Localization of the Locus Coeruleus in MRI via Coordinate Regression

Max Dünnwald[1,2], Matthew J. Betts[3,4,6], Emrah Düzel[3,4,5], Steffen Oeltze-Jafra[1,6]

[1]Department of Neurology, Otto von Guericke University Magdeburg (OVGU)
[2]Faculty of Computer Science, OVGU
[3]German Center for Neurodegenerative Diseases (DZNE), Magdeburg
[4]Institute of Cognitive Neurology and Dementia Research, OVGU
[5]Institute of Cognitive Neuroscience, University College London
[6]Center for Behavioral Brain Sciences (CBBS), OVGU
max.duennwald@med.ovgu.de

Abstract. The locus coeruleus (LC) is a small nucleus in the brain stem. It is gaining increasing interest of the neuroscientific community due to its potentially important role in the pathogenesis of several neurodegenerative diseases such as Alzheimer's Disease (AD). In this study, an existing LC segmentation approach has been improved by adding a preceding LC localization to reduce false positive segments. For the localization, we propose a network that can be trained using coordinate regression and allows insights into its function via attention maps.

1 Introduction

The LC is a small brain structure in the upper dorsolateral pontine tegmentum of the brainstem. It is involved in several important functions, such as memory, learning, attention, arousal and pain modulation[1]. The LC currently attracts increasing interest, since it may also play an important role in the pathogenesis of neurodegenerative diseases[1]. It has been found, that so-called neuromelanin sensitive Magnetic Resonance Imaging (MRI) allows the in-vivo visualization of the LC. Further investigations require a delineation of the LC for which manual segmentation methods have primarily been applied to date[2]. One of the few exceptions employed a Convolutional Neural Network for this task, namely an adapted version of the 3D-Unet[3], and was published recently[4].

LC segmentation is a challenging task that results in relatively low inter-rater agreements, which are characterized by a Dice Similarity Coefficient (DSC) in the range of 0.499 to 0.64[5,6], depending on the used acquisition and segmentation protocols as well as the experience of the rater. A reason for this is the substantial uncertainty in the measurements, which is mostly caused by the small size of the LC, requiring an appropriate resolution. Higher resolutions however, yield worse signal to noise ratios (SNRs) jeopardizing the relatively weak signal

(see Figure 3). Nonetheless, reasonable compromises can be found and the resulting hyperintense regions (or properties of them) were shown to correspond to LC properties obtained in post-mortem studies, such as anatomical position and dimensions, LC cell density[7] and age-related effects of neuromelanin aggregation[1].

In this work, we propose a pipeline for LC segmentation that outperforms the approach in [4]. Instead of a false positives removal requiring careful parameterization, we apply an initial localization network, that was trained to regress the coordinates of LCs centers of mass (COMs) to obtain a single relevant patch containing the LC. This patch is then, processed by the 3D-Unet of [4]. This pipeline is more efficient, as the number of inferences is substantially lower, since the application is no longer done in a sliding window manner, while reducing false positive segments outside of the relevant region. For the localization network, we propose a combination of 3D-Unet and a differentiable spatial to numerical transform (DSNT)[8]. This architecture offers two major advantages: it allows insights into the networks function via attention maps as well as the processing of input volumes of arbitrary shape.

2 Materials and Methods

2.1 Data

This work utilizes the same data set as in [9] and [4]. It contains T_1-weighted FLASH 3T MRI whole-brain acquisitions of 82 healthy subjects, of which 25 are younger (22-30 years old; 13 male, 12 female) and 57 are older adults (61-80 years; 19 male, 38 female). The LCs were manually segmented by two expert raters, however, for this study, we made use of the masks of just one of the raters. Prior to delineation, the data was upsampled from an isotropic voxel size of 0.75mm^3 to 0.375mm^3 by means of a sinc filter. Additionally, a bias field correction was applied. For more details on the data set, see [9].

