GNN-based structural information to improve DNN-based basal ganglia segmentation in children following early brain lesion

Patty Coupeau, Jean-Baptiste Fasquel, Lucie Hertz-Pannier, Mickael Dinomais

To cite this version:

Patty Coupeau, Jean-Baptiste Fasquel, Lucie Hertz-Pannier, Mickael Dinomais. GNN-based structural information to improve DNN-based basal ganglia segmentation in children following early brain lesion. Computerized Medical Imaging and Graphics, 2024, 115, pp.102396. 10.1016/j.compmedimag.2024.102396 . hal-04575146

Highlights

GNN-based structural information to improve DNN-based basal ganglia segmentation in children following early brain lesion

Patty Coupeau,Jean-Baptiste Fasquel,Lucie Hertz-Pannier,Mickaël Dinomais

- Due to their key role in brain function, the basal ganglia are often studied in children. However, they are difficult to study, especially after early brain injury, because current deep neural networks (DNN) cannot accurately segment them on MRI.
- A graph neural network is applied on top of a DNN to enhance the segmentation of the basal ganglia in healthy children and children suffering from neonatal arterial ischemic stroke (NAIS), using both the structural information and the preliminary segmentation map generated by the DNN.
- The structural information conveyed by graph edges (spatial relationships between regions detected by the DNN) and nodes (region probabilities from the DNN) can help to correct spatial inconsistencies produced by the DNN in its segmentation of injured brains.
- The method is generic regarding the nature of the deep neural network under consideration for preliminary segmentation.
- The method is efficient on small training datasets, addressing the issue of limited medical imaging data. Furthermore, it outperforms recent deep learning methods for basal ganglia segmentation on larger datasets, specifically in injured children.

GNN-based structural information to improve DNN-based basal ganglia segmentation in children following early brain lesion

Patty Coupeau[a,*], Jean-Baptiste Fasquel[a], Lucie Hertz-Pannier[b] and Mickaël Dinomais[a,c]

[a] *Universite d'Angers, LARIS, SFR MATHSTIC, F-49000 Angers, France*
[b] *UNIACT/Neurospin/JOLIOT/DRF/CEA-Saclay, and U1141 NeuroDiderot/Inserm, CEA, Paris University, France*
[c] *Departement de medecine physique et de readaptation, Centre Hospitalier Universitaire d'Angers, France*

ARTICLE INFO

Keywords:
basal ganglia
early brain lesion
semantic segmentation
structural information
graph neural network

ABSTRACT

Analyzing the basal ganglia following an early brain lesion is crucial due to their noteworthy role in sensory-motor functions. However, the segmentation of these subcortical structures on MRI is challenging in children and is further complicated by the presence of a lesion. Although current deep neural networks (DNN) perform well in segmenting subcortical brain structures in healthy brains, they lack robustness when faced with lesion variability, leading to structural inconsistencies. Given the established spatial organization of the basal ganglia, we propose enhancing the DNN-based segmentation through post-processing with a graph neural network (GNN). The GNN conducts node classification on graphs encoding both class probabilities and spatial information regarding the regions segmented by the DNN. In this study, we focus on neonatal arterial ischemic stroke (NAIS) in children. The approach is evaluated on both healthy children and children after NAIS using three DNN backbones: U-Net, UNETr, and MSGSE-Net. The results show an improvement in segmentation performance, with an increase in the median Dice score by up to 4% and a reduction in the median Hausdorff distance (HD) by up to 93% for healthy children (from 36.45 to 2.57) and up to 91% for children suffering from NAIS (from 40.64 to 3.50). The performance of the method is compared with atlas-based methods. Severe cases of neonatal stroke result in a decline in performance in the injured hemisphere, without negatively affecting the segmentation of the contra-injured hemisphere. Furthermore, the approach demonstrates resilience to small training datasets, a widespread challenge in the medical field, particularly in pediatrics and for rare pathologies.

1. Introduction

The basal ganglia are a set of deep gray matter nuclei that include the globus pallidus (or pallidum) and the striatum, subdivided into three subnuclei: the caudate, the putamen and the nucleus accumbens [1]. The caudate and putamen receive corticostriatal inputs related to the motor pathway, whereas the nucleus accumbens receives inputs from the emotional and reward pathways [2]. The pallidum is part of the subcortical nervous circuitry that is involved in motor control. It transmits information from the striatum, which originates in the motor cortex, to the thalamus [3]. Research on the cortico-basal ganglia-thalamic loop has largely viewed the thalamus as a relay that conveys basal ganglia output to the cerebral cortex, thereby enabling the control of movement [4]. In the context of the study of motor development in children, the basal ganglia (especially the caudate, putamen, and pallidum) and thalamus therefore play a crucial role in neuroimaging studies and clinical practice. This includes the diagnosis of brain diseases [5] and the planning of surgery [6]. Therefore, examining the basal ganglia and thalami in young children is especially important, given their key role in brain function [1] and growth after birth [7]. Consequently, these subcortical brain structures are frequently studied [5, 8], yet the challenge persists in automatically segmenting them from MRI. This is due to the difficulty of distinguishing ambiguous voxels near the boundaries of brain structures, which is further compounded by the low tissue contrast of brain MRI [9]. This difficulty is particularly pronounced in young children [10].

In recent years, deep learning techniques have achieved remarkable success in different tasks of image segmentation [11]. In the field of brain MRI segmentation, deep neural networks (DNN) have proven to be an effective alternative to atlas-based segmentation methods. As a result, several DNN-based algorithms [9, 12, 13] as well as automated software [14, 15] have been developed to segment subcortical brain structures. Although these algorithms have

---

*Corresponding author
✉ `patty.coupeau@univ-angers.fr` (P. Coupeau)
ORCID(s):

enhanced the accuracy of basal ganglia segmentation, they are not efficient enough to handle injured brains. In fact, a brain lesion induces alterations, such as variations in intensity and changes in the overall brain structure, causing artifacts and misclassifications when using existing DNN-based techniques. Recent work has focused on the segmentation of pathological and damaged brains. Thus, the patch-based method *MONSTR* [16], involving atlases of different MR acquisition protocols (e.g., T1-w, T2-w), is proposed to automatically skull-strip adult brains with mild to severe traumatic brain injury or brain tumors. Other works have suggested the use of a modified 3D U-Net to perform ventricular parcellation from MRI in adult brains with hydrocephalus, a disease which causes dementia [17]. Some methods have focused on the segmentation of brain structures in brains with traumatic brain injury, such as hippocampus segmentation in adult rats [18] or ROI parcellation in adult humans with a multi-atlas-based method [19]. However, no effective method has been proposed for the segmentation of the basal ganglia and thalamus in children with a cortical and localized lesion. Some studies for brain structures segmentation have addressed white matter lesions [15], even in children, but they have not proven their effectiveness for cortical lesions such as strokes. Therefore, manual segmentation of the basal ganglia is currently required in cases of brain injury (especially stroke), as evidenced by the work of Kirton et al. [8]. This task is time-consuming and tedious.

This paper proposes a method to improve basal ganglia segmentation in healthy and injured brains, complementing existing DNN-based algorithms. The study focuses on a population of 7-year-old children, some of whom suffer from neonatal arterial ischemic stroke (NAIS), the most common type of perinatal ischemic stroke [20]. The population selection is motivated by the aim to illustrate the method on a relatively complex and representative example (children with perinatal stroke). The subsequent stage of the study will examine the method's generalizability to older subjects and other types of early brain lesion.

The proposed method is based on the established uniform spatial organization of the basal ganglia and thalamus [1]. In fact, the basal ganglia are in the center of the brain. The caudate is characterized by its curved shape, with the head, body and tail moving from front to back. The putamen and pallidum are at the junction of the diencephalon and telencephalon and are characteristically banana-shaped in horizontal section. The two substructures together have been given the name lenticular core because of their triangular shape on a frontal section. It is also recognized that there is a substantial reduction in the volume of structures moving from the cortex (e.g., putamen) towards the deepest nuclei (e.g., pallidum) [21]. These structures are symmetrical in each hemisphere. The thalamus is a large, oval-shaped nucleus that is paired and symmetrical on either side of the third ventricle. It is connected to the caudate nucleus above, the head of the caudate nucleus in front, and the tail of the caudate nucleus behind. The current DNNs do not make explicit use of the structural organization of the basal ganglia and thalamus. In fact, the structural information assimilated by DNN is restricted to the receptive field dimension during convolution, despite attempts by recent models such as transformers [22] to extend beyond this constraint. Thus, segmenting injured brains using current DNN-based methods can sometimes lead to spatial inconsistencies. We propose the use of high-level structural information, which in our context corresponds to spatial relationships between regions (e.g., distances, relative directional position [23, 24]), to improve DNN-based segmentation of the basal ganglia by correcting for these spatial inconsistencies.

This type of high-level information is commonly represented with graphs, where nodes correspond to regions and edges carry the structural information between regions. The aim is to match each node of the graph with the corresponding class, i.e., the brain structure in our context. In a recent work, the combination of high-level structural information with the DNN output was proposed as a means of improving segmentation performance [25]. However, the approach relied on standard graph matching techniques (e.g., quadratic assignment problems [26]), which can be highly combinatorial in nature. To address this limitation, we propose using a graph neural network (GNN) to perform a node classification task [27] (i.e., classification of regions detected by the DNN) instead of graph matching approaches. When combining DNN and GNN for semantic image segmentation, there are numerous options available. Some proposals exploit superpixel information for graph construction, requiring additional over-segmentation algorithms such as SLIC [28] to generate superpixels [29, 30]. Other recent studies have attempted to incorporate spatial relationships through the use of a GNN in the feature space of neural networks [31, 32, 33, 34] to capture long-range spatial dependencies. We opted for an alternative approach, whereby the GNN acts as a post-processing layer that can be applied to any DNN that generates a segmentation map, without requiring a specific DNN-based architecture (as in [31, 32, 33, 34]) or the integration of additional over-segmentation algorithms to generate superpixels (as in [29, 30]). In a previous study, we demonstrated the enhancement of DNN-based segmentation by a GNN-based post-processing approach that combines DNN probabilities with structural information [23]. However, this was in a different context, namely face segmentation, and in a nonmedical domain.

