to normal brain tissue. For example, as shown in Fig. 1, the intensity of a chronic stroke lesion in the white matter can resemble that of normal gray matter, and even a state-of-the-art segmentation approach could confuse the lesion with normal tissue and produce low-quality segmentation results.

To address this problem, it is possible to incorporate prior knowledge about normal brain tissue, so that the lesions can be better discriminated even when their intensities are similar to those of normal brain tissue. For example, based on the assumption that normal tissue tends to appear symmetric, an image of interest can be registered to its reflected version, and the difference between them can indicate asymmetric regions that are possibly lesions [10,12]. In [3], based on the assumption that a CNN model trained to reconstruct images of healthy subjects is not capable of truthfully reconstructing the high frequency components of lesions, a scale-space autoencoder is proposed to estimate pixel-wise anomalies. The results of these methods can be used as anatomical priors and provide auxiliary information to improve brain lesion segmentation.

In this work, we continue to explore the incorporation of prior knowledge about normal brain tissue for brain lesion segmentation. Although the intensity of brain lesions can be similar to that of the normal tissue elsewhere, the locations of these lesions help to discriminate them. Based on this observation, we propose to use the scans of healthy subjects to provide voxelwise reference intensities of normal tissue, so that abnormal intensity at each voxel can be better identified. Specifically, we register a set of reference images of healthy subjects to a scan of interest with lesions. Then, intuitively, the difference map between the intensities of the scan of interest and the registered reference images can be computed to indicate anomaly. In addition, the intensities of the registered scans at each voxel can be considered samples drawn from a distribution of normal tissue intensities at that voxel. We fit the distribution with these samples and compute the likelihood of the intensities of the scan of interest, which gives a

different perspective of anomaly indication. These two types of anomaly score maps are combined with the image with lesions for network training and inference, so that knowledge about normal brain tissue is exploited to improve the segmentation. For evaluation, we integrated the proposed method with the nnU-Net segmentation pipeline [5] and performed experiments for segmenting chronic stroke lesions and ischemic stroke lesions. Results show that the proposed strategy of incorporating anatomical prior information improves the quality of brain lesion segmentation. The code of our method is publicly available at https://github.com/lchdl/NLL_anomaly_detection.

2 Method

2.1 Problem Formulation

Given an image $\boldsymbol{x}$ with brain lesions, we seek to perform CNN-based brain lesion segmentation for $\boldsymbol{x}$ aided by a set $\mathcal{R} = \{\boldsymbol{r}_i\}_{i=1}^{K}$ of K reference images of healthy subjects without lesions, where $\boldsymbol{r}_i$ ($i \in \{1, \ldots, K\}$) represents the i-th reference image. These reference images can allow the incorporation of anatomical prior information about normal brain tissue for improved discrimination of lesions. We hypothesize that by properly comparing $\boldsymbol{x}$ and $\mathcal{R}$, it is possible to suggest lesion areas that may be confused with normal tissue, and such anomaly information can be used as an auxiliary input to the segmentation network. To this end, we design two strategies to compute the anomaly map using image registration, which are based on purely the intensities of the reference images and a distribution fitted from these intensities, respectively. For each of the two strategies, the reference images $\mathcal{R}$ are aligned to $\boldsymbol{x}$ with deformable registration, leading to a new set $\mathcal{R}'$ of registered reference images $\phi_i(\boldsymbol{r}_i)$, where $\phi_i(\cdot)$ represents the spatial transformation for $\boldsymbol{r}_i$, and then $\boldsymbol{x}$ is compared with $\mathcal{R}'$. The detailed design of these strategies, which for convenience are referred to as the intensity-based strategy and the distribution-based strategy, respectively, is described below.

2.2 Intensity-Based Strategy

The intensity-based strategy directly calculates the intensity difference between the registered reference images and the image with lesions. To avoid noise outside brain tissue, brain masks of the reference images are also used for the computation. We denote the brain mask of the reference image $\boldsymbol{r}_i$ by $\boldsymbol{m}_i$, and then the anomaly score map $\boldsymbol{\alpha}$ is derived as

$$\boldsymbol{\alpha} = \frac{1}{K} \sum_{i=1}^{K} \phi_i(\boldsymbol{m}_i) \cdot (\boldsymbol{x} - \phi_i(\boldsymbol{r}_i)) \,, \tag{1}$$

where the average difference between the image with lesions and the registered reference images is computed for each voxel inside the brain. In this way, if at a voxel of $\boldsymbol{x}$ the intensity is very different from the reference intensities of