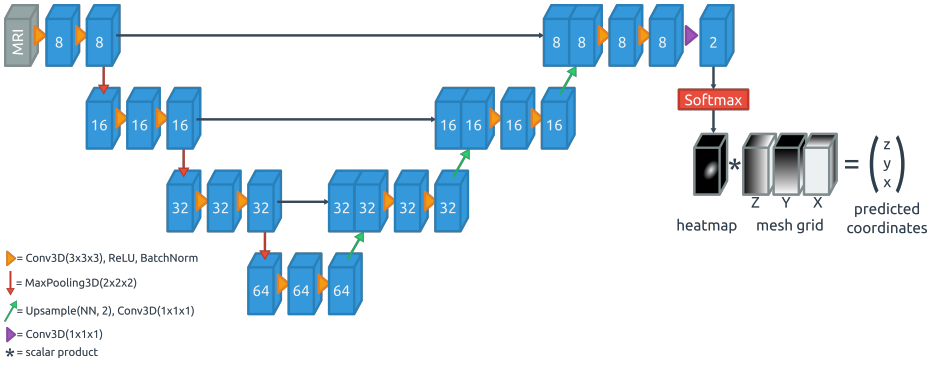


Fig. 1. Schematic illustration of CoRe-Unet, which directly predicts the coordinates of a voxel in the input volume. The white numbers denote the number of features.

2.2 Network architecture

Our network is based on the version of 3D-Unet[3] described in [4]. However, a major adaption was applied. The Unet is followed by a DSNT[8] that has been adapted to work with 3D data. It applies a softmax function, such that the sum over the Unet's output's spatial dimensions (a heatmap) equals 1 and, afterwards, calculates the scalar product between the result and a mesh grid of equal size, which encodes the coordinates of each of the volumes voxels. This adaption enables the model to directly predict coordinates and has three prime advantages. It's prediction is independent of the size and shape of the input volume, which is especially useful if both whole-brain as well as slab acquisitions shall be processed without any further adaptation. Furthermore, this model can be trained directly via a regression loss of the coordinates. And finally, the heatmap that can be obtained allows insight into the network's behaviour. In contrast to [8], no further regularization steps were performed on the heatmap. To take account of GPU memory limitations, the number of features of the Unet was reduced. We will refer to the network as coordinate regression Unet or CoRe-Unet for short. The CoRe-Unet is illustrated in Figure 1.

2.3 Evaluation scheme

The following scheme was used for the evaluation of the networks performance. First, a test set of 23 randomly selected subjects was held out. On the remaining 59 subjects, a 5-fold crossvalidation was performed. Hence, they were subdivided into 5 sets and during 5 trainings each subset was used as the validation set once, while the rest formed the training set for the respective training iteration. Every training lasted 500 epochs and converged without exception, while the final weights were chosen based on the validation set loss performance.

The network was trained with the data of the original isotropic resolution of 0.375mm^3. As the ground truth, we determined the COMs of the LCs based on the manual segmentations and propagated their coordinates into the lower, original resolution. Adam[10] (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) was used as the optimization scheme and the euclidean distance was chosen as a loss function. Two different versions of the network were trained. One with and one without randomly applied data augmentation. When applied, the augmentation comprised random combinations of the following transformations within the specified ranges: rotation around every axis (-15°, 15°), translation in every direction (-image_size/2, image_size/2) and scaling (-20%, 20%).

For determining the networks performance, we applied the resulting nets of each fold to the held-out test set. Afterwards, the euclidean distance and a dimension-wise mean squared error of all the predictions to the COMs (based on the manual raters mask) were calculated. Furthermore, the impact of combining the CoRe localization with the existing segmentation network of [4] was assed by computing and comparing the resulting DSC ($DSC = \frac{2TP}{2TP+FP+FN}$) and False Discovery Rates ($FDR = \frac{FP}{FP+TP}$) for three cases: First, applying the 3D-Unet[4] in a sliding window fashion without any post-processing, apart