The efficacy of the GNN for node classification is contingent upon the nature of the convolution operator considered. In recent years, numerous operators have been developed [35]. Graphs embed structural information both in the graph structure and in the edge attributes. Regarding edge attributes, n-dimensional information management is important [36, 37]. Thus, we have decided to consider the ECConv operator [36], which handles both n-dimensional node and edge attributes. In addition to edge attributes, graph architecture is an essential aspect [38]. Due to the permutation-invariant nature of convolution operators during the message passing (neighborhood aggregation), it may not be optimal to work on complete graphs (resulting from using all relations between all regions) to distinguish nodes for classification purposes. To tackle this problem while still considering complete graphs that regroup all available structural information between regions, we propose a function within the convolution operator itself to modify the weight of each edge attribute in the calculation of the new node attributes, in comparison to our previous work [23].
Moreover, our proposal aims to enhance the semantic segmentation of the basal ganglia based on DNN, particularly when the training dataset is sparse. The lack of medical imaging data is particularly important in pediatrics, where the frequent movement of the child during the MRI scan makes analysis impossible. Data is even more limited when addressing infrequent pathologies such as early brain lesion. Thus, this study also partly addresses a key limitation of deep learning: the need for a large and representative dataset for training purposes. The objective is to propose a method based on a GNN with few trainable parameters. This method should be efficient despite a limited amount of training data and capable of correcting DNN errors that are associated with the lack of training data. To evaluate this, we aim to investigate the performance of our method as the size of the training dataset decreases. It should be noted that a data augmentation strategy [39] is used, in addition to the proposed method, when training DNNs to address the lack of data. Other approaches, such as transfer learning [40] or few-shot learning [41], could have been considered. However, they require the definition of a suitable source domain with sufficient data available to pre-train the model, or the definition of the right examples to provide to the model.
After detailing the proposal in Section 2, the experiments are described in Section 3 and discussed in Section 4.

2. Method

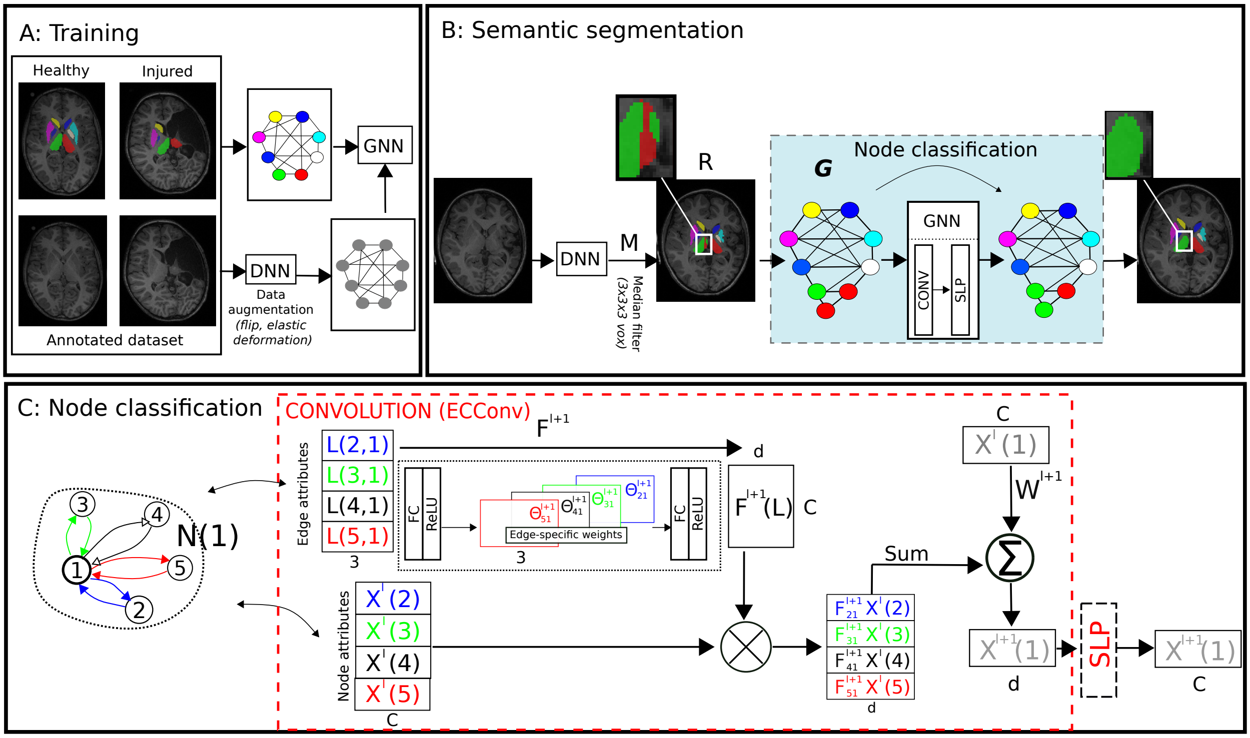

Figure 1: Overview of the proposed approach. For the sake of clarity, not all edges of the complete graphs are shown, nor are the biases of the convolution operator. DNN: Deep neural network, GNN: Graph neural network.

Figure 1 presents an overview of the proposed approach. A deep neural network (DNN) is trained to segment the basal ganglia using an annotated dataset (Figure 1-A) consisting of healthy and injured brains. Any deep neural network can be considered for this initial segmentation, as long as it provides a segmentation map (M). Regardless of the DNN under consideration, a data augmentation strategy is employed during the training phase (detailed in section 3.3). A complete graph is constructed from the segmentation map of the training data, as detailed in section 2.1. Each node in the graph corresponds to a region segmented by the DNN. The annotated dataset is used to train a graph neural network (GNN) to correctly assign the class associated with each graph node (i.e., each region segmented by the DNN) of the graphs constructed from the segmented training images. Further details regarding the GNN considered are provided in section 2.2.

Upon analysis of a novel MRI scan (Figure 1-B), the DNN provides a segmentation proposal M, which may contain errors (e.g., left thalamus within right thalamus). From this segmentation map M, a set of connected components is generated. A 3^3 median kernel filter is applied to eliminate small, connected components that are distant from the basal ganglia, leading to the updated set of connected components R. A complete graph is constructed based on this set of connected components R, as described in section 2.1. Subsequently, the trained graph neural network is used to perform node classification relying on the attributes of both nodes and edges, by using the ECConv graph convolution operator described in section 2.2 (Figure 1-C). Based on this node classification, spatially consistent regions are finally identified.

In section 2.1, the construction of the graphs from the MRI is detailed, while section 2.2 outlines the architecture of the considered GNN.

2.1. Image and graph

When segmenting an image, the deep neural network generates a segmentation map represented as a tensor $M \in \mathbb{R}^{P \times C}$, where P is the dimensionality of the 3D image and C is the total number of classes. At each voxel location p, the value $M(p, c) \in [0, 1]$ denotes the probability that the voxel p belongs to class $c \in \{1, \ldots, C\}$. These probabilities enable the creation of a set R comprising all resulting connected components (i.e., the set of connected voxels that a priori belong to the same class based on $M(p, c)$, as illustrated in Figure 1). From this set R, we construct a graph $G = (V, E, X, L)$, where V is the set of nodes (each $v \in V$ corresponds to a region $R_v \in R$) and E is the set of edges. X refers to the node attribute assignment function ($X : V \rightarrow \mathfrak{R}^C$) related to the average membership probability vector over the set of voxels $p \in R_v$. L is an edge attribute assignment function ($L : E \rightarrow \mathfrak{R}^3$) assigning to edge (i, j) the relative position (in mm) of the barycenters of the connected regions R_i and R_j in the three dimensions (left-right, anterior-posterior, inferior-superior), normalized by the child's head circumference D (in mm):

$$L((i, j)) = [\frac{d_x^{R_i,R_j}}{D}, \frac{d_y^{R_i,R_j}}{D}, \frac{d_z^{R_i,R_j}}{D}] \tag{1}$$

The normalization by D in $L((i, j))$ ensures that the edge attributes are of a comparable magnitude to the node attributes, which typically range between 0 and 1. This normalization is crucial for the neighborhood aggregation later performed by the GNN (eq.2). Normalizing the relative position (in mm) by the head circumference accommodates MRIs of varying dimensions or resolutions while accounting for differences in brain size. The graph G is directed with $L(j, i) = -L(i, j)$.

2.2. Graph neural network

As depicted in Figure 1, the GNN consists of two layers. The first layer performs convolution, while the second layer assigns a membership probability vector to each node. Importantly, the number of nodes to be classified is a priori unknown and exceeds the number of classes due to the multiple candidate regions generated by the DNN. This leads to a realistic hypothesis of over-segmentation [42]. Consequently, graphs can have arbitrary sizes.

The first layer comprises a convolution ($\mathfrak{R}^C \rightarrow \mathfrak{R}^d$) aimed at aggregating the neighborhood information related to each node (general notion of message passing [27]). The dimension d, which represents the attribute dimensions of nodes after convolution (as shown in Figure 1 - C), is a hyperparameter of the method explored in our experiments. We selected the ECConv convolution operator [36] due to its spatial-based nature, which allows it to handle graphs of arbitrary sizes better than spectral approaches (spectral graph theory) [43]. Furthermore, ECConv considers both node and edge attributes during neighborhood aggregation. Its efficacy in enhancing semantic segmentation tasks has been demonstrated previously [23, 37]. For a given node $i \in V$, this layer computes a new attribute at layer $l + 1$ (yielding

$X^{l+1}(i)$), by combining different information from layer l : the attribute $X^l(i)$ of node i, the attributes of the set $N(i)$ of neighboring nodes ($N(i) = \{j|(j,i) \in E\}$), and the attributes of the set of related edges (i.e. set $\{L((j,i))|j \in N(i)\}$). It can be formulated as follows:

$$X^{l+1}(i) = \sigma(W^{l+1}X^l(i) + \sum_{j \in N(i)} F^{l+1}(L(j,i))X^l(j) + b^{l+1}) \tag{2}$$

where σ denotes a component-wise non-linear function (e.g. ReLU), b^{l+1} is a bias, $W^{l+1} \in \Re^{d \times C}$ represents a matrix of trainable weights, and F^{l+1} is a differentiable function (a multi-layer perceptron, detailed in Figure 1-C). All of these entities are learned through training.

The mapping function $F^{l+1} : \Re^{3 \times |N(i)|} \to \Re^{d \times C}$ is the most important entity. This function consists of two fully connected layers, each followed by a ReLU activation function. The first connected layer, denoted as $F_1^{l+1} : \Re^{3 \times |N(i)|} \to \Re^{3 \times |N(i)|}$, adjusts the weights of each attribute for every related edge in the computation of the new node attribute. Subsequently, the second layer, $F_2^{l+1} : \Re^{3 \times |N(i)|} \to \Re^{d \times C}$, maps these edge-specific weights to the node dimensions. Note that, while theoretically supporting any F function, our preliminary work on ECConv [23] has primarily focused on the F_2 mapping function. This paper introduces the preliminary F_1 NN-compliant function, similar to that in [36], which acts as an edge-cutting operator [38] to mitigate the influence of irrelevant edges during the mapping process (F_2). The structural information carried by the edges $L((i,j))$ is thus harnessed by the differentiable function F^l described earlier, leading to $F^{l+1}(L(j,i)) \in \Re^{d \times C}$ (Figure 1-C). Then, the edge-conditioned operator handles the combination of the information embedded by neighboring nodes (weighted region properties) with that embedded by edges (weighted relationships between regions), through the product $F^{l+1}(L(j,i))X^l(j)$. The resulting matrix from the product is then combined with the information of the studied node $X^l(i)$. The sum operator, which is a permutation-invariant operator (as required in this context [27]), is used to compute the new representation $X^{l+1}(i)$ (Figure 1-C). The second layer of the GNN consists of a single layer perceptron ($\Re^d \to \Re^C$) that assigns a class membership probability vector to each node in the graph.