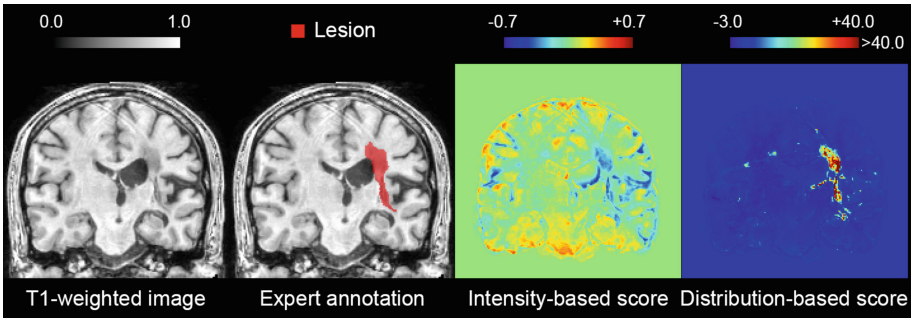

Fig. 2. An example of the anomaly score maps computed with the intensity-based strategy and distribution-based strategy for chronic stroke lesion segmentation on the T1-weighted image. The expert annotation is also shown for reference.

normal tissue, a much higher or lower value is observed on $\boldsymbol{\alpha}$, indicating the possible existence of brain lesions. Note that the difference may also be caused by registration error, especially at tissue interfaces where the reference intensities can have large variations. Figure 2 provides an example of the anomaly score map computed with the intensity-based strategy for chronic stroke lesion segmentation on the T1-weighted image.

2.3 Distribution-Based Strategy

To reduce the impact of registration error at tissue interfaces, in addition to the intuitive intensity-based strategy, it is also possible to treat the intensities of the registered reference images as samples drawn from a distribution of normal tissue intensities. Then, the distribution can be fitted from these samples, and the likelihood of an intensity being normal can be computed. For simplicity, we assume that the distribution of the normal tissue intensity at each voxel v is a Gaussian distribution $\mathcal{N}(\mu_v, \sigma_v^2)$ parameterized by the mean μ_v and variance σ_v^2. μ_v and σ_v^2 can be computed by fitting the registered reference intensities to the distribution at each voxel.

With the distribution $\mathcal{N}(\mu_v, \sigma_v^2)$, for the intensity x_v at each voxel v of an image with lesions, an anomaly score β_v can then be derived as the negative log-likelihood:

$$\beta_v = \log \sqrt{2\pi}\sigma_v + \frac{(x_v - \mu_v)^2}{2\sigma_v^2}. \tag{2}$$

We denote the anomaly score map comprising β_v by $\boldsymbol{\beta}$. When x_v deviates from $\mathcal{N}(\mu_v, \sigma_v^2)$, the negative log-likelihood is higher. In this way, at tissue interfaces where $\boldsymbol{\alpha}$ has a very large or small value, if the variance of the reference intensities is also large due to registration error, then a small anomaly score will be assigned. An example of $\boldsymbol{\beta}$ is shown in Fig. 2, where regions with higher β_v are consistent with the lesion location. Note that similar to the intensity-based strategy, brain masks are also applied to the anomaly score map.

2.4 Integration with CNNs and Implementation Details

To train a segmentation CNN, the anomaly score maps $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are concatenated with the image $\boldsymbol{x}$ with lesions as network input. For test scans, the anomaly score maps are also computed with the reference images and concatenated with $\boldsymbol{x}$ for brain lesion segmentation. The proposed method can be integrated with various CNN-based segmentation approaches, such as [5] and [7].

We use ANTs [1] for the registration of reference images, where an affine registration is followed by a deformable registration. The brain masks are extracted from the reference images with BET [14,16] for constraining the anomaly score map. $K = 10$ reference images are used to reach a compromise between the abundance of reference samples and the computational overhead of image registration. Note that since deformable registration is time-consuming, learning-based registration [2] may also be used to accelerate the registration.

3 Results

3.1 Dataset Description

The proposed method was evaluated on two brain lesion segmentation tasks, which are the segmentation of chronic stroke lesions and ischemic stroke lesions. The datasets associated with these tasks are introduced below.

