



Projected pooling loss for red nucleus segmentation with soft topology constraints

Guanghui Fu, Rosana El Jurdi, Lydia Chougar, Didier Dormont, Romain Valabregue, Stéphane Lehéricy, Olivier Colliot

► To cite this version:

Guanghui Fu, Rosana El Jurdi, Lydia Chougar, Didier Dormont, Romain Valabregue, et al.. Projected pooling loss for red nucleus segmentation with soft topology constraints. *Journal of Medical Imaging*, 2024, 11 (04), 10.1117/1.JMI.11.4.044002 . hal-04643094

HAL Id: hal-04643094

<https://hal.science/hal-04643094>

Submitted on 10 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Projected Pooling Loss for Red Nucleus Segmentation with Soft Topology Constraints

Guanghui Fu^a, Rosana El Jurdì^a, Lydia Chougar^{a,b,c,d}, Didier Dormont^{a,c}, Romain Valabregue^{b,e}, Stéphane Lehéricy^{b,c,e}, Olivier Colliot^{*a}, the ICEBERG Study Group^b

^aSorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

^bICM, Centre de NeuroImagerie de Recherche-CENIR, Paris, France.

^cAP-HP, Hôpital de la Pitié Salpêtrière, DMU DIAMENT, Dep. of Neuroradiology, F-75013, Paris, France

^dThe Neuro (Montreal Neurological Institute-MNI), McGill University, Montreal H3A 2B4, Canada

^eSorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

Abstract. **Purpose.** Deep learning is the standard for medical image segmentation. However, it may encounter difficulties when the training set is small. Also, it may generate anatomically aberrant segmentations. Anatomical knowledge can be potentially useful as a constraint in deep learning segmentation methods. In this paper, we propose a novel loss function based on projected pooling to introduce soft topological constraints. Our main application is the segmentation of the red nucleus from quantitative susceptibility mapping (QSM) which is of interest in parkinsonian syndromes.

Approach. This new loss function introduces soft constraints on the topology by magnifying small parts of the structure to segment to avoid that they are discarded in the segmentation process. To that purpose, we use projection of the structure onto the three planes and then use a series of MaxPooling operations with increasing kernel sizes. These operations are performed both for the ground-truth and the prediction and the difference is computed to obtain the loss function. As a result, it can reduce topological errors as well as defects in the structure boundary. The approach is easy to implement and computationally efficient.

Results. When applied to the segmentation of the red nucleus from QSM data, the approach led to a very high accuracy (Dice 89.9%) and no topological errors. Moreover, the proposed loss function improved the Dice accuracy over the baseline when the training set was small. We also studied three tasks from the medical segmentation decathlon challenge (MSD) (heart, spleen, and hippocampus). For the MSD tasks, the Dice accuracies were similar for both approaches but the topological errors were reduced.

Conclusions. We proposed an effective method to automatically segment the red nucleus which is based on a new loss for introducing topology constraints in deep learning segmentation. The code is publicly available at <https://github.com/GuanghuiFU/TopologyLoss/>.

Keywords: Segmentation, Loss function, Anatomical priors, Topology, Connected components, Deep Learning.

*Corresponding author: O. Colliot (olivier.colliot@cnrs.fr)

1 Introduction

Deep learning is the main approach for medical image segmentation.¹ However, deep learning models often need large labeled datasets for training. In the medical image segmentation, obtaining labeled data is an expensive and tedious process since it requires voxel-wise annotations.² Furthermore, anatomically aberrant segmentations may still be generated even with sufficient training data.^{3,4} For medical segmentation, plenty of anatomical information is available. Medical experts extensively use their anatomical knowledge to perform manual segmentations. This is particularly useful when the target boundary is not clearly visible in the image. Prior knowledge may include information about the shape, size, location, texture or topology of the target. In particular, topology is an example of important prior anatomical information as many anatomical structures have

fixed topological characteristics.⁵ Thus, incorporating prior knowledge into deep learning models has the potential to improve their performance. In particular, this may avoid anatomically aberrant segmentations, which are highly problematic since they reduce the confidence of the users even when the voxel-wise accuracy is high.

In deep learning, a natural way to use prior knowledge is to propose new loss functions.³ Various types of prior knowledge have been introduced as novel loss functions including bounding boxes,⁶ area and size,⁷ perimeter,⁸ topology^{9–11} or skeleton of the target structure.¹² Regarding topology, existing approaches rely on persistent homology (PH),¹³ to construct loss functions.^{9–11} This is a principled way to constrain the topology. However, calculating PH of the segmentation prediction on each epoch is particularly time-consuming, especially when dealing with 3D data. There is thus a need for computationally efficient loss functions.

Our target application is the segmentation of the red nucleus of the brain from quantitative susceptibility mapping (QSM) data. The red nucleus is a structure of the midbrain which plays an important role in parkinsonian syndromes.¹⁴ The red nucleus is rich in iron. QSM is therefore well adapted to study it since it provides an indirect measure of iron content from magnetic resonance imaging (MRI) acquisitions. Thus, automatic segmentation of the red nucleus from QSM can be useful for studying and assisting the diagnosis parkinsonian syndromes. Several approaches have been proposed for automatic segmentation of the red nucleus from QSM data.^{15–20} However, to the best of our knowledge, none of them has assessed the topological correctness of the results. Moreover, these papers only includes healthy participants and were thus not evaluated in patients with parkinsonism.

In this paper, we propose a novel approach to introduce soft topological constraints for segmentation of the red nucleus from QSM data. To that purpose, we use projection and pooling operations to obtain multiscale representations of both the ground-truth and the predicted segmentations. Specifically, we first perform maximum projections onto the three planes (axial, coronal and sagittal) which concentrate together the errors scattered in different slices. We then use Max-Pooling with different kernel sizes to construct a series of simplified representations of the ground-truth and predicted segmentations. This allows magnifying small errors, such as in particular small erroneous connected components. We first experimented on controlled synthetic data to validate the generality of our method, and then focused on our target application which is the segmentation of the red nucleus from QSM. We finally also performed an evaluation on three tasks from the medical segmentation decathlon (MSD).²¹

A preliminary version of this work was published as a conference paper.²² The present article extends the previous work with: i) a more detailed overview of related works; ii) a more extensive motivation and methodological description of the approach; iii) a heuristic to choose the kernel sizes; iv) experiments with varying the size of the training set; v) comparison of our method with a connected component preserving post-processing operation. The rest of this paper is organized as follows. In Section 2, we describe related work. Section 3 describes the proposed method and its implementation. Section 4 presents the experiments and results. The paper ends with a discussion (Section 5).

2 Related work

The standard approach for deep learning-based segmentation is to supervise at the voxel level, using a voxel-wise loss (typically cross-entropy²³ and Dice loss²⁴). However, this may gener-

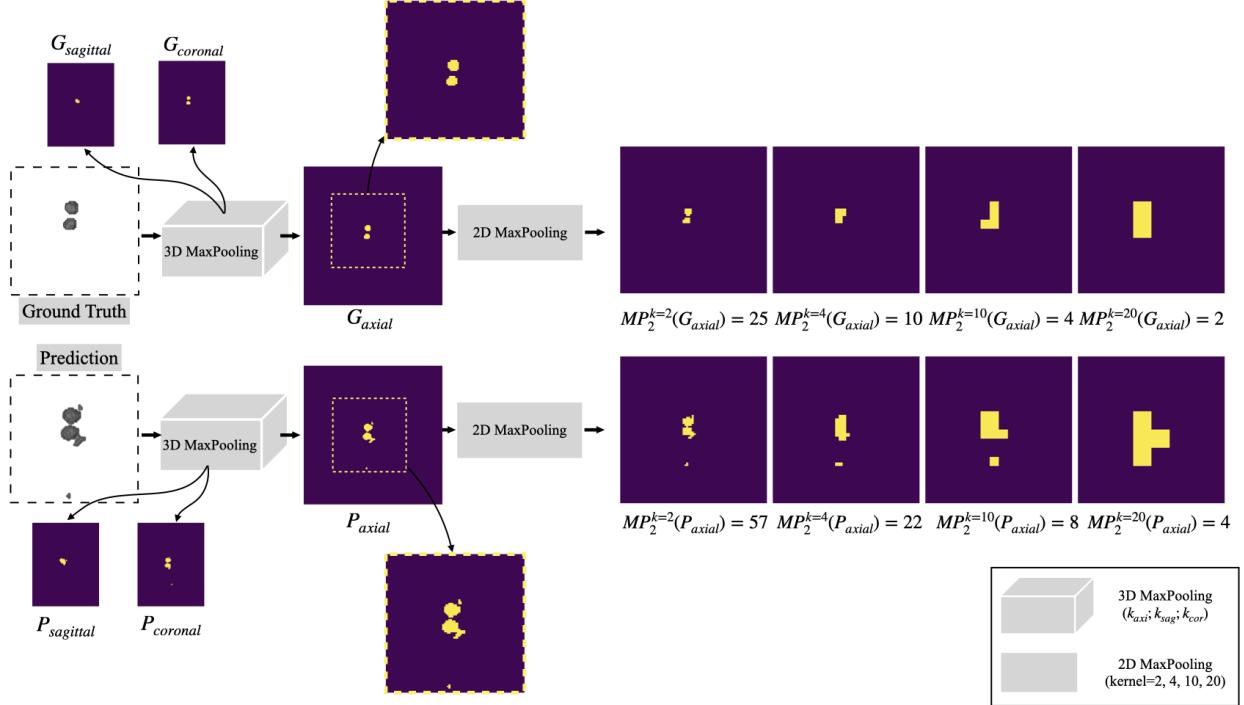


Fig 1 This diagram describes the construction of the topological loss function. For both the ground-truth segmentation and the prediction, we build a multiscale characterization using two steps. We first perform 3D MaxPooling to project the object onto the axial, sagittal and coronal planes. We then use 2D MaxPooling with different kernel size to build a multiscale representation (due to space limitation, we only show the 2D MaxPooling steps for the axial projection, but the process is the same for the sagittal and coronal projections). Counting the pixels in the output of each 2D MaxPooling provides a series of numbers (one for each scale) which constitute the characterization of the structure to segment. The topological loss function is then the average across scales of the differences between the numbers obtained for the ground-truth segmentation and that obtained for the prediction.

ate anatomically aberrant segmentations. To overcome such problems, one can introduce prior anatomical knowledge in the segmentation procedure. In particular, prior knowledge can be introduced through new loss functions. Note that the use of prior knowledge for segmentation is a classical topic in medical image computing that goes way beyond the design of alternative losses²⁵ or even deep neural networks.^{26–29} In the remainder of this section, we provide a non-exhaustive overview of existing loss functions for introducing prior knowledge. We first briefly deal with shape, size and texture (Section 2.1) and then focus specifically topological constraints (Section 2.2). More extensive surveys can be found in.^{3,5,30}

2.1 Loss functions for integrating prior knowledge

Prior knowledge includes various types of information including shape, size, location, texture or topology.

