Visual and structural feature combination in an interactive machine learning system for medical image segmentation

Gaëtan Galisot, Jean-Yves Ramel, Thierry Brouard, Elodie Chaillou, Barthélémy Serres

Machine Learning with Applications 8 (2022) 100294

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

Visual and structural feature combination in an interactive machine learning system for medical image segmentation

Gaëtan Galisot [a], Jean-Yves Ramel [a,*], Thierry Brouard [a], Elodie Chaillou [b], Barthélémy Serres [c]

[a] LIFAT, Université de Tours, 64 avenue Jean Portalis, Tours 37200, France
[b] CNRS, IFCE, INRAE, Université de Tours, 37380, Nouzilly, France
[c] ILIAD3/LIFAT, Université de Tours, 64 avenue Jean Portalis, Tours 37200, France

ARTICLE INFO

Keywords:
3D image segmentation
Machine learning
Interactive method
Spatial relationship
Atlas
Brain images

ABSTRACT

Currently, Convolutional Neural Networks achieve good performance in automatic image segmentation situations; however, they have not demonstrated sufficiently accurate and robust results in the case of more general and interactive systems. Also, they have been designed specifically for visual features and cannot integrate enough anatomical knowledge inside the learned models they produce. To address these problems, we propose a novel machine-learning-based framework for interactive medical image segmentation. The proposed method incorporates local anatomical knowledge learning capabilities into a bounding box-based segmentation pipeline. Region specific voxel classifiers can be learned and combined to make the model adaptive to different anatomical structures or image modalities. In addition, a spatial relationship learning mechanism is integrated to capture and use additional topological (anatomical) information. New learning procedures have been defined to integrate both types of information (visual features to characterize each substructure and spatial relationships for a relative positioning between the substructures) in a unified model. During incremental and interactive segmentation, local substructures are localized one by one, enabling partial image segmentation. Bounding box positioning within the entire image is performed automatically using previously learned spatial relationships or by the user when necessary. Inside each bounding box, atlas-based methods or CNNs that are dedicated to each substructure can be applied to automatically obtain each local segmentation. Experimental results show that (1) the proposed model is robust for segmenting objects with a small amount of training images; (2) the accuracy is similar to other methods but allows partial segmentation without requiring a global registration; and (3) the proposed method leads to accurate results with fewer user interactions and less user time than traditional interactive segmentation methods due to its spatial relationship learning capabilities.

1. Introduction

1.1. Motivations

Most current methods for medical image segmentation are fully automatic and specifically dedicated to pathology or to a model (human, small animal, brain, heart, ...). Convolutional Neural Networks (CNNs) achieve good performance in automatic image segmentation. However, they have demonstrated less accurate and robust results for more general and interactive tasks because they require a large amount of training data and they lack adaptability to variations inside the images. They have also been designed to focus primarily on visual features and are thus typically unable to incorporate anatomical knowledge inside the learned models.

Currently, fully automatic methods are usually based on a training step to select the visual features that drive segmentation. The more complex the segmentation is, the larger the size of the training database must be, which is not the case in the medical area. This problem is even more constraining in the case of animal imaging, particularly in large animal models or unusual species without templates or atlases, in which the numbers of studies and clinical routines are smaller. Another problem that is linked to these automatic methods is the inability of the user to control the segmentation result. If the quality of the delineation is insufficient, the user does not have any method to improve it.

Alternatively, interactive segmentation methods are widely used because integrating user knowledge can allow application requirements to be considered, making it easier, for example, to distinguish different tissues (Criminisi et al., 2008; Zhao & Xie, 2013). Thus, interactive segmentation remains the best method for existing commercial surgical planning and navigation products. Although leveraging user interactions often leads to more robust segmentation, a good interactive

* Corresponding author.
E-mail addresses: gaetan.galisot@univ-tours.fr (G. Galisot), ramel@univ-tours.fr (J.-Y. Ramel), thierry.brouard@univ-tours.fr (T. Brouard), elodie.chaillou@inrae.fr (E. Chaillou), barthelemy.serres@univ-tours.fr (B. Serres).

https://doi.org/10.1016/j.mlwa.2022.100294
Received 12 July 2021; Received in revised form 30 November 2021; Accepted 23 March 2022
Available online 30 March 2022

method should require as little user time as possible to reduce user burden.

Motivated by these observations, we investigate combining actual machine learning architectures with user interactions for medical image segmentation to achieve higher segmentation accuracy and robustness. We propose a novel machine-learning-based framework for interactive medical image segmentation. The proposed method incorporates local anatomical knowledge learning capabilities into a bounding box-based segmentation pipeline. Region-specific voxel classifiers can be learned and combined to make the model adaptive to different anatomical structures, pathologies or image modalities. In addition, a spatial relationship learning mechanism is integrated to capture and use additional topological (anatomical) information. New learning procedures have been defined to integrate both types of information (visual features to characterize each substructure and spatial relationships for a relative positioning between the substructures) in a unified model. During incremental and interactive segmentation, the local substructures are localized one by one, enabling partial image segmentation. Bounding box positioning within the entire image is performed automatically using previously learned spatial relationships or by the user when necessary. Inside each bounding box, atlas-based methods or CNNs that are dedicated to each substructure can be applied to obtain each local segmentation automatically.

An interactive process is integrated inside the segmentation by manually repositioning the borders of bounding boxes inside the image if necessary. With this characteristic, the process is no longer fully automatic and can be controlled by the user. Finally, spatial relationships can be defined, learned and used between the different regions of interest (ROI), allowing automatic positioning thereafter. Learned relationships are relative to the region and are applicable when the modality or the resolution is different, maintaining the general nature of the process. It is also remarkable that, unlike most other methods, the local nature of the proposed approach allows for structures to be segmented without registering the entire image.

To illustrate the generalizability of the proposed approach, we apply it to the human brain, a reference model for automatic segmentation, the sheep brain, a model for which few image processing tools exist, and the human heart. Moreover, we show that several image modalities can be segmented since the brain images are from 3D MRI and the heart images are from CT-SCAN images. Experimental results show that (1) the proposed model is more robust to segment various anatomical structures with a small amount of training images; (2) the accuracy is similar to other methods but allows partial segmentation without requiring a global registration; and (3) the proposed method leads to accurate results with fewer user interactions and less user time than traditional interactive segmentation methods due to its spatial relationship learning capabilities.

