

Automated Segmentation of the Locus Coeruleus from Neuromelanin-sensitive 3T MRI using Deep Convolutional Neural Networks

Max Dünnwald^{1,2}, Matthew J. Betts^{3,4}, Alessandro Sciarra¹, Emrah Düzel^{3,4,5},
Steffen Oeltze-Jafra^{1,6}

¹Department of Neurology, Otto-von-Guericke University Magdeburg (OVGU)

²Faculty of Computer Science, OVGU

³German Center for Neurodegenerative Diseases (DZNE), Magdeburg

⁴Institute of Cognitive Neurology and Dementia Research, OVGU

⁵Institute of Cognitive Neuroscience, University College London

⁶Center for Behavioral Brain Sciences (CBBS), OVGU

max.duennwald@med.ovgu.de

Abstract. The locus coeruleus (LC) is a small brain structure in the brainstem that may play an important role in the pathogenesis of Alzheimer’s Disease (AD) and Parkinson’s Disease (PD). The majority of studies to date have relied on using manual segmentation methods to segment the LC, which is time consuming and leads to substantial interindividual variability across raters. Automated segmentation approaches might be less error-prone leading to a higher consistency in Magnetic Resonance Imaging (MRI) contrast assessments of the LC across scans and studies. The objective of this study was to investigate whether a convolutional neural network (CNN)-based automated segmentation method allows for reliably delineating the LC in in vivo MR images. The obtained results indicate performance superior to the inter-rater agreement, i.e. approximately 70% Dice similarity coefficient (DSC).

1 Introduction

The LC is a nucleus located alongside the fourth ventricle in the pons. This small nucleus in the brainstem is the major source of noradrenaline modulation in the brain and is involved in a variety of important brain functions such as memory, learning, attention and arousal. Recently, it has attracted increasing interest since it may also play an important role in the pathogenesis of neurodegenerative disorders such as PD [1] and AD [2,3].

Similar to the substantia nigra (SN), the LC contains large amounts of neuromelanin (NM) - a pigmented polymer that results from the oxidation of catecholamines such as noradrenaline in the LC. This allows for in vivo visualization of the LC using neuromelanin sensitive MRI (NMs-MRI), e.g.,

T₁-weighted Fast Low Angle Shot (FLASH) imaging [4] and Magnetization Transfer (MT)-weighted imaging as recently reviewed [5].

A reliable segmentation is often the prerequisite for the extraction of (potential) biomarkers, e.g., contrast ratios between the LC and a reference region, which are used for further analyses of the characteristics of this structure in different cohorts (for instance subjects with AD, PD or healthy controls). For the analysis of larger cohorts, a manual segmentation approach is unfeasible, since it is tedious, time consuming and may be prone to errors [6]. Instead, automated segmentation approaches are required. To date, few methods have been published for the segmentation of the LC. The existing approaches are similar to those used for the segmentation of the SN [7] and sometimes have been applied to both structures [6,8].

Several methods dependent purely on the intensity of the voxels. For instance, the localization of a region of interest (ROI) based on landmarks found in histological analyses and the subsequent application of an intensity threshold within the specified volume [6,9] were proposed. This approach is inherently biased towards high intensity voxels which may pose a problem when applied to pathological cases where the signal intensity is typically decreased and the resulting contrast ratios in relation to a reference region might be overestimated. Alternatively, a fixed-numbered set of 10-connected brightest voxels can be chosen from within the ROI [10,11]. Besides, different atlas registration-based approaches have been developed [8,12]. However, they strongly depend on the quality of the registration output and are computationally expensive due to the required multiple co-registrations.

Recently, CNNs have been successfully applied to the automated segmentation of the SN [7]. Since SN and LC show similar imaging characteristics, we investigate here a CNN approach for LC segmentation.

2 Materials and Methods

2.1 Data and manual segmentations

For this study, the same data set as in [4] was employed. It consists of T₁-weighted FLASH 3T MRI scans of 82 healthy subjects: 25 younger (22-30 years old; 13 male, 12 female) and 57 older adults (61-80 years; 19 male, 38 female). The images were acquired with an isotropic voxel size of 0.75mm. Prior to delineation of the LC, they were upsampled by means of a sinc filter to an isotropic resolution of 0.375mm. Furthermore, a bias field correction was applied. We refer the reader to [4] for more details on the characteristics of the data set.

The inter-rater agreement was determined for this study based on manual segmentations from two independent raters, rater 1 (R1) and rater 2 (R2). The DSC is on average 0.495 ± 0.158 and 0.457 ± 0.155 for the left and right LC, respectively, which is comparable to the 0.499 reported in [8], but lower than the approximately 0.6 (left LC) and 0.55 (right LC) from [12]. These relatively low values underline the difficulty of the segmentation task. The comparison with the

inter-rater agreement of the SN of around 0.7 [8] indicates that LC segmentation might be even more challenging.

