



## RESEARCH ARTICLE

# Deep-learning based fully automatic segmentation of the globus pallidus interna and externa using ultra-high 7 Tesla MRI

Oren Solomon<sup>1</sup>  | Tara Palnitkar<sup>1,2</sup> | Re'mi Patriat<sup>1</sup> | Henry Braun<sup>1</sup> | Joshua Aman<sup>2</sup> | Michael C. Park<sup>2,3</sup> | Jerrold Vitek<sup>2</sup> | Guillermo Sapiro<sup>4,5,6,7</sup> | Noam Harel<sup>1,3</sup> 

<sup>1</sup>Department of Radiology, Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota

<sup>2</sup>Department of Neurology, University of Minnesota, Minneapolis, Minnesota

<sup>3</sup>Department of Neurosurgery, University of Minnesota, Minneapolis, Minnesota

<sup>4</sup>Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina

<sup>5</sup>Department of Biomedical Engineering, Duke University, Durham, North Carolina

<sup>6</sup>Department of Computer Science, Duke University, Durham, North Carolina

<sup>7</sup>Department of Mathematics, Duke University, Durham, North Carolina

## Correspondence

Oren Solomon, Department of Radiology, Center for Magnetic Resonance Research, University of Minnesota, 2021 6th Street S.E. Minneapolis, Minnesota, 55455, USA. Email: solom267@umn.edu

## Funding information

National Institution of Health, Grant/Award Numbers: P30 NS076408, P41 EB027061, P50 NS098753, R01 NS081118, R01 NS113746

## Abstract

Deep brain stimulation (DBS) surgery has been shown to dramatically improve the quality of life for patients with various motor dysfunctions, such as those afflicted with Parkinson's disease (PD), dystonia, and essential tremor (ET), by relieving motor symptoms associated with such pathologies. The success of DBS procedures is directly related to the proper placement of the electrodes, which requires the ability to accurately detect and identify relevant target structures within the subcortical basal ganglia region. In particular, accurate and reliable segmentation of the globus pallidus (GP) interna is of great interest for DBS surgery for PD and dystonia. In this study, we present a deep-learning based neural network, which we term GP-net, for the automatic segmentation of both the external and internal segments of the globus pallidus. High resolution 7 Tesla images from 101 subjects were used in this study; GP-net is trained on a cohort of 58 subjects, containing patients with movement disorders as well as healthy control subjects. GP-net performs 3D inference in a patient-specific manner, alleviating the need for atlas-based segmentation. GP-net was extensively validated, both quantitatively and qualitatively over 43 test subjects including patients with movement disorders and healthy control and is shown to consistently produce improved segmentation results compared with state-of-the-art atlas-based segmentations. We also demonstrate a postoperative lead location assessment with respect to a segmented globus pallidus obtained by GP-net.

## KEYWORDS

7 Tesla magnetic resonance imaging, convolutional neural network, deep brain stimulation, deep-learning, globus pallidus, Parkinson's disease, patient-specific, segmentation

## 1 | INTRODUCTION

In the past several decades, deep brain stimulation (DBS) therapy has shown clear clinical efficacy in the mediation of symptomatic motoric behavior associated with Parkinson's disease (PD), essential tremor

(ET), dystonia, and other conditions (Benabid et al., 1987; Deuschl et al., 2006; Hariz et al., 2008; Mueller et al., 2008; Obeso et al., 2001; Volkmann et al., 2012). One of the most prominent DBS targets for PD and dystonia is the globus pallidus (GP) (Patriat et al., 2018). The GP is divided into two compartments, the internal

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

segment (GPi) and the external segment (GPe), of which the former is typically the actual target for electrode placement. The GPe and GPi are separated by a thin layer, called the internal medullary lamina (Lozano & Hutchinson, 2002; Patriat et al., 2018). Several past studies have reported that lesions applied to the GPi led to improvement in motor function (Baron et al., 1996; Obeso et al., 2001; Vitek et al., 2003). The application of lesions (or pallidotomy) is associated with nonreversible risks of applying the lesion outside of the intended target. DBS surgery, on the other hand, has risen as an alternative with similar benefits, whose application can be reversed or even stopped if erroneously applied to the wrong region (Benabid et al., 1987; Obeso et al., 2001). Recent studies have shown that accurate placement of the DBS electrode within the sensorimotor region of the target (e.g., subthalamic nucleus [STN] or GPi) is directly correlated with the success of the DBS procedure and reduction of adverse effects (Ellis et al., 2008; Marks et al., 2009; Paek et al., 2013; Patel et al., 2015; Richardson et al., 2009; Rolston et al., 2016; Welter et al., 2014). Correct anatomical target identification is characterized not only by the target's center of mass, but also by its boundaries (Kim et al., 2019). Thus, precise identification of both GPe and GPi and their lamina boundary is of great importance.

A fully automated segmentation process of the GP (both the internal and external segments) has several clear advantages, among which are accurate and fast inference. From a clinical point of view, an automated process has the potential to streamline clinical workflow and increase patient throughput, both in preoperative surgery planning and postoperative assessment of the DBS lead location with respect to the target. Such a process can also eliminate human bias associated with the segmentation process, and provide more accurate and consistent segmentation results.

Since some anatomical structures are not easily identified or visualized on standard clinical images (e.g., 1.5 or 3 T MRI scanners), a common approach to localize brain structures, and in particular those located in the basal ganglia, is to rely on an atlas (Ewert et al., 2018; Horn et al., 2019; Horn & Kühn, 2015).<sup>i</sup> An atlas provides an average location of brain structures, often based on multiple inputs, such as different MRI scans and histology, merged from numerous subjects. For example, the authors of (Chakravarty et al., 2006) combined a histological atlas with a 3 T based multimodal subcortical atlas built from MRIs of PD patients (Xiao et al., 2017). In addition, they also combined high resolution multimodal MRIs and structural connectivity data as well (Ewert et al., 2018). Atlases can be deterministic (Xiao et al., 2017), that is, each pixel corresponds to a single brain structure, or probabilistic (Ewert et al., 2018), where each pixel is associated with a vector of probabilities which indicate how likely that pixel is associated with different brain structures. For example, (Horn, Kühn, et al., 2017) took a probabilistic approach to map DBS electrode locations onto the Montreal Neurological Institute (MNI) space.

Atlases, which are typically defined in a normalized space, have shown great importance in retrospective population studies (Horn, Kühn, et al., 2017; Horn, Neumann, et al., 2017; Horn, Reich, et al., 2017; Kim et al., 2019). However, recent works have shown that variability exists in the size and shape of deep brain structures

between different subjects (Abosch et al., 2010; Duchin et al., 2018; Lenglet et al., 2012; Patriat et al., 2018). This inherent interpatient variability is not fully captured by atlas-based segmentations, as they provide a single, averaged shape of brain structures, be it deterministic or probabilistic. Accounting for interpatient variability is typically done by registering the atlas to specific patient anatomy (i.e., via an MRI scan), although this approach cannot often account for the discrepancy between the template and the individual patient brain, and is also affected by registration errors (Dadar et al., 2018; Ewert et al., 2018; Kim et al., 2019). Several other approaches have been introduced in recent years. However, these approaches either segment the GP as an entire structure and do not provide distinction between the GPe and the GPi (Manjón & Coupé, 2016; Visser et al., 2016), or still rely on registration to a template for segmenting the GPe/GPi and other structures (Bazin et al., 2020). These drawbacks motivate the development of a truly patient-specific segmentation technique for GPe/GPi structures.

In recent years, the fields of image analysis and computer vision have undergone a monumental and profound change with the introduction of deep-learning (DL), and in particular deep fully convolutional neural networks (CNNs; Krizhevsky et al., 2012; Lecun et al., 2015). These types of networks consist of many aggregated layers of convolution filters of various sizes and nonlinear elementwise activation functions, such as the rectified linear unit (ReLU). In many computer vision tasks (e.g., segmentation/classification), CNNs are trained end-to-end in a fully supervised manner, over pairs of input images and (often manual) delineations or labels. Such trained CNNs have shown state-of-the-art performance in many tasks, such as semantic segmentation (Long et al., 2015), image classification (Krizhevsky et al., 2012; Zhong et al., 2015), and image registration (Balakrishnan et al., 2018; Dalca et al., 2018) to name a few. As opposed to iterative algorithms, CNNs perform inference in a single forward step without any need for time-consuming iterations. Moreover, since CNNs are composed from convolution operations and elementwise nonlinearities, they are implemented very efficiently, which leads to fast execution times during inference.

In this study, we present *GP-net*,<sup>ii</sup> a fully convolutional deep neural network for the efficient and accurate 3D segmentation of both the GPe and GPi. *GP-net* is based on a variant of one of the most prominent deep network architectures, the U-net (Ronneberger et al., 2015), which exploits skip connections in order to prevent loss of contextual information at multiple image scales. In particular, we exploit the attention gated (AG) U-net proposed in (Schlemper et al., 2019), which has previously shown improved segmentation performance in medical imaging. Attention mechanisms are able to automatically learn to focus, or direct attention, without additional supervision. This ability allows AGs to highlight salient features during inference time in the input images or intermediate feature maps, and have been applied successfully in numerous machine learning disciplines, such as natural language processing and machine vision (Bahdanau et al., 2015; Wang & Shen, 2018). For example, in (Schlemper et al., 2019) it was shown that the added AGs improve model sensitivity and accuracy in medical computerized tomography

(CT) and ultrasound segmentation, by suppressing irrelevant feature activations in irrelevant areas.

In addition, we augment the attention gated U-net with the recently introduced deformable convolutions (Dai et al., 2017; Pominova et al., 2019), by replacing some of the intermediate 3D convolution layers in the network with 3D deformable convolutions. Classical convolution filters rely on convolving the learned kernel with input which lies on a regular Cartesian grid. In (Dai et al., 2017) it was shown that by learning the grid sample offsets, instead of using a fixed grid, improved performance in vision-based tasks such as classification and segmentation can be achieved. In that sense, deformable convolutions can be thought of as another form of an attention mechanism.

Recent advances in ultra-high MRI machines and acquisition protocols have allowed 7 T imaging methods to directly visualize and identify small subcortical deep brain structures. Structures such as the STN or the lamina border between the GPe and GPi cannot be clearly visualized in lower field MRI machines, but can be more clearly identified with the use of 7 T due to improved contrast and resolution (Abosch et al., 2010; Patriat et al., 2018). Ultra-high field MR has already been used for deep brain structure identification; for example (Duchin et al., 2018) have shown that direct visualization of the STN is possible, and (Kim et al., 2019) have shown that by relying on 7 T manual delineations and machine learning techniques, direct segmentation of the STN on 3 T images is possible, with 7 T accuracy and precision. Atlas-based approaches have also exploited the benefits of 7 T imaging and superior contrast to construct state-of-the-art atlases.

One such example is the study of (Keuken et al. 2014) which constructed an atlas that relies on multi-contrast 7 T acquisitions to analyze the anatomical variability of subcortical structures.

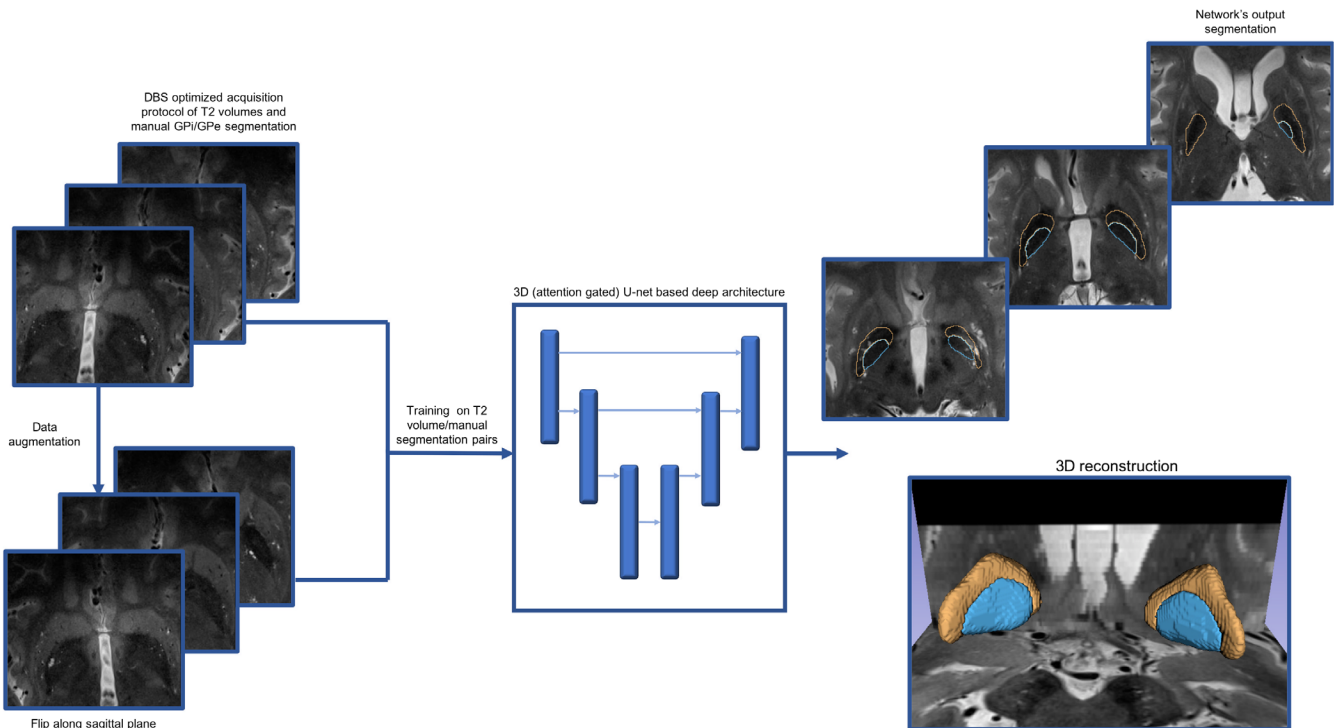
In this study, we rely on acquired T2 volumes from an ultra-high field 7 T MRI machine, specifically tailored to visualize the basal ganglia region. The training cohort consists of movement disorder patients as well as healthy control subjects, such that for each subject manual delineations of both left and right GPe and GPi are obtained by experts to train the network end-to-end. An overall illustration of the proposed process is given in Figure 1. In the next sections, we outline the mechanism behind GP-net, provide extensive experimental validation and finish with a discussion and concluding remarks.