We first used the publicly available ATLAS dataset [8] for evaluation, which comprises T1-weighted MRI scans of 220 subjects with chronic stroke lesions. The lesions in these scans have been annotated by experts. All images were affinely registered to the MNI152 template, intensity normalized, and defaced. The dimension of the preprocessed images is $197 \times 233 \times 189$ with an isotropic $1\,\text{mm}^3$ resolution. Since the ATLAS dataset only contains scans of patients with brain lesions, we further selected 10 T1-weighted scans of healthy subjects from the HCP dataset [15] as the reference images. The registration of each reference image to a scan in the ATLAS dataset took about 20 min. The histograms of the reference images were matched to an image randomly selected from the ATLAS dataset.

A second in-house dataset for ischemic stroke lesion segmentation was also used for evaluation, which comprises 21 normal-appearing *diffusion weighted images* (DWIs) and 219 DWIs with ischemic stroke lesions. Each DWI belongs to a different subject, and the DWIs were annotated in their native space by an experienced radiologist. All images were acquired on a 3T Siemens Verio scanner with a b-value of $1000\,\text{s/mm}^2$, together with a $b0$ image. The image dimension is $240 \times 240 \times 21$ and the resolution is $0.96\ \text{mm} \times 0.96\ \text{mm} \times 6.5\ \text{mm}$. We randomly picked 10 images from the 21 normal-appearing scans as the reference images. The registration of each reference image to a DWI scan with brain lesions took about 5 min.

3.2 Experimental Settings

For each segmentation task, we selected 50 scans as the test set, and different dataset configurations (numbers of labeled images) were used for training. Specifically, for chronic stroke lesion segmentation, a total number of 170, 80, or 15 labeled images were used for training, and for ischemic stroke lesion segmentation, the numbers were 169, 40, or 10. In addition, for each configuration, 20% of the training images were used as a validation set. Model selection was performed based on the best performance on the validation set.

We evaluated the proposed method based on the state-of-the-art medical image segmentation pipeline nnU-Net [5] with default settings. Given a dataset, nnU-Net automatically determines the optimal U-Net-based structure variant, pre- and post-processing strategy, and performs end-to-end image segmentation. Before training and inference, z-score normalization was applied to all input images. Also, nnU-net automatically performs data augmentation in the training and inference stages. During training, random elastic deformation, rotation, scaling, mirroring, and gamma intensity transform are applied to each training sample. During inference, mirroring is applied to each input sample, and the final prediction is obtained by averaging the predictions from eight differently mirrored inputs. For each test scan, the computation of anomaly score maps and network inference took about one minute in total.

The proposed method was compared with two competing methods. The nnU-Net model trained with the intensity image input only (without using any auxiliary information) was considered the baseline method for comparison. In addition, the proposed method was compared with nnU-Net integrated with the autoencoder-based anomaly detection method in [3], where the anomaly score was also concatenated with the intensity image and fed into nnU-Net. This competing method is referred to as 'baseline+[3]'.

3.3 Segmentation Performance

The proposed method was applied to perform the two brain lesion segmentation tasks. Since in the proposed method the anomaly maps are computed with the intensity-based and distribution-based strategies, our method is referred to as 'baseline+IB+DB'. For reference, the results achieved with the intensity-based or distribution-based strategy alone were also considered (referred to as 'baseline+IB' and 'baseline+DB', respectively). Both qualitative and quantitative evaluation was performed, and we used the *Dice similarity coefficient* (DSC) to quantitatively measure the segmentation quality. The detailed results are described below.

Coronal views of the results of chronic stroke lesion segmentation are shown in Fig. 3, together with the expert annotation. With the auxiliary information computed with both intensity-based and distribution-based strategies, the segmentation results reached a better consensus with the expert annotations than the results of the competing methods. Although using either baseline+IB or baseline+DB could also improve the segmentation performance, when the anomaly score maps were combined, the quality of the results was better.