The network parameters are optimized on the training dataset to maximize the node classification rate. This corresponds to minimizing the negative log-likelihood loss function: $\text{Loss}(Y, \hat{y}) = -\sum_{n=1}^{N} \sum_{c=1}^{C} Y_{n,c} \times \log(\hat{y}_{n,c})$ where N is the number of nodes, C the number of classes, $Y_{n,c}$ the actual class of node n (1 if node n belongs to class c, 0 otherwise), and $\hat{y}_{n,c}$ the probability that node n belongs to class c. Based on the node classification provided by the GNN, the voxels corresponding to the region associated with each node are assigned the value of the predicted class. This updates the tensor map M, which is the output of the DNN.

3. Experiments

In this section, we introduce the dataset (Section 3.1), the evaluation protocol (Section 3.2), the DNNs utilized for preliminary segmentation along with the graph construction (Section 3.3), and conclude with the results (Section 3.4).

3.1. Dataset

Our method is evaluated in a cohort of 69 children aged 7 years, comprising both healthy children and children after NAIS with either right or left hemisphere unilateral injury. Each child underwent MRI scanning on a 3.0 Tesla scanner (MAGNETOM Trio Tim system, Siemens, Erlangen, Germany, 12 channel head coil) at Neurospin, CEA-Saclay, France. High-resolution 3D T1-weighted volumes were acquired using a magnetization-prepared rapid acquisition gradient-echo sequence [176 slices, TR = 2300 msec, TE = 4.18 msec, FOV = 256 mm, flip angle = 9°, voxel size 1x1x1mm^3]. Imaging sequences also included a high-resolution 3D FLAIR sequence, which was solely used to guide manual segmentation of the basal ganglia in lesioned areas during ground truth definition. The head circumference of all children was measured by clinicians at the time of MRI acquisition.

Major nuclei within the basal ganglia complex include the putamen, the caudate nucleus, and the pallidum. To generate ground truths for segmenting these structures, along with the thalamus, on the 3D T1 MRI scans, we combined manual segmentation with atlas-based segmentation using the Hammersmith atlas [44]. The Hammersmith atlas, available for download[1], has been widely used in publications and research studies focusing on children [45]. Statistical Parametric Mapping version 12 (SPM12) software (Welcome Department of Imaging Neuroscience, University College, London,

---

[1] http://brain-development.org/brain-atlases/pediatric-brain-atlases/pediatric-brain-atlas-gousias/

UK) was used to obtain the tissue probability map (TPM) for each control subject (T1 MRI) in the native space. A group template for 7-year-old control children was constructed from their TPMs using the DARTEL tool. Basal ganglia and thalamus segmentation in the MNI space was performed using the Hammersmith atlas. The T1-weighted volume of each subject was registered to the DARTEL template of 7-year-old children using the Computational Anatomy Toolbox CAT12 (version 12.7) [46]. This process yielded the deformation fields from each subject's native space to the DARTEL template space, which were necessary for obtaining basal ganglia and thalamus segmentations in each child's native space. The segmentation of each MRI was manually corrected slice-by-slice using ITKSnap [47], with the aid of the FLAIR sequence to assist in identifying gliosis for segmenting the injured areas. Each segmentation was visually checked and validated by consensus with a specialist. The segmentation comprised nine classes: thalamus, caudate, putamen, pallidum (both left and right in all cases), and background.

One child was excluded from the study due to excessive movement during MRI acquisition, which compromised the quality of the acquired 3D T1-weighted volume, making segmentation insufficiently reliable. Consequently, the final dataset consists of 68 segmented T1 MRIs from 7-year-old children, comprising 31 healthy and 37 injured children as detailed in Table 1 - data.

3.2. Evaluation protocol

All experiments were conducted in a Python environment using the PyTorch Geometric library [48]. The graph neural network model was trained with the Adam optimizer for 600 epochs, with a dropout rate of 0.5. A learning rate reduction strategy on plateau was employed, starting with an initial learning rate $L_{r0} = 0.01$ and a reduction factor $\iota = 5e^{-4}$. The output dimension d of the graph convolution operator was set to 12. The performance of the model was evaluated as a function of this hyperparameter value.

To assess the segmentation quality, we computed the Dice score [49] and Hausdorff distance (HD) [50] for each class. While the Dice provides a general measure of overlap between the segmented and reference regions, the HD emphasizes cases where the segmented region may include small connected components located far from the reference annotations. In addition to the Hausdorff distance, we computed the 95^{th} percentile of the Hausdorff distance, measured in millimeters and denoted as HD95, using the *MedPy* library [51]. This metric was used to mitigate the impact of a small subset of outliers. The reported results are the averages of these metrics across the nine classes considered. Furthermore, substructure-wise performance was analyzed to assess the ability of the method to improve the semantic segmentation of each individual structure, independent of the others.

We further evaluated the added value of our method by comparing the segmentation results from the GNN with those obtained using only the DNN. To assess the efficiency of the method in accurately segmenting the basal ganglia in both healthy and injured brains, we differentiated the results obtained for each population. Additionally, we measured the segmentation quality per hemisphere in injured children, comparing the lesioned and contra-lesioned hemispheres. Checking segmentation separately in each hemisphere is important for conducting subsequent inter-hemispheric studies based on the segmentation results [8, 52].

To ensure robustness across different brain types, we trained and tested both the DNN and GNN on datasets comprising both healthy and injured brains (Table 1). To evaluate the resilience of the method to small datasets, a challenging issue in medical applications [53], we experimented with varying sizes of the training dataset. Specifically, we considered four configurations outlined in Table 1, referring to the proportion of the training dataset used (100%, 50%, 25%, 12%). For each training dataset size, the test set remained consistent (composed of 11 healthy subjects and 12 subjects with neonatal stroke). However, the subjects used for training and validation were randomly drawn from the remaining 45 children (healthy and after NAIS). To evaluate the generalizability of the DNN, two independent random draws of training and validation data were considered for each configuration (100%, 50%, 25% or 12%). The validation set is used to provide an unbiased evaluation of the DNN's performance while tuning hyperparameters. The reported results in this paper are the average performance across these two random draws.

To evaluate the impact of the structural information carried by edges, we compared our performance with that obtained without edge attributes (i.e. considering a unitary weight on edges). Additionally, we benchmarked our approach against results obtained using the widely used graph convolution operator GraphConv [54], which handles only scalar edge attributes (set to a unitary weight in our experiments).

In our experiments, we considered a 3D U-Net and two state-of-the-art DNNs (detailed hereafter), selected from recent studies that provide available implementations suitable for 3D medical images. Furthermore, we compared the performance of our proposal with two atlas-based methods: the fully automatic *vol2Brain* algorithm, which leverages

Table 1
Dataset size and training (validation)/test configurations (Conf.). The number of healthy and injured subjects in the training (validation) dataset is given for the two random selections of the training dataset.

	data	train (validation)				test
Conf.		100%	50%	25%	12%	
Total	68	35 (10)	17 (6)	8 (4)	4 (2)	23
Healthy	31	18 (2)	6 (3)	3 (1)	3 (0)	11
		16 (4)	10 (1)	1 (2)	2 (2)	
Injured	37	17 (8)	11 (3)	5 (3)	1 (2)	12
		19 (6)	7 (5)	7 (2)	2 (0)	

multi-atlas label fusion technology [15], and the atlas of 7-year-old children constructed from the control children of our dataset.

3.3. Graph construction from DNN output

We evaluated the proposed post-processing method with various DNN architectures. The evaluation encompassed a 3D U-Net combined with CRF [55], the recently introduced UNETr using transformers [22], and the MSGSE-Net [9], which has been shown to perform well in segmenting basal ganglia in healthy brains through multiscale image contexts and attention mechanisms.

We chose U-Net as it is a benchmark deep neural network for medical image segmentation [56], and employed a PyTorch implementation[2]. The network was trained over 60 epochs, with an early stopping strategy to prevent overfitting. A CRF-based post-processing is considered, acting as a spatial regularization technique, to remove some of the artifacts produced by the U-Net, particularly with limited training data (with 25% and 12%). The aim was to demonstrate the ability of the method to correct segmentation errors not addressed by standard spatial regularization methods such as CRF. For the CRF model, we considered the *crfseg* library[3] in PyTorch, based on a Gaussian filter of size 11. We set the number of iterations in mean field approximation to 5 and initialized the smoothness kernel weight and bandwidths for each spatial feature in the Gaussian smoothness kernel to 1. The CRF model was placed at the output of the pre-trained U-Net and trained with Adam over 30 epochs to fine tune the whole model. Given the highly unbalanced classes (i.e., small target region sizes compared to other brain tissue and background), we used a 3D patch-based technique [57]. Patches are volumes of size 32^3 voxels extracted around the centroid of each label (random selection) using the *Torchio* library [58]. For each MR image, 64 patches were selected, with selection frequency proportional to the inverse prior probability of the corresponding class. During inference, the segmentation map for each MRI was generated using a grid sampler based on 32^3 voxels patches with an overlap of 4^3 voxels.

We also evaluated our method on the recent UNETr architecture, specifically designed for medical image segmentation. UNETr incorporates attention mechanisms, enabling it to capture global context and model long-range dependencies [22]. The hyperparameters were set according to [22], with patches of size 32^3 voxels. To reduce computational time and memory requirements, the 256x256x176 voxel MRIs were downsampled to 128x128x128 voxel images. During inference, each MRI was similarly downsampled to a 128x128x128 voxel image. We then followed the process described in [22], adapting it for patches of size 32^3 voxels. The resulting segmentation map later used for graph construction was thus a volume of 128x128x128 voxels. After segmentation with our method, the images were resized back to their original dimensions using 3^{rd} order spline interpolation for segmentation evaluation.

Finally, we considered the MSGSE-Net network, recently proposed for MRI-based segmentation of subcortical brain structures in healthy subjects. We chose this network due to its demonstrated superior performance in segmenting basal ganglia in healthy adult brains [9] outperforming other state-of-the-art methods like Free Surfer [14], BrainSegNet [12], and Ψ-Net [13]. MSGSE-Net, designed for 2D patches, leverages multiscale image contexts and attention mechanisms to improve the ability of the network to learn discriminative feature representations for accurate segmentation. The network was trained with Adam over 60 epochs, using the hyperparameters and the loss function (entropy-weighted Dice loss function) described in [9], applied to the 256x256 pixels slices of the MRI. During inference, the whole segmentation map was generated by aggregating the segmentation maps of each 256x256 pixel 2D slice constituting the 3D MRI.