## 2 | METHODS

### 2.1 | Overview

GP-net is trained end-to-end in a fully supervised manner. Manual 3D delineations performed by domain experts for both the GPe and GPi are extracted per each patient's 7 T T2 scan. Thus, the network is fed with pairs of 7 T T2 volumes and corresponding 3D manual segmentations in the training procedure.

To evaluate the performance of GP-net, as well as the quality of its patient-specific segmentations, the resulting automatic delineations are compared against manual GP segmentations, performed on



**FIGURE 1** Proposed GP-net for the automatic and patient-specific segmentation of the GPe and GPi from acquired 7 T T2 volumes. GP-net is an attention gated U-net, trained end-to-end from pairs of 7 T T2 scans and manual 3D delineations of both GPe and GPi. To increase training size, each training pair is augmented by flipping the images along its sagittal plane. Resulting network's output are the automatically segmented GPe (orange) and GPi (blue) 3D volumes

7 T T2 scans by the same group of experts, from a test set, which was excluded from the training phase of the network. The following metrics against the manual delineations are compared: dice score, center of mass Euclidean distance, volume and mean surface differences. In addition, we provide extensive quantitative comparison between GP-net and four state-of-the-art atlases. GP-net shows significant improvement over existing atlas-based techniques. A test-retest experiment was performed to assess the consistency of GP-net in segmenting the same structure of the same patient over several different scans acquired over the course of days. Moreover, to evaluate the clinical utility of the proposed method we compared the DBS lead location postsurgery as determined based on manual segmentation and the automatic GP-net segmentations.

## 2.2 | Scanning protocol

Patients were scanned on a 7 T MRI scanner (Magnetom 7 T Siemens, Erlangen, Germany) using our previous published protocols (Abosch et al., 2010; Duchin et al., 2018). The scanner was equipped with SC72 gradients capable of 70 mT/m and a 200 T/m/s slew rate using a 32-element head array coil (Nova Medical, Inc., Burlington, MA). On the day of scanning, the patients were instructed to take their usual medication in order to optimize patient comfort and minimize motion. Whenever patient head size enabled enough space in the coil, dielectric pads were utilized in order to enhance signal in the temporal regions (Teeuwisse et al., 2012). The scan protocol consists of: T1-weighted whole brain scan (0.6 mm<sup>3</sup> isotropic) and T2-weighted axial slab covering from the top of the thalamus to the bottom of the substantia nigra with 0.39 × 0.39 × 1 mm<sup>3</sup> resolution. The T1 weighted scan was used only for atlas-based registration when comparing GP-net with different atlases and was not used in the network's training, inference or validation phases (details are given in Sections 2.6 and 3).

## 2.3 | Database and preprocessing

A cohort of 101 subjects, including 24 healthy controls and 77 movement disorder (PD and ET) patients participated in this study. All subjects were scanned on the 7 T scanner; patients were scanned prior to their DBS surgeries. Even though ET patients do not typically undergo GPi-DBS, we also chose to include their imaging data in our dataset.

Out of the 101 participants, images from 58 participants were used for training and 43 for testing; demographic details are given in Tables 1 and 2. In addition, one subject was scanned three times on 2 days with the same scanning protocol to assess the network's stability. For each subject in the cohort, manual delineations of both the left and right GPe and GPi were obtained by three independent experts from the scanned T2 volumes. Final manual delineations for training and testing per each subject were obtained by a consensus between all experts.

To increase the number of training pairs (T2 images and manual delineations), each training pair was mirrored along the sagittal plane (Figure 1). Thus, the number of training pairs was doubled and provided more training examples of spatially translated GP structures. The T2 volumes and manual segmentations were resampled to an isotropic grid of 0.39 × 0.39 × 0.39 mm<sup>3</sup> with a nearest neighbor interpolation kernel prior to training and inference. All of the quantitative analysis is performed on the resampled grid. This study was approved by the Institutional Review Board at the University of Minnesota and all participants gave their informed consent.

## 2.4 | Network architecture

As previously mentioned, GP-net is a fully convolutional 3D deep neural network. Its base architecture consists of an attention gated U-net, in which some of the inner 3D convolution layers were replaced by 3D deformable convolutions. The basic U-net architecture is composed of two paths, encoder and decoder paths, which consist of aggregated layers of convolutions, max-pooling, and ReLU activations. Each layer shrinks its input size by a factor of 2. Thus, at the end of the encoder stage, the feature dimensions are shrunk by a factor of 2<sup>m</sup>, where *m* is the number of encoder stages (assuming isotropic max-pooling in all input dimensions). The output of this stage is then fed into the decoder, which consists of the same number of *m* layers, each layer is built with convolution layers and upsampling by a factor of 2. The final output of the network has the same dimensions as the input. In this work, *m* = 4. Each encoder stage is composed of two consecutive blocks of 3D convolution (same parameters in both blocks), 3D batch normalization, ReLU activation followed by a max pooling layer of size 2 × 2 × 2 pixels. Deformable convolution layers consist of a 3D standard convolution kernel to estimate the offsets, a deformable convolution layer, followed by another standard 3D convolution (all convolutions of same kernel size and padding). Detailed parameters of the convolution layers are given in Table 3.

**TABLE 1** Subjects demographics

Group type	Ages 20–40		Ages 41–60		Ages 61–80	
	Training	Testing	Training	Testing	Training	Testing
Control	9	0	2	7	1	5
ET	0	0	7	0	8	8
PD	0	0	11	8	20	15

Note: Age was determined at the day of the scan.

Abbreviations: ET, essential tremor; PD, Parkinson's disease.

Control		ET		PD	
Training	Testing	Training	Testing	Training	Testing
34.83 ± 15.68	59 ± 8.95	63.2 ± 9.45	68.75 ± 4.33	63.93 ± 8.11	64.08 ± 9.71

Abbreviations: ET, essential tremor; PD, Parkinson's disease; SD, standard deviation.

**TABLE 2** Subjects mean (av) age ± SD, divided into training and testing

Stage	Block level	Type	Kernel size	Padding size	Dilation
Encoder	1	Regular 3D convolution	11	15	3
	2	Regular 3D convolution	9	8	2
	3	Deformable 3D convolution	5	2	-
	4	Deformable 3D convolution	3	1	-
Central	5	Deformable 3D convolution	3	1	-
Decoder	4	Deformable 3D convolution	3	1	-
	3	Deformable 3D convolution	3	1	-
	2	Regular 3D convolution	5	2	1
	1	Regular 3D convolution	7	3	1
Final		Regular 3D convolution	1	0	1

**TABLE 3** GP-net convolution layers parameters

Note: Kernel sizes are isotropic. All numbers are given in pixels. There is no dilation for the deformable convolution kernels.

In addition, as illustrated in Figure 1, each layer in the encoder stage is directly connected to its corresponding decoder stage, traditionally using skip connections. Skip connections allow more efficient gradient flow through the network during the training stage and prevent loss of contextual information at multiple image scales. Following (Schlemper et al., 2019), in this study the encoder stages are connected to the corresponding decoder stages via attention gates. A detailed description of the attention gate architecture, as well as the overall network architecture is given in (Schlemper et al., 2019). GP-net has three attention gates, for the second, third, and fourth layers of the decoder stage. Each block in the decoder stage is composed of a 3D convolution kernel, followed by an upsampling operator by a factor of  $2 \times 2 \times 2$ . At each decoder level, the resulting attention signal and corresponding upsampled feature map from the previous lower decoder stage are concatenated and convolved with a 3D convolution filter of kernel size 1, no padding and no dilation. Deformable convolutions in the decoder stage are structured the same way as deformable convolutions in the encoder stage. Between the encode and decoder there is another stage (central stage) which is used as the gating signal for the fourth AG.

## 2.5 | Training loss function

GP-net is trained end-to-end using pairs of T2 volumes and corresponding manual delineations. The training loss function used in this study is a combination of several loss functions, detailed below.

Tversky loss (Salehi et al., 2017): In the case of binary segmentation (e.g., network's output and manual delineation), the Tversky index between two groups  $A$  and  $B$  is written as:

**TABLE 4** Tversky loss parameters

Structure	$\alpha$	$\beta$
Background	0.7	0.3
GPe	0.4	0.6
GPI	0.4	0.6

$$\text{Tversky}_{\text{index}}(A, B) = \frac{\text{tp}(A, B)}{\text{tp}(A, B) + \alpha \cdot \text{fp}(A, B) + \beta \cdot \text{fn}(A, B)},$$

where tp stands for true positives (correctly classified voxels), fp stands for false positives (wrongly classified voxels), fn stands for false negatives (wrongly misclassified voxels), and  $\alpha$  and  $\beta$  are corresponding weights. The Tversky loss is taken as  $1 - \text{Tversky}_{\text{index}}$ , as we wish to minimize the loss function through gradient descent. The Tversky loss was reintroduced and utilized in the context of deep-learning based segmentation (Kim et al., 2020; Salehi et al., 2017) as an efficient tool for handling imbalanced class labels, since the parameters  $\alpha$  and  $\beta$  control the relative weights between false positives and false negatives. In this study, we choose these parameters according to Table 4.

The Tversky index can be considered as a generalization of the dice index (Dice, 1945), and indeed by taking  $\alpha = \beta = 0.5$  the Tversky index coincides with the dice index. Additionally and similarly to (Kim et al., 2019), to minimize possible overlap between two different classes in the segmentation, we penalize (minimize) the dice score between each prediction to each other label. This term is weighed by a factor of 0.01 relative to the Tversky loss.

Hausdorff distance (Karimi & Salcudean, 2020): The Hausdorff distance measures the largest (in our case Euclidean) distance

between two given contours. Thus, minimizing the Hausdorff distance can be thought of as minimizing the worst case, or largest outlier distance between the network's segmentation and the manual segmentation, which is indicative of the largest segmentation error. Given two point sets  $X$  and  $Y$ , the one sided Hausdorff distance is defined as (Karimi & Salcudean, 2020; Rockafellar & Wets, 2009)

$$hd(X, Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|_2,$$

and the bidirectional Hausdorff distance is given by

$$HD(X, Y) = \max(hd(X, Y), hd(Y, X)).$$

Since the Hausdorff metric is highly nondifferentiable, we require some differentiable proxy in order to use back-propagation. We rely on the proposed estimator given in equation (8) of (Karimi & Salcudean, 2020). This estimator is a smooth approximation of the Hausdorff distance, which allows back-propagation using gradient descent. In practice, we found it most efficient to minimize the Hausdorff distance for the entire GP (left and right sides together). Initial weight relative to the Tversky loss is 0.00001 and increases by a factor of 5 every 50 epochs.

We train the network of 94 epochs using stochastic gradient descent with learning rate of 0.0001 and momentum factor of 0.9. Batch size is 1. GP-net was implemented in Python 3.6 with PyTorch 1.4 and trained on a single Nvidia V100 GPU with 32 GBs of memory. It takes 3 days to train GP-net on this GPU (done only once).

## 2.6 | Validation

To evaluate GP-net, we performed an extensive quantitative analysis of its performance over 43 subjects. These subjects were not included in the training cohort and were only used for inference and evaluation of the network's performance.