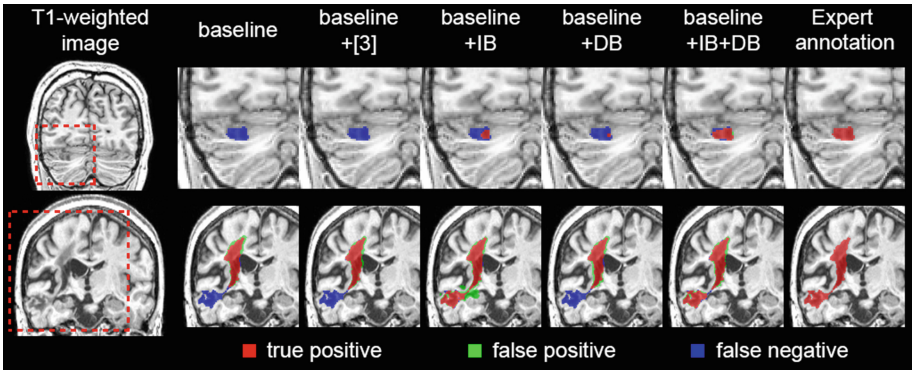

Fig. 3. Coronal views of representative segmentation results for chronic stroke lesions.

Table 1. The DSC results (mean ± std) for the chronic stroke lesion segmentation task. The best results are highlighted in bold. Asterisks indicate that the difference between Baseline+IB+DB and the competing method is significant using a paired student's t-test. (*: $p < 0.05$)

#labeled images	170	80	15
baseline	0.5870 ± 0.3254	0.5551 ± 0.3353	0.4362 ± 0.3315*
baseline+ [3]	0.5510 ± 0.3525*	0.5403 ± 0.3377	0.4850 ± 0.3242
baseline+IB	0.6075 ± 0.3018	0.5646 ± 0.3231	$\mathbf{0.5268 \pm 0.3308}$
baseline+DB	0.5996 ± 0.3051	0.5442 ± 0.3267	0.4831 ± 0.3288
baseline+IB+DB	$\mathbf{0.6256 \pm 0.2918}$	$\mathbf{0.5722 \pm 0.3076}$	0.5245 ± 0.3274

Table 2. Method ranking for the chronic stroke lesion segmentation task. The best results are highlighted in bold.

#labeled images	170	80	15
baseline	3.22	3.32	3.56
baseline+ [3]	3.08	3.26	3.16
baseline+IB	2.94	2.84	2.58
baseline+DB	3.14	3.06	3.14
baseline+IB+DB	**2.62**	**2.52**	**2.56**

Quantitatively, the means and *standard deviations* (stds) of the DSCs of the segmentation results on the test set for the ATLAS dataset are shown in Table 1 for each data configuration. Baseline+IB+DB has better segmentation accuracy than the two competing methods. Its performance is also better than baseline+IB and baseline+DB for the cases of 170 and 80 training images, and close to the best result of baseline+IB given 15 labeled images for training. We

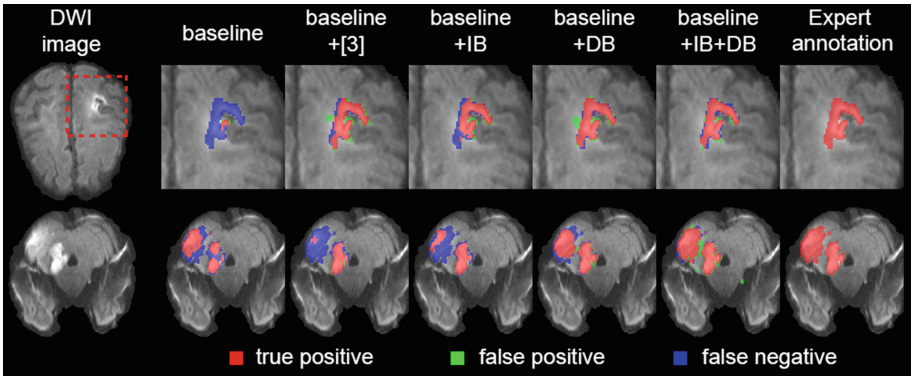

Fig. 4. Axial views of representative segmentation results for ischemic stroke lesions.

Table 3. The DSC results (mean ± std) for the ischemic stroke lesion segmentation task. The best results are highlighted in bold. Asterisks indicate that the difference between Baseline+IB+DB and the competing method is significant using a paired student's t-test. (**: $p < 0.01$, ***: $p < 0.001$)

#labeled images	169	40	10
baseline	0.7788 ± 0.1936***	0.7477 ± 0.1938	0.6481 ± 0.2195
baseline+ [3]	0.7985 ± 0.1904	0.7390 ± 0.2105	0.5981 ± 0.2702**
baseline+IB	0.8038 ± 0.1842	0.7481 ± 0.2099	0.6275 ± 0.2399
baseline+DB	0.8016 ± 0.1899	$\mathbf{0.7625 \pm 0.1885}$	$\mathbf{0.6897 \pm 0.1941}$
baseline+IB+DB	$\mathbf{0.8089 \pm 0.1826}$	0.7538 ± 0.2014	0.6733 ± 0.2048

Table 4. Method ranking for the ischemic stroke lesion segmentation task. The best results are highlighted in bold.