1.2. Contributions

First, we propose an operational and easy-to-use framework for incremental and interactive 3D medical image segmentation. The proposed framework can be directly used by non-experts in machine learning and computer science. Second, the proposed framework does not require a large amount of training data and annotations for each of the substructures of an organ for training. The amount of training images can be different for each substructure. Thus, the proposed framework can be directly applied to new (sub)structures or new segmentation protocols. Third, we propose incorporating spatial relationship information into a unified bounding box-based segmentation pipeline. The integration of the spatial relations allows either unsupervised (without additional user interactions) or supervised (user-provided positions) guidance of the segmentation process. Fourth, different types of voxel classifiers (atlas methods and CNNs have been tested) can be combined to segment each desired substructure (or pathology) inside an organ automatically. This also eliminates the need for whole-brain registration (improving in the same way the resistance to variability).

1.3. Organization of the paper

After a brief review of related research, Section 3 describes how the proposed framework allows us to learn anatomical models including the two types of information (visual and spatial) from a small training database. First, the learning of the local voxel classifiers modeling the visual information (shape and intensities) of each substructure of interest is described in detail. Second, the learning of the spatial relationships modeling the relative position and size between the substructures is explained. Section 4 analyzes the influence of different parameters on the segmentation results. A user experiment is also presented to describe the usability of the proposed approach.

2. Related works

Different methods that segment anatomical structures inside medical images fall into different categories. The following sections present the categories that are related to the proposed approach.

2.1. Global models that use visual features

2.1.1. Atlas based methods

Atlas-based methods are among the primary types of techniques that are used to segment 3D medical imaging. Training images are registered to the image to be segmented, providing information to drive segmentation in the unknown image. Such methods are known to be efficient in cases of brain segmentation (Landman et al., 2012; Wang et al., 2013). However, they are fully automatic and deeply linked to the registration step. This registration is often applied to the entire image, resulting in a long computation time. Segmenting MRI brain images is an important task, but segmenting heart or liver images can also be analyzed with atlas-based methods (Bai et al., 2013; Zhuang et al., 2010; Zhuang & Shen, 2016). Animal images have also been described using atlas methods, particularly small animals (Kovacevic et al., 2005) and other specific species (Ella & Keller, 2015; Nitzsche et al., 2015). These methods have also been used for intermodal segmentation (Iglesias et al., 2013; Voronin et al., 2013). To improve the ability of an atlas to represent a population, probabilistic atlases have been proposed (Ashburner & Friston, 2005; Fischl, 2012; Pohl et al., 2006; Yeo et al., 2008). Gaussian mixture models, Markov random chains and fields are often used to combine information from the test image with the membership probability obtained from the registered atlas.Thus, this type of method is difficult to use when the inter-image variability is important because they require a global registration step.

2.1.2. Deep learning

In recent years, deep learning methods have been used to segment medical images (Litjens et al., 2017). For 2D biomedical image segmentation, efficient networks such as U-Net (Ronneberger et al., 2015), Nabla-net (McKinley et al., 2016) and DCAN (Chen et al., 2016) have been proposed. For 3D volumes, patch-based CNNs (Konstantinos et al., 2017) and more powerful end-to-end 3D architectures such as V-Net (Milletari et al., 2016), High-Res3DNet (Li et al., 2017), and 3D deeply supervised networks (Dou et al., 2017) have been proposed to segment brain tumors. These methods are still primarily used for tumor (Zhao & Jia, 2016) or tissue (Moeskops et al., 2016) segmentation more than for complex anatomical structures. Today, only a few systems (de Brébisson & Montana, 2015; Mehtal et al., 2017) perform entire brain segmentation with a deep artificial neural network.

As discussed in the following section, few studies have used CNNs for interactive segmentation because current CNNs are not adaptive to different test images, due to model parameters being learned from training images and then fixed during testing without any possibility of image-specific adaptation. For medical images, annotations are often expensive to acquire because both expertise and time are required to produce accurate annotations. Furthermore, to make a CNN-based

interactive segmentation method effective, in addition to the large amount of training data needed, fast computation and memory efficiency are required. The new method should work on a standard computer, enabling the system to respond quickly to user interactions. As examples, the interactive system DeepIGeoS (Wang et al., 2019), which has a lack of adaptability to unseen image contexts; HighRes3DNet (Li et al., 2017), which needs a large amount of GPU memory; and DeepMedic (Konstantinos et al., 2017), which works on local patches to reduce memory requirements but results in long computation times, have been reported in the literature.

2.2. Combination of local models that use structural features

Even if more rarely used, some previous methods use more structural approaches to segment images. The global structure of an anatomical organ is preserved between subjects but also between species. The spatial relationship between ROI is another type of modeling that can store information about structures. Some models have been proposed to describe the object structures. These methods provide confidence values regarding some propositions; for example, the region A is to the left of the region B, which is the case for angle or force histograms (Matsakis & Wendling, 1999; Wang & Keller, 1999). Other models aim to position the region directly in the coordinate space of the image with the region that has already been localized. The fuzzy spatial relationships presented in Bloch (2005) are good examples of this model. Distance and orientation relationships are learned to provide probability maps of the region's membership, and regions that have already been localized in the image are used as references for these relationships. A probability map can be used to drive another segmentation process, and this information is typically directly used by static and image-specific segmentation methods but is not integrated into an adaptive, interactive and incremental segmentation framework.

2.3. Interactive segmentation methods

Several interactive segmentation methods have been proposed in the literature. The first representative methods were based on active contours (Yushkevich et al., 2006), graph cuts (Boykov & Jolly, 2001; Wang et al., 2016), 2D or 3D livewire (Poon et al., 2008) or geodesic distance transforms (Criminisi et al., 2008). However, most of these methods rely on low-level features and require a relatively large number of user interactions to deal with complex images (*e.g.* ambiguous boundaries). Then, machine learning methods (Scherrer et al., 2009; Wang et al., 2016) were introduced but were limited by hand-crafted features that depend on user experience.