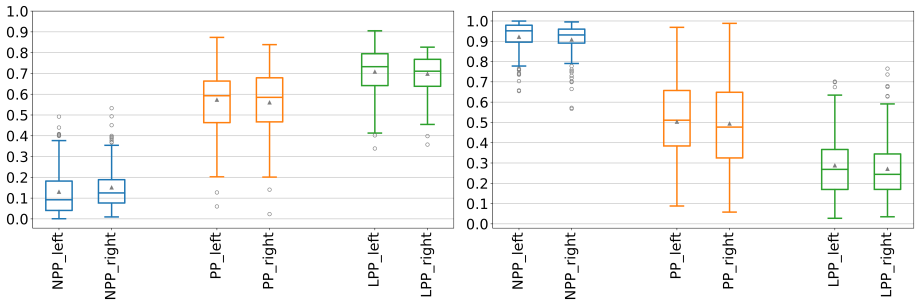
Table 1. Errors measured in euclidean distance of the respective networks predicted coordinates and the COM of the manually created LC masks. Every cell contains the values for left and right LC in Millimeters (mm) and voxels (vx).

	no augmentation	with augmentation
mean	2.97mm (3.96vx), 3.13mm (4.18vx)	1.34mm (1.79vx), 1.23mm (1.64vx)
std	1.56mm (2.08vx), 1.52 (2.02vx)	0.86mm (1.15vx), 0.67mm (0.89vx)

from a tresholding with a value of 0.5 (labeled "no post-processing" - NPP). Second, the same with the post-processing steps suggested by [4] (labeled "post-processing" - PP), which requires setting several parameters that had been determined empirically. Third, using the CoRe localizer for pre-processing, i.e. passing only one patch, generated with the localizers predicted coordinates as the center (labeled "localizer pre-processing" - LPP), to the 3D-Unet without any further post-processing. The median and maximum contrast ratios (CRs) between LC and a reference region in the pons represent popular LC biomarkers[2]. To gain insights into how the addition of the localizer affects them, they were determined for the aforementioned three variants as well and the intra-class coefficients (ICCs) (confidence 0.95) were calculated for each fold to compare them to the CRs determined using the manual masks.

3 Results

The errors measured as euclidean distances, which are reported in Table 1, indicate better performance for the network that was trained with augmentation (AUG), as its errors are often less than half than those of the version, that was trained without augmentation (NoAUG). For AUG, we determined the mean squared errors of each dimension and found that in axial direction



(a) DSC (b) FDR

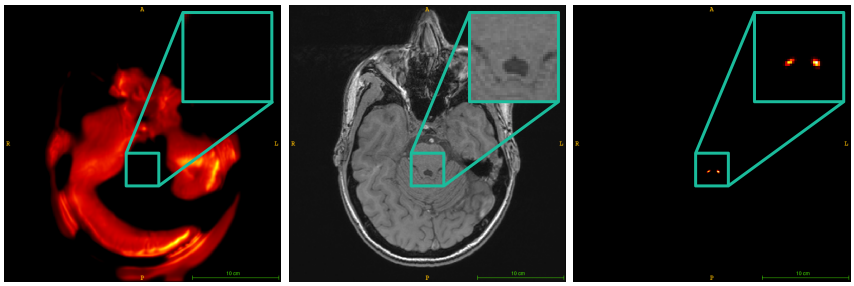
Fig. 2. Boxplots of DSCs and FDRs calculated for different scenarios: NPP: applied the segmentation net from [4] without its post-processing step; PP: same as NPP, but with the post-processing; LPP: using the proposed localizer (CoRe-Unet + AUG) to extract a patch that is passed to the net of [4]. "_left" and "_right" encode the respective LC.

(left LC: 2.27mm ±2.99mm, right LC: 1.68mm ±1.85mm), the errors are substantially larger than in both sagittal (left LC: 0.13mm ±0.16mm, right LC: 0.12mm ±0.17mm) and coronal (left LC: 0.14mm ±0.18mm, right LC: 0.15mm ±0.18mm) direction. This trend was not observed in NoAUG.