GP-net is compared to the manual segmentation performed by domain experts and against four publicly available state-of-the-art atlases to quantify its performance and validate its reliability for patient-specific GPe/GPi segmentation: (a) DBS intrinsic template atlas (DISTAL, referred to as Ewert 2017) (Ewert et al., 2018); (b) California Institute of Technology (CIT) probabilistic high-resolution in vivo atlas of the human amygdala, also known as the CIT168 reinforcement learning atlas (referred to as Pauli 2020) (Pauli et al., 2018); (c) population-averaged atlas that was made with 3 T MRI of 25 PD patients (referred to as Xiao 2017) (Xiao et al., 2017); and (d) atlas of the basal ganglia and thalamus (referred to as He 2020) (He et al., 2020). All atlases were taken from the lead-DBS software package, in which they are registered to the MNI ICMB2009b template (Horn et al., 2019; Horn & Kühn, 2015).

Since GP-net is patient-specific and operates on the patient's T2 (isotropically resampled) volumetric scan, to perform a fair comparison, we first register the atlases to the same T2 space. This

registration process starts with a registration of the T1 MNI ICBM2009b (3 T) template to a 0.39 mm isotropic MNI template (3 T) using the Advanced Normalization Toolbox (ANTs) (Avants et al., 2011), via the command `antsApplyTransforms` and the `LanczosWindowedSinc` interpolation kernel. Next, the registered template is registered to the patient-specific (0.39 mm isotropically resampled) T1 scan (7 T). This is done via a combination of FLIRT and FNIRT modules from the FSL toolbox (Jenkinson et al., 2012) (implemented via HCP pipelines), using the `ApplyWarp` command and a spline interpolation kernel. The final registration stage includes registering the patient's T1 scan into his/her (isotropically resampled) T2 scan (7 T) using ANTs with a B-spline interpolation kernel and linear registration. The same transformations are applied to the atlases with a nearest-neighbor interpolation kernel. All registrations were verified visually.

We note that some of the atlases, such as the DISTAL (Ewert 2017, Ewert et al., 2018) and CIT168 (Pauli 2017, Pauli et al., 2018) are probabilistic, while GP-net provides a deterministic segmentation map. To make a fair comparison, we have thresholded these atlases with a value of 0.001, meaning that for each class (GPe/GPi), voxels with values below 0.1% probability are zeroed out, while voxels with higher probability are given a value of 1 (GPe) or 2 (GPi). This threshold value was validated visually for each atlas and compared with the T2 ICBM2009b template. We also tried different values (such as 0.01), which yielded similar quantitative results. All registrations which were applied to the atlases, were applied after the thresholding of the probabilistic atlases, in order to make a fair comparison with the non-probabilistic segmentations of GP-net.

On a final note, in some segmentation cases, we observe that the output of GP-net might contain small segmented regions ("islands") which are clearly not related to either the GPe or GPi. These regions are easily removed automatically with a small postprocessing step which removes all segmentation regions except for the largest four (left and right GPe and GPi) through the use of the connected components algorithm (Rosenfeld & Pfaltz, 1966).

## 2.7 | Metrics and statistics

We use the following metrics to compare between the performance of GP-net and the different atlases: (a) Dice score, (b) Center of Mass distance (CoM) (mm), (c) Mean (Euclidean) Surface Distance (MSD) (mm), (d) volume estimation ( $\text{cm}^3$ ), and (e) precision versus recall rates. Dice, CoM, and MSD are also calculated against the manual GPe and GPi segmentations. CoM distance describes how well the segmented structure is being localized in space (i.e., inside the brain), while the dice, MSD, and volume measurements describe how well the shape of the structure is being captured by the different segmentation techniques.

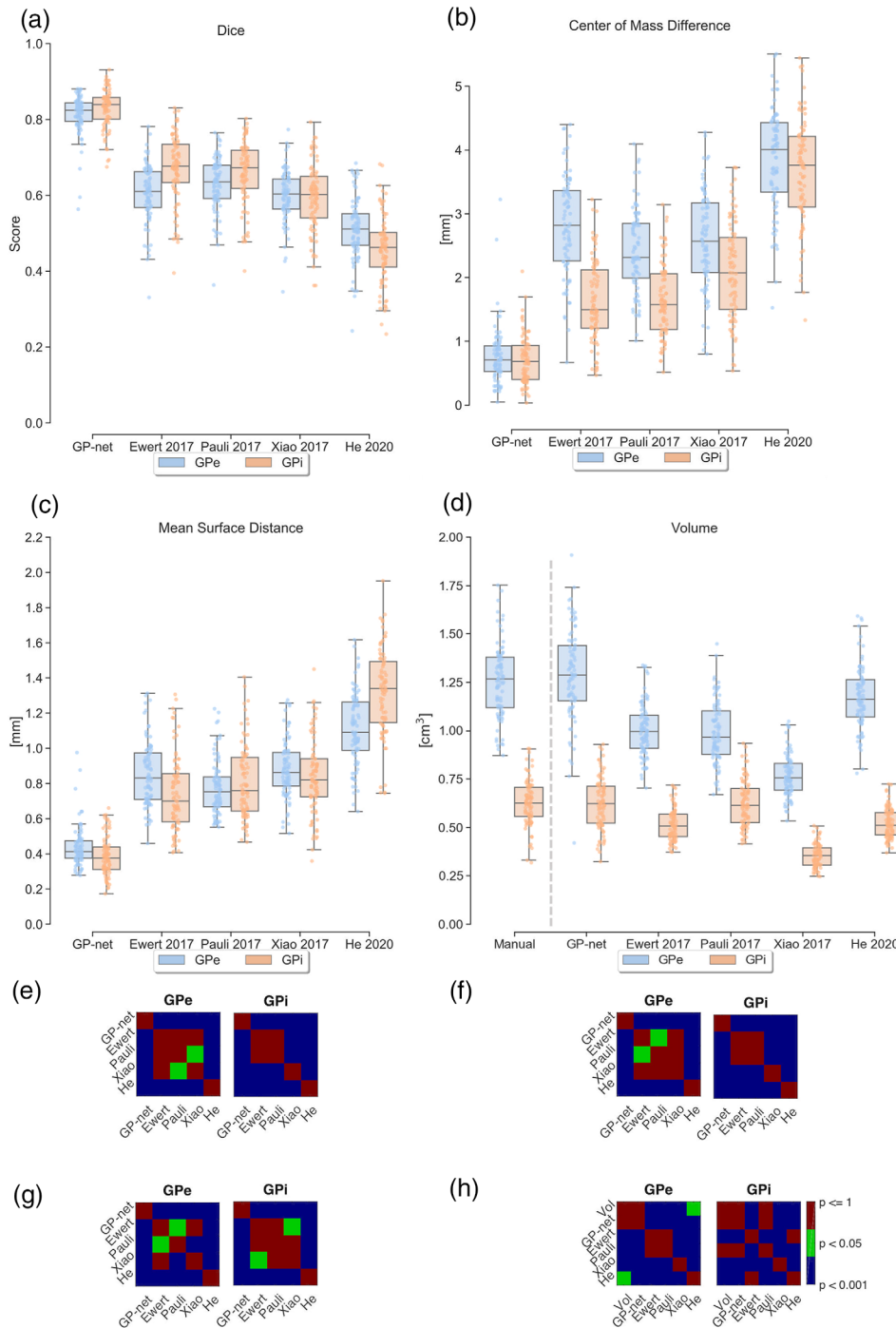
A one-way analysis of variance (ANOVA) was calculated for each metric, followed by a multiple comparison correction and post hoc tests with Tukey's honest significant difference to determine statistical significance between the different methods.

The matrices (Figures 2 and 3) indicate the statistical significance between each method. Each cell in the matrices corresponds to the  $p$  value that the method written in its corresponding row is statistically different from the method written in its corresponding column. A blue cell indicates  $p$  value lower than 0.1%, a green cell indicates  $p$  value lower than 5% and a red cell indicates  $p$  value higher than 5%.

Figure 2d compares between GPe and GPi volume estimates. The two leftmost boxplots (left of the dashed gray vertical line) correspond to the volume estimates measured from the manual delineations, while the other columns correspond to the estimates measured by the

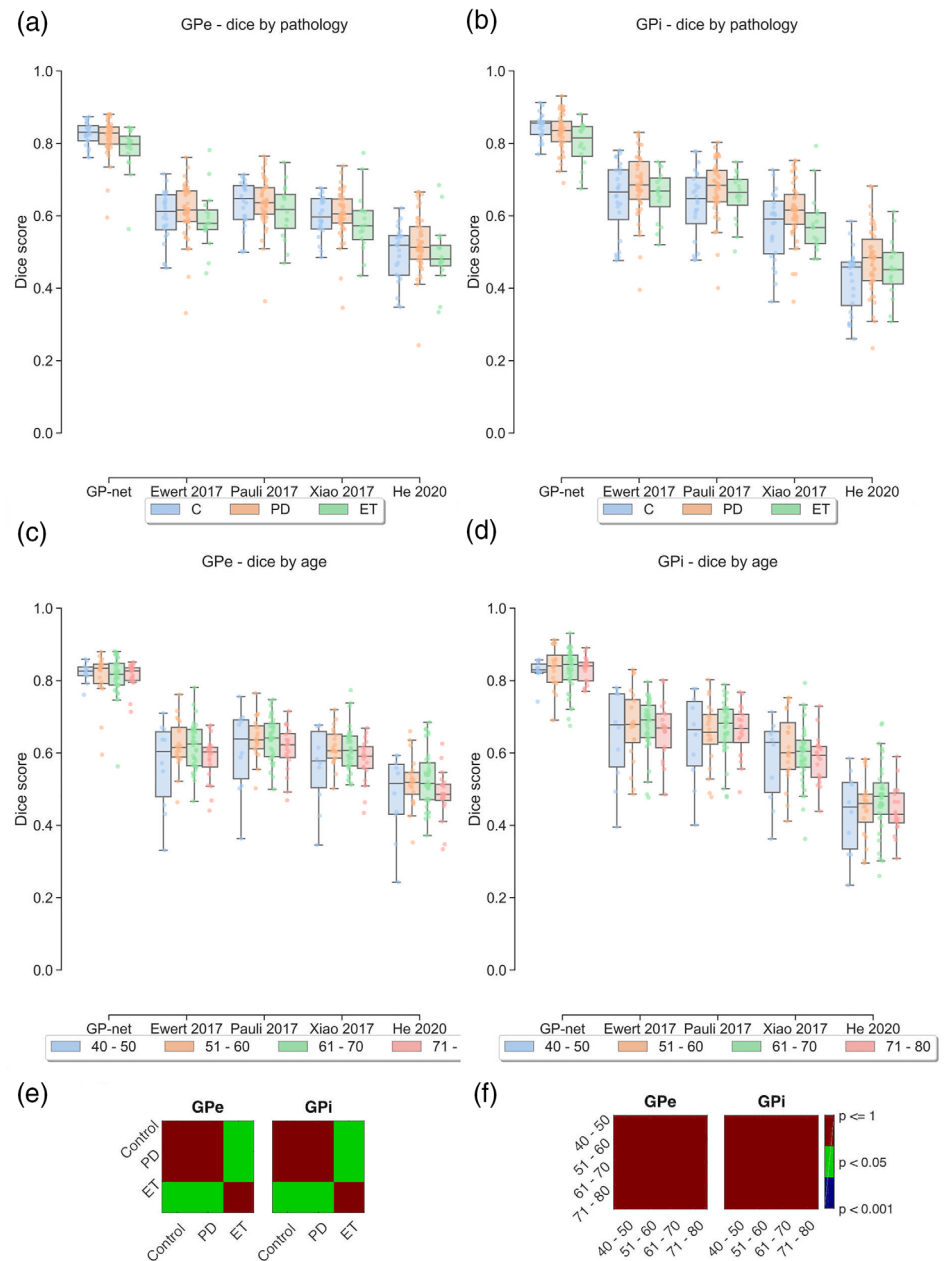
different segmentation techniques. GP-net exhibits the closest similarity between the GPe and GPi volume distributions. Note that corresponding  $p$  values between the manual estimates and GP-net's estimates (both for the GPe and GPi) indicate no statistical differences. Most of the atlas-based estimations show statistically significant volume estimates errors, compared with the manual segmentations, and thus do not fully capture the entire volume distribution.