---

[2] https://doi.org/10.5281/zenodo.3522306

[3] https://github.com/mishgon/crfseg

Table 2
Mean number of nodes for the different DNNs and training configurations.

Training data ratio	100%	50%	25%	12%
U-Net+CRF	11.57	15.39	20.39	18.78
UNETr	9.10	10.96	12.04	27.91
MSGSE-Net	12.04	15.43	16.26	23.78

Note that, in all cases, the whole volume was normalized during a pre-processing step. To compensate for the scarcity of training data, we implemented a data augmentation strategy that included a random flip along the left-right axis (probability of 0.2) and a random elastic deformation with five control points and linear image interpolation (probability of 0.2) to generate "realistic" data. Data augmentation was also performed using the *Torchio* library.

From the DNN output, a graph G was constructed as detailed in section 2. Prior to graph construction, a median filter with a 3^3 voxel kernel was applied to the DNN output to perform a smoothing aimed at eliminating small connected components and reducing the number of nodes in the graph. A 26-connectivity was considered to extract connected components from the smoothed segmentation map generated by the DNN. Each node was associated with an attribute of dimension 9, corresponding to the DNN-based probabilities of belonging to the 9 classes. Each edge has a three-dimensional attribute that corresponds to the relative positions of the barycenters of the connected regions across the three dimensions, as described in section 2.

3.4. Results

Table 2 reports the graph sizes for each neural network architecture across the four training configurations. As expected, we face arbitrary graph sizes. Specifically, the number of nodes tends to increase as the training dataset size decreases because the DNN becomes less efficient and identifies more candidate regions (i.e., nodes) per class.
Table 3 compares the segmentation performance with the DNN alone and with the proposed method. Bold values highlight the best method. For instance, with 50% of the training data, our proposal with MSGSE-Net outperforms the use of MSGSE-Net alone (shown in bold for our proposal). We present results for all three neural networks across the four training configurations. Whatever the DNN and the training configuration, the proposal consistently improves both the median Dice and the median Hausdorff distance (HD and HD95). Improvement with GNN, illustrated in Figure 2, results in a slight enhancement in median Dice (up to 4% with MSGSE-Net for both healthy and injured children when considering 12% of the training data). The improvement is particularly notable for the Hausdorff distance, (including HD95). This underscores the ability of the method to remove artifacts generated by the DNN that are either distant from or close to the target regions. Indeed, even with UNETr, which provides fewer artifacts due to its attention mechanisms (capturing long-range spatial dependencies and thereby avoiding distant artifacts - Table 2), our proposal demonstrably reduces the mean HD from 4.19 to 3.85 (average of healthy and injured children) when considering 100% of the training dataset (a 9% reduction), from 5.39 to 4.45 (an 18% reduction) with 50% of the training dataset, from 11.41 to 8.33 (a 27% reduction) when considering 25% of the training dataset, and from 23.44 to 10.14 (a 57% reduction) with only 12% of the training dataset. Even when dealing with a large lesion close to the motor pathway, where basal ganglia segmentation tends to be less accurate, our method outperforms the DNN in Dice, HD and HD95 (Figure 3 - correction of spatial inconsistencies in the injured hemisphere).
In Table 3, we observe that the performance of the deep neural network decreases with the reduction of the number of training data in terms of Dice and Hausdorff distance (visually observed in Figure 2). Nevertheless, our GNN-based post-processing appears to mitigate this decline, maintaining a relatively consistent and small median HD95 across all training configurations, as illustrated in Figure 4. As the number of training data decreases, the improvement provided by our method becomes more pronounced (see Figure 2). This underscores the efficacy of our method in addressing the challenges posed by limited training data: the relationships learned during training enable corrections of structural mistakes made by the DNN.
We also evaluated the potential progress in GNN-based node classification by analyzing the maximum achievable performance for each DNN across all configurations (Upp. B. in Table 3). This upper bound is determined by assigning each region detected by the DNN to the class that maximizes the Dice score, using the ground truth to define appropriate

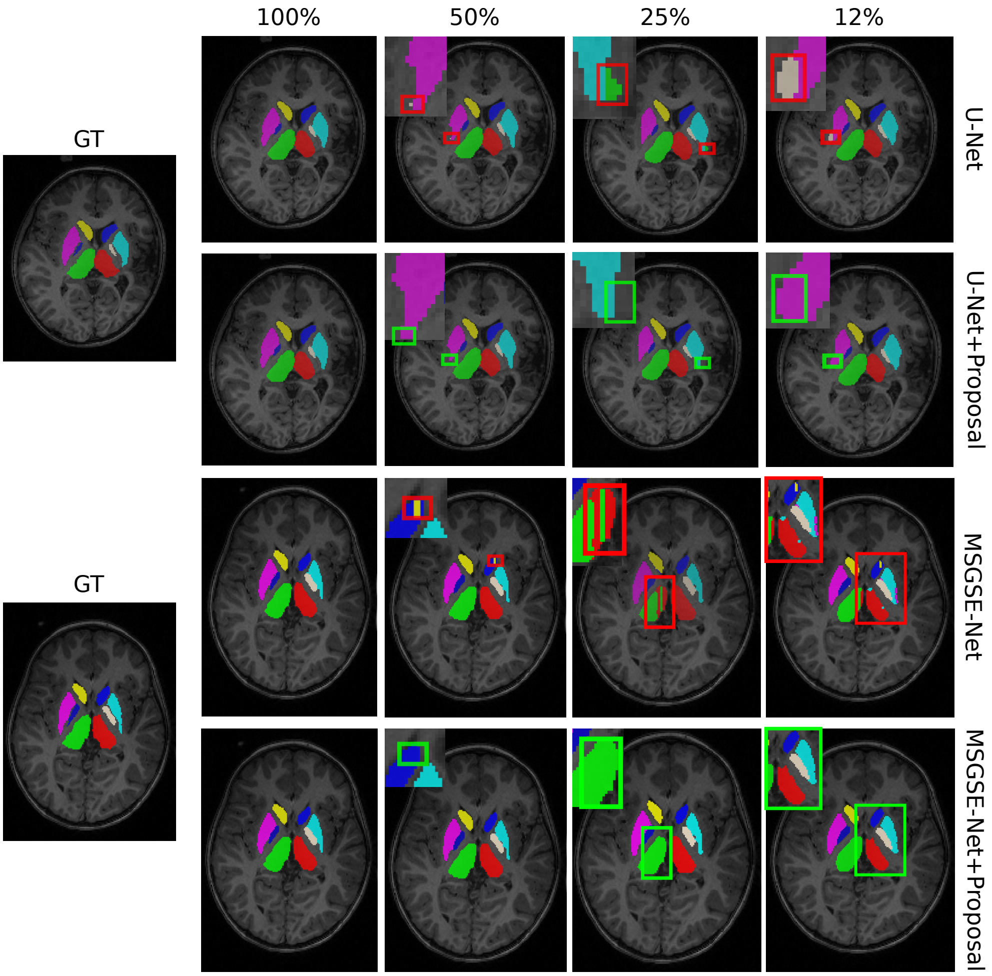

Figure 2: Segmentation of basal ganglia in two injured children using several DNNs (U-Net+CRF, MSGSE-Net) across different training data ratios. Results with and without our proposal are illustrated. GT: ground truth.

associations. We observe that our method often achieves performance very close to the maximum possible limit for both healthy and injured subjects (e.g., for 50% of the training data with all DNNs when considering median results, and for 12% of the training data with U-Net+CRF and UNETR). However, when the training dataset size is very small, the outcomes occasionally deviate from optimal performance (e.g., injured subjects with MSGSE-Net - configurations 25% and 12%) due to some misclassifications. Importantly, even in these cases where our method falls short of the optimal threshold, the median results still surpass those of the DNN alone, indicating that GNN-based post-processing enhances DNN-based segmentation.

In Table 3, it is also evident that the results achieved for children after NAIS are inferior to those obtained for healthy children, even with our post-processing method (e.g., with U-Net+CRF and 25% of the training data: Dice of 0.86 vs 0.94 for healthy brains, HD of 16.56 vs 2.93 and HD95 of 6.31 vs 0.91. This can be explained by substantial variations in the brain, sometimes observed, due to the lesion that disrupts the DNN (Figure 3). This leads to a much higher standard deviation in injured children for all metrics (standard deviation values are not indicated in Table 3 for sake of clarity). For a few injured subjects, particularly when the training dataset is very small, our mean HD and HD95 (but not the median HD and HD95) are lower than those of the DNN due to some outliers (e.g., MSGSE-Net with 25% and 12% of the training dataset, and the HD95 for UNETr with 25% of the training dataset).

Table 3

Comparison of basal ganglia segmentation in healthy children and children after NAIS with DNN alone (U-Net+CRF, UNETr, or MSGSE-Net) and with the proposed combination (Prop.) of DNN and GNN. Mean and median results across the four training configurations are presented. An upper bound (Upp. B.), referring to a classification of the regions identified by the DNN maximizing the Dice score, is provided to assess the potential enhancement of the proposed method. Bold values indicate improvement by our method. Underlined values highlight when our proposal achieves perfect node classification, reaching the maximum Dice performance achievable with DNN-based segmentation for all the test subjects.