Huang et al.³¹ proposed a loss function based on a probabilistic atlas which includes prior information on location and shape of the organs. Kervadec et al.⁶ proposed a weakly supervised learning segmentation approach that introduces several global constraints derived from box annotations. They assume that there must be a vertical and horizontal line across the foreground region

in the bounding box. Chen et al.⁷ proposed a loss function which can incorporate area and size information. Such a loss function may be beneficial when the dimension of the target object is stable. However, it may not be adapted to targets with a very variable size such as brain tumors or anatomical structures which are affected by a pathology (e.g. atrophy due to a degenerative process). Shit et al.¹² proposed a loss function named cIDice which adds supervision from the skeleton of the target structure. It is beneficial for accurate segmentation of tubular structures such as blood vessels. El Jundi et al.⁸ proposed a perimeter-based medical image segmentation loss function. This soft optimization of contour boundaries allows the network to take into consideration border irregularities. However, it is not yet adapted to the segmentation of multiple connected components.

2.2 Topology-based loss functions and models

Several topology-based loss functions are based on persistent homology (PH). PH is an algebraic method for characterizing the topology of shapes and functions.¹³ The basic idea of PH is to represent the data with a family of complexes, and encode the change of the topological features (such as the number of connected components, holes, voids) across different scales, with a barcode or persistence diagram.³² More details on PH can be found in Edelsbrunner et al.^{13,33} and Pun et al.³⁴

Clough et al.¹⁰ proposed a topological loss function based on PH. They treat 2D images or 3D volumes as cubical complexes, and use the topology layer⁹ to extract the birth and death of topological features of different dimensions from the predictions for supervision. Although the approach can provide substantial gains in segmentation accuracy, its computational cost may become prohibitive, at least in 3D.¹⁰ Hu et al.³⁵ proposed a topology-preserving loss function based on PH. This method adjusts the prediction by PH-based thresholding. However, it does not work if the model fails to detect the target region because the adjustment of the threshold does not compensate for the missing region. In other words, this method can only remove erroneous connected components and cannot cope with the missing ones. Byrne et al.³⁶ designed a new loss function for multi-class segmentation based on PH, extending from binary setting to consider a richer set of topological priors, including hierarchical class containment and adjacency. It is an extension of the work of Clough et al.¹⁰ to the multi-class setting. Nevertheless, its computational cost is also very high. Another method based on PH is that of Santhirasekaram et al.³⁷ who proposed a method called Topology Preserving Compositionality. The method constrains the model's latent space to a dictionary of components, ensuring topological accuracy through PH. All the aforementioned works are based on PH.

However, the involvement of PH makes them computationally expensive, in particular in 3D. Moreover, PH is applicable when the topology of the target object is precisely defined. For instance, one needs to know the exact number of connected components and this needs to be reliable across subjects. It may not be adapted for objects with imperfections such as random cuts that can be for example found in blood vessel ensembles. For such cases, a soft topological constraint may be beneficial.

Mosinska et al.³⁸ proposed a new loss function to control the higher-order topological features of linear structures. They use pretrained VGG19 to extract feature maps at several layers, and minimise the difference between the ground-truth and the prediction. But the so-called high-level features are not clearly defined, and the backbone network pre-trained from ImageNet³⁹ may

not generalize well to medical images. BennTaieb et al.⁴⁰ introduced a loss function which can represent topological (containment and exclusion) and geometric priors (boundary smoothness). However, it is not clear whether these constraints can be encoded in a binary segmentation task or whether it can be applied to 3D data. He et al.⁴¹ proposed a cascaded fully convolutional network (FCN) framework to segment eight retina layers while preserving the topological relationships between the layers. However, the idea of considering layer nesting as constraints may not generalize to other tasks. Hu et al.⁴² proposed a network which based on discrete Morse theory (DMT)⁴³ for topological accuracy by identify global structures, including 1D skeletons and 2D patches. However, they only dealt with 2D images.

2.3 Positioning of our contribution

Our aim was to propose a simple and computationally efficient way to introduce constraints in deep learning based segmentation. It aims at being applicable with any deep learning architecture. The constraints are soft in the sense that we don't ensure but only favor topological correctness. Moreover, as we will see, the proposed approach not only constrains the topology but also the boundary of the regions. Compared to previous works based on PH, the main limitation is that we do not guarantee an exact topology: the number of connected components are only approximately constrained. Furthermore, we do not deal with other topological features such as voids or holes. However, it has the advantage of being simple to implement and computationally efficient. Moreover, it does not require to have a fixed and predefined topology.

3 Proposed method

The idea underlying our loss function is to characterize the structure to segment at different scales. We compute a series of numbers (one for each scale) that is the characterization of the segmentation. Characterizations are computed for both the ground-truth and the predicted segmentation. The characterizations of the ground-truth and the prediction are compared to produce the loss function. The characterization is computed using two steps. We first perform a maximum projection of the 3D segmentation into 2D space according to three views (axial, sagittal and coronal). We then characterize the result using 2D pooling layers with different kernel sizes (corresponding to different scales). This produces the characterization which contains a number for each scale. These two steps are summarized on Figure 1. The topological loss is then computed as the average across scales of the differences between the ground-truth and the prediction.

3.1 Projection by 3D-MaxPooling

The first step is to create 2D representations of the 3D structure to segment. Specifically, we perform a maximum projection onto three planes (axial, coronal, and sagittal) by 3D-MaxPooling. The projections of the prediction onto the axial, sagittal and coronal planes are denoted respectively as P_{axial} , P_{sagittal} and P_{coronal} and computed as follows:

$$\begin{aligned} P_{\text{axial}} &= MP_3^{k_{\text{axi}}}(V_{w,h,s}) \\ P_{\text{sagittal}} &= MP_3^{k_{\text{sag}}}(V_{w,h,s}) \\ P_{\text{coronal}} &= MP_3^{k_{\text{cor}}}(V_{w,h,s}) \end{aligned} \tag{1}$$

where $V_{w,h,s}$ denotes the 3D volume, w , h and s are the width, height and number of slices of the volume respectively, MP_3^k denotes the 3D-MaxPooling with kernel k and $k_{axi} = (1, 1, s)$, $k_{sag} = (w, 1, 1)$ and $k_{cor} = (1, h, 1)$ are the kernels for projecting onto the axial, sagittal and coronal planes respectively. Note that P_{axial} is of dimension (w, h) , $P_{sagittal}$ is of dimension (h, s) and $P_{coronal}$ is of dimension (w, s) . The projections of the ground-truth are denoted respectively as G_{axial} , $G_{sagittal}$ and $G_{coronal}$ and are computed in the same way as for the prediction.

The above procedure concentrates the overall object along each axis. The idea is to be able to magnify the effect of small errors on the overall loss. Indeed, a small segmentation error (such as for example a small extra connected component) has little effect on the overall loss. On the other hand, if we project it to 2D, its impact will be magnified. For example, the proportion of a voxel mispredicted in the volume $V_{w,h,s}$ is $\frac{1}{w \times h \times s}$, while if we project it onto the axial plane, the error proportion is $\frac{1}{w \times h}$. Thus, projecting the prediction onto axial planes augments the weight of this error by s times.

Note that using these projections does not give a theoretical guarantee to achieve a correct topology. For instance, a hole within an object would not be detected. Another example is that one erroneous connected component can in principle be hidden by true connected components along each of the projection axes.

3.2 Topology characterization using 2D-MaxPooling across multiple scales

We characterize the topology of the projected view by using a series of 2D-MaxPooling with different kernel sizes (and different strides as well, the strides value being equal to the kernel size value).

By doing so, we increasingly magnify the importance of small areas. In the ground-truth, these can be genuine but tiny anatomical details. In the prediction they can be small errors.

We perform the series of 2D-MaxPooling for the projections onto axial, sagittal and coronal plane and for both the ground-truth (G_{axial} , $G_{sagittal}$, $G_{coronal}$) and the prediction (P_{axial} , $P_{sagittal}$, $P_{coronal}$), as follows:

$$\begin{aligned} P_{topo}^k &= |MP_2^k(P_{axial})| + |MP_2^k(P_{sagittal})| + |MP_2^k(P_{coronal})| \\ G_{topo}^k &= |MP_2^k(G_{axial})| + |MP_2^k(G_{sagittal})| + |MP_2^k(G_{coronal})| \end{aligned} \quad (2)$$

where MP_2^k is the 2D-MaxPooling operation with kernel size k (and stride k) and $|I|$ denotes the number of non-zero pixels in image I .

We then compute the absolute difference between P_{topo}^k and G_{topo}^k for each kernel size. The topology loss L_{topo} is obtained by averaging over all kernels and over the three projection planes.

$$L_{topo} = \frac{|\sum_{k \in K} G_{topo}^k - P_{topo}^k|}{3 \times |K|}. \quad (3)$$

where K is the set of kernels and $|K|$ is the number of kernels.

In the implementation, we chose the different kernel sizes based on the characteristics of the task. The details of the kernel selection strategies are introduced in Section 3.4.

3.3 Final loss

The loss function is a combination of the voxel-wise Dice loss and the proposed topological loss weighted by a hyper-parameter λ :

$$L_{total} = L_{topo} + \lambda L_{dice}. \quad (4)$$

λ can be a fixed or dynamic hyperparameter. In our experiments, we used a fixed value of $\lambda = 1$.

Similar to reference,¹⁰ pre-training with a standard loss is necessary before using our loss function. This is because if the network is randomly initialized, excessive topological errors (which can be in the hundreds) will make the network unable to fit. So we first pre-train with the Dice loss and then fine-tune with our loss function which combines Dice and topological losses.