More recently, CNNs have attracted increasing attention due to their automatic feature-learning abilities and high performances. Some neural network methods require the user to draw a bounding box around the desired object (Castrejón et al., 2017; Ling et al., 2019). ScribbleSup trains CNNs for semantic segmentation supervised by scribbles and combined bounding box annotations with a graph convolutional network (GCN) (Acuna et al., 2018; Lin et al., 2016). The selected bounding box is cropped from the image and fed to a GCN to predict the polygon/spline around the object. The polygon surrounding the object can be iteratively adjusted by refining the deep model. All these methods employ user interactions as sparse annotations for the training set rather than as guidance for refinement or for the segmentation of unknown images.

Similar to the proposed approach, BIFSeg (Wang et al., 2018) is a bounding box-based interactive framework that includes a CNN voxel classifier. An image-specific fine-tuning step is added, to be able to classify unseen objects (zero shot learning). In this framework, only one CNN is learned to segment unseen objects in test images. Additionally, DeepIGeoS (Wang et al., 2019) uses geodesic distance transforms of scribbles as additional channels of CNNs for interactive segmentation of a specific structure inside medical images. A P-Net is used to obtain an initial automatic segmentation, and an R-Net is used to refine the result based on user interactions.

As explained in Koohbanani et al. (2020) for 2D images and even more surely in 3D images, manual selection of boundary points or drawing a bounding box is still difficult and time-consuming when many different objects must be segmented. When structures have a complex shape, bounding boxes also do not provide sufficient guidance to delineate boundaries. Therefore, for most of these methods, the quantity of interactions is critical to provide the initial approximated position of the region to segment. Additionally, none of these methods take advantage of contextual information (*e.g.* spatial relationships between substructures). Then, the segmentation of an organ composed of a large number of structures remains laborious.

To conclude, to our knowledge, this study is the first to combine visual and spatial information in a unified interactive machine learning framework. This specificity allows (i) an easy definition of several different local or substructure specific voxel classifiers (atlases, CNNs, etc.), (ii) partial learning from a few images, and (iii) in addition, unlike the methods described above, an automatic initial positioning of the 3D bounding boxes corresponding to the desired substructures is implemented to reduce and simplify the manual interactions, which are required only for the optimization of the automatic positioning.

3. Method

This section describes how the proposed framework allows us to learn anatomical models including the two types of information (visual and spatial) from a small training database. First, the learning of the local voxel classifiers modeling the visual information (shape and intensities) of each substructure of interest is described in detail. Second, the learning of the spatial relationships modeling the relative position and size between the substructures is explained. We then describe the unified model that we defined to store both types of information simultaneously. The segmentation method using this new unified model incrementally and interactively is then presented. The automatic or manual positioning of the bounding boxes corresponding to the desired substructures within the entire image is explained, and the classification process provides the final local segmentation.

The proposed framework, SILA 3D, is shown in Figs. 1 and 2. To manage variability inside object classes, image modalities, or organs, we use a combination of specific voxel classifiers that receive a bounding box corresponding to a substructure as input and generate the corresponding binary segmentation. During testing, bounding boxes are provided by the user or are automatically positioned from the learned spatial relationships confronted with the position and size of previously segmented regions.

3.1. Anatomical graph model to combine visual and structural features

This section describes how the required visual and structural information can be learned and stored in a unified model. A small set of labeled images (training set) is required for this learning step. This model, which is called the Anatomical Graph Model (AGM), can represent the entire anatomical structure composed of several substructures that we would like to segment.

In the following, the N training images are denoted as I^n with $n = \{1, \dots, N\}$ and the corresponding label maps are denoted as L^n. The set of labeled regions on the label map L^n is denoted as R_{L^n}, and the set of all the regions available from the whole set of label maps is denoted as $\mathrm{SR} = \bigcup R_{L^n}$ with $r = \{1, \dots, R\}$ the total number of regions. It is noticeable that each label map L^n can be different from the other in terms of content (available labeled regions).

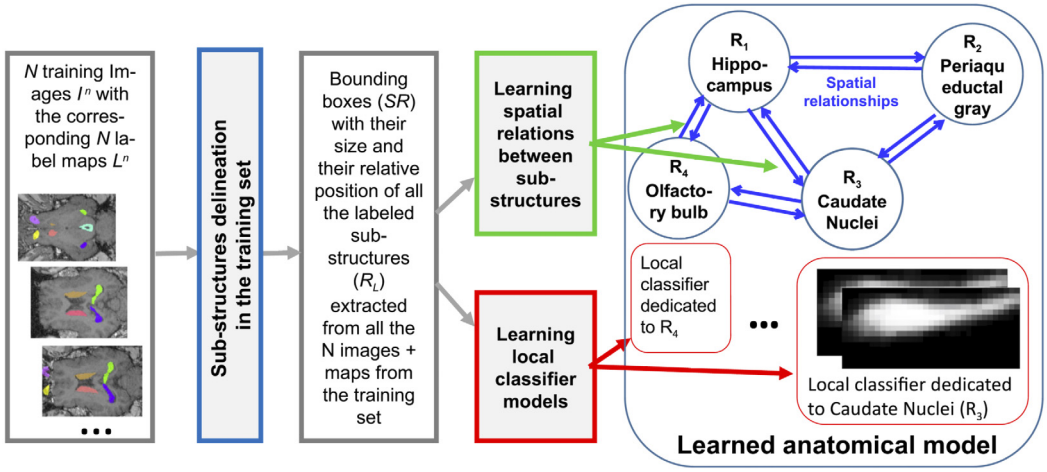

Fig. 1. The learning step in SILA 3D with 2 main steps: 1/ Learning of local classifier models for each substructure; 2/ Learning of spatial relationships between substructures.