2.2 Network model

Inspired by results obtained for SN segmentation [7], we applied a U-Net-like [13] network. However, to address the even smaller size of the LC (as compared to the SN), we employed an architecture similar to 3D-U-Net [14], which exploits information in neighboring slices. The network is composed of blocks that each contain two convolutional layers (kernel size of $3 \times 3 \times 3$) which are each followed by a ReLU activation and a batch normalization layer. Three of such blocks form the contracting path, one is used for the bottom and another three blocks make up the expansive path. The downsampling is carried out by MaxPooling ($2 \times 2 \times 2$) and the upsampling using transposed convolutional layers.

The subjects were separated into 10 subsets and a 10-fold cross validation has been carried out in order to assess the potential of our network for LC segmentation. In every iteration, one of the 10 subsets was left out as test set while the others composed the training set. This was repeated until every subset was once used for testing. All evaluation values in the results section have been computed based on the segmentation masks resulting from the test iteration of each subset. For the training, patches of the size $64 \times 64 \times 64$ voxels were chosen as the input to the network. These were extracted randomly from all over the whole-brain volumes. However, the probability of a patch containing the LC was set to 50% to avoid a negative bias. Other forms of data augmentation such as rotation and scaling were not applied to avoid further interpolation of the data. The fuzzy DSC [15] was used as a loss function and Adam (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) was chosen as the optimization scheme. The networks were trained for 250 epochs where for each epoch, 10 random patches were extracted from each subject in the training subsets. The manual segmentations of R1 were used as ground truth for the training. When applying the network to the subjects in the test set, overlapping (by half of the patch size) patches were extracted and processed by the network. The resulting masks were combined by creating their union.

An investigation of preliminary results revealed that the network tends to segment very small regions outside of the brainstem in some of the subjects.

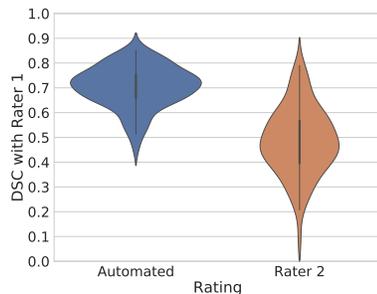


Fig. 1. Violinplots comparing the obtained DSCs of the proposed method with the masks of R1 and the inter-rater agreement (between R1 and R2).

Due to the tiny size of these false positive regions, they did not have a noticeable impact on the DSC or the median intensity contrast ratios. However, the maximum intensity was mostly located in these voxels and therefore, caused erroneous maximum intensity contrast ratios as will be shown later. To address this issue, we applied two simple post-processing steps to the network’s masks: First, some of the most outer slices of the volume were removed (fixed number across all subjects to roughly narrow down the region) and second, all connected segments with a size smaller than 50 voxels were disregarded. An alternative post-processing could be the application of a mask of the brainstem, which can be generated by popular tools like FSL [16] or using another CNN as demonstrated in a method for SN segmentation [7].

3 Results

The average DSCs of the network’s masks and the segmentations of R1 for the left, right LC and the combination of both are 0.711 ± 0.096 , 0.697 ± 0.091 and 0.705 ± 0.082 . These values are higher than both, the inter-rater agreement (left LC: 0.501 ± 0.148 , right LC: 0.463 ± 0.147 , both: 0.485 ± 0.135), which can be seen in Figure 1, as well as the results of a recently proposed atlas registration-based approach [8] (0.404 ± 0.141). However, latter comparison is to be taken with a grain of salt as [8] was evaluated on a different dataset. When compared to the masks created by R2, the network performs comparable to the inter-rater

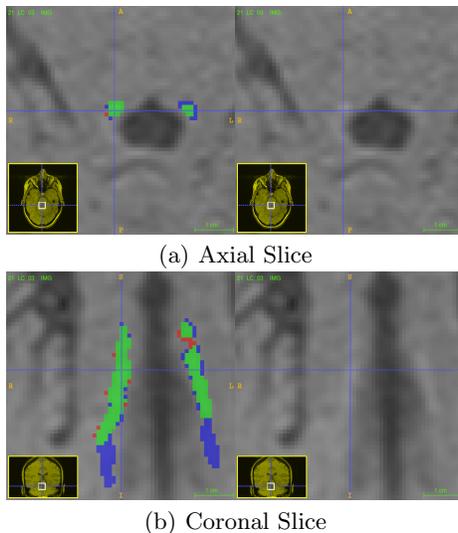


Fig. 2. Selected slices of an example of average network performance (DSC for the left LC: 0.706, right LC: 0.683, combined: 0.696). Green color indicates agreement between network and R1, red voxels were segmented by R1 but not the network (false negatives) and blue voxels were segmented by the network but not R1 (false positives). On the right, the respective section is shown without the masks.

agreement: 0.499 ± 0.119 (left LC), 0.467 ± 0.126 (right LC) and 0.486 ± 0.108 (combined). Figure 2 shows the example of an average network result.

Furthermore, the contrast ratios between the masks of the LC and a reference region within the pons were calculated. Contrast ratios are an important element in analyzing LC properties [4], which is why the influence of the automatic LC segmentations on these values was evaluated. Hence, the intra-class correlation (ICC) with a confidence level of 0.95 was determined between the contrast ratios based on manual (R1) and automatic LC masks (each in relation to the same reference region [4]). For the median intensity contrast ratios, ICCs of 0.893 (left LC) and 0.871 (right LC) could be achieved. The maximum intensity contrast ratios are in a similar range with ICCs of 0.842 (left LC) and 0.974 (right LC). Without the previously described post-processing, the ICCs of the maximum ratios were -0.008 (left LC) and -0.014 (right LC).