3.4 Kernel size selection

In our loss function, the kernel sizes of the 2D-Max-Pooling layer are hyperparameters, which can be adjusted according to the characteristics of the target structure. Here, we propose a simple heuristic to choose the kernel sizes. More sophisticated approaches could be used in the future. The largest kernel size k_1 is equal to the width of the image projected onto the axial plane divided by four and by the number of connected components in the target. Then, we iteratively define smaller kernels k_j as $\frac{k_{j-1}}{2}$ if this results in an even number or $\frac{k_{j-1}}{2} - 1$ otherwise. We stop the process when k_j is strictly smaller than a threshold T defined as follows. T approximates the size of the projection smallest structure among all participants in the training set. (More precisely, it is equal to $T = \min_{i=1 \dots p}(\min(G_{\text{axial}}^{(i)}, G_{\text{sagittal}}^{(i)}, G_{\text{coronal}}^{(i)}))$ where p is the number of participants in the training set, $G_{\text{axial}}^{(i)}$ (resp. $G_{\text{sagittal}}^{(i)}$ and $G_{\text{coronal}}^{(i)}$) is the axial (resp. sagittal and coronal) projection of the target structure of participant i . The kernel size results obtained are as follows. For red nucleus: 2, 4, 10, 20. For spleen: 10, 20, 40. For heart: 10, 20, 40. For hippocampus: 4, 8, 16.

4 Experiments and results

4.1 Implementation

Our loss function is compatible with any model architecture. In our experiments, we used the standard 3D-UNet.⁴⁴ Since the training sets are very small in some experiments, we used early stopping in the training process to avoid over-fitting. The early stopping was done using only the validation set in order to obtain an unbiased performance on the test set. We did not use data augmentation or hyperparameter search techniques. In all experiments, the hyperparameters were identical.

As mentioned in Section 3.3, training is done in two parts: i) pre-training with the Dice loss; ii) fine-tuning with the proposed combined loss. For pre-training, the learning rate was 10^{-3} , for fine tuning it was 10^{-4} . For all real datasets, pre-training and fine-tuning comprised each 150 epochs. When comparing the proposed loss to the standard Dice loss, in order to have a fair comparison, the total number of training epochs was the same in both cases (300 epochs for Dice loss, 150+150 epochs for the proposed loss). We always used the Adam optimizer.⁴⁵

We performed the following preprocessing steps: reshaping images to the same size for each task and min-max normalization. To that purpose, we used the open-source Python library TorchIO⁴⁶¹. Our code was developed based on the PyTorch framework.⁴⁷

4.2 Datasets

We evaluate the approach on synthetic data and then on several medical imaging datasets. The first medical imaging dataset is our main application where the task is to segment the red nucleus from QSM (Quantitative Susceptibility Mapping) data. The others are three publicly available datasets which are part of the Medical Segmentation Decathlon (MSD)²¹² (heart, hippocampus and spleen).

4.2.1 Synthetic datasets

To test the performance of our loss function in different situations and explore the limitations, we synthesize a series of parameter controllable datasets. The controlled synthetic datasets aim at studying the performance as a function of different characteristics of the target: size, distance, number of components, and noise. The process for synthesizing data is as follows: first, we generate the 3D volume containing different numbers of 3D spheres following the strategies described below. Each of the spheres is treated as the target segmentation region (i.e., ground truth) of the synthesized data. Then, we used the TorchIO⁴⁶ package to generate synthetic images by adding random noise and blurring the 3D volumes under default settings (excluding the Noise strategy generation). For the test set we replaced the starting position of the spheres' center to avoid repeating the same information as found in the training set. The next steps are the same as before. All synthetic images are standardized to a size of $160 \times 160 \times 48$. Specifically, the 5 strategies to generate the synthetic datasets are as describe as follows: we generated five controlled synthetic datasets as follows:

- 1) **Scale.** The target is composed of two spheres. The scale of one of the spheres varies from 0.2x to 3.9x, providing insight into how scale variations impact performance.
- 2) **Shift.** Here, the target also consists of two spheres. We systematically modify the distance between the spheres' boundaries from 0 to 39 voxels in single increments of 1, examining the effect of spatial separation on our model.
- 3) **Scale+shift.** In this configuration, the target remains with two spheres. As one sphere's size changes, the gap between the two spheres also increases, allowing us to evaluate the combined effects of scaling and shifting..
- 4) **Components.** The target in this dataset varies from 1 to 8 components, randomly placed. To account for the potential merging of components as their number grows, we reduced the initial radius compared to other datasets. The scale ratio is randomly set between 1x and 2x, challenging the model with a diverse component structure.
- 5) **Noise.** We introduce Gaussian noise to the Shift dataset volumes, varying the noise's standard deviation across five levels: 0.01, 0.05, 0.1, 0.3, 0.5. This dataset tests the model's robustness against varying degrees of noise.

¹<https://torchio.readthedocs.io/>

²<http://medicaldecathlon.com/>

The examples of the different controlled synthetic datasets are depicted in Figure 2, with the noise task showcasing training samples at different noise levels. The characteristics of these synthetic datasets are summarized in Table 1.

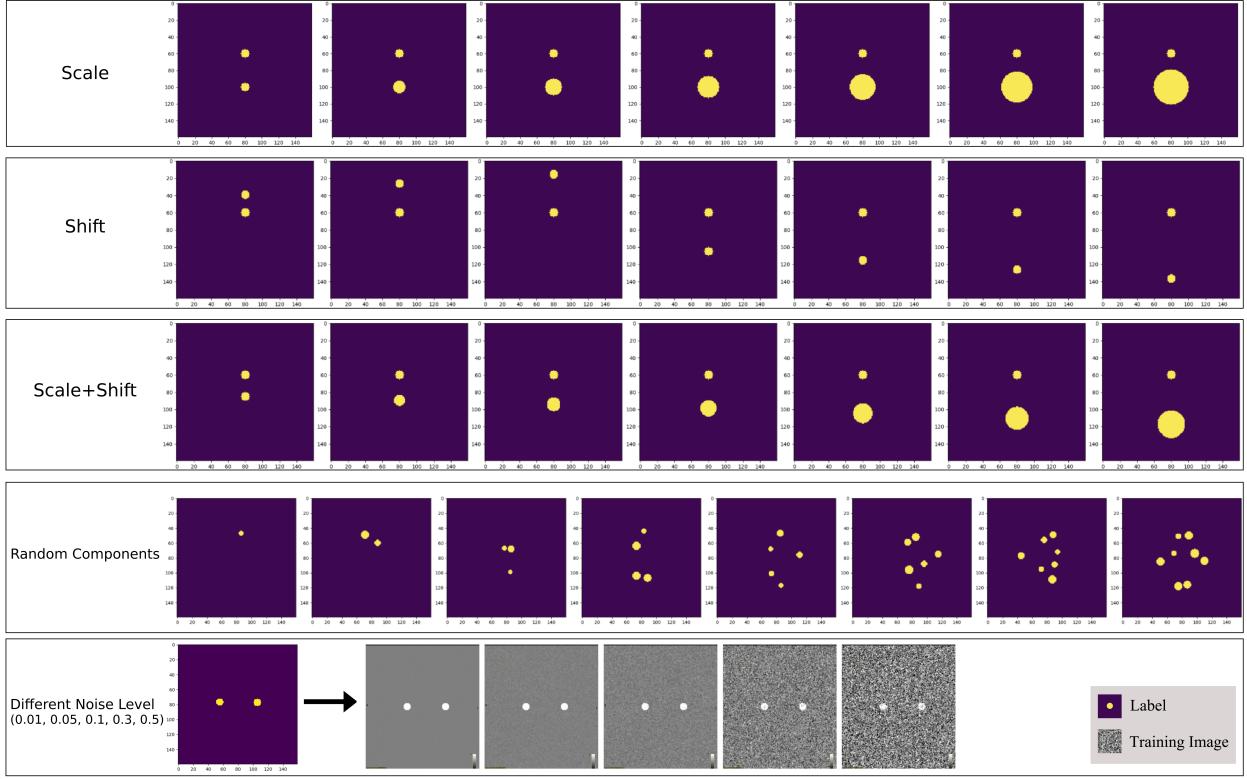


Fig 2 Controlled synthetic datasets. Projection in the axial plane for the different targets of the synthetic datasets. For noise category, we present training data to represent different levels of noise.

Table 1 Synthetic datasets characteristics.

Task	Train	Validation	Test	Initial Radius
Scale	25	7	40	10
Shift	25	7	40	5
Shift+scale	25	7	40	5
Components #C: 1-8 (4/class)	32	32 (4/class)	80 (10/class)	3
Noise-0.01	25	7	40	5
Noise-0.03	25	7	40	5
Noise-0.1	25	7	40	5
Noise-0.3	25	7	40	5
Noise-0.5	25	7	40	5

4.2.2 Red nucleus dataset

The task was to segment the red nucleus from Quantitative Susceptibility Mapping (QSM) brain images.⁴⁸ QSM is an indirect method for imaging iron levels in vivo.⁴⁹ The red nucleus has ferromagnetic properties and QSM is thus an imaging technique that is well-adapted to its visualization and segmentation. The dataset includes a total of 80 participants, including 18 healthy controls, 46 patients with early Parkinson's disease (i.e. with a disease duration below 4 years), and 16 patients with prodromal parkinsonism (idiopathic rapid eye movement sleep behavior disorder-iRBD), recruited between May 2015 and January 2019. The institutional ethical standard committee approved the study (CPP Paris VI/RCB: 2014-A00725-42). All participants gave written informed consent. The dimensions of the images are: $160 \times 160 \times 128$, and the voxel size is: $1 \times 1 \times 1\text{mm}^3$. QSM images were manually segmented by a trained neuroradiologist (L.C.). Figure 3 shows an example of QSM data with the red nucleus segmentation.