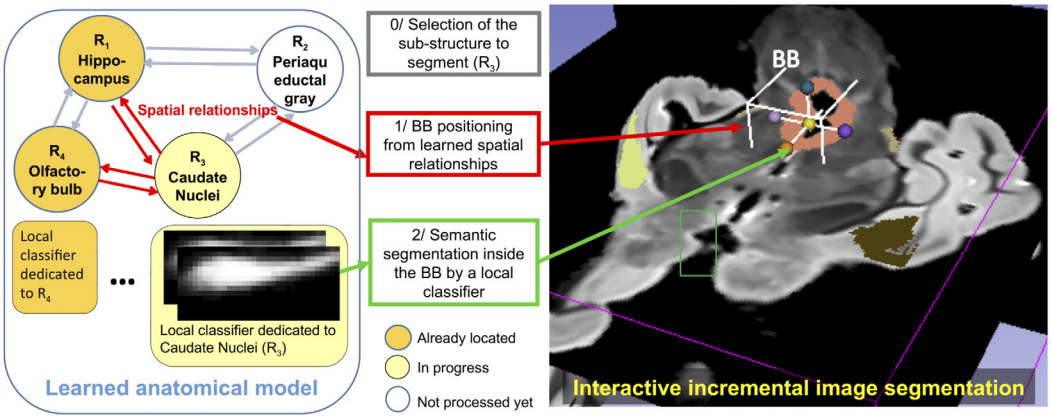

Fig. 2. The Segmentation step in SILA 3D with 2 main steps: 1/ The positioning of the bounding box according to the learned spatial relations; 2/ The segmentation using the local model to classify the voxels inside the bounding box.

3.1.1. Learning the local voxel classifier dedicated to each substructure

The visual features (*e.g.* shape, voxel intensity) associated with each subpart of an anatomical structure are learned locally. The resulting set of learned models LM_r encodes the *a priori* knowledge of the corresponding substructures. The local classifiers LM_r are built independently for each subpart of the anatomical structure. The following paragraphs show the different steps necessary to build the models required to classify the voxels inside each bounding box. It is easy to adapt the method to build different types of voxel classifiers, such as CNN models and local atlases (Galisot et al., 2019).

As already mentioned, the set of labeled regions in each training image is not necessarily identical. The consequence is that the number of training images N_r used to create each local model is also not necessarily the same for each substructure.

(a) Region delineation: As shown in Fig. 1, the process of local model creation is initialized by the delineation of the bounding box associated with each region $r \in$ SR. From the ground truth, the subvolume contained inside the bounding box of $r \in L^n$ is extracted and denoted as L^n_r, and the corresponding subvolume in the images I^n is denoted as I^n_r. In practice, a margin around the bounding box is used to limit brutal variation in membership probability into the border of the region. This margin is a percentage of the real size of the bounding box (*e.g.* 10% of the size of the region). Another strength of this margin is that it improves model robustness to the inaccurate positioning of a bounding box across the entire image during segmentation. In this study, the bounding box refers to this extended bounding box. Each region (substructure) is described by N_r couples of images $\{I^n_r, L^n_r\}$ with $N_r = card(R_{L^n})$ is the amount of training image maps L^n_r, where region r is labeled.

(b) Local model construction: A model can now be learned based on these subimages in the same way that it can be learned based on global images. Although it is possible to learn various types of classifiers, atlas-based methods and neural network models are the two primary classes of techniques for the anatomical segmentation of 3D medical images. We describe in the following a particular example of each type: one based on U-Unet3D (LM_r), and one based on probabilistic local atlases (LA_r).

U-Net3D: For each substructure r, it is possible to train a 3D U-Net model from the set of available couples of subimages $\{I^n_r, L^n_r\}$ in the training set. Because training images can come from different sources, a histogram-based intensity correction (Nyul et al., 2000) is used to match the intensities of the subimages I^q_r together. The intensities are rescaled between -1 and 1. In this study, we use a small 3D U-net network (Çiçek et al., 2016) that consists of 3 convolution layers and 3 deconvolution layers with a kernel of size $3 \times 3 \times 3$. As proposed in the original architecture, the size of the image gradually reduces while the depth (number of filters) gradually increases (from 16 to 64) in the encoder part. Pooling step is performed between each layer. The CNN has been trained with 100 epochs. The training subimages $\{I^n_r, L^n_r\}$ are resized to a multiple of 2 dimensions depending on the size of the bounding boxes of a specific region. This size is chosen as the closest multiple of two of the average of the bounding box sizes in the training data set. A linear augmentation is applied using random translation, rotation, shear, and scaling deformations. The final model is denoted as LM_r.

Probabilistic atlas: Instead of U-Net3D, the construction of local probabilistic atlases can be performed iteratively. The atlas is defined by a *Template image* and a corresponding *Probability map* denoted as $\{T_r, P_r\}$. During an iterative learning procedure, the subimages $\{I_r^{(q)}, L_r^{(q)}\}$ are added to a current atlas $\{T_r^{(q)}, P_r^{(q)}\}$ at iteration q. For each image, intensity normalization, image registration and averaging steps are performed. We use the same intensity normalization procedure as used with U-Net3D (Nyul et al., 2000). Then, a linear and a nonlinear registration based on B-spline, denoted as τ, are applied to warp the image and the corresponding label map to the current template. Finally, a weight average is applied to merge the information of the new registered couple of images to the current atlas with the following procedure:

$$\begin{aligned} T_r^{(q)} &= \frac{T_r^{(q-1)}+(q-1)+I_r^{(q-1)}(\tau(v))}{q} \\ P_r^{(q)} &= \frac{P_r^{(q-1)}+(q-1)+L_r^{(q-1)}(\tau(v))}{q} \end{aligned} \tag{1}$$

These steps are applied for the N_r couples of subimages available to build the final local probabilistic atlas, compound of a template and a probability map $\{T_r, P_r\}$, which is denoted as LA_r and can be used later to segment a specific anatomical structure.

The primary criterion to decide between both types of classifiers to use is the available amount of training images (Galisot et al., 2019). Also, the main difference between these two approaches is the possibility to update in an iterative manner the probabilistic atlases but not the CNN models that have to be retrained from scratch when we want to update them with new data. Also, it is important to keep in mind that many other methods could be implemented (*e.g.* multiatlas method, graph cut, various types of CNN, etc.) depending on the needs and the type of images to analyze.