An additional metric which we utilize to compare between the different segmentations is the precision versus recall rates of each



**FIGURE 2** (a) Dice score comparison between GP-net and the different atlases for both GPe (blue) and GPi (orange) against manual segmentations. Values are compared against the manual delineations for both left and right sides of each patient and presented both as box plots and individual points. (b) Center of mass (CoM) distance relative to the manual delineations. (c) Mean surface distance (MSD) relative to the manual delineations. (d) Estimated volumes. Dashed vertical gray line in (d) separates between the manual volume estimates (left) and the different segmentation techniques' estimates. (e–h) Statistical significance matrices between the different techniques for panels (a–d), respectively. Note that the matrix of panel (h) has  $6 \times 6$  cells, since the first column/row refers to the manual volume estimates. Vol stands for volume

**FIGURE 3** (a, b) Dice scores divided by pathology type: healthy control in blue, PD in orange and ET in green for the GPe and GPi, respectively. (c, d) Dice scores divided by age group (measured in years). (e) Statistical significance matrices for panels (a) and (b) between the healthy control, PD and ET cohorts for GP-net only. (f) Statistical significance matrices for panels (c) and (d) between the different age groups for GP-net only



method. Given a binary segmentation scenario, precision is defined as the ratio

$$Precision = \frac{tp}{tp + fp},$$

while recall (sensitivity) is defined as the ratio

$$Recall = \frac{tp}{tp + fn}.$$

Precision measures a technique's rate of correctly segmenting only GPe/GPi voxels as such, as opposed to classifying background voxels

as either GPe or GPi (hence, high precision rate implies low false positives rate), while recall measures a technique's rate of correctly capturing and segmenting GPe/GPi voxels as such, without misclassifying them as background (hence, high recall rate implies low false negatives rate).

### 3 | RESULTS

In this section, we present a detailed quantitative and qualitative analysis of the performance of GP-net, compared against state-of-the-art atlases and experts' segmentation, as well as a stability test for GP-net.



**TABLE 5** Mean scores and SDs of dice scores, CoM, MSD, and volume (Vol) estimate comparisons between GP-net, atlas-based segmentations, and 7 T manual segmentation (applicable only for volume estimate)

		7 T manual	GP-net	Ewert 2017	Pauli 2017	Xiao 2017	He 2020
Dice	GPe	-	<b>0.81 ± 0.05</b>	0.6 ± 0.08	0.62 ± 0.07	0.59 ± 0.06	0.51 ± 0.08
	GPI	-	<b>0.83 ± 0.05</b>	0.66 ± 0.08	0.66 ± 0.08	0.6 ± 0.08	0.45 ± 0.09
CoM (mm)	GPe	-	<b>0.76 ± 0.45</b>	2.82 ± 0.77	2.42 ± 0.64	2.57 ± 0.76	3.84 ± 0.79
	GPI	-	<b>0.7 ± 0.38</b>	1.63 ± 0.64	1.64 ± 0.57	2.1 ± 0.75	3.66 ± 0.83
MSD (mm)	GPe	-	<b>0.44 ± 0.11</b>	0.85 ± 0.18	0.77 ± 0.15	0.88 ± 0.16	1.11 ± 0.2
	GPI	-	<b>0.38 ± 0.1</b>	0.74 ± 0.21	0.8 ± 0.2	0.84 ± 0.21	1.32 ± 0.25
Vol (cm <sup>3</sup> )	GPe	1.25 ± 0.19	<b>1.29 ± 0.24</b>	1.0 ± 0.13	1.0 ± 0.16	0.76 ± 0.1	1.17 ± 0.16
	GPI	0.63 ± 0.12	<b>0.62 ± 0.13</b>	0.5 ± 0.08	0.62 ± 0.12	0.35 ± 0.06	0.52 ± 0.07

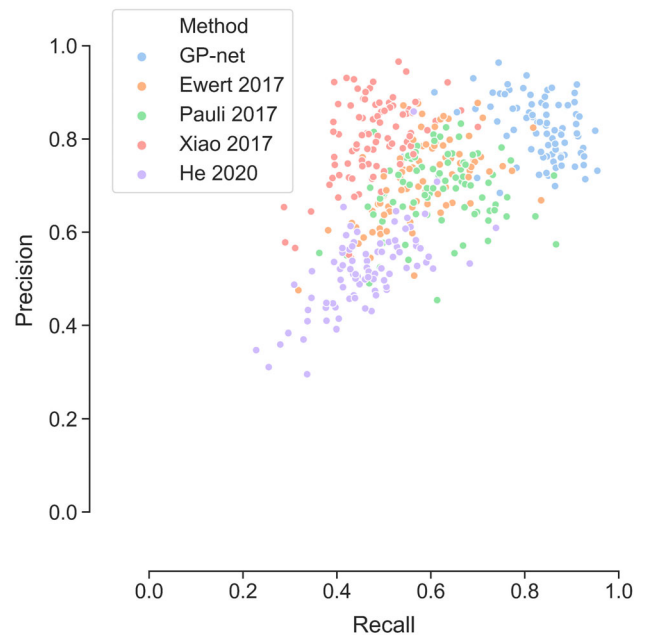
Note:  $n = 86$ .

### 3.1 | Quantitative analysis

We start by presenting comparative results between GP-net and atlas-based segmentation with respect to the manual delineations, as can be seen in Figures 2 and 3, comparing the dice score, CoM (mm), MSD (mm), and volume (cm<sup>3</sup>). In all of these figures, the metrics are calculated per patient, per structure (GPe and GPI), and per side (left and right). Figure 2a shows dice scores for GP-net and the selected atlas-based segmentations, both for the GPe and GPI, individually. Panels (b), (c), and (d) show CoM differences, MSD differences, and volume estimates based on the different segmentation techniques, respectively. Matrices (e–h) indicate statistical significance ( $p$  values) between each pair of the different methods for panels (a–d), respectively.

GP-net has notably superior performance relative to the atlas-based segmentations, with distinctly higher dice scores (for both GPe and GPI) and lower dispersity, as well as lower values for the average CoM difference, and MSD. In both cases,  $p$  value matrices show clear statistical significant difference between GP-net and the other atlas-based metrics. These results are further supported in Table 5 by lower mean values and standard deviations of the CoM and MSD measurements for GP-net compared with the atlas-based segmentations. We provide further insight to the MSD metric in the Figure S1.

Figure 3 presents dice scores divided by disease and age groups. Panels (a) and (b) present the dice scores categorized according to condition (healthy control, PD, and ET), for the GPe and GPI, respectively, while panels (c) and (d) present the dice scores categorized according to four age groups. The two matrices in panel (e) indicate statistical significance ( $p$  values), only for GP-net, between the different pathologies, for panels (a) and (b), respectively. Likewise, the two matrices in panel (f) indicate statistical significant difference ( $p$  values), only for GP-net, between the different age groups for panels (c) and (d), respectively. Panels (a) and (b) show that GP-net consistently achieves higher dice scores compared with the different atlas based segmentations, for healthy controls as well as PD and ET patients. Panels (c) and (d) demonstrate that GP-net's performance is not affected by the large distribution of age amongst the scanned subjects (40 – 80 years). Atlas-based segmentations are not affected by age;



**FIGURE 4** Precision versus recall rates between the different methods (each point represents the rates of either the GPe or the GPI of a single patient, left and right sides combined). Notice that GP-net does not exhibit a trade-off between the two rates and systematically achieves higher values than all of the other atlas-based segmentations

however, they exhibit lower dice scores than GP-net and higher distribution variance. GP-net is shown to consistently produce higher dice scores, without much variation between the different age groups. Consistency of performance, as demonstrated for the proposed GP-net, is critical for DBS and for real deployment.

Lastly, we compare the precision versus recall rates of each of the methods. Figure 4 presents the precision versus recall rates of GP-net and atlas-based segmentations (both GPe and GPI). GP-net's recall and precision rates are both approaching 1 (ideal case) and are higher than all other atlas-based segmentation rates. The average precision rates are  $0.82 \pm 0.06$ ,  $0.71 \pm 0.08$ ,  $0.68 \pm 0.07$ ,  $0.80 \pm 0.08$  and  $0.51 \pm 0.08$  for GP-net, Ewert 2017, Pauli 2017, Xiao 2017 and He 2020, respectively. The average recall rates are  $0.82 \pm 0.08$ ,  $0.58 \pm 0.09$ ,

$0.61 \pm 0.09$ ,  $0.47 \pm 0.07$ , and  $0.46 \pm 0.09$  for GP-net, Ewert 2017, Pauli 2017, Xiao 2017, and He 2020, respectively. Moreover, the recall rates of GP-net are statistically significant from the recall rates of all other atlas-based segmentations ( $p$  value  $<.001$ ). Likewise, GP-net's precision rates are statistically significant than all other atlas-based segmentations' recall rates ( $p$  value  $<.001$ ), except from the precision rates of Xiao 2017.

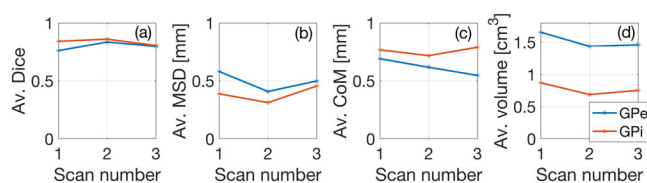
### 3.2 | Stability test

We further test the stability of the network to assess GP-net's systematic behavior. We acquired three independent and repeated 7 T scans of a single healthy control subject over a period of 2 days, and performed inference using GP-net. Figure 5 presents the averaged (left and right hemispheres) dice, MSD, CoM, and volume estimations for both GPe and GPI over the three scans.

The average deviations from the mean dice values in panel (a) are 3.09%/2.44% for the GPe and GPI, respectively. The average deviations for the average MSD (panel [b]) are 11.76%/12.56% for the GPe and GPI, respectively, and for the average CoM (panel [c]) are 7.78%/3.57% for the GPe and GPI, respectively. Finally, the average deviations from the mean value for the volume estimation (panel [d]) are 6.06%/8.62% for the GPe and GPI, respectively. These ranges are well inside the deviations reported in the literature for easier to compute brain characteristics, such as total volume. This analysis demonstrates GP-net's ability to produce reliable and consistent segmentations, not only between different patients, as was previously shown, but also between different scans of the same subject over time.

### 3.3 | Qualitative examples

Figure 6 illustrates segmentation results for a PD patient. The first column shows the 3D reconstructions of both left and right GPe (green for manual/ground truth [GT] and orange GP-net/atlas) and GPI (yellow for manual/GT and blue GP-net/atlas). The second and third columns show selected axial and coronal views (respectively) of T2 slices, superimposed with the manual segmentation and GP-net/atlas segmentation, as well as the different metrics per each method. Judging visually, GP-net is able to produce the most accurate segmentation of both GPe and GPI, presenting excellent agreement with the manual delineation.



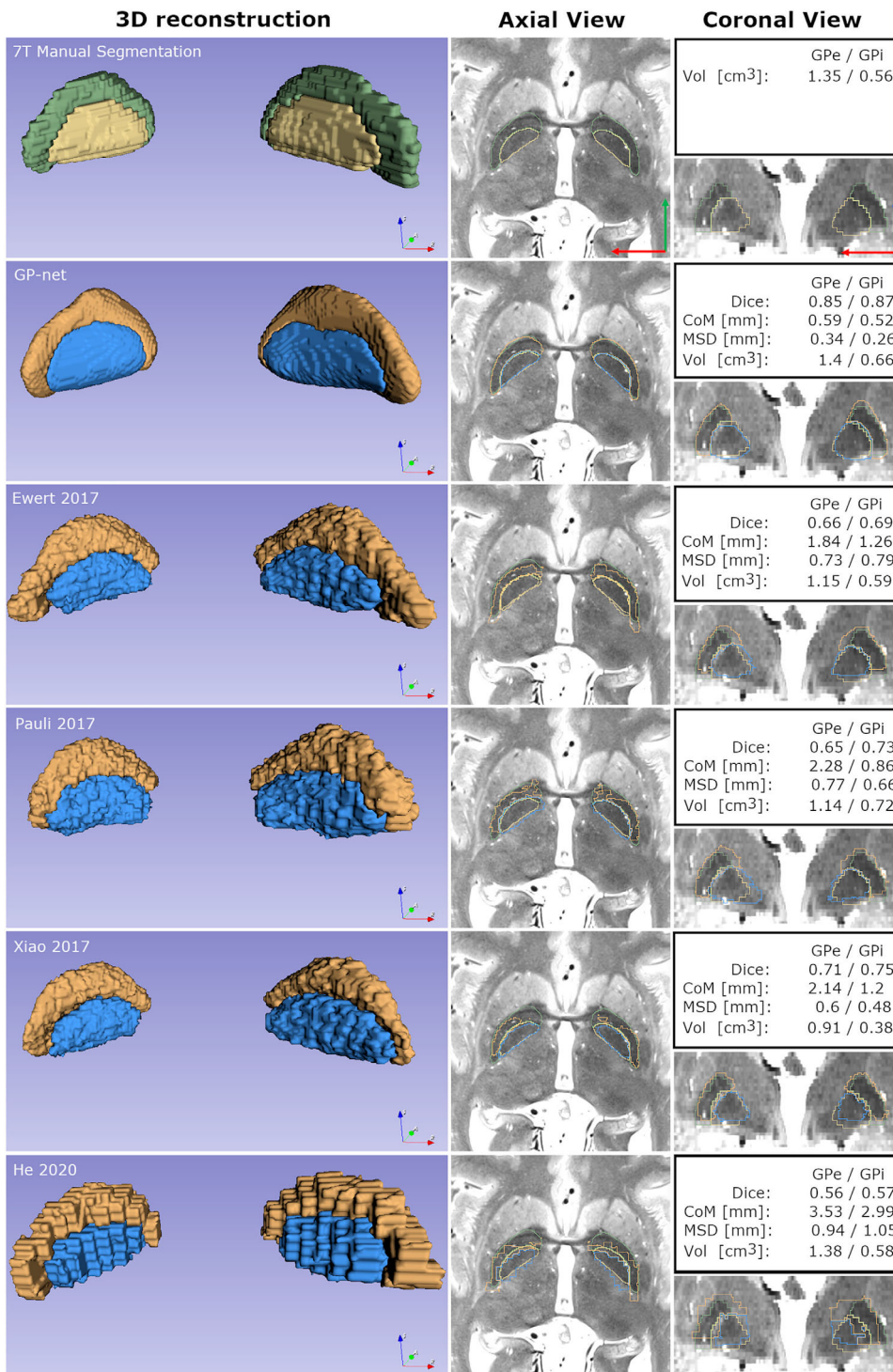
**FIGURE 5** Test-retest scores. Av. indicates average over the left and right GPe (blue)/GPI (red). (a) Av. dice score, (b) av. MSD, (c) av. CoM, and (d) av. estimated volume; GPe in blue and GPI in red