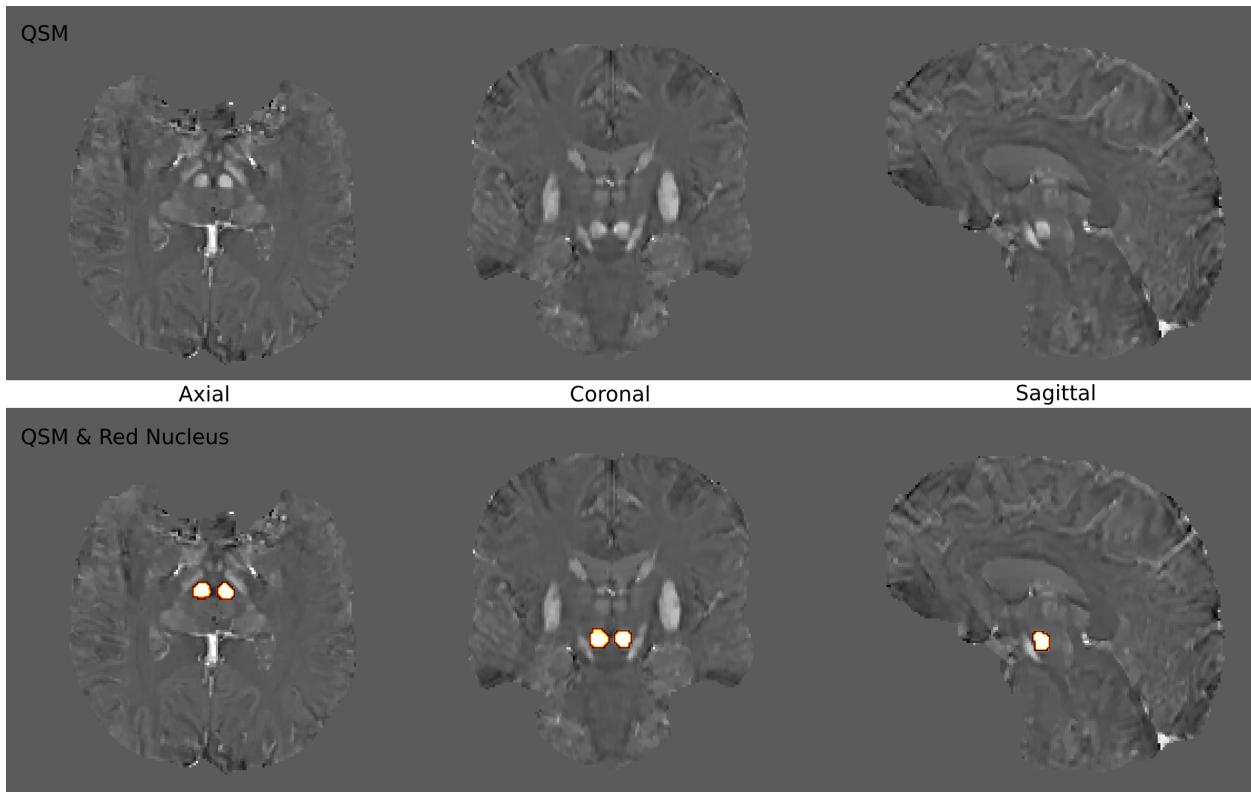


Fig 3 A QSM image displayed from axial, coronal, and sagittal planes together with the manual segmentation of the red nucleus.

4.2.3 Datasets from the medical segmentation decathlon

We used three datasets from the medical segmentation decathlon:²¹ spleen, heart and hippocampus. The spleen task is to segment the spleen in the portal phase in a CT scan. The heart task is to segment the left atrium in an MRI scan. The hippocampus task is to segment the anterior and posterior parts of the hippocampus in an MRI scan.

4.3 Dataset splits

Each dataset was split into training, validation and test sets. The splits were done at the participant level to avoid any data leakage.⁵⁰ The number of images in each split is reported in Table 2. Moreover, the models were trained with different percentages of the training set to explore the performance with varying training set sizes, while the validation and test sets were left unchanged.

Table 2 Splits of the medical imaging datasets into training, validation and test sets. MSD: Medical Segmentation Decathlon.

Dataset	Task	Train+val	Test	Image Size
Red nucleus	Red Nucleus	51+13	16	160,160,128
MSD	Spleen	25+7	9	160, 160, 128
	Heart	12+4	4	160, 160, 128
	Hippocampus	166+42	52	64, 64, 48

4.4 Evaluation

We characterized the performance at both the voxel level and the topological level. For each metric, we report the mean and the 95% confidence interval computed using bootstrap (with 5000 resamplings).

4.4.1 Voxel level evaluation

Voxel-level performance was evaluated using the Dice score and the 95 percentile Hausdorff distance (95%HD) which are defined as follows. Let X (resp. Y) be a segmentation and x (resp. y) be a voxel in X (resp. Y).

The Dice score is: $\frac{2|X \cap Y|}{|X| + |Y|}$.

The 95%HD is defined as:

$$95\%HD(X, Y) = \max_{x \in X} (\text{Perc}_{95} d(x, Y), \text{Perc}_{95} d(y, X))$$

where $d(x, Y) = \min_{y \in Y} d(x, y)$.

4.4.2 Topological level evaluation

Topological errors may have only a small impact on voxel-level metric if they are of limited size. However, topological errors make the predictions anatomically inconsistent. We thus evaluated the connected components (CC) at the 3D level. The 3D CC error is simply defined as the absolute value of the difference between the number of CC in the ground-truth and that in the prediction. We used scikit-image³ to compute the number of connected components.

³<https://scikit-image.org/docs/dev/api/skimage.measure.html#skimage.measure.label>

4.5 Compared approaches

In the following, we compare four approaches. The first two are the standard Dice loss (referred to as baseline) and the proposed approach. In addition, we performed the following post-processing to remove small erroneous connected components: when the target object has n connected components in the ground-truth, we kept only the largest n connected components. The four approaches are referred to as: i) Baseline; ii) Baseline+Post (when post-processing is applied); iii) Proposed; iv) Proposed+Post (when post-processing is applied).

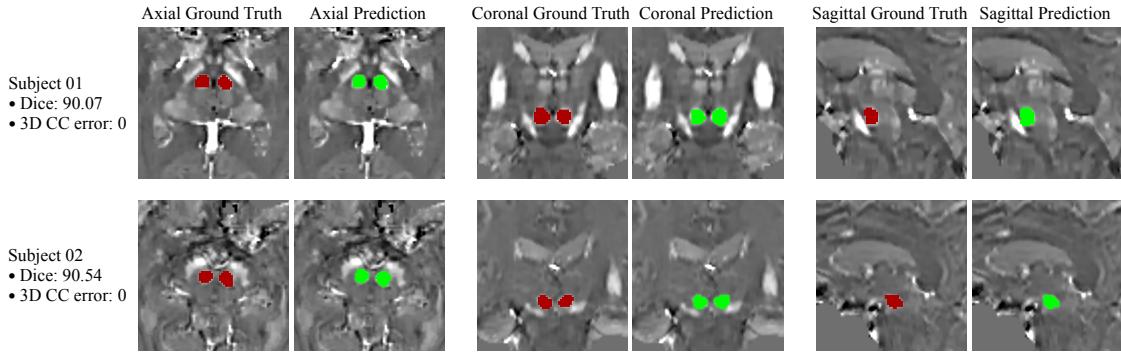


Fig 4 Examples of automatic and ground-truth segmentation for the red nucleus using the proposed approach for two subjects. The figure also displays the evaluation metrics: Dice and 3D connected components (CC) error.

4.6 Results

4.6.1 Synthetic datasets

Results on the controlled synthetic datasets are shown in Table 3. One can observe that the average performance is high for the different approaches. This is not surprising given that the targets are simple shapes (spheres). More interestingly, we can study the performance as a function of the different controlled parameters. As shown in Figure 5 (a), our model is stable for components of different sizes. When the scaling ratio is 0.2, the radius of the sphere is only 2 voxels (start radius = 10, scale ratio= 0.2), and a Dice score of 0.969 can still be achieved. Our performance also trended upwards as the components was scaled up, with sphere radius between 15 and 27 and dice score have greater than 0.99 at scales in the 1.5-2.7 range. This is because each layer of pooling kernels (kernel size = [4, 5, 8, 10, 20]) starts to provide useful topology information at different scales; while at scale 0.2, components with a radius of 2 voxels will be correspond to the same kernel size, in which case the multi-level kernel combination will only provide a single information. However, at small scales we can still achieve a Dice score of 0.97. Conversely, when the components are too large, the small kernel plays a limited role, which gradually masks the advantage of our loss function. In general, our loss function has stable performance and works with different element sizes.

We can find the performance under the “Shift” category in Figure 5 (b), our model is stable under different distance between target regions. Positive and negative axes mean that one of the element moves to the right or left with respect to the other one. The performance degradation for a distance of zero is because the two components are so close that they merge into one. The 1-component case is not included in the training set, so our topological loss tends to separate the

Table 3 Experiments on test set of synthetic dataset.

Task	Loss	Dice	95%HD	3D CC
Shift	Baseline	96.53	0.80	0.0
	Topology	95.65	1.000	0.03
Scale	Baseline	98.36	1.081	0.03
	Topology	98.35	0.95	0.03
Shift+Scale	Baseline	92.85	0.850	0.35
	Topology	94.13	0.950	0.08
Components	Baseline	96.30	0.8131	0.09
	Topology	95.90	0.9750	0.05
Noise-0.01	Baseline	98.87	0.00	0.0
	Topology	99.19	0.00	0.0
Noise-0.05	Baseline	97.77	0.225	0.0
	Topology	98.42	0.00	0.0
Noise-0.1	Baseline	94.00	1.000	0.0
	Topology	97.14	0.9737	0.0
Noise-0.3	Baseline	92.28	1.000	0.0
	Topology	94.27	1.000	0.0
Noise-0.5	Baseline	0.86	1.353	0.0
	Topology	0.886	1.000	0.03

two components. Although we have an error of 0.03 in the 3D topology loss, this error proves the reliability of our loss function. Because in the case of actual medical data, there is almost no case that the number of organs is different between populations. Learning by using our loss function, the prior knowledge that the dataset contains 2 components can be summarized from the training set. This kind of prior knowledge summary does not imply that our loss function is inflexible, as our loss function still performs better with a random number of components, and detailed analysis can be seen below.

Our above conclusion is verified again by the scale+shift experiment. When we change the two variables of distance and size at the same time, the difficulty of the task increases, and our advantages can be reflected at both the voxel and topology level.

For shift, scale, and noise-related tasks, since the shape and topology of the target are fixed, all models perform well at the voxel level for these tasks. Difficulty at the topology level increases with the number of components, as this examines the flexibility of the algorithm. As shown in the experimental results, all loss functions have topological errors. But the advantage of our loss function can be reflected in this task that topology loss will nearly have no unexpected predictions that cause topology error. We have traded negligible performance loss for topological correctness, which is often more important for medical tasks.