Conf.	100%						50%					
	Dice		**HD**		**HD95**		**Dice**		**HD**		**HD95**	
Healthy	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Upp. B. U-Net+CRF	0.96	0.96	2.04	2.00	0.89	0.89	0.95	0.95	2.12	2.19	0.89	0.89
U-Net+CRF	0.96	0.96	8.14	5.72	1.78	0.89	0.95	0.95	17.72	17.77	3.21	0.89
Prop.+U-Net+CRF	0.96	0.96	**2.80**	**2.13**	**0.89**	0.89	0.95	0.95	**3.34**	**2.19**	**0.89**	0.89
Upp. B. UNETr	0.92	0.92	2.70	2.65	0.97	0.98	0.89	0.90	3.17	3.06	1.15	1.14
UNETr	0.92	0.92	3.18	2.86	0.97	0.98	0.89	0.90	3.95	3.35	1.15	1.14
Prop. + UNETr	0.92	0.92	**2.99**	**2.66**	0.97	0.98	0.89	0.90	**3.78**	**3.06**	1.15	1.14
Upp. B. MSGSE-Net	0.95	0.95	2.46	2.39	0.93	0.89	0.94	0.94	2.83	2.84	1.03	0.97
MSGSE-Net	0.95	0.95	4.25	2.64	0.93	0.89	0.91	0.92	7.20	6.62	2.54	1.85
Prop. + MSGSE-Net	0.95	0.95	**3.26**	**2.51**	0.93	0.89	**0.93**	**0.94**	**3.99**	**3.04**	**1.45**	**1.02**
Injured	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Upp. B. U-Net+CRF	0.89	0.95	5.77	2.46	2.36	0.89	0.88	0.94	5.92	2.76	2.63	0.97
U-Net+CRF	0.89	0.94	11.11	7.30	2.61	0.93	0.87	0.93	20.52	21.87	5.36	4.82
Prop.+U-Net+CRF	0.89	**0.95**	**6.43**	**2.60**	**2.36**	**0.91**	**0.88**	**0.94**	**7.58**	**4.84**	**2.49**	**0.99**
Upp. B. UNETr	0.85	0.89	3.60	3.14	1.51	1.13	0.82	0.86	4.24	3.70	2.00	1.43
UNETr	0.85	0.89	5.21	3.86	1.51	1.13	0.82	0.86	6.84	4.66	**2.00**	1.43
Prop. + UNETr	0.85	0.89	**4.71**	**3.49**	1.51	1.13	0.82	0.86	**5.12**	**3.90**	2.01	1.43
Upp. B. MSGSE-Net	0.88	0.92	3.57	3.01	1.74	1.02	0.86	0.92	11.64	3.12	4.30	1.07
MSGSE-Net	0.86	0.87	8.40	5.85	4.12	2.63	0.79	0.87	16.64	13.28	7.75	5.47
Prop. + MSGSE-Net	**0.87**	**0.89**	**7.54**	**5.75**	**3.57**	2.63	**0.85**	**0.91**	**10.14**	**3.91**	**4.03**	**1.14**
Conf.	25%						12%					
	Dice		**HD**		**HD95**		**Dice**		**HD**		**HD95**	
Healthy	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Upp. B. U-Net+CRF	0.94	0.94	2.46	2.51	0.91	0.89	0.94	0.94	2.67	2.64	1.00	1.00
U-Net+CRF	0.92	0.92	35.33	36.45	14.72	13.15	0.92	0.92	33.72	33.47	12.41	12.87
Prop.+U-Net+CRF	**0.94**	**0.94**	**2.93**	**2.57**	**0.91**	**0.89**	**0.94**	**0.94**	**4.08**	**3.01**	**1.00**	**1.00**
Upp. B. UNETr	0.80	0.81	5.04	5.22	2.06	2.04	0.75	0.77	5.82	5.94	2.75	2.54
UNETr	0.80	0.80	8.12	7.5	2.07	2.05	0.74	0.75	19.69	16.67	4.40	3.24
Prop. + UNETr	0.80	**0.81**	**5.04**	**5.22**	**2.06**	**2.04**	**0.75**	**0.76**	**6.36**	**5.94**	**2.80**	**2.54**
Upp. B. MSGSE-Net	0.90	0.91	4.08	4.06	2.13	2.26	0.90	0.90	3.78	3.90	1.51	1.24
MSGSE-Net	0.88	0.88	10.13	9.26	3.84	3.63	0.85	0.86	17.33	15.47	5.84	4.47
Prop. + MSGSE-Net	**0.90**	**0.91**	**7.12**	**4.84**	**2.40**	**2.31**	**0.86**	**0.90**	**12.17**	**4.90**	**5.05**	**1.53**
Injured	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Upp. B. U-Net+CRF	0.86	0.93	9.19	2.87	3.55	1.00	0.86	0.92	9.32	3.50	3.57	1.13
U-Net+CRF	0.85	0.91	35.80	34.13	13.75	14.18	0.84	0.90	37.52	40.64	14.98	14.25
Prop.+U-Net+CRF	**0.86**	**0.93**	**16.56**	**5.46**	**6.31**	**1.03**	**0.86**	**0.92**	**10.45**	**3.50**	**3.70**	**1.13**
Upp. B. UNETr	0.72	0.73	13.76	5.45	5.30	2.67	0.68	0.69	9.27	6.04	4.15	3.17
UNETr	0.72	0.73	14.71	10.15	**5.00**	2.81	0.67	0.68	27.20	25.23	9.32	6.80
Prop. + UNETr	0.72	0.73	**11.63**	**5.43**	7.07	**2.71**	**0.68**	**0.69**	**13.92**	**7.15**	**6.81**	**3.37**
Upp. B. MSGSE-Net	0.84	0.90	20.12	3.46	7.01	1.21	0.82	0.85	9.93	4.26	3.77	1.67
MSGSE-Net	0.75	0.80	22.42	14.42	**10.69**	7.81	0.74	0.72	**18.27**	18.25	**7.87**	7.83
Prop. + MSGSE-Net	**0.78**	**0.81**	**20.83**	**12.31**	11.54	**1.93**	**0.76**	**0.76**	20.66	**7.95**	8.78	**4.43**

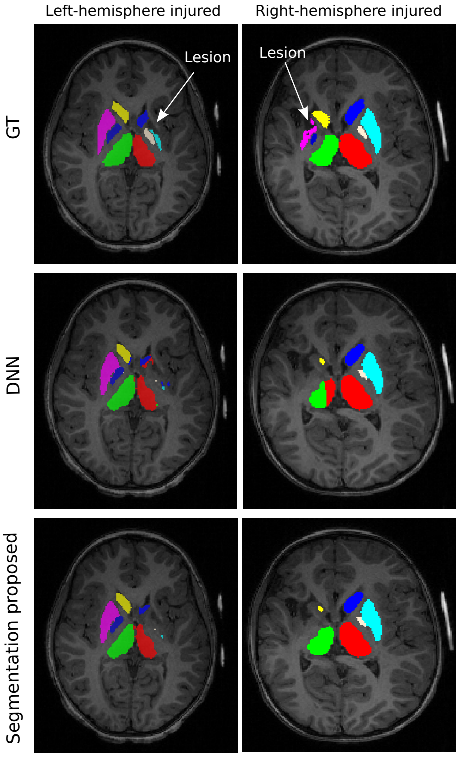

Figure 3: Examples of segmentation in subjects with a large lesion close to the basal ganglia. Results are obtained with 12% (left image) and 25% (right image) of the training dataset using MSGSE-Net. GT: ground truth. DNN: MSGSE-Net.

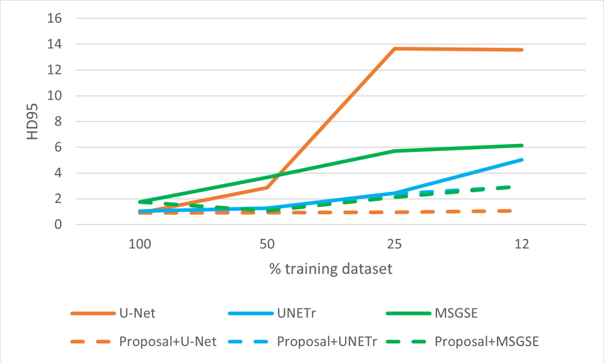

Figure 4: Evolution of the median Hausdorff distance 95 as a function of the number of training data for the three DNNs (U-Net+CRF, UNETr and MSGSE-Net), both with (dotted lines) and without (solid lines) our proposal.

Our method enhances the median HD of the DNN by up to 93% in healthy children (decreasing from 36.45 to 2.57 in configuration 25% with U-Net+CRF) and by up to 91% in injured children (decreasing from 40.64 to 3.50 in configuration 12% with U-Net+CRF). This demonstrates the ability of our post-processing method for segmenting the basal ganglia in both healthy and injured brains, as depicted in Figure 2.

To better understand the lower segmentation performance in children with neonatal stroke, we detail in Table 4 the results per hemisphere obtained for children suffering from NAIS, across the four training configurations, when applying our method to the U-Net+CRF output (the trend is similar for the other DNNs, but the results are not reported for clarity). It is evident that the segmentation of the injured hemisphere degrades the results, even when considering the whole training dataset (mean Dice of 0.83 for the injured hemisphere versus 0.96 for the contra-injured

Table 4
Mean (median) performance of basal ganglia segmentation in injured children separated by hemisphere, using U-Net+CRF across the 4 training configurations. A comparison without the proposed GNN-based post-processing is provided.

	Dice				HD				HD95			
Injured	100%	50%	25%	12%	100%	50%	25%	12%	100%	50%	25%	12%
Contra-injured Hem. w/o GNN	0.95 (0.96)	0.95 (0.95)	0.94 (0.94)	0.93 (0.93)	3.73 **(2.31)**	6.58 (4.74)	13.38 (13.42)	16.34 (21.90)	1.22 (0.80)	1.93 (0.80)	3.13 (1.02)	4.73 (1.18)
Contra-injured Hem.	**0.96** (0.96)	0.95 (0.95)	0.94 (0.94)	**0.94** **(0.94)**	**2.20** (2.34)	**2.33** **(2.43)**	**2.76** **(2.77)**	**2.92** **(2.9)**	**0.80** (0.80)	**0.82** (0.80)	**0.85** **(0.81)**	**0.93** **(0.85)**
Injured Hem. w/o GNN	0.82 (0.88)	0.80 (0.89)	0.79 (0.86)	0.81 (0.90)	11.01 (7.39)	9.00 (9.04)	24.16 (15.17)	29.19 (23.42)	4.33 (1.20)	4.65 (4.19)	11.73 (9.99)	14.09 (15.22)
Injured Hem.	**0.83** **(0.91)**	**0.82** **(0.90)**	**0.80** **(0.91)**	**0.82** (0.90)	**7.67** **(2.67)**	**8.41** **(3.02)**	**22.41** **(2.97)**	**13.26** **(3.81)**	**3.10** **(0.99)**	**3.63** **(1.00)**	**8.16** **(1.05)**	**4.85** **(1.28)**

Table 5
Mean performance of basal ganglia segmentation separated by sub-structures and study groups (healthy children, children with a lesion in the left hemisphere, children with a lesion in the right hemisphere). Results are obtained with MSGSE-Net (DNN) using 50% of the training dataset. A comparison without the proposed GNN-based post-processing is provided (DNN).