4 Discussion

We applied a 3D-U-Net-like CNN for segmenting the LC in T_1 -weighted FLASH NMs-MRI data. The obtained DSC and ICC values indicate a good correlation between the network and R1 as well as a performance in the range of the inter-rater agreement when compared to R2. Together, this indicates the high potential of the proposed method as an alternative to manual delineation. Small-sized false positive regions outside of the brainstem produced by the network can be addressed using different simple post-processing techniques.

However, several interesting aspects remain to be investigated in future work, for instance, an extensive evaluation of different network architectures, 2D instead of 3D input, and a direct comparison to the current state-of-the-art atlas registration-based LC segmentation methods. Furthermore, methods for the avoidance of the above-mentioned erroneous regions outside of the brainstem need to be addressed in more detail. To fully automate the LC analysis process as a whole, automatic delineation of the reference regions is also needed.

Considering that the performance of the proposed approach is already in the range or above the inter-rater agreement, we plan to set the focus of further investigations to the evaluation and improvement of the robustness of this method rather than to the increase of the performance. In particular, the robustness with respect to cohorts of subjects that show pathologies such as AD and PD and the resulting changes in the LC are of interest. Their investigation poses a challenge since they are characterized by a loss in signal intensity due to less NM in the LC region. Also, the robustness against domain shift in terms of different acquisition techniques such as the different MRI sequences suitable for LC imaging as well as methods for handling and leveraging the ambiguities of the masks created by different raters will be investigated in our future work.

Acknowledgement. This work received funding from the federal state of Saxony-Anhalt, Germany (Project I 88). Matthew J. Betts is supported by the Human Brain Project (SP3 WP 3.3.1).

References

1. Braak H, Tredici KD, Rüb U, et al. Staging of brain pathology related to sporadic Parkinson’s disease. *Neurobiology of Aging*. 2003;24(2):197 – 211.
2. Braak H, Thal DR, Ghebremedhin E, et al. Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years. *Journal of Neuropathology & Experimental Neurology*. 2011;70(11):960–969.
3. Stratmann K, Heinsen H, Korf HW, et al. Precortical Phase of Alzheimer’s Disease (AD)-Related Tau Cytoskeletal Pathology. *Brain Pathology*. 2016;26(3):371–386.
4. Betts MJ, Cardenas-Blanco A, Kanowski M, et al. In vivo MRI assessment of the human locus coeruleus along its rostrocaudal extent in young and older adults. *NeuroImage*. 2017;163:150 – 159.
5. Liu KY, Marijatta F, Hämmerer D, et al. Magnetic resonance imaging of the human locus coeruleus: A systematic review. *Neuroscience & Biobehavioral Reviews*. 2017;83:325 – 355.
6. Chen X, Huddleston DE, Langley J, et al. Simultaneous imaging of locus coeruleus and substantia nigra with a quantitative neuromelanin MRI approach. *Magnetic Resonance Imaging*. 2014;32(10):1301 – 1306.
7. Le Berre A, Kamagata K, Otsuka Y, et al. Convolutional neural network-based segmentation can help in assessing the substantia nigra in neuromelanin MRI. *Neuroradiology*. 2019;61(12):1387–1395.
8. Ariz M, Abad RC, Castellanos G, et al. Dynamic Atlas-Based Segmentation and Quantification of Neuromelanin-Rich Brainstem Structures in Parkinson Disease. *IEEE Transactions on Medical Imaging*. 2019;38(3):813–823.
9. Langley J, Huddleston DE, Liu CJ, et al. Reproducibility of locus coeruleus and substantia nigra imaging with neuromelanin sensitive MRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2017;30(2):121–125.
10. García-Lorenzo D, Longo-Dos Santos C, Ewencyk C, et al. The coeruleus/subcoeruleus complex in rapid eye movement sleep behaviour disorders in Parkinson’s disease. *Brain*. 2013;136(7):2120–2129.
11. Olivieri P, Lagarde J, Lehericy S, et al. Early alteration of the locus coeruleus in phenotypic variants of Alzheimer’s disease. *Annals of Clinical and Translational Neurology*. 2019;6(7):1345–1351.
12. Rong Ye, Claire O’Callaghan, Catarina Rua, et al.. Imaging the Locus Coeruleus in Parkinson’s Disease with Ultra-high 7T MRI; 2019. Website. Available from: <https://ww5.aievolution.com/hbm1901/index.cfm?do=abs.viewAbs&abs=4579>.
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*. 2015; p. 234–241.
14. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. 2016; p. 424–432.
15. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *CoRR*. 2016;abs/1606.04797. Available from: <http://arxiv.org/abs/1606.04797>.
16. Woolrich MW, Jbabdi S, Patenaude B, et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*. 2009;45(1, Supplement 1):S173 – S186.