Adding more noise means increasing the difficulty of the task, making its boundaries indistinguishable. As can be seen from Figure 5(d), as the noise increases, the performance of all the models gradually decreases. The Dice loss is more sensitive to noise, which is reflected in the performance degradation at noise levels of 0.3 and 0.5. For our loss function, although the performance is also degraded, it performs better than the dice loss function, across all noise levels.

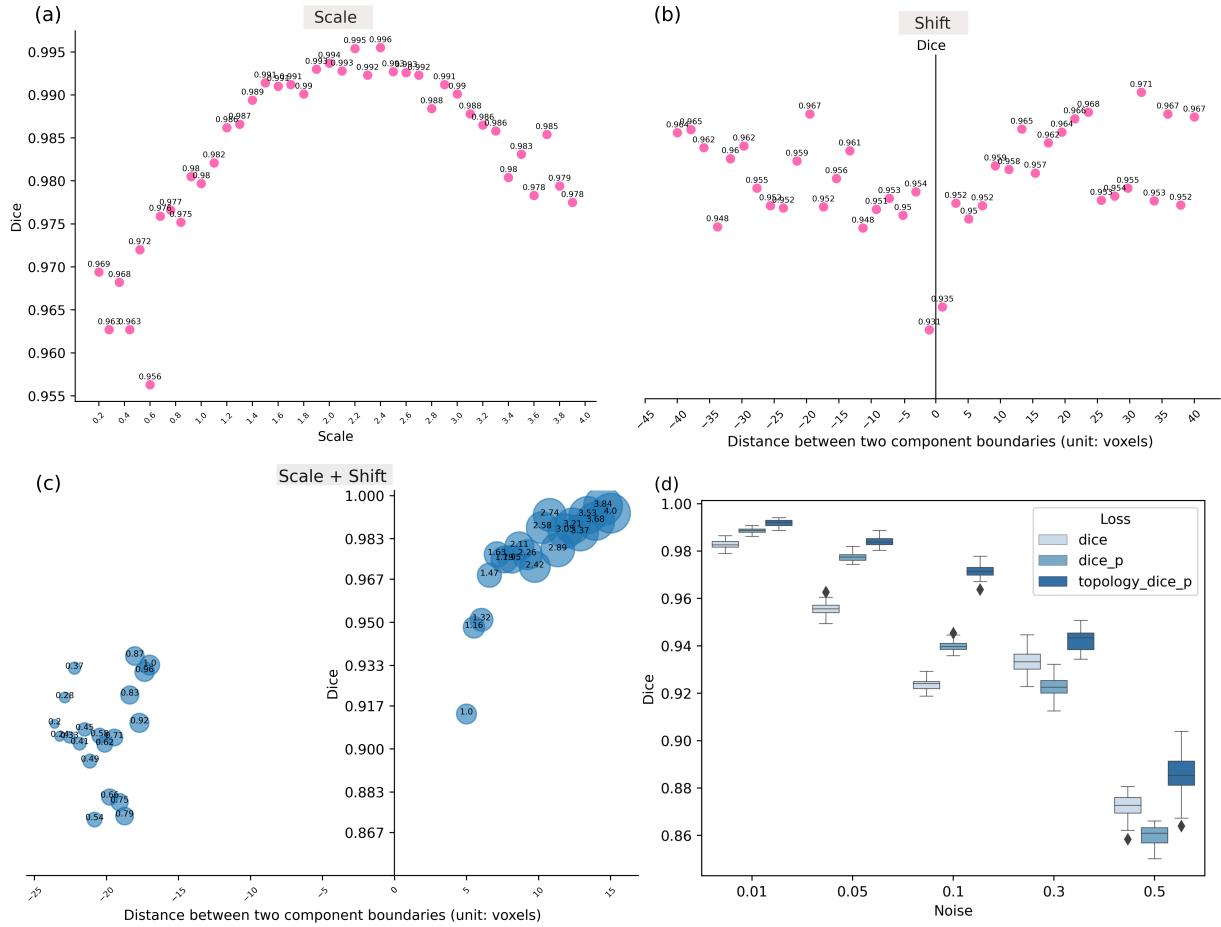


Fig 5 Voxel level performance per volume on 4 subtasks: (a) Scale, (b) Shift (c) Scale+shift (d) Noise. For (d), the *dice* means the pretrained model, *dice_p* and *topology_dice_p* mean fine-tuning models.

4.6.2 Red nucleus dataset

Results are presented in Table 4. Some examples of automatic segmentations with the proposed approach are shown in Figure 4. The voxel-level performance is very high when training with the full dataset of 51 participants (Dice around 90%, 95%HD of 1) for both the proposed approach and the baseline. However, our approach increased the topological correctness: there was no 3D CC error while there was an erroneous CC in about one third of the participants for the baseline. Moreover, when training only on a subset (from 7.5% to 30% of the original training set), the voxel-level performances are substantially higher for the proposed approach which allows to improve the Dice score by 6 to 16 percent points. One can also appreciate that the confidence intervals are non-overlapping, indicating that the improvement is statistically significant. Furthermore, the 95%HD error is also substantially reduced. 3D renderings obtained with baseline and proposed approaches are shown in Figure 6. One can observe that the proposed approach removes erroneous connected components with the need for post-processing. When post-processing is applied, erroneous connected components are removed in both cases but the Dice is higher for the proposed method.

The effect of the topological loss is illustrated on Figure 7. To ensure the consistency of the experiments, we systematically trained for 150+150 epochs. However, from this figure, we can see

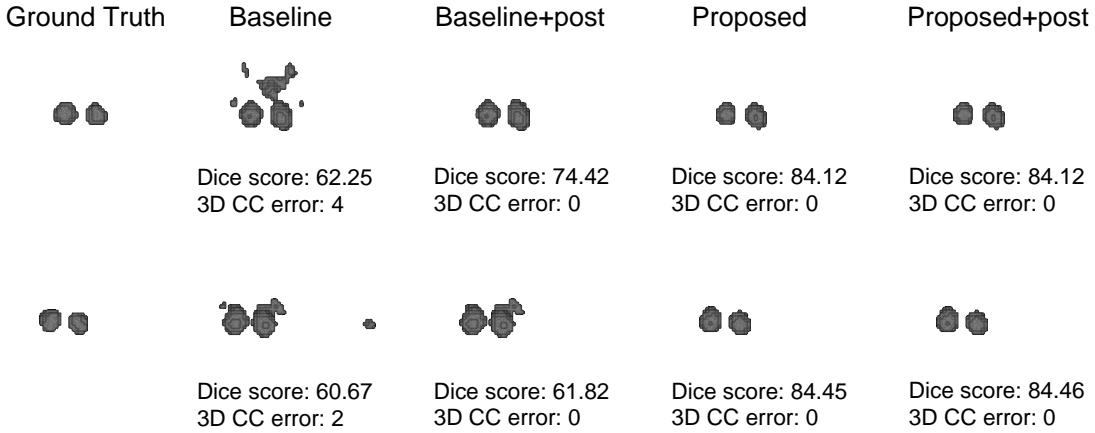


Fig 6 Comparison between the baseline and the proposed approach for red nucleus segmentation when training from a small sample (7.5% of the training set). One can observe that the proposed approach allows to remove most erroneous connected components without the need of post-processing. After post-processing, the proposed approach still allows to substantially improve the Dice metric.

that our loss function can improve the prediction results with only a few epochs.

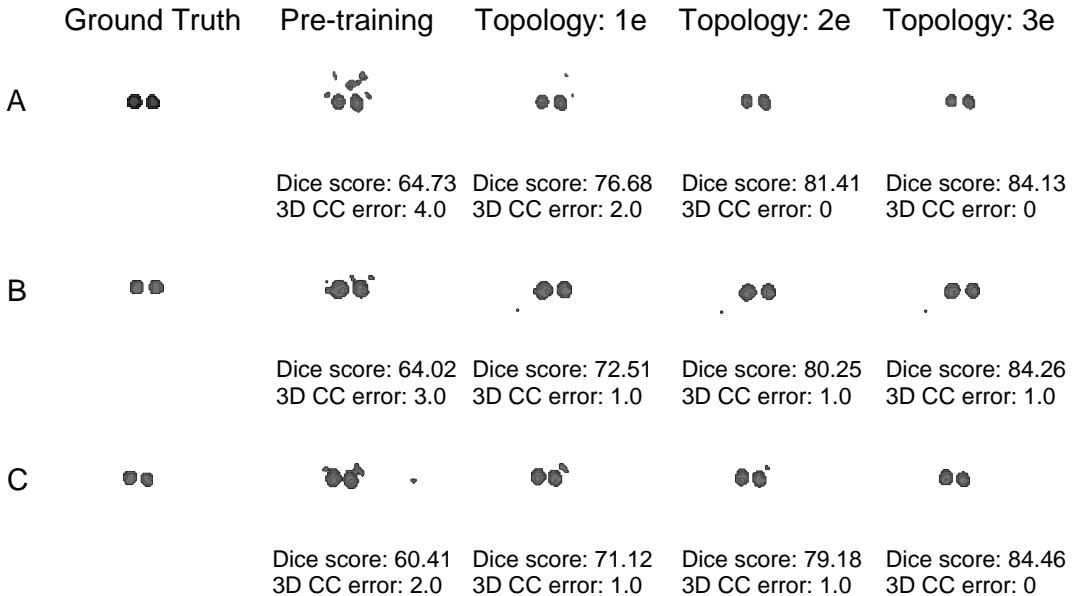


Fig 7 Some examples of the training process using the proposed loss function. One can observe that after pretraining with the Dice loss, many topological errors remain. Fine-tuning with the topological loss can often remove these errors with only a few epochs (here 1, 2 or 3 epochs). The example shown is the segmentation of the red nucleus from QSM using 7.5% of the training set.

4.6.3 Datasets from the medical segmentation decathlon

Results are presented in Table 4. Unlike for the red nucleus dataset, the proposed loss and the baseline Dice loss achieved similar Dice accuracies and 95%HD, across all training sample sizes.