Study group	**Healthy (11 children)**				**Left injured (7 children)**				**Right injured (5 children)**			
	DNN		**Proposal**		**DNN**		**Proposal**		**DNN**		**Proposal**	
Structures	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95	Dice	HD95
Left thalamus	0.90	3.63	**0.92**	**2.25**	0.74	6.50	**0.85**	**3.74**	0.85	3.59	**0.87**	**2.43**
Right thalamus	0.90	2.68	**0.92**	**2.42**	0.85	6.49	**0.91**	**1.39**	0.85	5.00	**0.86**	5.00
Left caudate	0.88	2.62	**0.95**	**1.00**	0.54	10.41	**0.69**	**5.83**	0.70	5.09	**0.91**	**3.43**
Right caudate	0.91	7.57	**0.95**	**1.00**	0.88	9.70	**0.94**	**1.00**	0.68	16.80	**0.83**	**2.51**
Left putamen	0.94	1.00	0.94	1.00	0.63	16.73	**0.65**	**16.55**	0.92	1.08	**0.94**	**1.00**
Right putamen	0.94	1.00	**0.95**	1.00	0.93	6.50	**0.94**	**1.00**	0.72	13.47	**0.74**	**5.10**
Left pallidum	0.88	1.18	**0.89**	1.18	0.54	16.47	**0.63**	**15.89**	0.84	1.16	**0.89**	**1.00**
Right pallidum	0.87	3.22	0.87	3.22	0.84	5.56	**0.88**	**1.04**	0.73	11.46	**0.77**	**1.56**

hemisphere, mean HD of 7.67 versus 2.20, mean HD95 of 3.10 versus 0.80). This degradation can be attributed to the poorer performance of the DNN in the injured hemisphere, as illustrated in Figure 3 and Table 4 (w/o GNN). Indeed, segmentation errors are observed in the left hemisphere for the child with a left hemisphere injury (loss of the left putamen and pallidum and part of the left caudate), while errors occur in the right hemisphere for the child with a right hemisphere injury. Regions lost by the DNN (classified as background) cannot be recovered by the GNN-based post-processing. However, Figure 3 and Table 4 show that the method improves the segmentation performance in both hemispheres, as evidenced by both mean and median values, regardless of the number of training data. Moreover, our method rebalances the performance of inter-hemispheric segmentation compared to DNN segmentation when considering median values. For example, with 50% of the training dataset, our method slightly improves the median HD of the contra-injured hemisphere (4.74 to 2.43), and more demonstrably improves the median HD of the injured hemisphere (9.04 to 3.02).The proposed method yields a segmentation with a median HD95 of approximately 0.8 for the contra-lesioned hemisphere and less than 1.3 for the lesioned hemisphere (compared to over 10 without GNN), whatever the training configuration. Regardless of the training dataset size, our method reduces the median inter-hemispheric performance difference to less than 1 for HD and less than 5% for the Dice score. Again, we notice that the benefit of our proposal in each hemisphere increases as the number of training data decreases.

To evaluate the impact of our method on the delineation of individual structures, we present the performance for each structure with and without our proposal when combined with MSGSE-Net (Table 5). The results were obtained by training both the DNN and GNN with 50% of the training data. The trends observed are similar across other training configurations. To analyze the contribution of our method to the segmentation of each structure, we decomposed the results on the test dataset (see Table 1) into three study groups: healthy children (11 children), children after NAIS with a lesion in the left hemisphere (7 children) and children after NAIS with a lesion in the right hemisphere (5 children).

Table 6

Comparison of basal ganglia segmentation in healthy and injured children using MSGSE-Net trained on 50% of the training dataset in several cases, assessing the influence of structural information on model performance. Reference: segmentation performed by MSGSE-Net, GraphConv: MSGSE-Net combined with a GNN based on the graph convolution operator GraphConv, No edge attr.: MSGSE-Net combined with our proposed GNN but ignoring attributes on edges, Proposal: MSGSE-Net with our proposal.

	Dice		**HD**		**HD95**	
	Mean	Median	Mean	Median	Mean	Median
Reference	0.85	0.90	11.92	9.95	5.14	3.66
GraphConv	0.70	0.90	59.22	9.39	21.12	5.11
No edge attr.	0.85	0.90	13.81	7.70	7.03	2.74
Proposal	**0.89**	**0.92**	**7.06**	**3.47**	**2.74**	**1.08**

Regardless of the study group or structure considered, our method consistently improves average segmentation performance in terms of Dice and HD95. Improvement with our method is more notable for children suffering from NAIS. In fact, the proposed method effectively enhances the segmentation of structures in both contra-lesioned and lesioned hemispheres. Notably, the caudate nucleus of the lesioned hemisphere exhibited substantial improvements in segmentation: a 15% increase in Dice score and a 44% reduction in HD95 for the left caudate in left-lesioned children, and a 15% increase in Dice score and an 85% reduction in HD95 for the right caudate in right-lesioned children. Although improvements in segmentation of the putamen and pallidum are smaller, they are still notable (left-injured children: +2% Dice and -1% HD95 for the left putamen, +9% Dice and -4% HD95 for the left pallidum; right-injured children: +2% Dice and -62% HD95 for the right putamen, +4% Dice and -86% HD95 for the right pallidum). Furthermore, the basal ganglia (caudate, putamen, and pallidum) pose greater segmentation challenges compared to the thalamus. It is important to note that average segmentation results are lower in subjects with a lesion in the left hemisphere compared to those with a lesion in the right hemisphere. This difference arises from a larger number of children with a large lesion affecting the basal ganglia in the left lesioned group. A comparison of the performance of our proposal for structures of the same type is now to be considered. In the control group, our method segments structures of the same nature with equivalent performance, except for the pallidum, where results for the left pallidum are slightly superior to those for the right pallidum. This observation does not hold for children after NAIS. In fact, segmentation performance of structures on the contra-lesioned side is consistently superior to that of symmetrical structures in the lesioned hemisphere, corroborating earlier inter-hemispheric studies (Table 4).

We studied the influence of n-dimensional structural information (i.e., edge attributes) on segmentation. Table 6 presents the mean and median performance achieved using MSGSE-Net (with 50% of the training dataset) in four cases: MSGSE-Net only (Reference), MSGSE-Net+GNN with GraphConv (a graph convolution operator that doesn't handle multidimensional edge attributes), MSGSE-Net+ECConv without edge attributes, and MSGSE-Net combined with our proposal. We first observe that the widely used convolution operator GraphConv is not suited for our segmentation task, resulting in notable classification errors and thus poor average results. Comparing lines 3 (ECConv without edge attributes) and 4 (Proposal), we clearly see the positive impact of structural information (i.e., edge attributes) on classification accuracy, leading to improvements across all metrics, both median and mean. For clarity, results from the other DNNs, although similar, are not reported. This study confirms our hypothesis that the specific spatial organization of the basal ganglia can be used to improve its segmentation.

In Figure 5, we analyze the impact of the hyperparameter d (dimension of the node attributes at the output of the graph convolution operator) on segmentation quality (mean HD). Several values of d are compared for the three DNNs and when considering 50% and 12% of the training dataset. For conciseness, results with other training configurations (100% and 25%) are not reported, but the influence of d was found to be similar. The impact of d on HD is generally limited, with one exception: MSGSE-Net trained with only 12% of the training dataset. These results confirm that setting d to 12 is a reasonable choice, because this value strikes a balance by minimizing the number of trainable GNN parameters, which increases with d, while optimizing model performance across most cases.

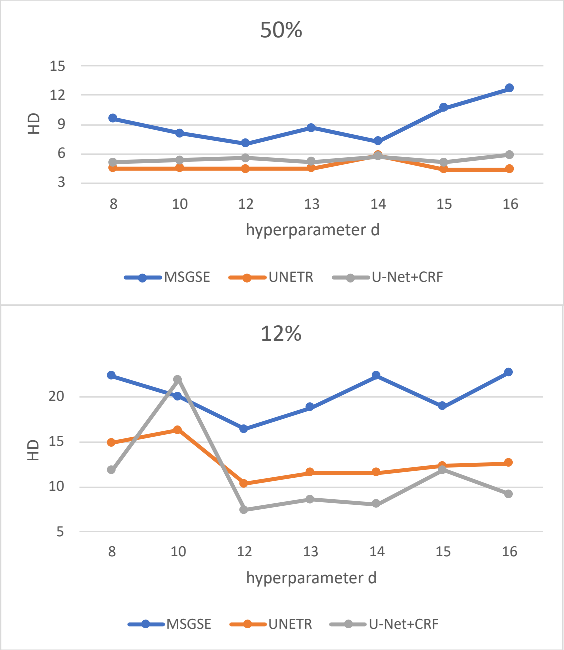

Figure 5: Evolution of the average Hausdorff distance as a function of the hyperparameter *d* (dimension of the node attributes at the output of the graph convolution). Results are shown for the three DNN backbones and for the use of 50% and 12% of the training dataset

.

A major challenge in the medical field is to develop efficient methods when data is scarce due to limited availability. Table 7 compares the results obtained by applying our proposal to U-Net+CRF considering only 6 images (4 for training and 2 for validation - configuration 12%) against the best outcomes achieved by the three DNNs when using the whole training dataset (100% - 35 images for training and 10 for validation). While UNETr and MSGSE-Net demonstrate superior HD performance compared to U-Net+CRF, the latter marginally outperforms them in Dice scores. Our post-processing method, adding a mere 345 trainable parameters, reduces the median HD of U-Net+CRF becoming lower than with UNETr and MSGSE-Net alone, despite its considerably smaller training dataset and fewer trainable parameters (15M vs. 94M for UNETr and 33M for MSGSE-Net). Additionally, Table 7 shows that that our approach, when combined with a U-Net+CRF network, achieves performance comparable to that of the 7-year-old atlas in terms of HD for both injured and healthy subjects using only 6 training subjects. This is a commendable performance, as the atlas-based method requires more data to build the atlas. Furthermore, the 7-year-old atlas was constructed using some of the healthy subjects included for evaluation. Hence, it is reasonable to assume that this leads to overly optimistic values (Dice and HD), compared to the DNNs not trained on these healthy subjects. Thus, while the HD is slightly inferior to that of the atlas, our method's performance remains compelling. Further supporting this, Table 7 demonstrates our method's superiority over the recent online multi-atlas-based *vol2Brain* algorithm [15]. This highlights the relevance of considering high-level relationships, not only for segmentation efficiency, but also concerning data requirements and trainable parameters, which is a challenging issue [59].

4. Discussion

In this paper, we introduce a graph neural network that leverages a DNN prediction and the spatial relationships between the regions segmented by the DNN to improve the semantic segmentation of the basal ganglia and thalami, key structures in motor function. Our proposal aims to automate the accurate segmentation of the basal ganglia and thalamus, even in the presence of a cortical lesion, aiming to facilitate studies related to early brain lesion. We focused on basal ganglia segmentation in both healthy children and children with neonatal arterial ischemic stroke (NAIS) to demonstrate the method's efficacy in a challenging situation (children with a perinatal stroke).

Table 7
Comparison of the proposal with an atlas-based method and state-of-the-art methods for basal ganglia segmentation in healthy and injured brains (the number of data for the deep neural networks corresponds to the configuration with the best performance). M: million, N.A.: not available.