3.1.2. Learning spatial relationships

Structural features (*i.e.* spatial relations for relative positioning and size between subparts) within the organ are required during segmentation and must be learned and stored. Information about the spatial relationships between substructures is also learned from the training data set. The learned information will be used during segmentation of a new image to position the bounding box of a target structure automatically throughout the image from the localization of one or several previously segmented structures inside the same test image. The term *position* refers to the detection of the bounding box of an anatomical structure.

The modeling of these spatial relationships are composed of distances between the borders of each couple of structures. In 3D, twelve distances are learned between one structure and another to connect the borders of the same plan (cf. Fig. 3b). These distances are relative to the size of the source region, allowing the relations to be independent of some feature of the image (resolution and size of the image). On the other hand, it is necessary to have a similar orientation for all the training and test images. It can be easily done with a simple rigid registration. We denote the source region as r_1 and the target region as r_2, and $(0, \vec{x}, \vec{y}, \vec{z})$ defines an orthonormal set. This process is the same for each direction, and we now consider the direction $\vec{x}$. The dimension (length) of the region r_1 in $\vec{x}$ is denoted as dim_x. The values $d^{x11}_{r_1r_2}$, $d^{x12}_{r_1r_2}$, $d^{x21}_{r_1r_2}$ and $d^{x22}_{r_1r_2}$ denote the distances between the first and the second borders of r_1 and the first and the second borders of r_2 in the direction $\vec{x}$. The relative distances are computed as:

$$dr^{xij}_{r_1r_2} = \frac{d^{xij}_{r_1r_2}}{dim_x} \qquad i=(1,2), \quad j=(1,2) \tag{2}$$

This relative distance is computed for all the training images on which the pair of regions (r_1, r_2) is labeled. The minimum and maximum relative distances among those observed in these training images are stored to describe the spatial relationships between regions r_1 and r_2.

$$\begin{cases} Max_{dr^{xij}_{r_1r_2}} &= \max\limits_{I_n, n=1,\dots,N}(dr^{xij}_{r_1r_2}) \\ Min_{dr^{xij}_{r_1r_2}} &= \min\limits_{I_n, n=1,\dots,N}(dr^{xij}_{r_1r_2}) \end{cases} \tag{3}$$

Given twelve different intervals, the same process is applied to the other directions $\vec{y}$ and $\vec{z}$. The relative character of the distances implies that the information from r_1 to r_2 is different than that from r_2 to r_1. In total, 24 intervals can be learned and used between two regions.

3.1.3. Storing the learned information inside a graph representation

When the local classifier models and the spatial relationships are learned from the training data set, all information is stored in a graph structure named *Anatomical Graph Model* (AGM), as described in Fig. 3a.

Each node of the graph represents a subpart of the entire anatomical structure and stores the associated local classifier model. The edges represent the spatial relationships between the subparts and store the intervals of relative distances. Edges are oriented, and the graph is complete if all pairs of regions (r_1, r_2) are labeled in at least one of the training images. The labels (*i.e.* information required for the local classifier model updating) associated with each node (substructure) r in the AGM are:

$$\ell_{n_r} = \begin{cases} N_r & Number\ of\ images\ associated\ to\ region\ r\ used\ for\ the\ training \\ LBB_r & Training\ data\ for\ region\ r\ (extracted\ bounding\ boxes\ with\ labels) \\ LA_r & Local\ atlas: template\ and\ probability\ map \\ & \textbf{or} \\ LM_r & Local\ model\ computed\ from\ the\ training\ data \end{cases} \tag{4}$$

The label associated with the edge between r_1 and r_2 in the AGM is defined as:

$$l_{a_{r_1r_2}} = \begin{cases} [Min_{dr^{xij}_{r_1r_2}}, Max_{dr^{xij}_{r_1r_2}}] & (i,j) \in (1,2)^2 \\ [Min_{dr^{yij}_{r_1r_2}}, Max_{dr^{yij}_{r_1r_2}}] & (i,j) \in (1,2)^2 \\ [Min_{dr^{zij}_{r_1r_2}}, Max_{dr^{zij}_{r_1r_2}}] & (i,j) \in (1,2)^2 \end{cases} \tag{5}$$

Fig. 3a shows an example of an anatomical graph model with three anatomical structures of a sheep brain.

3.2. Incremental and interactive segmentation

3.2.1. Overview

In this section, we explain the proposed incremental and interactive segmentation scheme. This pipeline requires an anatomical graph model corresponding to the anatomical (sub)structures that we would like to extract from the test images.

We denote as Y the 3D images that the user wants to segment. The desired substructures are extracted one by one according to the decision of an expert (*i.e.* a user) or using a heuristic method. The order of extraction of the substructures is not fixed and is a parameter of this approach. The same process is applied to the segmentation of each substructure and is composed of several steps: (1) the selection of the substructure to be segmented; (2) the positioning of the corresponding bounding box inside the test image; and (3) the voxel classification inside the positioned bounding box. The outline of the global process is shown in Fig. 2. The positioning of the bounding box and the voxel classification step are presented more specifically in the next subsections.