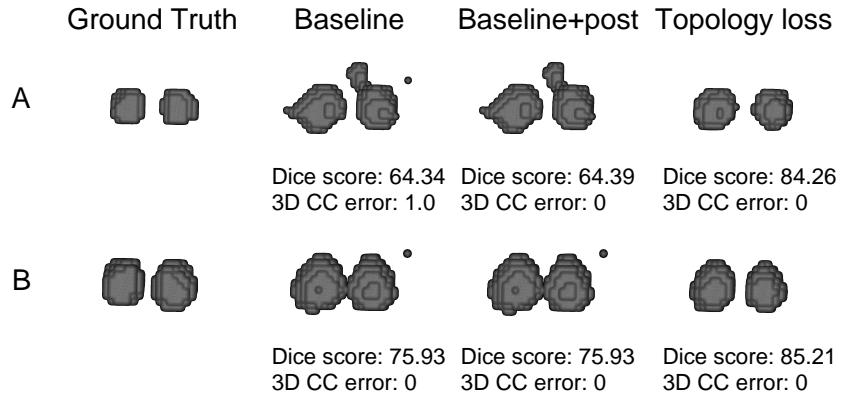


Fig 8 Comparison of the effects of post-processing and topological loss.

However, without topological post-processing, the proposed loss substantially decreased the topological errors over the baseline. In particular, the 3D CC error was lower for all tasks with the proposed loss. When applying the topological post-processing, the topological errors were similar for both approaches. As shown in Figure 9, the proposed approach allows to remove some topological errors without the need of post-processing. However, after post-processing, the results are similar.

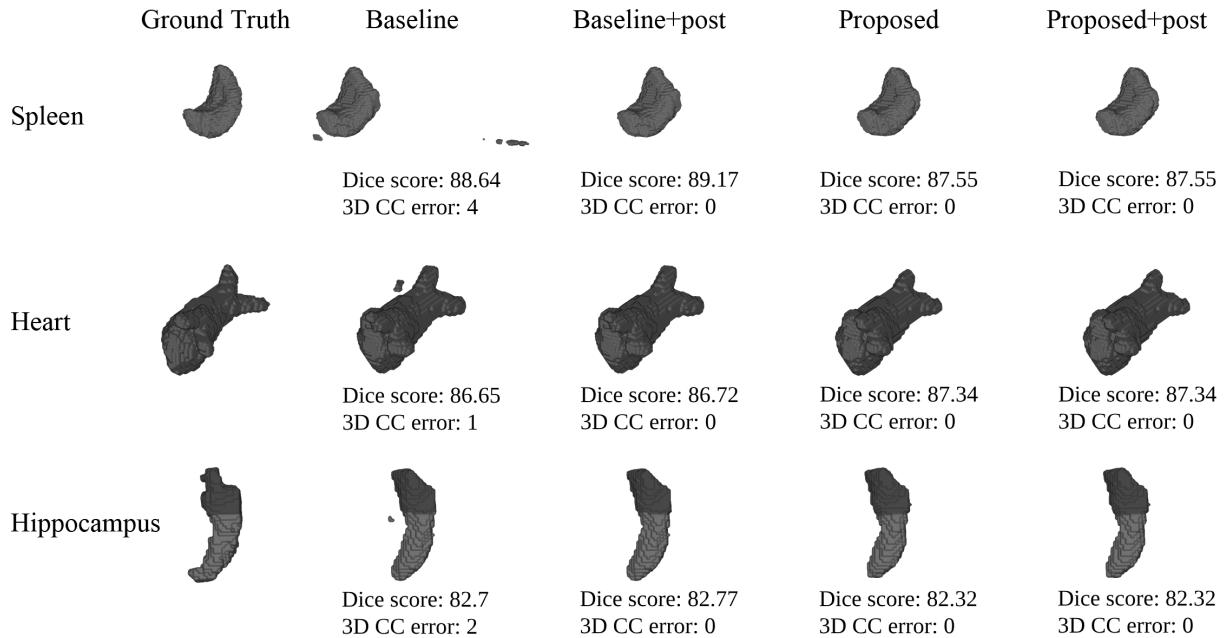


Fig 9 The figure shows the predictions of the models trained with baseline (Dice Loss) and with our proposed loss (Topological Loss) on three tasks: heart, spleen and hippocampus segmentation. One can observe that the proposed approach avoids topological errors without the need of post-processing. However, post-processing allows removing the errors.

Table 4 Results on medical datasets: red nucleus and three public datasets from the Medical Segmentation Decathlon (MSD). The training set size varies: size is indicated both as a percentage (%) of the total training set and as the absolute number of samples (n). Baseline refers to the Dice loss. “+Post” indicates that the ad-hoc post-processing was applied.

Dataset	Training set % (n)	Loss	Dice	95%HD	3D CC
Red nucleus	7.5% (4)	Baseline	64.6 [60.7, 68.2]	6.5 [4.1, 9.4]	2.06 [1.25, 3]
		Baseline+Post	66.9 [63.4, 70.1]	3.3 [2.4, 4.4]	0 [0, 0]
		Proposed	81.6 [78.8, 83.7]	2.1 [1.0, 4.3]	1.3 [0.69, 2.06]
		Proposed+Post	82.8 [80.9, 84.5]	1.1 [1.0, 1.2]	0 [0, 0]
	15% (8)	Baseline	77.4 [75.6, 79.3]	15.8 [4.9, 28.8]	4.94 [3.56, 6.44]
		Baseline+Post	80.3 [78.7, 82.1]	1.1 [1.0, 1.1]	0 [0, 0]
		Proposed	86.6 [84.7, 88.3]	1.0 [1.0, 1.0]	0.38 [0, 0.81]
		Proposed+Post	86.7 [84.8, 88.6]	1.0 [1.0, 1.0]	0 [0, 0]
	30% (16)	Baseline	73.4 [69.7, 76.8]	13.2 [4.9, 23.7]	4.19 [2.81, 5.63]
		Baseline+Post	76.2 [71.0, 79.9]	4.8 [1.1, 12.1]	0 [0, 0]
		Proposed	86.0 [83.0, 88.2]	1.3 [1.0, 1.7]	0.5 [0.19, 0.88]
		Proposed+Post	86.4 [83.5, 88.7]	1.1 [1.0, 1.2]	0 [0, 0]
	100 % (51)	Baseline	90.3 [89.4, 91.2]	1.0 [1.0, 1.0]	0.31 [0.06, 0.63]
		Baseline+Post	90.5 [89.5, 91.4]	1.0 [1.0, 1.0]	0 [0, 0]
		Proposed	89.8 [88.5, 90.9]	1.0 [1.0, 1.0]	0 [0, 0]
		Proposed+Post	89.8 [88.5, 90.9]	1.0 [1.0, 1.0]	0 [0, 0]
Spleen	15% (4)	Baseline	65.1 [53.0, 76.6]	32.9 [21.4, 43.9]	1 [0, 2.22]
		Baseline+Post	65.2 [53.2, 76.8]	32.6 [21.1, 43.9]	0 [0, 0]
		Proposed	65.6 [52.2, 77.6]	29.2 [18.9, 40.2]	0.67 [0, 1.56]
		Proposed+Post	66.0 [52.1, 78.4]	27.7 [16.9, 38.8]	0 [0, 0]
	30% (8)	Baseline	70.3 [57.5, 82.0]	29.0 [17.8, 40.2]	1 [0.22, 2.22]
		Baseline+Post	70.7 [57.5, 82.8]	27.0 [15.0, 38.9]	0 [0, 0]
		Proposed	69.7 [55.9, 81.4]	28.4 [18.5, 38.7]	0.33 [0, 0.78]
		Proposed+Post	70.2 [56.4, 82.3]	26.4 [15.4, 37.7]	0 [0, 0]
	100% (25)	Baseline	84.6 [78.1, 89.9]	15.1 [7.9, 22.8]	0.78 [0.11, 1.67]
		Baseline+Post	85.1 [78.5, 90.8]	12.6 [6.0, 20.2]	0 [0, 0]
		Proposed	85.1 [78.8, 90.4]	10.2 [5.1, 16.3]	0.11 [0, 0.33]
		Proposed+Post	85.3 [78.9, 90.6]	10.1 [4.9, 16.4]	0 [0, 0]
Heart	30% (4)	Baseline	72.6 [65.3, 79.9]	13.1 [7.7, 18.6]	0.5 [0, 1]
		Baseline+Post	72.9 [65.9, 79.9]	12.1 [7.7, 16.5]	0 [0, 0]
		Proposed	74.1 [68.1, 80.1]	10.8 [7.2, 14.4]	0.25 [0, 0.75]
		Proposed+Post	74.2 [68.2, 80.2]	10.8 [7.2, 14.3]	0 [0, 0]
	100% (12)	Baseline	82.0 [75.5, 88.4]	8.3 [3.7, 12.9]	1 [0.25, 1.75]
		Baseline+Post	82.3 [76.2, 88.4]	8.3 [3.8, 12.9]	0 [0, 0]
		Proposed	82.1 [75.2, 89.0]	8.2 [3.3, 13.1]	0.5 [0, 1]
		Proposed+Post	81.8 [74.6, 88.9]	8.9 [3.7, 14.2]	0 [0, 0]
Hippocampus	2.5% (4)	Baseline	75.8 [74.1, 77.3]	2.1 [1.9, 2.3]	0.13 [0.07, 0.22]
		Baseline+Post	75.8 [74.1, 77.4]	2.1 [1.9, 2.3]	0 [0, 0]
		Proposed	75.1 [73.4, 76.7]	2.3 [2.1, 2.5]	0.06 [0.02, 0.11]
		Proposed+Post	75.7 [73.8, 77.4]	2.5 [2.2, 2.7]	0 [0, 0]
	5% (8)	Baseline	78.4 [76.7, 79.8]	1.9 [1.8, 2.1]	0.08 [0.03, 0.13]
		Baseline+Post	78.4 [76.7, 79.8]	1.9 [1.8, 2.1]	0 [0, 0]
		Proposed	76.4 [74.9, 77.8]	2.2 [1.9, 2.4]	0 [0, 0]
		Proposed+Post	76.4 [74.9, 77.8]	2.2 [1.9, 2.4]	0 [0, 0]
	10% (16)	Baseline	80.5 [79.2, 81.7]	1.8 [1.6, 1.9]	0.04 [0.01, 0.08]
		Baseline+Post	80.5 [79.1, 81.7]	1.8 [1.6, 1.9]	0 [0, 0]
		Proposed	78.6 [77.3, 79.8]	1.9 [1.8, 2.1]	0 [0, 0]
		Proposed+Post	78.6 [77.3, 79.8]	1.9 [1.8, 2.1]	0 [0, 0]
	30% (50)	Baseline	84.8 [83.9, 85.5]	1.4 [1.3, 1.5]	0.04 [0.01, 0.08]
		Baseline+Post	84.8 [83.9, 85.6]	1.4 [1.3, 1.5]	0 [0, 0]
		Proposed	84.2 [83.3, 84.9]	1.4 [1.3, 1.5]	0 [0, 0]
		Proposed+Post	84.2 [83.3, 84.9]	1.4 [1.3, 1.5]	0 [0, 0]
	100% (166)	Baseline	86.1 [85.4, 86.8]	1.3 [1.2, 1.4]	0.02 [0, 0.05]
		Baseline+Post	86.1 [85.4, 86.8]	1.3 [1.2, 1.4]	0 [0, 0]
		Proposed	85.3 [84.6, 86.0]	1.4 [1.3, 1.5]	0 [0, 0]
		Proposed+Post	85.3 [84.6, 86.0]	1.4 [1.3, 1.5]	0 [0, 0]

5 Discussion

In this paper, we proposed a new loss function that integrates soft topologically constraints. It is simple and computationally efficient and can be applied with any segmentation model. When applied to the segmentation of the red nucleus, it resulted in high voxel-level accuracy and topological correctness. Moreover, when training with few samples, it resulted in improved performance over the baseline.