#labeled images	169	40	10
baseline	3.96	3.44	3.08
baseline+ [3]	3.04	3.20	3.48
baseline+IB	2.88	3.26	3.24
baseline+DB	2.60	2.76	**2.46**
baseline+IB+DB	**2.52**	**2.34**	2.74

also computed the average rank of each method for each data configuration using the ranking strategy in [9], where all the methods were ranked together based on the DSC of each test scan. The results are shown in Table 2. In all cases, baseline+IB+DB achieves the best ranking.

The qualitative results for ischemic stroke lesion segmentation are shown in Fig. 4, where the results of baseline+IB+DB better agree with the expert annotations than the competing methods. The quantitative DSC results are shown

in Table 3, where baseline+IB+DB outperforms the two competing methods baseline and baseline+[3]. Compared with baseline+IB and baseline+DB, baseline+IB+DB achieves either the best or the second best result. Method ranking was also performed and the results are shown in Table 4, where baseline+IB+DB has better ranking than the two competing methods. Its rank is also better than baseline+IB and baseline+DB except when 10 training images were used, where baseline+IB+DB is the second best. Together with the results for chronic stroke lesion segmentation, these results suggest that baseline+IB+DB outperforms the existing methods, and the combination of the intensity-based and distribution-based strategies is recommended over the use of a single one of them.

4 Conclusion

We have developed a method for improving CNN-based brain lesion segmentation, where prior knowledge about normal brain tissue is incorporated in the segmentation. The prior knowledge is obtained from reference images of healthy subjects via image registration and suggests spatial anomaly maps. The anomaly information is used as auxiliary network inputs to aid the segmentation. The experimental results on two brain lesion segmentation tasks show that our method improves the segmentation performance.

Acknowledgement. This work is supported by Beijing Natural Science Foundation (L192058 & 7192108).

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging **38**(8), 1788–1800 (2019)
3. Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: Scale-space autoencoders for unsupervised anomaly segmentation in brain MRI. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 552–561. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_54
4. Churchill, N., et al.: The effects of chronic stroke on brain function while driving. Stroke **47**(Supplement 1), 150 (2016)
5. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
6. Jongbloed, L.: Prediction of function after stroke: a critical review. Stroke **17**(4), 765–776 (1986)
7. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. **36**, 61–78 (2017)
8. Liew, S.L., et al.: A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Sci. Data **5**, 180011 (2018)

9. Maier, O., et al.: ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. **35**, 250–269 (2017)
10. Martins, S.B., Telea, A.C., Falcão, A.X.: Investigating the impact of supervoxel segmentation for unsupervised abnormal brain asymmetry detection. Comput. Med. Imaging Graph. **85**, 101770 (2020)
11. Nakling, A.E., et al.: Cognitive deficits in chronic stroke patients: neuropsychological assessment, depression, and self-reports. Dementia Geriatric Cogn. Disord. Extra **7**(2), 283–296 (2017)
12. Raina, K., Yahorau, U., Schmah, T.: Exploiting bilateral symmetry in brain lesion segmentation. arXiv preprint arXiv:1907.08196 (2019)
13. Riley, J.D., et al.: Anatomy of stroke injury predicts gains from therapy. Stroke **42**(2), 421–426 (2011)
14. Smith, S.M.: Fast robust automated brain extraction. Human Brain Mapp. **17**(3), 143–155 (2002)
15. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K.: The WU-Minn human connectome project: an overview. NeuroImage **80**, 62–79 (2013)
16. Woolrich, M.W., et al.: Bayesian analysis of neuroimaging data in FSL. NeuroImage **45**(1), S173–S186 (2009)
17. Zhu, L.L., Lindenberg, R., Alexander, M.P., Schlaug, G.: Lesion load of the corticospinal tract predicts motor impairment in chronic stroke. Stroke **41**(5), 910–915 (2010)