**Reviewer 1**

Notes:

* Interesting work comparing deep learning method for segmenting subthalamic nuclei on 3T and 7T data, compared against ground truth of hand traced nuclei by a neuroradiologist.
* Additional finding of specific subthalamic nuclei volume loss in multiple sclerosis (MS) patients, consistent with volume loss noted in MS pathologic studies suggests clinical utility for this technique.

Requests

1. Stating the source of ground truth within the abstract would strengthen the manuscript.

Thanks for pointing this out. We have added a sentence about manual segmentation ground truth labels to the Abstract.

1. Manuscript would be strengthened by discussion of how this technique and associated image acquisitions could be translated into clinical practice.

We thank the reviewer for bringing this to our attention. The most immediate clinical applications that this method can impact are deep brain stimulation and ultrasound guided thalamotomy for essential tremor. We have added some sentences in the Discussion adding this important point including the availability of pulse sequences esp. WMn-MPRAGE on vendor platforms. That is critical for clinical adoption.

1. Additionally, discussing aspects of results that would speak towards the necessity of 7T imaging, or the sufficiency of 3T imaging to aid in this clinical translation would also strengthen the manuscript.

We did a preliminary comparison of 3T vs. 7T WMn-MPRAGE segmentation performance in a small cohort of ET patients in this paper and found the segmentation comparable. More studies are needed to establish the validity of our segmentation method at 3T and especially using conventional MPRAGE. A recent study of Zahr et al. (HBM 2019) found that THOMAS at 3T can detect atrophy of the VLp nucleus. This is promising as our technique performs comparably or better than THOMAS. We have added a paragraph in Discussion on this.

1. Discussing rationale for, and impact of including patient with alcoholism for network initialization would also strengthen the manuscript.

The AUD data was used only to derive weights to initialize the WMn and CSFn networks to improve convergence rate (the alternative is random weights which can result in lower accuracy and convergence rate as is pointed in Supporting Figure S1). The networks were trained using MS, ET, and healthy subject data. We have clarified this in the Methods section. To test the ability of the network to segment AUD images was beyond the scope of this paper but is the scope of an ongoing project. AUD images are almost identical in appearance to healthy subject images except for some atrophy (i.e. no lesions) and we expect the network to also perform well on AUD images.

**Reviewer 2**

The authors developed a convolutional neural network and trained on an existing dataset to put segmentation of subnuclei of the thalamus, accuracy was evaluated against manual segmentation. Segmentation was performed on both white matter nulled MPRAGE and conventional MPRAGE images. The method was used to study patients with multiple sclerosis. The authors reported better accuracy of the neural network based method to a previously proposed registration based method from the same group.

In general  this is an interesting study. Although both the network architecture or the image acquisition technique are not novel, it represents a novel application of neural network on an interesting topic. I however have the following concerns:

1. The "gold standard" manual segmentation of the subnuclei is not contiguous. Please provide rationale.

Thanks for pointing out this lapse. The “gaps” are very small nuclei (centromedian, ventral posteromedial, and the interlaminar nuclei) that were in the Morel atlas but could not be reliably segmented manually by the neuroradiologist even at 7T due to lack of contrast. These gaps were considered as “background” during the neural network training process. We have added a line to Methods clarifying this.

1. Also, the authors only perform manual segmentation of the left hemisphere and flip the segmentation to the right side. This approach doesn't seem to match the segmentation with signal contrast as shown on Figure 4 for the right thalamus.

This is possibly a misunderstanding of how our method segments the right side, stemming from our inadequate description. We do not flip the left labels but rather **flip the original image**, then perform the segmentation (it treats the right side as a new subject essentially) and then flip both the segmented labels and images back to the original orientation. This makes the right side segmentation completely independent of the left side and its labels. We have clarified this point in the Methods section. Regarding Fig 4 and the segmentation not matching the contrast contours, a different window-level selected to reveal the image contrast better. Figure 5 was similarly modified.

1. The manuscript is slightly long. The authors might want to consider shorten it.

Our page counts are as follows

* Introduction: 3 pages
* Method: 7 pages
* Results: ~3.5 pages
* Discussion & Conclusion: ~2.5 pages

We agree that it is indeed long and had already trimmed it down from the original version by making use of Supplemental Figures and Tables. We have further cut several sentences from the Introduction which we thought were not germane. We have also moved the loss function formulae etc. from Methods to the Appendix. These changes make the manuscript more compact whilst preserving the flow.