Since GP-net performs inference directly on the isotropically resampled grid ( $0.39 \text{ mm}^3$ ), its output is smooth, as can be seen in the 3D reconstruction, as opposed to the manual delineation (first row), which was segmented on the original  $0.39 \times 0.39 \times 1 \text{ mm}^3$  grid. To make a fair comparison, the atlas-based segmentations were registered from ICBM2009b to the isotropically resampled T2 grid through a  $0.39 \text{ mm}^3$  isotropic MNI template. However, even though we use the same grid spacing, these segmentation results seem to be more pixelated.

Figure 7 presents a clinical scenario, a retrospective analysis of an actual GPI-targeted DBS case. Manual 7 T (panel [a]) and GP-net (panel [b]) 3D reconstructions of the GPe and GPI are shown. Electrode locations were extracted from a postsurgery CT acquired 4 weeks after the implant and registered with the MRI scan (Duchin et al., 2018). Excellent agreement of the electrode locations, with respect to the structures' volumes and borders, as compared with the manual GPe and GPI delineations is observed between both panels. This image exemplifies the clinical potential of GP-net to produce high quality GPe and GPI segmentations from 7 T T2 scans for accurate and reliable visualization of image-based DBS targeting and postsurgery lead localization.

We finish this section by presenting two further unique examples. The first is an example which clearly demonstrates the efficacy of patient-specific segmentation over atlas-based segmentation. A subject with irregular blood vessel bifurcations which traverse into the lower GP region is presented. This subject was not part of the previous statistical analysis and was not part of the training set. Figure 8 demonstrates the manual segmentation (upper row), GP-net segmentation (middle row), and a selected atlas, Ewert 2017 (Ewert et al., 2018) (bottom row). Traversing blood vessels were manually segmented (shown in red), while the manual segmentation and automatic segmentations follow the previously used color convention. As the MNI template does not contain the irregular blood vessels, the atlas-based segmentation is stretched above the vessels in the registration process and misses the true GP region. This can further be observed in the corresponding 3D reconstructions, where the manual and GP-net segmentations near the blood vessels are correct, while the GP atlas-based segmentation is posed above them with a wide gap between, missing sections of the GPI and GPe. Table 6 summarizes the different metrics for this example.

The second unique example is illustrated in Figure 9. In this example, two PD patients (panels a and b, respectively) had to be rescanned, as the first T2 scan was noticeably blurry. For each patient a second, free of motion ("sharp") image was acquired to allow accurate and reliable manual delineation of both the GPe and GPI. In each of the panels a and b, each row corresponds to a 3D reconstruction, a selected axial slice and a zoomed-in region of the same slice from the manual segmentation (first row in both panels), GP-net segmentation based on the sharp acquisition (second row in both panels) and GP-net segmentation based on the blurry acquisition (third row in both panels), respectively. Corresponding dice scores, CoM, MSD and volume estimates are presented in the leftmost column. Interestingly, in the patient of Figure 9a, while notable motion was present in the image, GP-net performed similarly to the sharper image input with equivalent performance while in the second case (Figure 9b), GP-net's



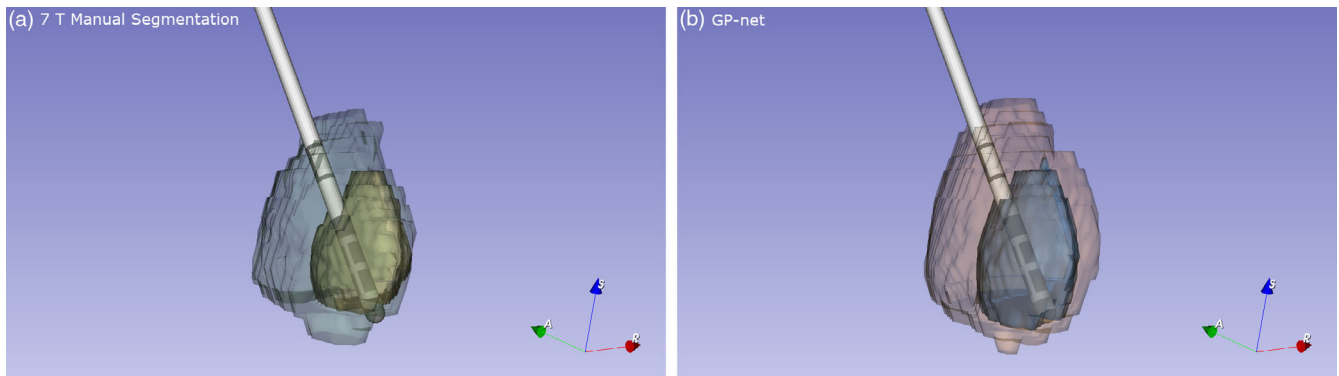
**FIGURE 6** Segmentation and 3D reconstruction of the GP of a representative PD patient. Leftmost column shows 3D reconstruction of both the GPe (green in the manual segmentation and orange in the different reconstructions) and GPI (yellow in the manual segmentation and blue in the different reconstructions). No smoothing was applied at any stage to the reconstructions or manual delineations. Middle column shows a selected T2 axial slice and superimposed outlines of manual delineation, the corresponding segmentations of GP-net (second row) and the different atlases (third to sixth rows). Rightmost column similarly illustrates a selected coronal view. Blue arrow points in the superior direction, green arrow points in the anterior direction and red arrow points in the right direction

output was affected. This example serves to show that GP-net can achieve reasonable performance, even in the presence of nonideal acquisition conditions.

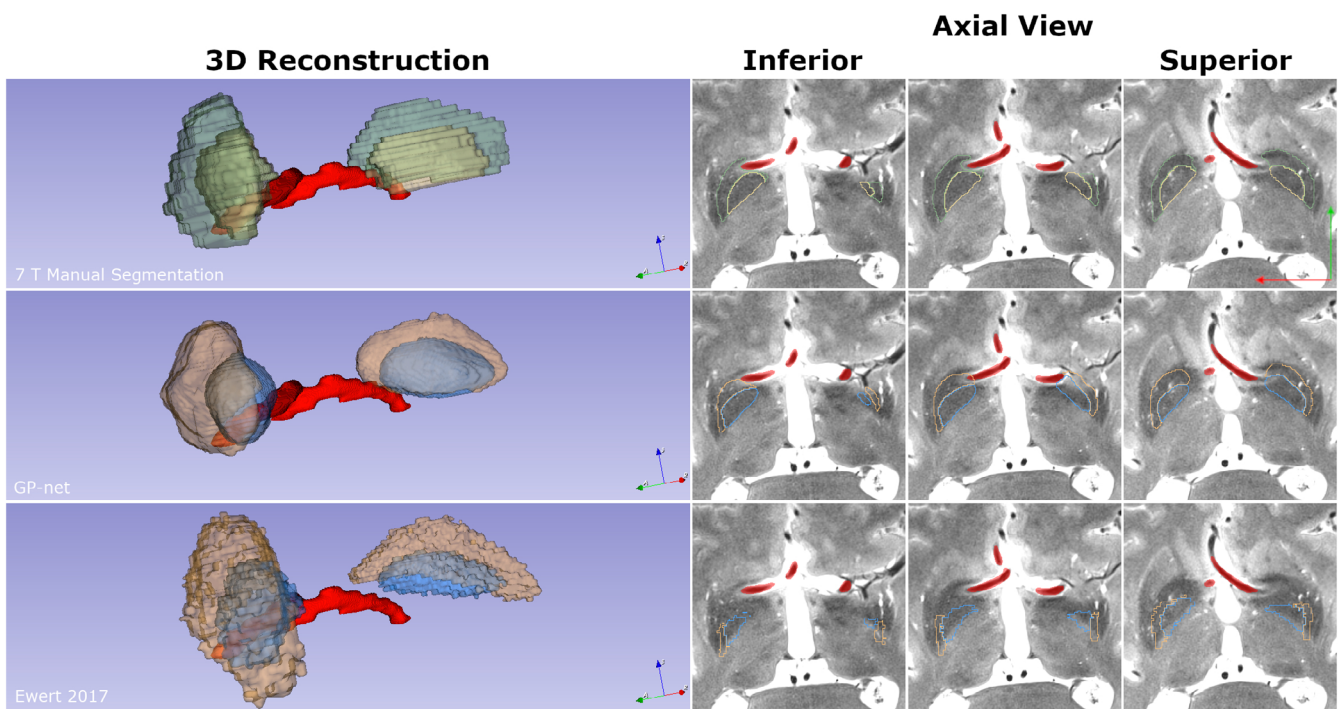
## 4 | DISCUSSION

Alterations in neuronal activity in the internal segment of the GP has been shown to be correlated with motor symptoms of PD. For

example, animal models of PD have shown a characteristic increase in neuronal activities in both the STN as well as in the GPI (Obeso et al., 2001; Wichmann et al., 1994). Lesions applied to these regions have shown striking improvement in motor function. Moreover, the creation of lesions in the GPI of PD patients have been reported to improve contralateral dyskinesia and provide moderate anti-parkinsonian benefits (Baron et al., 1996, 2000; Obeso et al., 2001; Vitek et al., 2003). However, adverse effects which may be caused by lesions cannot be averted once surgery is performed (Kringelbach



**FIGURE 7** (a) 3D reconstruction of a DBS electrode placement with respect to the manually delineated GPe (green)/GPi (yellow) for a specific PD patient (age: 54). (b) 3D reconstruction of the same DBS electrode with respect to GP-net's segmentation of the GPe (orange)/GPi (blue)



**FIGURE 8** GPe/GPi segmentation of a PD patient with irregular blood vessels. Upper row, left to right: 3D reconstruction of the manual segmentation of the GPe (green) and GPi (yellow). Blood vessels are in red (smoothed only for visualization purposes, no other smoothing was applied to any structure). Different panels correspond to selected axial T2 slices, going from the inferior side to the superior side of the brain. Middle row: GP-net reconstruction (GPe in orange and GPi in blue). Bottom row: Segmentation based on the DISTAL atlas (Ewert 2017; Ewert et al., 2018). Other atlases produced similar results to the DISTAL atlas, and were thus omitted for brevity. Blue arrow points in the superior direction, green arrow points in the anterior direction, and red arrow points in the right direction