	# required data	# trainable param.	Dice (median)	HD (median)
Reference (atlas)	31	–	0.99	3.09
Vol2Brain [15]	N.A.	N.A.	0.84	5.02
U-Net+CRF [55]	45 (35+10)	**15M**	**0.95**	6.51
UNETr [22]	45 (35+10)	94M	0.91	3.36
MSGSE-Net [9]	45 (35+10)	33M	0.91	4.24
U-Net+CRF + Proposal	**6** (4+2)	**15M**	0.93	**3.25**

In this way, we trained both the DNN and GNN on a dataset comprising healthy children as well as children with NAIS. Our approach aims to rectify the segmentation errors generated by the DNN (artifacts and spatial inconsistencies often related to lesion-induced structural differences) by relying on the known structural relationships between these brain structures. We have shown that our GNN-based post-processing layer enhanced the segmentation performance (Dice, Hausdorff distance and HD95) for both healthy subjects and those with NAIS (Table 3). Specifically, the proposal corrected most segmentation errors produced by the DNN in the injured hemisphere, resulting in equivalent median segmentation performance between hemispheres (Table 4). A comprehensive structure-by-structure performance analysis demonstrated that our method effectively enhances the segmentation of each structure in both healthy children and children with NAIS, including the injured cerebral hemisphere (Table 5). The thalamus was found to be more easily segmented than the basal ganglia, likely due to its larger size and higher tissue contrast. We observed a clear improvement in the segmentation of the caudate nucleus across all groups, with particularly noteworthy enhancements for the caudate nucleus of the lesioned hemisphere.

Improvements were observed with various DNNs widely used in this domain, including the recently developed UNETr and MSGSE-Net. We effectively leveraged all spatial information between regions by considering complete graphs to enrich the prediction of each node. The incorporation of the F_1^l function into the mapping function F^l of ECConv enabled easier distinction of the influence of each neighboring node and edge attribute during message passing. Experiments reproduced considering only the F_2^l mapping function, as in [36], yielded inferior results. Moreover, we demonstrated the beneficial effect of structural information, specifically edge attributes, on segmentation compared to GNNs that rely only on node information (Table 6).

The proposed method also addresses a well-known challenge in the medical field, particularly in pediatric studies: the lack of training data. To illustrate the strength of the proposal, we considered the example of a population of children with a rare pathology (NAIS [60]). Despite having a small dataset, it is comparable in size to other studies dealing with perinatal stroke (31 healthy and 37 injured children vs 20 healthy and 44 injured children in [8]). To evaluate the efficiency of the method on smaller datasets, which is common in the medical field when dealing with rare pathologies, we studied the influence of the size of the training dataset on the performance, ranging from 100% to 12% of the training dataset. We managed to accurately segment (median HD $\leq$ 13, median HD95 $\leq$ 5) the basal ganglia and thalamus in both healthy and injured children across all training configurations, even with a training dataset as small as 6 images (Table 1). Our GNN-based post-processing demonstrably enhanced the segmentation results achieved by any DNN trained on a small dataset (Table 3 and Figure 4 - training dataset ratio of 12%), by correcting the structural inconsistencies introduced by the under-trained DNN (Figure 2). In addition, Table 7 shows that applying our method to the output of a U-Net (+CRF) trained with only 6 images yielded superior results across the entire test dataset (including healthy children and children after NAIS) compared to the multi-atlas based method *vol2Brain* and more intricate deep learning networks like UNETr and MSGSE-Net, both trained with 45 images.

A limitation of our approach is its reliance on the connected components partitioning provided by the DNN. If the DNN fails to detect a region, our proposal can't locate it either (Figure 3). Likewise, if the DNN identifies a set of voxels associated with distinct classes as a single region, our method will be unable to subdivide the identified region. The GNN-based post-processing therefore relies on the DNN-based proposal, which is usually accurate, to classify each

cluster of voxels. To reduce this reliance on the DNN segmentation and potentially increase the upper limit specified in Table 3, we could exploit DNN uncertainty [61] to more finely subdivide the regions generated by the DNN.
Another limitation worth noting is that if the GNN misclassifies a region, it can affect numerous voxels and reduce performance (Table 3: MSGSE-Net applied on injured children with 12% of the training dataset). However, this situation is rarely encountered using our method and typically only arises when the training dataset is extremely limited. Overall, our method outperforms the DNN in all cases (Table 3), with superior Dice, HD and HD95 scores in median.
We evaluated the potential for progress in GNN-based node classification by analyzing the upper bound of performance achievable from the DNN output across all configurations (Table 3). Our method often reached this upper limit, reflecting a high rate of correct classification. However, we also noticed that as the number of training data decreases, especially in the case of injured children, the performance can deviate from this upper bound due to some misclassifications (e.g., U-Net+CRF - 25%, MSGSE-Net - 12%). Note that DNN alone perform even further from this upper limit. To increase the GNN classification rate, enriching the information carried by the graphs with additional edge attributes (e.g., angles) and node attributes (e.g., volumes, elongations or other data from morphometric analyses) would be interesting.
In the case of a large lesion close to basal ganglia, the brain structure is highly modified (Figure 3). In our sense, any DNN-based approach will struggle to accurately segment brain structures without exhaustive and representative examples in the training dataset (high variability of lesions in NAIS: localization, shape, and size). To face this issue, one alternative could be, in such rare cases, to consider third-party techniques that focus on lesion segmentation (lesion masking) [62]. While this approach could guide basal ganglia segmentation, it would introduce added complexity to the segmentation process. Such a situation could also be addressed through manual correction of some basal ganglia regions (semi-automatic segmentation) for use in clinical routines. Indeed, as observed in the experiments, such instances were relatively rare in our dataset.
Although we have considered a dataset similar in size to those used in works related to NAIS, these results need to be confirmed with more data. The method is currently tailored for 7-year-old children with a specific type of early brain lesion (NAIS). If the approach is to be extended to adult populations, where MRI offers higher tissue contrast, it will be necessary to test it in younger age groups to enable longitudinal studies of basal ganglia development, particularly following early brain lesion. Additionally, the efficacy of the proposal should be confirmed with other types of early brain lesions, such as white matter lesions or other forms of neonatal stroke, whether unilateral or bilateral.

5. Conclusion

To address the challenging but crucial task of segmenting basal ganglia and thalami in MRI scans of children, especially following early brain lesion, we propose a GNN-based framework as a post-processing layer to improve DNN-based segmentation. The proposal relies on the well-established spatial organization of the basal ganglia. The GNN we employ can be applied to any deep neural network that provides a segmentation map. It performs node classification based on both the vector probabilities from the DNN output (node attributes) and the spatial relationships between structures (edge attributes) to correct structural inconsistencies provided by the DNN. The weights of the edge attributes in the node classification task are fitted with a NN-compliant function prior to a mapping in the graph convolution operator of the GNN. The proposed GNN post-processing enhances the segmentation of each structure, including more complex delineations in the injured hemisphere resulting from the presence of a lesion. Our proposal shows resilience against the lack of large, representative training datasets, which is a challenge in the medical field, especially in pediatric studies and rare pathologies. It outperforms multi-atlas based methods as well as more complex deep learning architectures (UNETr), including a state-of-the-art method for the segmentation of subcortical brain structures (MSGSE-Net) trained on larger datasets. This study focuses on neonatal arterial ischemic stroke, the most common type of perinatal ischemic stroke. Nonetheless, the method's efficiency in both healthy children and children after NAIS suggests potential applicability to brains affected with other types of early unilateral or bilateral brain lesions. Future work will aim to improve segmentation across all subjects, with a particular focus on the injured hemisphere of subjects with a large centrally located lesion altering the structural organization, based on DNN uncertainty and lesion segmentation. We plan to extend the study to earlier stages of development and other types of early brain lesion.

Acknowledgment

We would like to warmly express our gratitude to the AVCnn group for their contribution to this ongoing project. This research was supported by the University hospital of Angers (EudraCT number 2010-A00976-33), the *Ministry of Solidarity and Health* (EudracT number 2010-A00329-30), and the *Fondation de l'Avenir* (ET0-571). Supports of the study had no role in the study design data collection, data analysis, writing of the paper, or decision to submit for publication.

References

[1] M. Arsalidou, E. Duerden, M. Taylor, The centre of the brain: topographical model of motor, cognitive, affective, and somatosensory functions of the basal ganglia, Human Brain Mapping 34 (11) (2013) 3031–3054. `doi:10.1002/hbm.22124`.

[2] J. Martin, The limbic system and cerebral circuits for reward, emotions, and memory, Neuroanatomy Text and Atlas, 4th Edn,(New York, NY: McGraw-Hill Publishing) (2012) 385–413.

[3] J. Lanciego, N. Luquin, J. Obeso, Functional neuroanatomy of the basal ganglia, Cold Spring Harbor Perspectives in Medicine 2 (12). `doi:10.1101/cshperspect.a009621`.

[4] J. Smith, Y. Smith, L. Venance, G. Watson, Editorial: Thalamic interactions with the basal ganglia: Thalamostriatal system and beyond, Frontiers in Systems Neuroscience 16 (2022) 883094. `doi:10.3389/fnsys.2022.883094`.

[5] B. T. Craig, H. L. Carlson, A. Kirton, Thalamic diaschisis following perinatal stroke is associated with clinical disability, Neuroimage: Clinical 21 (2019) 101660. `doi:10.1016/j.nicl.2019.101660`.

[6] R. Kikinis, M. Shenton, R. Donnino, F. Jolesz, D. Iosifescu, R. McCarley, P. Saiviroonporn, H. Hokama, A. Robatino, D. e. a. Metcalf, A digital brain atlas for surgical planning, model-driven segmentation, and teaching, IEEE Transactions on Visualization and Computer Graphics 2 (3) (1996) 232–241. `doi:10.1109/2945.537306`.

[7] A. Makropoulos, P. Aljabar, R. Wright, B. Hüning, N. Merchant, T. Arichi, N. Tusor, J. Hajnal, A. Edwards, S. Counsell, D. Rueckert, Regional growth and atlasing of the developing human brain, NeuroImage 125 (2016) 456–478. `doi:10.1016/j.neuroimage.2015.10.047`.

[8] J. Hassett, H. Carlson, A. Babwani, A. Kirton, Bihemispheric developmental alterations in basal ganglia volumes following unilateral perinatal stroke, NeuroImage: Clinical 35. `doi:10.1016/j.nicl.2022.103143`.

[9] X. Li, Y. Wei, L. Wang, S. Fu, C. Wang, MSGSE-Net: Multi-scale guided squeeze-and-excitation network for subcortical brain structure segmentation, Neurocomputing 461 (2021) 228–243. `doi:10.1016/j.neucom.2021.07.018`.

[10] C. Devi, A. Chandrasekharan, V. Sundararaman, Z. Alex, Neonatal brain mri segmentation: A review, Computers in Biology and Medicine 64 (2015) 163–178. `doi:10.1016/j.compbiomed.2015.06.016`.

[11] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, J. Garcia-Rodriguez, A survey on deep learning techniques for image and video semantic segmentation, Applied Soft Computing 70 (2018) 41–65. `doi:10.1016/j.asoc.2018.05.018`.

[12] R. Mehta, A. Majumdar, J. Sivaswamy, BrainSegNet: a convolutional neural network architecture for automated segmentation of human brain structures, Journal of Medical Imaging 4 (2017) 024003. `doi:10.1117/1.JMI.4.2.024003`.