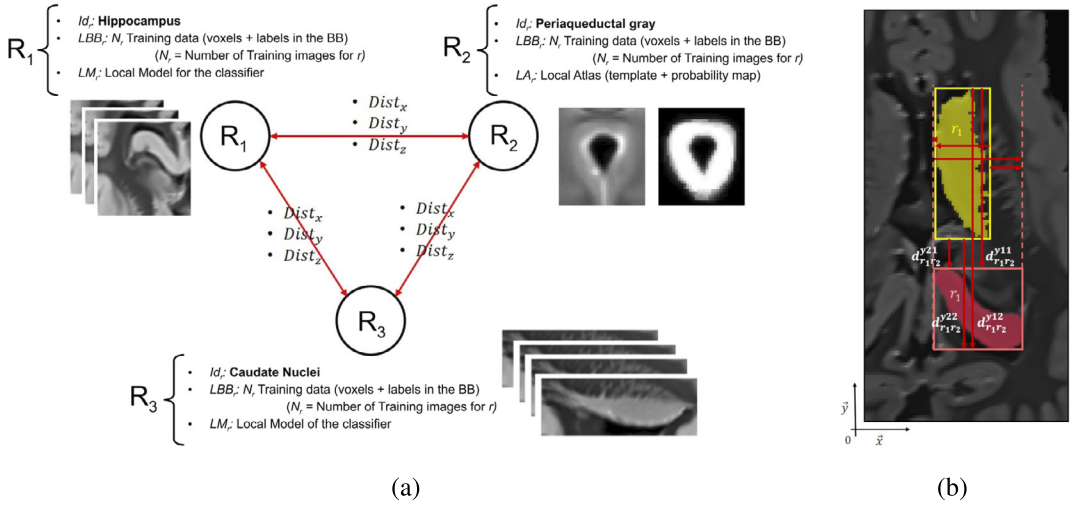

Fig. 3. (a) Example of a learned Anatomical Graph Model with three anatomical structures of the sheep brain. (b) Illustration of the spatial relationship in the 2D case, where 8 different distances are learned and stored in these relationships between structure r_1 and structure r_2.

3.2.2. Automatic positioning of the bounding boxes

The position of the bounding box can be processed automatically or manually. In some cases, it can be interesting to allow the user to define or refine the position of the bounding box of a substructure to attain more efficient segmentation. The user can also allow the algorithm to use the spatial relationships learned previously (cf. Section 3.1.2) to compute the position of the bounding box of the desired substructure automatically because the user always has the ability to correct incorrect positioning. As during training, a margin is determined around the bounding box. The selected extended volume inside the test image is denoted as Y_r, and the regions that are already localized in the test image Y are denoted as R' and are used as references. In the AGM, the edges from the nodes associated to the regions R' and the node associated to region r contain the intervals of relative distance that can be used.

The goal of this method is to determine the localization of the 6 borders of the bounding box, denoted as B_r^j with $j \in \{1, \ldots, 6\}$. The positioning is performed independently for each border of the bounding box.

Next, we explain the process for the first border of r in the direction $\vec{x}$. The same methodology is applied for the other directions and other borders. Each region r' in R' provides two intervals for the positioning of a border. The interval $[Min_{dr'_r}^{x11}, Max_{dr'_r}^{x11}]$ presented in the previous section describes the relative distance between border No. 1 of r' and border No. 1 of r in the direction $\vec{x}$ (same as the border No. 2 of r with $[Min_{dr'_r}^{x21}, Max_{dr'_r}^{x21}]$). Then, the intervals are transformed in absolute positions on the target image. The position and the size of the regions R' are known, and the intervals of absolute position denoted as $[Xmin_{r'}^{i1}, Xmax_{r'}^{i1}]$ can be computed as:

$$\begin{aligned} Xmin_{r'}^{i1} &= x_{r'}^{i} + dim_{r'} * Min_{dr'_r}^{xi1} \quad i \in (1,2) \\ Xmax_{r'}^{i1} &= x_{r'}^{i} + dim_{r'} * Max_{dr'_r}^{xi1} \quad i \in (1,2) \end{aligned} \tag{6}$$

Next, all position intervals are merged to obtain the final position of the borders of region r. Each interval is described by a rectangular function, denoted as. The rectangular function is equal to 1 between $[Xmin_{r'}^{i1}, Xmax_{r'}^{i1}]$ and 0 outside. Then, they are weighted with the weight inversely proportional to the length. We assume that the intervals with a short size represent some relations that were more stable in the different images of the training data set:

$$\mathcal{W}_{r'}^{i1} = \frac{1}{Xmax_{r'}^{i1} - Xmin_{r'}^{i1} + \gamma} \tag{7}$$

where γ is a positive parameter to keep the value different from zero, which can occur when a region occurs only once in the training set. Finally, the merging is performed by summing the different weighted intervals. The expected value of the sum provides the position of the border of the bounding box and is computed as follows:

$$B_r^j = E\Big(\sum_{r' \in R'} \mathcal{W}_{\Pi_{r'}^{1j}} \Pi_{r'}^{1,j}(x) + \sum_{r' \in R'} \mathcal{W}_{\Pi_{r'}^{2j}} \Pi_{r'}^{2,j}(x) \Big) \tag{8}$$

This result is the position of border No. 1 in the direction $\vec{x}$, but this process is applied six times to set the position of each border. The bounding box positioning can then be exposed to the users for validation or adjustment. The subvolume inside this bounding box is then extracted and used during the voxel classification step.

The information provided by the spatial relationships can also be used to choose the automatic order of segmentation of the regions. Due to the incremental nature of the proposed segmentation scheme, the quality of each segmentation is impacted by the quality of previous segmentations. It appears important to initially segment the regions for which we can be sure that the segmentation quality is sufficiently correct. The spatial relationships from such regions could be considered more accurate. As with every segmentation method, the quality depends on the anatomical structures (*e.g.* some structures are more stable than others), but in the proposed approach, it also depends on the quality of the learned relationship information.

As explained below, a stability score can be computed for all remaining regions to segment by analyzing the stability (variance) in the position estimations. At the end of the segmentation of a region, all non-segmented regions $R \setminus R'$ with at least one spatial relationship with an already segmented region R' are considered. The process described previously to position the bounding box automatically is then used. The sum of the intervals described in Eq. (8) is computed and used to obtain a stability score for the position of the bounding box for the remaining regions. The standard deviation of this function is used to describe how well the spatial information coming from all source regions matches. If all available information is accurate and consistent, the standard deviation is small. The standard deviation is computed for each edge, and the sum of the 6 standard deviations defines the final stability score. The region with a smaller score is considered the region where we have the most confidence and is proposed to the user or automatically selected as the next region to segment. After segmentation of this region, the same process can be applied again but with the information from an additional region.