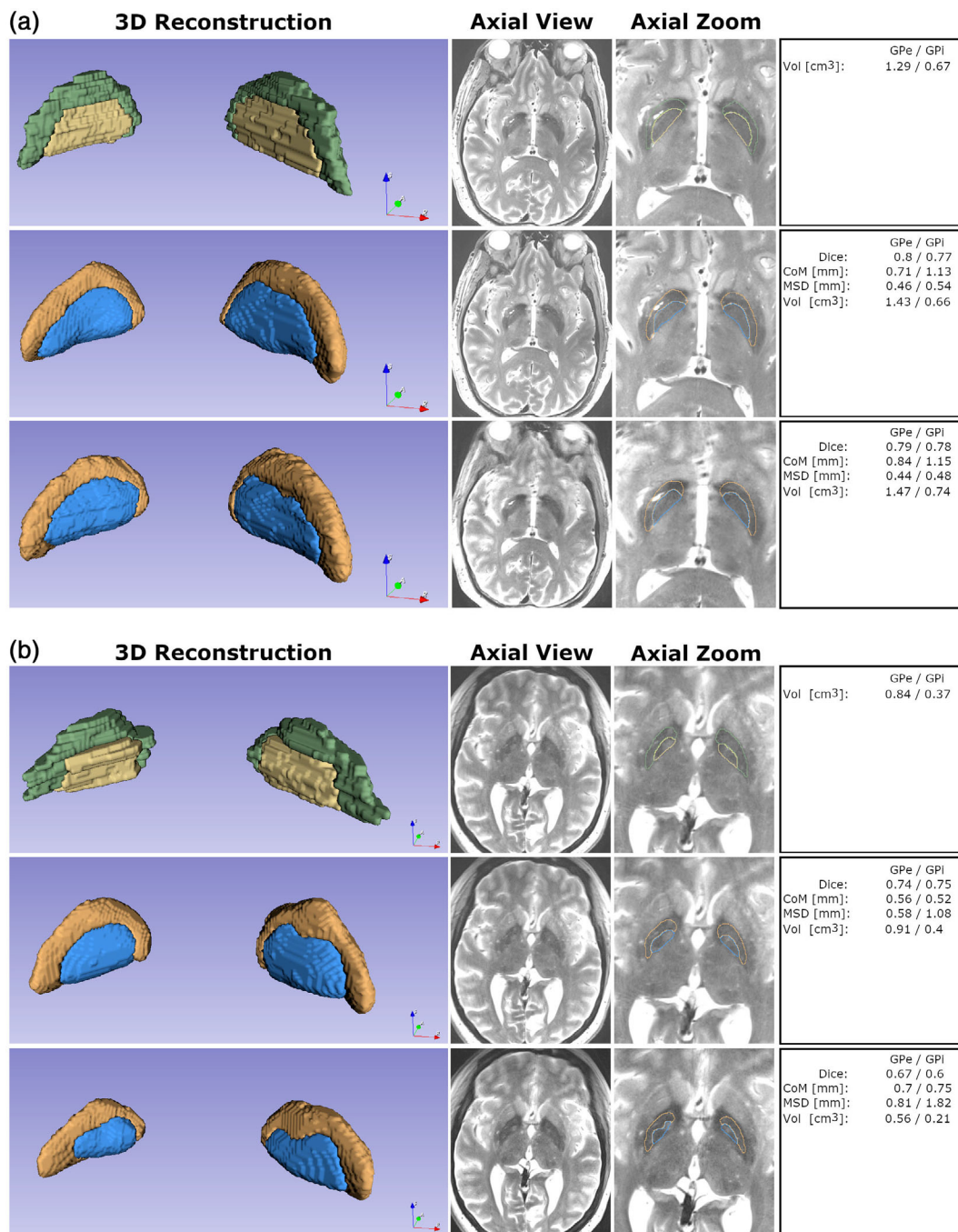
**TABLE 6** Measured metrics (GPe/GPi) for the example in Figure 8

	Dice	MSD (mm)	CoM (mm)	Vol (cm <sup>3</sup> )
GP-net	0.61/0.67	0.83/0.77	0.82/1.4	0.86/0.4
Ewert 2017 (DISTAL)	0.41/0.44	1.27/1.31	4.06/3.34	0.79/0.4

Note: Each value is an average for both the left and right segments. Manually measured volume estimates are 1.07/0.63 (cm<sup>3</sup>) for the GPe and GPi, respectively.

et al., 2007; Munro-Davies et al., 1999). On the other hand, DBS was shown to provide similar clinical benefit compared to a lesion-based therapy while avoiding permanent brain damage which might occur

with lesioning (Benabid et al., 1987; Obeso et al., 2001; Vitek et al., 2020). The success of DBS surgery directly relates to the accurate identification of target regions (Paek et al., 2013; Patel



**FIGURE 9** Robustness of GP-net segmentation under motion condition. Panels (a) and (b) correspond to two different PD patients for which the first scan suffered significant motion blur. 3D Reconstruction column shows 3D reconstruction of both the GPe (green in the manual segmentation and orange in the two GP-net reconstructions) and GPI (yellow in the manual segmentation and blue in the two GP-net reconstructions). No smoothing was applied at any stage to the reconstructions or manual delineations. Axial view column presents selected T2 axial slice of the brain. Apparent motion blurring can be seen in the lowest panel. Axial zoom column illustrates the same slice, zoomed in, and with superimposed outlines of the manual delineation, the corresponding segmentations of GP-net from the sharp image (second row) and from the blurred image (third row). Rightmost column provides the metrics values

et al., 2015; Patriat et al., 2018; Richardson et al., 2009; Rolston et al., 2016; Welter et al., 2014), thus greatly motivating the need for an accurate, robust and reliable identification of these brain areas in an automated manner.

GP-net is a deep-learning based segmentation technique specifically tailored to an accurate and robust segmentation of both the GPe and GPI. Although GP-net was described here in the context of DBS surgery, all clinical procedures which require presurgery GP trajectory

planning, such as DBS surgery and magnetic resonance guided focused ultrasound (Ebani et al., 2020; Miller et al., 2020; Zaaroor et al., 2018) can benefit from this method. In this study, we have utilized recent advances in ultra-high MRI scanner and acquisition protocols and train the network in an end-to-end manner on pairs of 7 T T2 acquisitions and manual delineations produced by domain experts. GP-net is based on several key components: it is a U-net structure which relies on 3D convolutions (Goodfellow et al., 2016) as well as the recently introduced 3D deformable convolutions. It also relies on data augmentation to effectively increase the size of the training set. By mirroring each T2 scan around the anterior–posterior axis, for example, left to right (as well as the corresponding manual delineation), we effectively double the amount of training data. The trained network is able to produce fast (a few seconds per subject), accurate and reliable GPe and GPi segmentations from new 7 T T2 images.

The results presented in this study imply two key observations. First, for all the metrics considered in this study, deep-learning based GP-net was found to be superior to all the atlas-based segmentations tested. GP-net has exhibited improved average dice (scores above 0.8) indicating its ability to more accurately capture the shape of the structure. A higher mean value along with reduced variance indicates that not only does GP-net perform better on average, it is also more stable and has far fewer outlier segmentations. This conclusion is also supported by the *p* value matrices below panel (a) of Figure 2. The first row clearly indicates that GP-net is statistically different than the atlas-based segmentations. On the other hand, some atlas-based segmentations do not significantly differ from one another (e.g., Ewert, Ewert et al., 2018 and Pauli, Pauli et al., 2018), which is indicative of similar performance. Different atlases are constructed from different datasets, each with its own possible bias (e.g., an atlas based on PD patients versus an atlas based on healthy subjects), modalities and reconstruction technique, which may account for the variation in performance between them.

Moreover, the average CoM difference, an indication of how well the structure can be localized in the brain, was measured to be on the order of  $\sim 0.7$  mm, which corresponds to a difference of less than two voxels on the resampled grid (0.39 mm). This CoM localization error (with respect to the manual delineations) is below the actual slice thickness of the acquired 7 T T2 volumes (1 mm). The average mean surface distance for both GPe and GPi was measured to be  $\sim 0.4$  mm, which is on the order of a single voxel on the resampled T2 grid. These numerical results present a significant improvement compared to the atlas-based registrations. All atlas-based registrations achieve an average dice score of about 0.45 – 0.66 and an average CoM error larger than 1.6 mm which corresponds to 4 pixels on the resampled grid. The average MSD for the atlas-based approaches is between 0.74 – 1.32 mm, higher than the MSD reported for GP-net.

In many cases, there exists a trade-off between the precision and recall rates. As exemplified in Figure 4, the precision rates of the atlas-based segmentation of Xiao 2017 (Xiao et al., 2017) is notably higher than its recall rates. However, such a trade-off does not seem to exist for GP-net. The rates of the other atlas-based segmentations do not exhibit such a trade-off, however, these rate values are clearly lower

than those of GP-net. GP-net exhibits both the highest precision and recall rates, both approaching the maximum value of 1, which further validates the superior performance of GP-net over the atlas-based segmentations.

Variability between patients has been previously reported, for example, in (Duchin et al., 2018; Kim et al., 2019; Lenglet et al., 2012; Patriat et al., 2018). This intrinsic variability between different subjects motivates the need for patient-specific care, and suitably tailored algorithms to address this need. Standard atlases, which are typically defined in a normalized space, have shown great importance in retrospective population studies (Horn, Kühn, et al., 2017; Horn, Neumann, et al., 2017; Horn, Reich, et al., 2017; Kim et al., 2019). However, atlas-based registrations, which are ultimately based on a single depiction of the target structure (even though this depiction can be based on multiple inputs from multiple patients and modalities), cannot fully capture the intra and interpatient variability. Intra and interpatient variability, as well as the need for patient-specific imaging tools is clearly exemplified by panel (d) of Figure 2. It is sufficient to consider the volume estimations for both GPe and GPi, obtained from the manual delineations, to see that profound variability exists between patients. The ratios between the smallest and the largest measured volume estimates correspond to an increase of 200% and 284% for the GPe and GPi, respectively. From a statistical point of view, all volume estimates of the atlas-based approaches have a consistently smaller distribution. Only GP-net exhibits a distribution which is similar to the manually measured volume distribution, for both parts of the GP. This dramatic variability in GPe/GPi volume estimates further motivates the need for a stable and consistent patient-specific, accurate segmentation tool, especially for any invasive procedure such as DBS surgery.

GP-net is trained on healthy subjects as well as PD and ET patients and with a large age distribution. This training process allows the network to encounter and learn from a large variety of different pathologies (Figure 3a,b) and across varying age groups (Figure 3c,d), thus allowing stable and robust segmentation performance. Robust performance across different pathologies is further exemplified by the *p* value matrices below Figure 3a,b. These conclusions are further supported by the average and standard deviation values presented in Table 5. *p* value matrices below Figure 3c,d show no statistical significance between the different age groups (calculated for GP-net). This ability of GP-net to generate accurate segmentations over different clinical pathologies and age groups suggests that GP-net might play an important role in clinical DBS targeting or similar applications for a variety of patient populations.

One of the key advantages of the proposed method is that GP-net relies solely on 7 T T2 scans to perform its inference for visualization of the GPe/GPi. No registrations are involved in its segmentation process and, therefore, GP-net performs segmentation directly in the patient's system coordinates. Atlas-based segmentations on the other hand, are given in a normalized space (i.e., MNI space) and must be registered into the patient's unique space prior to any trajectory planning. Thus, they are greatly affected by the registration process, which can adversely affect the outcomes of DBS surgery. Unfortunately,

registration errors cannot be modeled easily, are typically unpredictable, and often have large variance (Kim et al., 2019). Moreover, the registration process often involves several successive registrations, which tend to accumulate and increase the registration errors with each step. Due to these factors, registrations often must be verified manually, as was done in this study for the atlas-based registrations. This dramatically prolongs the processing time per patient.

Although the atlas-based registrations and final segmentations were performed on the same grid as GP-net, they are visually more pixelated, as no spatial filtering is performed on the acquired T2 volume to infer the segmentations. Contrary to this, GP-net operates directly on the  $0.39 \text{ mm}^3$  isotropically resampled grid of the T2 volume, and by the use of 3D convolution filters it is able to produce smooth segmentations, and higher quality structure detection of both parts of the GP.

Since atlas-based segmentations rely on a single (often average) depiction of the target structure, such a process cannot truly account for different interpatient variabilities in a true patient-specific manner, which is clearly exemplified by the segmentation results depicted in Figure 8. The middle row of Figure 8 shows that GP-net is able to segment the GP, while accounting for the blood vessels, even though GP-net was not trained on such irregular cases. GP-net is able to account for this irregular vessel shape and still be able to produce clear GP segmentations which align with the underlying 7 T scan anatomy, and is relatively close to the manual segmentation. On the other hand, the atlas-based segmentation presented in the lower row of Figure 8 is greatly affected by the registration process of the MNI template (which does not contain such vessels) and thus erroneously segments a large portion of the GP. The large CoM difference for the atlas based segmentation, presented in Table 6, further validates and quantifies this misalignment. This example indicates the true potential of deep-learning based approaches such as GP-net, in being a patient-specific automatic segmentation technique, even for irregular and unique cases as well as nonideal scanning conditions (Figure 9), without any additional training.

As was mentioned before, the internal and external segments of the GP are separated by a thin lamina layer. Often when using MER during DBS surgery, the lamina layer is characterized by the absence of somatodendritic action potentials, which characterize both the GPe and GPi, each with its own characteristic firing pattern (Baron et al., 1996, 2000; Lozano & Hutchinson, 2002). Thus, a clear and reliable visual depiction of the lamina can also be of invaluable importance to DBS surgery. Even when using high contrast 7 T T2 scans, this border is not always clearly presented in the scans. However, the overall GPe and GPi structure can be inferred, as presented in this study. A possible way to achieve lamina depiction is to accurately detect the two GP compartments, and indirectly infer the lamina border from both segmentations. This direction is a matter of future research and extension for GP-net.