[13] L. Liu, X. Hu, L. Zhu, C.-W. Fu, J. Qin, P. Heng, Ψ-Net: Stacking Densely Convolutional LSTMs for Sub-Cortical Brain Structure Segmentation, IEEE Transactions on Medical Imaging 39 (2020) 2806–2817. `doi:10.1109/TMI.2020.2975642`.

[14] B. Fischl, FreeSurfer, NeuroImage 62 (2012) 774–781. `doi:10.1016/j.neuroimage.2012.01.021`.

[15] J. V. Manjón, R. Vivo-Hernando, G. Rubio, F. Aparici, M. de la Iglesia-Vaya, P. Coupé, vol2Brain: A New Online Pipeline for Whole Brain MRI Analysis, Frontiers in neuroinformatics 16 (2022) 862805. `doi:10.3389/fninf.2022.862805`.

[16] S. Roy, J. Butman, D. Pham, Robust skull stripping using multiple MR image contrasts insensitive to pathology, NeuroImage 146 (2017) 132–147. `doi:10.1016/j.neuroimage.2016.11.017`.

[17] M. Shao, S. Han, A. Carass, X. Li, A. Blitz, J. Shin, J. Prince, L. Ellingsen, Brain ventricle parcellation using a deep neural network: Application to patients with ventriculomegaly, NeuroImage: Clinical 23. `doi:10.1016/j.nicl.2019.101871`.

[18] R. De Feo, E. Hämäläinen, E. Manninen, R. Immonen, J. Valverde, X. Ndode-Ekane, O. Gröhn, A. Pitkänen, J. Tohka, Convolutional Neural Networks enable robust automatic segmentation of the rat hippocampus in MRI after traumatic brain injury, Frontiers in Neurology 13 (2022) 820267.

[19] C. Ledig, R. Heckemann, A. Hammers, J. Lopez, V. Newcombe, A. Makropoulos, J. Lötjönen, D. Menon, D. Rueckert, Robust whole-brain segmentation: Application to traumatic brain injury, Medical Image Analysis 21 (2015) 40–58. `doi:10.1016/j.media.2014.12.003`.

[20] J. Fluss, M. Dinomais, S. Chabrier, Perinatal stroke syndromes: Similarities and diversities in aetiology, outcome and management, European Journal of Paediatric Neurology 23 (3) (2019) 368–383. `doi:10.1016/j.ejpn.2019.02.013`.

[21] B. Pidoux, Physiologie des noyaux gris centraux, Sorbonne University (2011).
URL `https://physio.sorbonne-universite.fr/wp-content/uploads/2014/01/PhysiologieNoyauxGrisCentrauxPidoux.pdf`

[22] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, D. Xu, UNETR: Transformers for 3D medical image segmentation, in: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022, pp. 1748–1758. `doi:10.1109/WACV51458.2022.00181`.

[23] P. Coupeau, J.-B. Fasquel, M. Dinomais, On the relevance of edge-conditioned convolution for GNN-based semantic image segmentation using spatial relationships, in: 2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA), 2022, pp. 1–6. `doi:10.1109/IPTA54936.2022.9784143`.

[24] J. Chopin, J.-B. Fasquel, H. Mouchère, R. Dahyot, I. Bloch, Improving semantic segmentation with graph-based structural knowledge, in: Pattern Recognition and Artificial Intelligence - Third International Conference, ICPRAI 2022, Paris, France, Vol. 13363, 2022, pp. 173–184.

doi:10.1007/978-3-031-09037-0_15.

[25] J. Chopin, J.-B. Fasquel, H. Mouchère, R. Dahyot, I. Bloch, Model-based inexact graph matching on top of DNNs for semantic scene understanding, Computer Vision and Image Understanding 235 (2023) 103744. doi:10.1016/j.cviu.2023.103744.

[26] Z. Drezner, The quadratic assignment problem, in: Location Science., Springer, Cham., 2015, p. 345–363. doi:10.1007/978-3-319-13111-5_13.

[27] D. Bacciu, F. Errica, A. Micheli, M. Podda, A gentle introduction to deep learning for graphs, Neural Networks 129 (2020) 203–221. doi:10.1016/j.neunet.2020.06.006.

[28] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (2012) 2274–2282. doi:10.1109/TPAMI.2012.120.

[29] S. Ouyang, Y. Li, Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery, Remote Sensing 13. doi:10.3390/rs13010119.

[30] J. Zhang, Z. Hua, K. Yan, K. Tian, J. Yao, E. Liu, M. Liu, X. Han, Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images, Medical Image Analysis 73. doi:10.1016/j.media.2021.102183.

[31] R. Xu, Y. Li, C. Wang, S. Xu, W. Meng, X. Zhang, Instance segmentation of biological images using graph convolutional network, Engineering Applications of Artificial Intelligence 110 (2022) 104739. doi:10.1016/j.engappai.2022.104739.

[32] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, H. Liu, Spatial pyramid based graph reasoning for semantic segmentation, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8947–8956. doi:10.1109/CVPR42600.2020.00897.

[33] Y. Chen, M. Rohrbach, Z. Yan, S. Yan, J. Feng, Y. Kalantidis, Graph-based global reasoning networks, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 433–442.

[34] Q. Liu, M. Kampffmeyer, R. Jenssen, A.-B. Salberg, Scg-net: Self-constructing graph neural networks for semantic segmentation (2020). doi:10.48550/ARXIV.2009.01599.
URL https://arxiv.org/abs/2009.01599

[35] Z. Zhang, P. Cui, W. Zhu, Deep learning on graphs: A survey, IEEE Transactions on Knowledge and Data Engineering 34 (2020) 249–270. doi:10.1109/TKDE.2020.2981333.

[36] M. Simonovsky, N. Komodakis, Dynamic edge-conditioned filters in convolutional neural networks on graphs, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3693–3702. doi:hal-01576919.

[37] G. Renton, M. Balcilar, P. Héroux, B. Gaüzère, P. Honeine, S. Adam, Symbols detection and classification using graph neural networks, Pattern Recognition Letters 152 (2021) 391–397. doi:10.1016/j.patrec.2021.09.020.

[38] F. Diehl, Edge contraction pooling for graph neural networks, ArXiv arXiv:1905.10990.

[39] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6. doi:10.1186/s40537-019-0197-0.

[40] K. Weiss, T. Khoshgoftaar, D. Wang, A survey of transfer learning, Journal of Big Data 3. doi:10.1186/s40537-016-0043-6.

[41] J. Ding, N. Xue, G. Xia, D. Dai, Decoupling zero-shot semantic segmentation, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11583–11592. doi:10.1109/CVPR52688.2022.01129.

[42] J.-B. Fasquel, N. Delanoue, A graph based image interpretation method using a priori qualitative inclusion and photometric relationships, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2019) 1043–1055. doi:10.1109/TPAMI.2018.2827939.

[43] Z. Zhang, P. Cui, W. Zhu, Deep learning on graphs: A survey, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 249–270. doi:10.1109/TKDE.2020.2981333.

[44] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, J. S. Duncan, Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe, Human Brain Mapping 19 (2003) 224–247.

[45] I. S. Gousias, D. Rueckert, R. A. Heckemann, L. E. Dyet, J. P. Boardman, A. D. Edwards, A. Hammers, Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest, NeuroImage 40 (2008) 672–684. doi:10.1016/j.neuroimage.2007.11.034.

[46] C. Gaser, R. Dahnke, P. Thompson, F. Kurth, E. Luders, A. D. N. Initiative, CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. doi:10.1101/2022.06.11.495736.

[47] P. Yushkevich, Y. Gao, G. Gerig, ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images, in: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 3342–3345. doi:10.1109/EMBC.2016.7591443.

[48] M. Fey, J. Lenssen, Fast Graph Representation Learning with PyTorch Geometric, in: ICLR 2019 Workshop on Representation Learning on Graphs and Manifolds, 2019.
URL https://arxiv.org/abs/1903.02428

[49] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, A. C. Palmer, Morphometric analysis of white matter lesions in MR images: method and validation, IEEE Transactions on Medical Imaging 13 (1994) 716–724. doi:10.1109/42.363096.

[50] M. Beauchemin, K. P. B. Thomson, G. Edwards, On the hausdorff distance used for the evaluation of segmentation results, Canadian Journal of Remote Sensing 24 (1998) 3–8. doi:10.1080/07038992.

[51] O. Maier, A. Rothberg, P. Raamana, R. Bèges, F. Isensee, M. Ahern, Z. Vincent XWD, J. Joshi, et al., loli/medpy: Medpy 0.4. 0, Zenododoi:10.5281/zenodo.2565940.

[52] M. Al Harrach, F. Rousseau, S. Groeschel, X. Wang, L. Hertz-pannier, S. Chabrier, A. Bohi, J. Lefevre, M. Dinomais, A. group, Alterations in cortical morphology after neonatal stroke: compensation in the contralesional hemisphere?, Developmental Neurobiology 79 (4) (2019) 303–316. doi:10.1002/dneu.22679.

[53] M. I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, Classification in BioApps (2018) 323–350doi:10.1007/978-3-319-65981-7_12.

[54] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, M. Grohe, Weisfeiler and leman go neural: Higher-order graph neural networks, in: Conference on Artificial Intelligence (AAAI), 2019, p. 4602–4609. doi:10.1609/aaai.v33i01.33014602.

[55] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, P. H. S. Torr, Conditional random fields as recurrent neural networks, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1529–1537. `doi:10.1109/ICCV.2015.179`.

[56] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science() 9351. `doi:10.1007/978-3-319-24574-4\_28`.

[57] B. Lee, N. Yamanakkanavar, J. Y. Choi, Automatic segmentation of brain MRI using a novel patch-wise U-Net deep architecture, PLOS ONE 15 (2020) 1–20. `doi:10.1371/journal.pone.0236493`.

[58] F. Pérez-García, R. Sparks, 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 (2021) 106236.

[59] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A ConvNet for the 2020s, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11976–11986. `doi:10.48550/arXiv.2201.03545`.

[60] A. Kirton, G. deVeber, Paediatric stroke: pressing issues and promising directions, Lancet Neurology 14 (2015) 92–102. `doi:10.1016/S1474-4422(14)70227-3`.

[61] A. A. Abdullah, M. M. Hassan, Y. T. Mustafa, A review on bayesian deep learning in healthcare: Applications and challenges, IEEE Access 10 (2022) 36538–36562. `doi:10.1109/ACCESS.2022.3163384`.

[62] A. M. Pagnozzi, Y. Gal, R. N. Boyd, S. Fiori, J. Fripp, S. Rose, N. Dowson, The need for improved brain lesion segmentation techniques for children with cerebral palsy: A review, International Journal of Developmental Neuroscience 47 (2015) 229–246. `doi:10.1016/j.ijdevneu.2015.08.004`.