3.2.3. Voxel classification

After the bounding box has been positioned inside the entire image, voxel classification is applied to the extracted subvolume to obtain

the final segmentation of the region inside the bounding box Y_r. From its nodes, the AGM (Anatomical Graph Model) provides information about the local model to use to classify the voxels inside region r. The following process depends on the model.

U-Net3D: With CNN, an inference computation must be performed on the voxels inside the bounding box Y_r using the local network LM_r stored in the corresponding node of the AGM. Intensity normalization is applied between Y_r and a reference images. We decided to use a template T_r computed as for the probabilistic atlases as reference (see Section 3.1.1. This step is followed by an intensity rescaling between -1 and 1. The image Y_r is also resized to match the dimension of the local network LM_r. Finally, LM_r is used to get the final segmentation of the region r. The binary result is then resized to the original shape.

Probabilistic atlas: With probabilistic atlas-based methods, the probability map LA_r and the associated template $\{T_r, Pr_r\}$ are used. First, as during the learning step, an intensity matching method is applied between the bounding box Y_r and the template of the corresponding atlas T_r. Then, the template is registered to Y_r. The transformation obtained with the registration process is also applied to the probability map P_r associated with the template. Finally, a Hidden Markov Random Field (HMRF) is applied to obtain the final segmentation and takes the registered probability map Pr_r as an external field to drive the segmentation process. More details about the use of the local atlas and HMRF for this step can be found in Galisot et al. (2019).

The results computed on the subvolume are propagated on the coordinates of the entire image Y. The proposed approach is incremental, and some voxels could have been classified during previous segmentation. If the voxels were already classified as part of a previous region of R', the labeled voxels were not updated. If the voxels v are not already classified, the voxels are labeled as r when it belongs to the class defined as *"region"*, and it is not labeled when it belongs to one of the classes as a *"non-region"* by the local binary classifier dedicated to region r. This decision process is summarized in Eq. (9) with S_{new} the new current segmentation after the segmentation of the region r and the segmentation obtained during the previous step denoted as $S_{R'}$.

$$S_{new_i} = \begin{cases} S_{R'_i} & if\ S_{R'_i} \neq 0 \\ 0 & if\ S_{R'_i} = 0\ and\ z_{r_i} = \text{“}non - region\text{”} \\ r & if\ S_{R'_i} = 0\ and\ z_{r_i} = \text{“}region\text{”} \end{cases} \tag{9}$$

where 0 is the label of the voxels that have not been classified and still belong to the background.

The user can validate this result and choose to segment a new region, or he can perform the segmentation again to obtain better results by positioning the bounding box differently.

As explained before, the proposed method is incremental (*i.e.* the regions are segmented one after another), and the user has the ability to reposition the bounding box of each region before running the voxel classification instead of using the learned spatial relationships. Therefore, during the experiments, the proposed approach is denoted as *LIS* (Local and Incremental Segmentation), and when all regions are well positioned with the ground truth, *LIS* is denoted as LIS_{GT}. LIS_x denotes the segmentation when the first x regions are well positioned according to the groundtruth (GT).

4. Experiments

After a brief description of the implementation of SILA 3D, this section analyzes the influence of different parameters on the segmentation results. The size of the training database and the impact of the order of segmentation are evaluated. Experiments with sheep brain and human brain images highlight the effects of the method by providing satisfactory results within a short computation time compared to other methods. A user experiment is also presented to describe the usability of the proposed approach. Finally, the generic nature of the method is demonstrated by providing segmentations of heart images and intermodal cases.

4.1. Implementation details

4.1.1. Global architecture description

The described framework named SILA 3D has been used in several studies and different kinds of projects.[1] SILA 3D is available online.[2] Readers can refer to this web site for videos demonstrating how to use the proposed framework.

4.1.2. Core kernel C++/ITK

The main piece of software (Core/Kernel) has been designed and implemented in C++, in a modular and extensible way. This kernel includes not only a set of computational tasks, but also loading and saving functions. The main functionalities are exposed in a binary library to help further development and integration. For instance, the kernel includes *DataInput*, *Registration*, *Segmentation* and *Results* classes that are available to implement new registration methods, new segmentation methods, or new data input formats in order to build a versatile software tool.

The current implemented registration methods are based on the ITK (Insight Toolkit Software Consortium) framework, which is powerful and well-known in the medical image community.

4.1.3. Standalone app, command-line interface (CLI)

This kernel has been designed to be usable through a CLI tool, allowing integration in any kind of scripting. This makes easy to design and build computationally intensive test campaigns.

4.1.4. Web service integration (Image server, API)

Web services integration for computational intensive tasks are a good trade-off between extensiveness, community usage and ease of use for neuroscientists researchers. It also makes possible to run segmentation tasks or atlas building tasks in a relative low-end computer from everywhere.

The Fig. 4 provides an overview of the Software as a Service (SaaS) architecture. The REST API, which is exposed to clients through HTTP, is currently a subset of the main functionalities. A multi-user access and session is implemented to allow resource sharing on CPU clusters or GPU clusters.

4.1.5. Client app, graphical user interface (GUI)

Thanks to the wide open source framework provided by 3DSlicer (Fedorov et al., 2012), and its python programming interface (SDK), 3DSlicer was chosen to build a standalone client GUI to communicate with the SILA 3D core functionalities. This client GUI reuses Slicer visual widgets to allow users to navigate through image slices and to drive the interactive segmentation process. The GUI provides access to all needed core functionalities such as loading modalities images, loading a new model, choosing structures and launching process on servers, multi-users session management (saving/exporting segmentation results).

For instance, the proposed GUI allows a precise and easy positioning of the borders of the bounding boxes through few intuitive moves of 2D boxes in each image plans with the help of an additional global 3D preview (see Fig. 2).

4.1.6. Data packaging

The user data and models are packaged to allow data loading, saving, and sharing through the web. Basically, data packages contain image portions, models, and processed data compressed into archive files with a specifically designed format. Our data model package usually includes the created and used AGM, the segmentation states and precise logging information. This allows the system and users to keep track of any changes, updates of shared data and models in case of multi-user environment.