GP-net is currently optimized to segment the GP structures from 7 T T2 acquisitions only. Moreover, the use of different MRI hardware, software and acquisition protocols may also affect its overall

performance. One such example is the use of dielectric pads. In our cohort, out of a total of 101 subjects, 72 were scanned with dielectric pads (training set, 42/58 and the test set 30/43 patients were scanned with pads). When comparing the GP-net performance between these two groups (with and without dielectric pads) we found differences for the dice and MSD metrics (dice: GPe  $0.82 \pm 0.03$  vs.  $0.79 \pm 0.07$ , GPi  $0.83 \pm 0.04$  vs.  $0.81 \pm 0.06$  and for MSD: GPe  $0.42 \pm 0.08$  vs.  $0.48 \pm 0.16$  and GPi  $0.36 \pm 0.1$  vs.  $0.43 \pm 0.1$  with and without pads, respectively). While a small difference exists, it's not clear if it's due to the pads signal enhancement or whether it's due to the bias in the training data that had more cases with the pads. Different parameters, such as the acquisition sequence, resolution and field strength may also affect the performance of GP-net, since, at this stage, GP-net was only trained on 7 T T2 data, which was acquired with the protocol described in this manuscript. Thus, GP-net is currently not optimized for inference based on other modalities and/or data acquired with different field strength (for example, see Figure S2). This is in-fact a common limitation to many deep-learning based architectures and achieving increased robustness in the face of changing datasets (e.g., varying signal to noise ratios, resolution, etc.) is a matter of ongoing research in the machine learning community (often denoted as domain shift or domain transfer). Extending this study to operate on standard clinical images (1.5–3 T) is the subject of future study. Additionally, deep-learning frameworks currently lack interpretability. In some cases, the inference might result in sub-optimal performance (e.g., in panel a of Figure 2, two GPe dice scores for GP-net are below 0.6). In these cases, it is not always easy to understand why the network reacted the way it did. Fortunately, as was statistically verified in this study, such occurrences seem to be rare. Exploiting standard CNN visualization tools can potentially educate the user on the internal workings of GP-net.

The clinical potential of GP-net is clearly exemplified in Figure 7, which presents an excellent match between the implanted DBS electrode and the manual versus GP-net's segmentations of both GPe/GPi, for a representative PD patient. GP-net will provide the clinical team the ability to have a much better understanding of the correlation between lead locations and outcomes, and might help reduce the extra operative time invested in planning of the surgical procedure. For DBS surgery, both the center of mass of the DBS target, as well as the accurate identification of its borders are of great importance. This figure, supported by the reported results for CoM distance and MSD, show that an accurate depiction of the DBS target clearly achieves this goal. Accurate identification can also contribute to reliable and accurate DBS lead placement, which has been associated with improved clinical outcomes in leads placed in the STN (Richardson et al., 2009). This often results in a reduced amount of programming time in the clinic necessary for optimizing symptom reduction and an overall improvement in quality of life for the patient (Hell et al., 2019). GP-net is currently being used under a research protocol to assist with DBS surgery preplanning and postoperative lead location assessment based on 7 T T2 scans at the University of Minnesota medical school.

## 5 | CONCLUSIONS AND FUTURE STUDY

In this study, we presented GP-net, a deep-learning based neural network for the segmentation of both the GPe and GPi. GP-net is able to produce accurate and reliable segmentations in a fully automated manner from 7 T T2 MR acquisitions, both for healthy subjects as well as PD and ET patients and with a large age distribution. The network is trained end-to-end on pairs of acquired 7 T T2 scans and corresponding manual delineations of both GPe and GPi. We have shown, both qualitatively and quantitatively, that GP-net outperforms state-of-the-art atlas-based segmentations and produces stable and consistent high quality patient-specific segmentations, while reducing potential biases.

GP-net is tailored for the segmentation of the GP (both internal and external segments). However, it can be extended to segment additional subcortical structures which are of interest for DBS surgery, such as the STN, Red Nucleus, and Substantia Nigra. This extension is currently being investigated in our group. With the ability to reliably segment all DBS-related targets, we further plan to investigate the relationship between patients' properties (such as volume), as ascertained from MRI images, and their clinical characteristics. For example, in (Patriat et al., 2020) it was shown that correlation exists between STN volumes and the Unified Parkinson's Disease Rating Scale (UPDRS) III scores.

With the incorporation of advanced DBS presurgery targeting and postsurgery lead localization tools and software, 7 T MRI based approaches, either for training or for deployment, have great potential in becoming clinical standards, especially now that the 7 T MRI is FDA approved for standard clinical applications. In this scenario, the use of fully automated segmentation software may prove to be very advantageous, leading to an accurate, fast, easy and reliable visualization tool, contributing to an improved surgical procedure and patient experience.

### ACKNOWLEDGMENTS

This study was funded by the following National Institution of Health Grants: R01 NS081118, R01 NS113746, P50 NS098753, P30 NS076408, and P41 EB027061. Additional support by NSF (GS) and Department of Defense (GS) is also acknowledged.

### CONFLICT OF INTERESTS

Tara Palnitkar—consultant for Surgical Information Sciences, Inc. Remi Patriat—consultant for Surgical Information Sciences, Inc. Michael C. Park—Listed faculty for University of Minnesota Educational Partnership with Medtronic, Inc., Minneapolis, MN, Consultant for: Zimmer Biomet, Synerfues, Inc., NeuroOne, Boston Scientific. Grant/Research support from: Medtronic, Inc., Boston Scientific, Abbott. Jerrold Vitek—Consultant for: Medtronic, Inc., Boston Scientific, Abbott, Surgical Information Sciences, Inc. Guillermo Sapiro—consultant and a shareholder for Surgical Information Sciences, Inc. Consultant for Apple and Volvo. Noam Harel—consultant and a shareholder for Surgical Information Sciences, Inc.

### AUTHOR CONTRIBUTIONS

Oren Solomon designed and wrote the algorithm, performed the analysis and wrote the manuscript. Tara Palnitkar, Remi Patriat, and Henry

Braun acquired the 7 T MRI data and performed the manual segmentations and participated in the writing of the manuscript. Joshua Aman, Michael C. Park, and Jerrold Vitek provided clinical insight, performed the manual segmentations as well as participated in the writing of the manuscript. Guillermo Sapiro participated in designing the algorithm and writing of the manuscript. Noam Harel participated in designing the algorithm, performing the analysis, acquiring the 7 T MRI data, obtaining the manual segmentations and writing the manuscript.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### ORCID

Oren Solomon  <https://orcid.org/0000-0003-0240-0852>

Noam Harel  <https://orcid.org/0000-0002-6034-3801>

### ENDNOTES

<sup>i</sup> Intraoperative microelectrode recording (MER) could be considered as well. MER for the GP is significantly less understood than for the STN for example, is time consuming, and requires expertise.

<sup>ii</sup> Code is available at: <https://github.com/KrakenLeaf/Attention-Gated-Networks>

### REFERENCES

- Abosch, A., Yacoub, E., Ugurbil, K., & Harel, N. (2010). An assessment of current brain targets for deep brain stimulation surgery with susceptibility-weighted imaging at 7 Tesla. *Neurosurgery*, 67(6), 1745–1756. <https://doi.org/10.1227/NEU.0b013e3181f74105>
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3), 2033–2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations*.
- Balakrishnan, G., Zhao, A., Sabuncu, M. R., Dalca, A. V., & Guttag, J. (2018). An unsupervised learning model for deformable medical image registration. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 9252–9260. <https://doi.org/10.1109/CVPR.2018.00964>
- Baron, M. S., Vitek, J. L., Bakay, R. A. E., Green, J., Kaneoke, Y., Hashimoto, T., ... DeLong, M. R. (1996). Treatment of advanced Parkinson's disease by posterior GPI pallidotomy: 1-year results of a pilot study. *Annals of Neurology*, 40(3), 355–366. <https://doi.org/10.1002/ana.410400305>
- Baron, M. S., Vitek, J. L., Bakay, R. A. E., Green, J., McDonald, W. M., Cole, S. A., & DeLong, M. R. (2000). Treatment of advanced parkinson's disease by unilateral posterior GPi pallidotomy: 4-year results of a pilot study. *Movement Disorders*, 15(2), 230–237. [https://doi.org/10.1002/1531-8257\(200003\)15:2<230::AID-MDS1005>3.0.CO;2-U](https://doi.org/10.1002/1531-8257(200003)15:2<230::AID-MDS1005>3.0.CO;2-U)
- Bazin, P. L., Alkemade, A., Mulder, M. J., Henry, A. G., & Forstmann, B. U. (2020). Multi-contrast anatomical subcortical structures parcellation. *eLife*, 9, 1–23. <https://doi.org/10.7554/ELIFE.59430>
- Benabid, A. L., Pollak, P., Louveau, A., Henry, S., & de Rougemont, J. (1987). Combined (thalamotomy and stimulation) stereotactic surgery