Our loss function imposes soft constraints in the sense that we don't have absolute guarantees of topological correctness. Moreover, we mainly control connected components and not other topological characteristics (holes, tunnels...). This contrasts with approaches based on persistent homology (PH) ^{10,35,36} which are advantageous in that respect. However, PH-based approaches are computationally expensive, in particular in 3D. One can also note that our approach not only introduces constraints on the topology but also more generally will magnify small errors, even if they don't constitute isolated connected components. This is due to the use of the pooling process. This can be advantageous as it can also favor the correct reconstruction of anatomical details in general, beyond connected components.

Across the vast majority of experiments, our approach substantially decreased topological errors. By contrast, the average 3D CC errors of the baseline are higher. For instance, on average one third of the subjects have an erroneous CC for the red nucleus segmentation while there are not 3D CC errors for the proposed approach. It is true that removing erroneous connected components (CC) can be achieved through ad-hoc post-processing operations. The tested post-processing was efficient in the sense that it removes the 3D CC errors. However, it is more satisfying to be able to achieve the same results without any ad-hoc post-processing. Moreover, the proposed approach still provided better Dice accuracies when training with small samples, even after post-processing.

When using the full training sets, the Dice accuracy was comparable between our approach and the baseline. Nevertheless, an interesting feature of our approach is that, in the case of the segmentation of the red nucleus, it led to considerable improvements in Dice accuracy (from 6 to 15 percent points) when using only a fraction of the training set. Note that this improvement still held when the ad-hoc post-processing was applied. However, this was not the case for other target structures (spleen, heart, hippocampus). It is unclear where this difference comes from. One can only speculate that it comes from the differences in terms of shapes of the target structures. Further work will be needed to investigate if this improvement is specific to the red nucleus or can be also obtained for other tasks.

While the developed approach is general, our motivating target application was the segmentation of the red nucleus from QSM images. There are several existing approaches for automatic segmentation of the red nucleus.^{15–20,51,52} Among these, some are based on QSM data.^{15–18,20} Our results are in line with those of the best performing approaches which reach around 90% of Dice accuracy.^{15,17,18} However, these three studies only included healthy controls. Thus, it is unclear how they would perform on patients. On the contrary, we included patients with parkinsonism at different stages. Furthermore, we report narrow confidence intervals which indicate that our performance estimates are precise.⁵³ This provides confidence on the applicability of our results.

Our work has the following limitations. First, as mentioned above, our loss does not impose strict topological guarantees. This can be a limitation if such guarantees are critical for a given application and in that case, PH-based method are likely to be more adapted. However, restricting the number of connected components at the topology level reduces the flexibility of the method.

Our loss function achieves better performance in these two cases. In Figure 8 (A), although an additional error can be removed by post-processing, the performance of Dice cannot be improved due to the poor performance of the main part. In Figure 8 (B), the two main parts are connected as a single component, so the post-processing cannot remove the error region.

It would have been good to compare our approach with that of the topology loss proposed by Clough et al.¹⁰ and the Hausdorff distance loss proposed by Karimi et al.⁵⁴ But, the topological loss code provided in the paper¹⁰ does not work in 3D. For the Hausdorff distance loss,⁵⁴ it is unclear if it can deal with the case of incorrectly merged regions as we found for the red nucleus. We agreed that these losses could in principle help remove erroneous components. But as mentioned above, removing erroneous connected components can also be achieved through post-processing. However, we could see that the proposed loss still has benefits, even after post-processing is applied. In particular, it lead to improvements when training with small samples.

Another interesting approach is the loss function based on distance weighting.⁵⁵ The distance transform map-based weighting (DTM) loss,^{54,56} which multiplies prediction errors by the DTM, could potentially solve the problem of erroneous distant connected components by assigning higher weights to pixels further from the boundary of ground-truth labels. While the DTM-based weighting approach has advantages, particularly in penalizing errors in separate topological components, there are specific scenarios where it might not fully address the issues. For instance, in the case where connected components of the nucleus are merging (as illustrated in Figure 8.B.), DTM-based weighting may not adequately resolve the ambiguities. Our proposed method can target these touching components by supervising from a topological angle. We plan to explore this comparison in future work to further understand the benefits and limitations of both methods.

In this work, we chose to use a single label for both left and right red nuclei. Another possibility would be to use separate labels for each organ as done for instance by Milletari et al.⁵⁷ and Dolz et al.⁵⁸ This could simplify the topology, resulting in single connected components for each label. However, our loss function also takes into account the relationship between the two regions, as shown in Figure 8.B. If these regions are treated as independent, the error situation arising from merging of the two prediction regions into a single connected component may not be effectively resolved. Additionally, it is uncertain whether this approach would address all segmentation challenges, particularly under conditions of very small training sets, where we observed that our approach resulted in substantial improvements in performance metrics for the red nucleus, even after post processing.

In our study, our loss function combines 3D projection with 2D max-pooling, but this method may have a drawback, i.e., the transition from 3D to 2D projection may lead to a loss of spatial information, which may potentially affect the algorithm’s ability to preserve the complex 3D structure of the data. Another solution would be to directly perform the max-pooling in 3D. However, our solution also benefits from the subtle capture of orientation changes by the 2D projection for better control of details and increased sensitivity. We have not compared the two solutions in this research, which is a limitation of our current study, and is left for future work.

Furthermore, the heuristic proposed for the choice of kernel sizes is still ad-hoc. Future work could focus on dynamically learning the optimal kernel size during the training process. Another limitation of the study is the fact that we only included patients with typical parkinsonism. Further studies will be needed to asses the performance on patients with other parkinsonian syndromes such as progressive supranuclear palsy or multiple system atrophy. Finally, our study only included a relatively small number of patients and future validation on larger cohorts will be needed.

In conclusion, we proposed a novel loss function for integrating soft topological constraints in deep learning based segmentation. The approach is simple to implement, has a low computational cost and is generic as it can be used with any segmentation architecture. We demonstrated that it can accurately segment the red nucleus from QSM images, even when training with small samples.

Disclosure of interests

Competing financial interests related to the present article: none to disclose for all authors. Competing financial interests unrelated to the present article: OC reports having received consulting fees from AskBio and Therapanacea and that his laboratory has received grants (paid to the institution) from Qynapse. Members from his laboratory have co-supervised a PhD thesis with Qynapse. OC's spouse is an employee and holds stock-options of myBrainTechnologies. OC holds a patent registered at the International Bureau of the World Intellectual Property Organization (PCT/IB2016/0526993, Schiratti J-B, Allassonniere S, Colliot O, Durrleman S, A method for determining the temporal progression of a biological phenomenon and associated methods and devices) (2017).

Data availability statement

The red nucleus dataset is available upon reasonable request to the ICEBERG Study Group. The Medical Segmentation Decathlon data is publicly available at <http://medicaldecathlon.com/>.

Acknowledgments

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program (reference ANR-19-P3IA-0001, PRAIRIE 3IA Institute and reference ANR-10-IAIHU-06, Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), from Agence Nationale de la Recherche (reference ANR-23-CE17-0054 - LYMP-PD), and was supported by the Fonds Recherche Neurosciences under the project "NEIMO - Monitoring Neuroinflammation: a Paris Brain Institute-Yale Neurology Collaborative program. Guanghui Fu is supported by a Chinese Government Scholarship provided by the China Scholarship Council (CSC). Lydia Chougar is supported by a Poste d'accueil from Inria and Assistance Publique-Hôpitaux de Paris (AP-HP). The ICEBERG study is supported by the European Research Council (ERC) under grant agreement No 678304, the European Union's Horizon 2020 research and innovation program under grant agreement No 826421 (TVB-Cloud), Agence Nationale de la Recherche (ANR) under grant agreements ANR-10-IAIHU-06 (IHU ICM), ANR-11-INBS-0006, and ANR-19-JPW2-000 (JPND E-DADS), association France Parkinson (PRECISE-PD project), the Fondation d'Entreprise EDF, Biogen Inc, Fondation Thérèse and René Planiol, Fondation Saint Michel. It received unrestricted support for Research on Parkinson's disease from Energipole (M. Mallart), M. Villain and the Société Française de Médecine Esthétique (M. Legrand).

References

- 1 M. H. Hesamian, W. Jia, X. He, *et al.*, “Deep learning techniques for medical image segmentation: achievements and challenges,” *Journal of digital imaging* **32**(4), 582–596 (2019).