Furthermore, the boxplots in Figure 2 show, that the scenario including the CoRe-Unet as a pre-processing step to determine the relevant patch first (LPP), yielded preferable segmentation performance of the subsequent 3D-Unet in terms of both reported metrics as compared to the other tested options (NPP, PP[4]). This trend continues when assessing the ICCs of the CRs. The average ICC of NPP was close to zero (for median CRs: 0.06, for maximum CRs: 0.01). PP slightly improved to averages of 0.37 and 0.16 for median and maximum CRs respectively. LPP obtained average ICCs of 0.89 (median CRs) and 0.66 (maximum CRs).

4 Discussion

The calculation of the errors and their standard deviations reported in Table 1 indicated, that using augmentation during the training increases the performance of the network. Although the statistical significance of the performance difference remains to be determined, further investigation of possible reasons for this was carried out. We examined the heatmaps that are produced as a byproduct and that allow insight into the attention of the network. A clear trend, which is illustrated in Figure 3 by an example could be found. While the heatmaps of AUG have focussed on the actual position of LC, the NoAUG version merely highlighted structures that, when the scalar product is calculated, result in a coordinate roughly in the center of the pons. The latter is undesirable behaviour that indicates that the network did not learn to localize the LC itself, but relied on correlating, more prominent structures. Similar to the additional weighted term in the loss function proposed by [8] that forces similarity of the heatmap to a certain distribution, we presume that the augmentation had a regularizing effect. We conclude from the measured dimension-wise errors, that it is most



(a) without augmentation (b) input (c) with augmentation

Fig. 3. An exemplary slice of a testset sample (b) and the corresponding heatmaps of the network once trained without augmentation (a) and once with augmentation (c). The heatmaps for left and right LC have been added for better visualization.

challenging for the network to determine the rostrocaudal extent of the LC. The experiments wrt. the different LC segmentation pipeline versions (NPP, PP, LPP) suggest that the option utilizing the CoRe localizer performs preferable to the others. We found quite substantial differences between the results of our PP option and the values reported in [4], which we assume have their origin in the different evaluation scheme applied in this study.

The results of this work suggest that including a CoRe-Unet localizer as a pre-processing step in the established LC segmentation pipeline can improve the performance. This improvement is mostly achieved by reducing false positive regions outside the relevant vicinity of the LC. Visual inspection of the heatmaps found that the method shows potential for generating segmentation masks. In this case, instead of laborously delineating a complete mask, it would be sufficient to merely select single points per LC as the ground truth for training. Therefore, we will investigate this approach as a weakly supervised segmentation method in future work.

Acknowledgement. This work received funding from the federal state of Saxony-Anhalt, Germany (Project I 88).

References

1. Betts MJ, Kirilina E, Otaduy MCG, et al. Locus coeruleus imaging as a biomarker for noradrenergic dysfunction in neurodegenerative diseases. Brain. 2019;142(9):2558–2571.
2. 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.
3. Çiçek O, 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.
4. Dünnwald M, Betts MJ, Sciarra A, et al. Automated Segmentation of the Locus Coeruleus from Neuromelanin-Sensitive 3T MRI Using Deep Convolutional Neural Networks. Bildverarbeitung für die Medizin 2020. 2020; p. 61–66.
5. 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.
6. Tona KD, Keuken MC, de Rover M, et al. In vivo visualization of the locus coeruleus in humans: quantifying the test–retest reliability. Brain Structure and Function. 2017;222(9):4203–4217.
7. Keren NI, Lozar CT, Harris KC, et al. In vivo mapping of the human locus coeruleus. NeuroImage. 2009;47(4):1261–1267.
8. Nibali A, He Z, Morgan S, et al. Numerical Coordinate Regression with Convolutional Neural Networks. arXiv:180107372 [cs]. 2018;.
9. 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.
10. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv:14126980 [cs]. 2017;.