- of the VIM thalamic nucleus for bilateral Parkinson disease. *Stereotactic and Functional Neurosurgery*, 50(1–6), 344–346. <https://doi.org/10.1159/000100803>
- Chakravarty, M. M., Bertrand, G., Hodge, C. P., Sadikot, A. F., & Collins, D. L. (2006). The creation of a brain atlas for image guided neurosurgery using serial histological data. *NeuroImage*, 30(2), 359–376. <https://doi.org/10.1016/j.neuroimage.2005.09.041>
- Dadar, M., Fonov, V. S., & Collins, D. L. (2018). A comparison of publicly available linear MRI stereotaxic registration techniques. *NeuroImage*, 174(March), 191–200. <https://doi.org/10.1016/j.neuroimage.2018.03.025>
- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable convolutional networks. Proceedings of the IEEE international conference on computer vision, 2017–October, 764–773. <https://doi.org/10.1109/ICCV.2017.89>
- Dalca, A. V., Balakrishnan, G., Guttag, J., & Sabuncu, M. R. (2018). Unsupervised learning for fast probabilistic diffeomorphic registration. International conference on medical image computing and computer-assisted intervention, 11070, 729–738. [https://doi.org/10.1007/978-3-030-00928-1\\_82](https://doi.org/10.1007/978-3-030-00928-1_82)
- Deuschl, G., Schade-Brittinger, C., Krack, P., Volkmann, J., Schäfer, H., Bötzel, K., ... Voges, J. (2006). A randomized trial of deep-brain stimulation for Parkinson's disease. *New England Journal of Medicine*, 355(9), 896–908. <https://doi.org/10.1056/NEJMoa060281>
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297–302. <https://doi.org/10.2307/1932409>
- Duchin, Y., Shamir, R. R., Patriat, R., Kim, J., Vitek, J. L., Sapiro, G., & Harel, N. (2018). Patient-specific anatomical model for deep brain stimulation based on 7 Tesla MRI. *PLoS One*, 13(8), 1–23. <https://doi.org/10.1371/journal.pone.0201469>
- Ebani, E. J., Kaplitt, M. G., Wang, Y., Nguyen, T. D., Askin, G., & Chazen, J. L. (2020). Improved targeting of the globus pallidus interna using quantitative susceptibility mapping prior to MR-guided focused ultrasound ablation in Parkinson's disease. *Clinical Imaging*, 68, 94–98. <https://doi.org/10.1016/j.clinimag.2020.06.017>
- Ellis, T. M., Foote, K. D., Fernandez, H. H., Sudhyadhom, A., Rodriguez, R. L., Zeilman, P., ... Okun, M. S. (2008). Reoperation for suboptimal outcomes after deep brain stimulation surgery. *Neurosurgery*, 63(4), 754–760. <https://doi.org/10.1227/01.NEU.0000325492.58799.35>
- Ewert, S., Plettig, P., Li, N., Chakravarty, M. M., Collins, D. L., Herrington, T. M., ... Horn, A. (2018). Toward defining deep brain stimulation targets in MNI space: A subcortical atlas based on multimodal MRI, histology and structural connectivity. *NeuroImage*, 170, 271–282. <https://doi.org/10.1016/j.neuroimage.2017.05.015>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. <http://www.deeplearningbook.org>
- Hariz, G.-M., Blomstedt, P., & Koskinen, L.-O. D. (2008). Long-term effect of deep brain stimulation for essential tremor on activities of daily living and health-related quality of life. *Acta Neurologica Scandinavica*, 118(6), 387–394. <https://doi.org/10.1111/j.1600-0404.2008.01065.x>
- He, X., Chaitanya, G., Asma, B., Caciagli, L., Bassett, D. S., Tracy, J. I., & Sperling, M. R. (2020). Disrupted basal ganglia-thalamocortical loops in focal to bilateral tonic-clonic seizures. *Brain*, 143(1), 175–190. <https://doi.org/10.1093/brain/awz361>
- Hell, F., Palleis, C., Mehrkens, J. H., Koeglsperger, T., & Bötzel, K. (2019). Deep brain stimulation programming 2.0: Future perspectives for target identification and adaptive closed loop stimulation. *Frontiers in Neurology*, 10(APR), 314. <https://doi.org/10.3389/fneur.2019.00314>
- Horn, A., & Kühn, A. A. (2015). Lead-DBS: A toolbox for deep brain stimulation electrode localizations and visualizations. *NeuroImage*, 107, 127–135. <https://doi.org/10.1016/j.neuroimage.2014.12.002>
- Horn, A., Kühn, A. A., Merkl, A., Shih, L., Alterman, R., & Fox, M. (2017). Probabilistic conversion of neurosurgical DBS electrode coordinates into MNI space. *NeuroImage*, 150, 395–404. <https://doi.org/10.1016/j.neuroimage.2017.02.004>
- Horn, A., Li, N., Dembek, T. A., Kappel, A., Boulay, C., Ewert, S., ... Kühn, A. A. (2019). Lead-DBS v2: Towards a comprehensive pipeline for deep brain stimulation imaging. *NeuroImage*, 184(July 2018), 293–316. <https://doi.org/10.1016/j.neuroimage.2018.08.068>
- Horn, A., Neumann, W.-J., Degen, K., Schneider, G.-H., & Kühn, A. A. (2017). Toward an electrophysiological “sweet spot” for deep brain stimulation in the subthalamic nucleus. *Human Brain Mapping*, 38(7), 3377–3390. <https://doi.org/10.1002/hbm.23594>
- Horn, A., Reich, M., Vorwerk, J., Li, N., Wenzel, G., Fang, Q., ... Fox, M. D. (2017). Connectivity predicts deep brain stimulation outcome in Parkinson disease. *Annals of Neurology*, 82(1), 67–78. <https://doi.org/10.1002/ana.24974>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Karimi, D., & Salcudean, S. E. (2020). Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39(2), 499–513. <https://doi.org/10.1109/TMI.2019.2930068>
- Keuken, M. C., Bazin, P. L., Crown, L., Hootsmans, J., Laufer, A., Müller-Axt, C., ... Forstmann, B. U. (2014). Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. *NeuroImage*, 94, 40–46. <https://doi.org/10.1016/j.neuroimage.2014.03.032>
- Kim, J., Duchin, Y., Shamir, R. R., Patriat, R., Vitek, J., Harel, N., & Sapiro, G. (2019). Automatic localization of the subthalamic nucleus on patient-specific clinical MRI by incorporating 7 T MRI and machine learning: Application in deep brain stimulation. *Human Brain Mapping*, 40(2), 679–698. <https://doi.org/10.1002/hbm.24404>
- Kim, J., Patriat, R., Kaplan, J., Solomon, O., & Harel, N. (2020). Deep cerebellar nuclei segmentation via semi-supervised deep context-aware learning from 7 T diffusion MRI. *IEEE Access*, 8, 101550–101568. <https://doi.org/10.1109/ACCESS.2020.2998537>
- Kringelbach, M. L., Jenkinson, N., Owen, S. L. F., & Aziz, T. Z. (2007). Translational principles of deep brain stimulation. *Nature Reviews Neuroscience*, 8(8), 623–635. <https://doi.org/10.1038/nrn2196>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. <http://code.google.com/p/cuda-convnet/>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lenglet, C., Abosch, A., Yacoub, E., de Martino, F., Sapiro, G., & Harel, N. (2012). Comprehensive in vivo mapping of the human basal ganglia and thalamic connectome in individuals using 7 T MRI. *PLoS One*, 7(1), e29153. <https://doi.org/10.1371/journal.pone.0029153>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431–3440.
- Lozano, A. M., & Hutchinson, W. D. (2002). Microelectrode recordings in the pallidum. *Movement Disorders*, 17, 150–154. <https://doi.org/10.1002/mds.10157>
- Manjón, J. V., & Coupé, P. (2016). Volbrain: An online MRI brain volumetry system. *Frontiers in Neuroinformatics*, 10, 1–14. <https://doi.org/10.3389/fninf.2016.00030>
- Marks, W. A., Honeycutt, J., Acosta, F., & Reed, M. (2009). Deep brain stimulation for pediatric movement disorders. *Seminars in Pediatric Neurology*, 16(2), 90–98. <https://doi.org/10.1016/j.spn.2009.04.001>
- Miller, T. R., Guo, S., Melhem, E. R., Eisenberg, H. M., Zhuo, J., Kelm, N., ... Gandhi, D. (2020). Predicting final lesion characteristics during MR-guided focused ultrasound pallidotomy for treatment of Parkinson's disease. *Journal of Neurosurgery*, 1, 1–8. <https://doi.org/10.3171/2020.2.jns192590>

- Mueller, J., Skogseid, I. M., Benecke, R., Kupsch, A., Trottenberg, T., Poewe, W., ... Volkmann, J. (2008). Pallidal deep brain stimulation improves quality of life in segmental and generalized dystonia: Results from a prospective, randomized sham-controlled trial. *Movement Disorders*, 23(1), 131–134. <https://doi.org/10.1002/mds.21783>
- Munro-Davies, L. E., Gregory, R., Squires, W., Radatz, M., Silburn, P., Scott, R., ... Stein, J. F. (1999). Lateral pallidotomy exacerbates akinesia in the parkinsonian patient. *Journal of Clinical Neuroscience*, 6(6), 474–476. [https://doi.org/10.1016/S0967-5868\(99\)90003-6](https://doi.org/10.1016/S0967-5868(99)90003-6)
- Obeso, J. A., Olanow, C. W., Rodriguez-Oroz, M. C., Krack, P., Kumar, R., & Lang, A. E. (2001). Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. *New England Journal of Medicine*, 345(13), 956–963. <https://doi.org/10.1056/NEJMoa000827>
- Paek, S. H., Yun, J. Y., Song, S. W., Kim, I. K., Hwang, J. H., Kim, J. W., ... Jeon, B. S. (2013). The clinical impact of precise electrode positioning in STN DBS on three-year outcomes. *Journal of the Neurological Sciences*, 327(1–2), 25–31. <https://doi.org/10.1016/j.jns.2013.01.037>
- Patel, D. M., Walker, H. C., Brooks, R., Omar, N., Ditty, B., & Guthrie, B. L. (2015). Adverse events associated with deep brain stimulation for movement disorders. *Operative Neurosurgery*, 11(1), 190–199. <https://doi.org/10.1227/NEU.0000000000000659>
- Patriat, R., Cooper, S. E., Duchin, Y., Niederer, J., Lenglet, C., Aman, J., ... Harel, N. (2018). Individualized tractography-based parcellation of the globus pallidus pars interna using 7 T MRI in movement disorder patients prior to DBS surgery. *NeuroImage*, 178, 198–209. <https://doi.org/10.1016/j.neuroimage.2018.05.048>
- Patriat, R., Niederer, J., Kaplan, J., Amundsen Huffmaster, S., Petrucci, M., Eberly, L., ... MacKinnon, C. (2020). Morphological changes in the subthalamic nucleus of people with mild-to-moderate Parkinson's disease: A 7 T MRI study. *Scientific Reports*, 10(1), 8785. <https://doi.org/10.1038/s41598-020-65752-0>
- Pauli, W. M., Nili, A. N., & Michael Tyszka, J. (2018). Data descriptor: A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei. *Scientific Data*, 5, 1–13. <https://doi.org/10.1038/sdata.2018.63>
- Pominova, M., Kondrateva, E., Sharaev, M., Bernstein, A., Pavlov, S., & Burnaev, E. (2019). 3D deformable convolutions for MRI classification. Proceedings: 18th IEEE international conference on machine learning and applications, ICMLA 2019, 1710–1716. <https://doi.org/10.1109/ICMLA.2019.00278>
- Richardson, R. M., Ostrem, J. L., & Starr, P. A. (2009). Surgical repositioning of misplaced subthalamic electrodes in Parkinson's disease: Location of effective and ineffective leads. *Stereotactic and Functional Neurosurgery*, 87(5), 297–303. <https://doi.org/10.1159/000230692>
- Rockafellar, R. T., & Wets, R. J.-B. (2009). *Variational analysis*. Berlin/Heidelberg, Germany: Springer science & business media.
- Rolston, J. D., Englot, D. J., Starr, P. A., & Larson, P. S. (2016). An unexpectedly high rate of revisions and removals in deep brain stimulation surgery: Analysis of multiple databases. *Parkinsonism and Related Disorders*, 33, 72–77. <https://doi.org/10.1016/j.parkreldis.2016.09.014>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention, 9351, 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Rosenfeld, A., & Pfaltz, J. L. (1966). Sequential operations in digital picture processing. *Journal of the ACM (JACM)*, 13(4), 471–494. <https://doi.org/10.1145/321356.321357>
- Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10541 LNCS, 379–387. [https://doi.org/10.1007/978-3-319-67389-9\\_44](https://doi.org/10.1007/978-3-319-67389-9_44)
- Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., & Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53, 197–207. <https://doi.org/10.1016/j.media.2019.01.012>
- Teeuwisse, W. M., Brink, W. M., & Webb, A. G. (2012). Quantitative assessment of the effects of high-permittivity pads in 7 Tesla MRI of the brain. *Magnetic Resonance in Medicine*, 67(5), 1285–1293. <https://doi.org/10.1002/mrm.23108>
- Visser, E., Keuken, M. C., Douaud, G., Gaura, V., Bachoud-Levi, A. C., Remy, P., ... Jenkinson, M. (2016). Automatic segmentation of the striatum and globus pallidus using MIST: Multimodal image segmentation tool. *NeuroImage*, 125, 479–497. <https://doi.org/10.1016/j.neuroimage.2015.10.013>
- Vitek, J. L., Bakay, R. A. E., Freeman, A., Evatt, M., Green, J., McDonald, W., ... DeLong, M. R. (2003). Randomized trial of pallidotomy versus medical therapy for Parkinson's disease. *Annals of Neurology*, 53(5), 558–569. <https://doi.org/10.1002/ana.10517>
- Vitek, J. L., Jain, R., Chen, L., Tröster, A. I., Schrock, L. E., House, P. A., ... Starr, P. A. (2020). Subthalamic nucleus deep brain stimulation with a multiple independent constant current-controlled device in Parkinson's disease (INTREPID): A multicentre, double-blind, randomised, sham-controlled study. *The Lancet Neurology*, 19(6), 491–501. [https://doi.org/10.1016/S1474-4422\(20\)30108-3](https://doi.org/10.1016/S1474-4422(20)30108-3)
- Volkmann, J., Wolters, A., Kupsch, A., Müller, J., Kühn, A. A., Schneider, G. H., ... Benecke, R. (2012). Pallidal deep brain stimulation in patients with primary generalised or segmental dystonia: 5-year follow-up of a randomised trial. *The Lancet Neurology*, 11(12), 1029–1038. [https://doi.org/10.1016/S1474-4422\(12\)70257-0](https://doi.org/10.1016/S1474-4422(12)70257-0)
- Wang, W., & Shen, J. (2018). Deep visual attention prediction. *IEEE Transactions on Image Processing*, 27(5), 2368–2378. <https://doi.org/10.1109/TIP.2017.2787612>
- Welter, M. L., Schüpbach, M., Czernecki, V., Karachi, C., Fernandez-Vidal, S., Golmard, J. L., ... Agid, Y. (2014). Optimal target localization for subthalamic stimulation in patients with Parkinson disease. *Neurology*, 82(15), 1352–1361. <https://doi.org/10.1212/WNL.0000000000000315>
- Wichmann, T., Bergman, H., & DeLong, M. R. (1994). The primate subthalamic nucleus. III. Changes in motor behavior and neuronal activity in the internal pallidum induced by subthalamic inactivation in the MPTP model of parkinsonism. *Journal of Neurophysiology*, 72(2), 521–530. <https://doi.org/10.1152/jn.1994.72.2.521>
- Xiao, Y., Fonov, V., Chakravarty, M. M., Beriault, S., al Subaie, F., Sadikot, A., ... Collins, D. L. (2017). A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort. *Data in Brief*, 12, 370–379. <https://doi.org/10.1016/j.dib.2017.04.013>
- Zaaroor, M., Sinai, A., Goldsher, D., Eran, A., Nassar, M., & Schlesinger, I. (2018). Magnetic resonance-guided focused ultrasound thalamotomy for tremor: A report of 30 Parkinson's disease and essential tremor cases. *Journal of Neurosurgery*, 128(January), 202–210. <https://doi.org/10.3171/2016.10.JNS16758>
- Zhong, Z., Jin, L., & Xie, Z. (2015). High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. Proceedings of the international conference on document analysis and recognition, ICDAR, 2015–November, 846–850. <https://doi.org/10.1109/ICDAR.2015.7333881>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Solomon O, Palnitkar T, Patriat R, et al. Deep-learning based fully automatic segmentation of the globus pallidus interna and externa using ultra-high 7 Tesla MRI. *Hum Brain Mapp*. 2021;42:2862–2879. <https://doi.org/10.1002/hbm.25409>