- 2 N. Tajbakhsh, L. Jeyaseelan, Q. Li, *et al.*, “Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” *Medical Image Analysis* **63**, 101693 (2020).
- 3 R. El Jurdi, C. Petitjean, P. Honeine, *et al.*, “High-level prior-based loss functions for medical image segmentation: A survey,” *Computer Vision and Image Understanding* **210**, 103248 (2021).
- 4 O. Bernard, A. Lalande, C. Zotti, *et al.*, “Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?,” *IEEE Transactions on Medical Imaging* **37**(11), 2514–2525 (2018).
- 5 L. Liu, J. M. Wolterink, C. Brune, *et al.*, “Anatomy-aided deep learning for medical image segmentation: a review,” *Physics in Medicine & Biology* **66**(11), 11TR01 (2021).
- 6 H. Kervadec, J. Dolz, S. Wang, *et al.*, “Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision,” in *MIDL*, 365–381 (2020).
- 7 X. Chen, B. M. Williams, *et al.*, “Learning active contour models for medical image segmentation,” in *Proc. CVPR*, 11632–11640 (2019).
- 8 R. E. Jurdi, C. Petitjean, P. Honeine, *et al.*, “A surprisingly effective perimeter-based loss for medical image segmentation,” in *MIDL*, 158–167 (2021).
- 9 R. B. Gabrielsson, B. J. Nelson, A. Dwaraknath, *et al.*, “A topology layer for machine learning,” in *International Conference on Artificial Intelligence and Statistics*, 1553–1563, PMLR (2020).
- 10 J. R. Clough, N. Byrne, I. Oksuz, *et al.*, “A topological loss function for deep-learning based image segmentation using persistent homology,” *IEEE transactions on pattern analysis and machine intelligence* **44**(12), 8766–8778 (2020).
- 11 N. Byrne, J. R. Clough, G. Montana, *et al.*, “A persistent homology-based topological loss function for multi-class cnn segmentation of cardiac MRI,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 3–13, Springer (2020).
- 12 S. Shit, J. C. Paetzold, A. Sekuboyina, *et al.*, “clDice - a Topology-Preserving Loss Function for Tubular Structure Segmentation,” in *Medical Imaging Meets NeurIPS 2019 Workshop*, (2019).
- 13 H. Edelsbrunner, J. Harer, *et al.*, “Persistent homology-a survey,” *Contemporary mathematics* **453**, 257–282 (2008).
- 14 D. R. Williams and A. J. Lees, “Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges,” *The Lancet Neurology* **8**(3), 270–279 (2009).
- 15 E. Visser, M. C. Keuken, B. U. Forstmann, *et al.*, “Automated segmentation of the substantia nigra, subthalamic nucleus and red nucleus in 7 t data at young and old age,” *Neuroimage* **139**, 324–336 (2016).
- 16 T. Guo, Y. Song, J. Li, *et al.*, “Seed point discontinuity-based segmentation method for the substantia nigra and the red nucleus in quantitative susceptibility maps,” *Journal of Magnetic Resonance Imaging* **48**(4), 1112–1119 (2018).
- 17 B. Garzón, R. Sitnikov, L. Bäckman, *et al.*, “Automated segmentation of midbrain structures with high iron content,” *Neuroimage* **170**, 199–209 (2018).
- 18 W. Zhao, Y. Wang, F. Zhou, *et al.*, “Automated segmentation of midbrain structures in high-resolution susceptibility maps based on convolutional neural network and transfer learning,” *Frontiers in Neuroscience* **16** (2022).

- 19 V. Beliveau, M. Nørgaard, C. Birkl, *et al.*, “Automated segmentation of deep brain nuclei using convolutional neural networks and susceptibility weighted imaging,” tech. rep., Wiley Online Library (2021).
- 20 X. Li, L. Chen, K. Kutten, *et al.*, “Multi-atlas tool for automated segmentation of brain gray matter nuclei and quantification of their magnetic susceptibility,” *Neuroimage* **191**, 337–349 (2019).
- 21 M. Antonelli, A. Reinke, S. Bakas, *et al.*, “The medical segmentation decathlon,” *Nature communications* **13**(1), 4128 (2022).
- 22 G. Fu, R. El Jundi, L. Chougar, *et al.*, “Introducing soft topology constraints in deep learning-based segmentation using projected pooling loss,” in *SPIE Medical Imaging 2023*, (2023).
- 23 O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. MICCAI*, 234–241, Springer (2015).
- 24 F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *Proc. 3DV*, 565–571, IEEE (2016).
- 25 O. Oktay, E. Ferrante, *et al.*, “Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation,” *IEEE transactions on medical imaging* **37**(2), 384–395 (2017).
- 26 N. Paragios, “A level set approach for shape-driven segmentation and tracking of the left ventricle,” *IEEE Transactions on medical imaging* **22**(6), 773–776 (2003).
- 27 J. E. Iglesias and M. R. Sabuncu, “Multi-atlas segmentation of biomedical images: a survey,” *Medical image analysis* **24**(1), 205–219 (2015).
- 28 T. Heimann and H.-P. Meinzer, “Statistical shape models for 3d medical image segmentation: a review,” *Medical image analysis* **13**(4), 543–563 (2009).
- 29 O. Camara, G. Delso, O. Colliot, *et al.*, “Explicit incorporation of prior anatomical information into a nonrigid registration of thoracic and abdominal CT and 18-FDG whole-body emission PET images,” *IEEE transactions on medical imaging* **26**(2), 164–178 (2007).
- 30 J. Ma, J. Chen, M. Ng, *et al.*, “Loss odyssey in medical image segmentation,” *Medical Image Analysis* **71**, 102035 (2021).
- 31 H. Huang, H. Zheng, *et al.*, “Medical image segmentation with deep atlas prior,” *IEEE Transactions on Medical Imaging* **40**(12), 3519–3530 (2021).
- 32 R. Ghrist, “Barcodes: the persistent topology of data,” *Bulletin of the American Mathematical Society* **45**(1), 61–75 (2008).
- 33 H. Edelsbrunner, D. Letscher, and A. Zomorodian, “Topological persistence and simplification,” in *Proceedings 41st annual symposium on foundations of computer science*, 454–463, IEEE (2000).
- 34 C. S. Pun, S. X. Lee, and K. Xia, “Persistent-homology-based machine learning: a survey and a comparative study,” *Artificial Intelligence Review* **55**(7), 5169–5213 (2022).
- 35 X. Hu, F. Li, D. Samaras, *et al.*, “Topology-preserving deep image segmentation,” *Advances in Neural Information Processing Systems* **32** (2019).
- 36 N. Byrne, J. R. Clough, I. Valverde, *et al.*, “A persistent homology-based topological loss for CNN-based multi-class segmentation of CMR,” *IEEE transactions on medical imaging* **42**(1), 3–14 (2022).

- 37 A. Santhirasekaram, M. Winkler, A. Rockall, *et al.*, “Topology preserving compositionality for robust medical image segmentation,” in *Proc. CVPR*, 543–552 (2023).
- 38 A. Mosinska, Marquez-Neila, *et al.*, “Beyond the pixel-wise loss for topology-aware delineation,” in *Proc. CVPR*, 3136–3145 (2018).
- 39 J. Deng, W. Dong, R. Socher, *et al.*, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, 248–255, Ieee (2009).
- 40 A. BenTaieb and G. Hamarneh, “Topology aware fully convolutional networks for histology gland segmentation,” in *Proc. MICCAI*, 460–468, Springer (2016).
- 41 Y. He, A. Carass, Y. Yun, *et al.*, “Towards topological correct segmentation of macular OCT from cascaded FCNs,” in *Fetal, Infant and Ophthalmic Medical Image Analysis*, M. J. Cardoso, Ed., 202–209, Springer (2017).
- 42 X. Hu, Y. Wang, L. Fuxin, *et al.*, “Topology-aware segmentation using discrete morse theory,” *arXiv preprint arXiv:2103.09992* **0** (2021).
- 43 R. Forman, “Morse theory for cell complexes,” *Advances in mathematics* **134**(1), 90–145 (1998).
- 44 Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, *et al.*, “3d U-Net: learning dense volumetric segmentation from sparse annotation,” in *Proc. MICCAI*, 424–432, Springer (2016).
- 45 D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980* **0** (2014).
- 46 F. Pérez-García, R. Sparks, and S. Ourselin, “TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,” *Computer Methods and Programs in Biomedicine* **208**, 106236 (2021).
- 47 A. Paszke, S. Gross, F. Massa, *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” *Proc. NeurIPS* **32** (2019).
- 48 L. De Rochefort, T. Liu, *et al.*, “Quantitative susceptibility map reconstruction from MR phase data using bayesian regularization: validation and application to brain imaging,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **63**(1), 194–206 (2010).
- 49 H. Sjöström, T. Granberg, E. Westman, *et al.*, “Quantitative susceptibility mapping differentiates between parkinsonian disorders,” *Parkinsonism & related disorders* **44**, 51–57 (2017).
- 50 E. Thibeau-Sutre, M. Diaz, R. Hassanaly, *et al.*, “ClinicaDL: an open-source deep learning software for reproducible neuroimaging processing,” *Computer Methods and Programs in Biomedicine* **220**, 106818 (2022).
- 51 Y. Xiao, P. Jannin, T. d’Albis, *et al.*, “Investigation of morphometric variability of subthalamic nucleus, red nucleus, and substantia nigra in advanced parkinson’s disease patients using automatic segmentation and PCA-based analysis,” *Human brain mapping* **35**(9), 4330–4344 (2014).
- 52 J. V. Manjón, A. Bertó, J. E. Romero, *et al.*, “pBrain: A novel pipeline for parkinson related brain structure segmentation,” *NeuroImage: Clinical* **25**, 102184 (2020).
- 53 R. El Jurdi and O. Colliot, “How precise are performance estimates for typical medical image segmentation tasks?,” in *Proc. IEEE ISBI*, (2023).
- 54 D. Karimi and S. E. Salcudean, “Reducing the hausdorff distance in medical image segmentation with convolutional neural networks,” *IEEE Transactions on medical imaging* **39**(2), 499–513 (2019).

- 55 T. Sugino, T. Kawase, S. Onogi, *et al.*, “Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks,” in *Healthcare*, **9**(8), 938, MDPI (2021).
- 56 H. Kervadec, J. Bouchtiba, C. Desrosiers, *et al.*, “Boundary loss for highly unbalanced segmentation,” in *International conference on medical imaging with deep learning*, 285–296, PMLR (2019).
- 57 F. Milletari, S.-A. Ahmadi, C. Kroll, *et al.*, “Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound,” *Computer Vision and Image Understanding* **164**, 92–102 (2017).
- 58 J. Dolz, C. Desrosiers, and I. B. Ayed, “3d fully convolutional networks for subcortical segmentation in MRI: A large-scale study,” *NeuroImage* **170**, 456–470 (2018).