[1] https://www.echosciences-centre-valdeloire.fr/communautes/un-cerveau-dans-toutes-les-tetes/articles/neuro2co-projet-de-recherches-en-sciences-participatives-centre-inra-val-de-loire

[2] http://www.rfai.li.univ-tours.fr/PublicData/3D_Brain_Seg/home.html

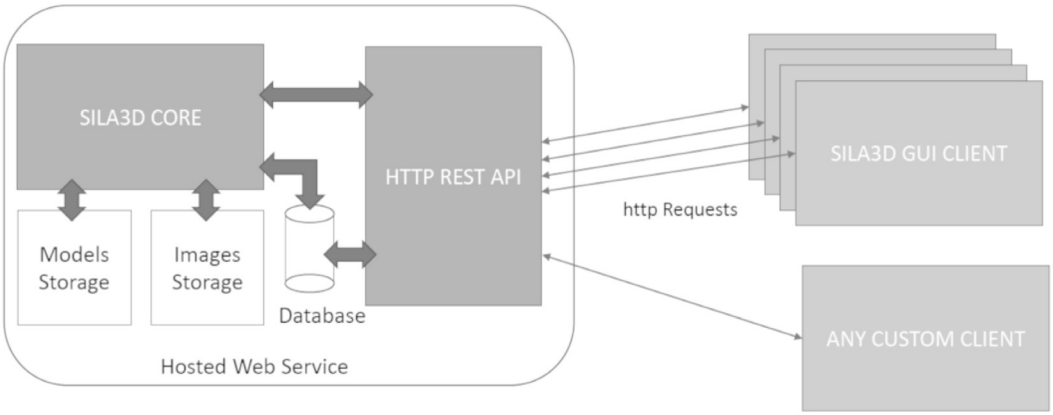

Fig. 4. SILA 3D Saas architecture. Used in the SILA 3D GUI Client, the REST API make it possible custom tools to be built.

4.2. Protocol

4.2.1. Datasets

Two image databases were used during the experiments. The MICCAI'12 image dataset (Landman et al., 2012) used for the multiatlas segmentation challenge dealing with brain images and coming from the OASIS project. This database consisted of 35 MRI T1 images of human brains, and each image was manually segmented with more than 140 anatomical structures; however, only 13 structures were used in this study, including 6 bilateral structures: the caudate nucleus, pallida, thalami, putamens, ventricles, hippocampi and brainstem. Fifteen images are used as training images, and 20 are used as test images. An image from the training database is shown in Fig. 5 (left).

NeuroGeoEx is an image database of T2 *ex vivo* sheep brains provided by Neurospin company (Leprince et al., 2015) as part of the NeuroGeo[3] project. Six brains were acquired with 7T 50 mT/m MRI with a spatial resolution of $0.3 \times 0.3 \times 0.3$ mm. Each image was manually labeled with 13 anatomical structures composed of 6 bilateral structures: olfactory bulbs, caudate nucleus, amygdala, optic superior colliculi, motor superior colliculi, hippocampus and periaqueductal gray (PAG). Leave-one-out process was applied to the NeuroGeoEx dataset; five images were used to build the Anatomical Graph Model used to segment the remaining image. A typical image is shown in Fig. 5 (right).

4.2.2. Metric

To evaluate the quality of the different segmentations, the Dice ratio is used and is defined as follows:

$$Dice = \frac{2.TP}{2.TP + FP + FN} \tag{10}$$

where TP is the number of true positives, FP is the number of false positives and FN is the number of false negative voxels.

4.2.3. User workshop

To evaluate SILA 3D, we asked nineteen people with different skills to participate in different experiments to evaluate and validate the efficiency of the proposed method and framework. The experiments were performed on both databases described in Section 4.2.1. Different tasks were performed:

- We asked users to segment images by manually positioning all bounding boxes without using spatial relationships.
- We asked users to segment images at their discretion with all available functionalities in SILA 3D (*e.g.* automatic positioning according to the learned relationships, manual positioning, order of segmentation, etc.)
- We asked users to segment an image by choosing only the order of segmentation (*e.g.* sequential selection of the substructure to segment). For each substructure, the bounding boxes were positioned according to the ground truth inside the entire image.

In all cases, the voxel classification inside the bounding box is achieved automatically using the corresponding learned local model stored inside the AGM.

Based on the observations, measurements, and feedback made by participants during this workshop, Table 4 provides a qualitative comparison between the main features of SILA 3D and some of the systems discussed in the related works.

4.3. Quantitative evaluation

As our local approach requires the positioning of the bounding boxes of each the different sub-structures, we evaluate two different configurations. In the first case (simulating ideal case with perfect manual positioning), all the bounding boxes are perfectly positioned by using the values coming from the ground truth (LIS_{GT}) and, in the second case (realistic semi-automatic positioning) by well positioning two regions (LIS_2) before using the learned spatial relationships to automatically position the next regions.

4.3.1. Impact of the selected local classifier

Our framework allows us to easily change the type of the voxel classifier used to segment the substructure inside each bounding box. The global framework based on bounding box positioning with user interaction or spatial relationship is kept the same. In order to evaluate the impact of the local classifier into our framework, a perfect positioning of each bounding box as well as a positioning based on spatial relationship are simulated.

We evaluated three types of local classifiers: HMRF (LIS_{HMRF}), CNN (LIS_{CNN}) as described in Section 3.2.3 and a simple voting method with multi-atlas (LIS_{SV}) (Klein et al., 2005). The results of the experiment applied to the human brain images are displayed on the Table 1.

CNN local classifiers provide better segmentation when the bounding box is perfectly positioned but they do not work when the bounding boxes are positioned according to the learned spatial relationships. On the contrary, the registration based methods and especially when it is coupled with HMRF segmentation have more stable results independently of the quality of the bounding box position. Because of this robustness as well as the possible incremental learning, we decide to use the local probabilistic atlases and HMRF as local classifiers rather than CNN in the rest of our experiments (and in the final implementation of SILA 3D framework available online). We will refer the method using local probabilistic atlases and HMRF as LIS in the rest of the experiments.

[3] French Regional project - see acknowledgment.

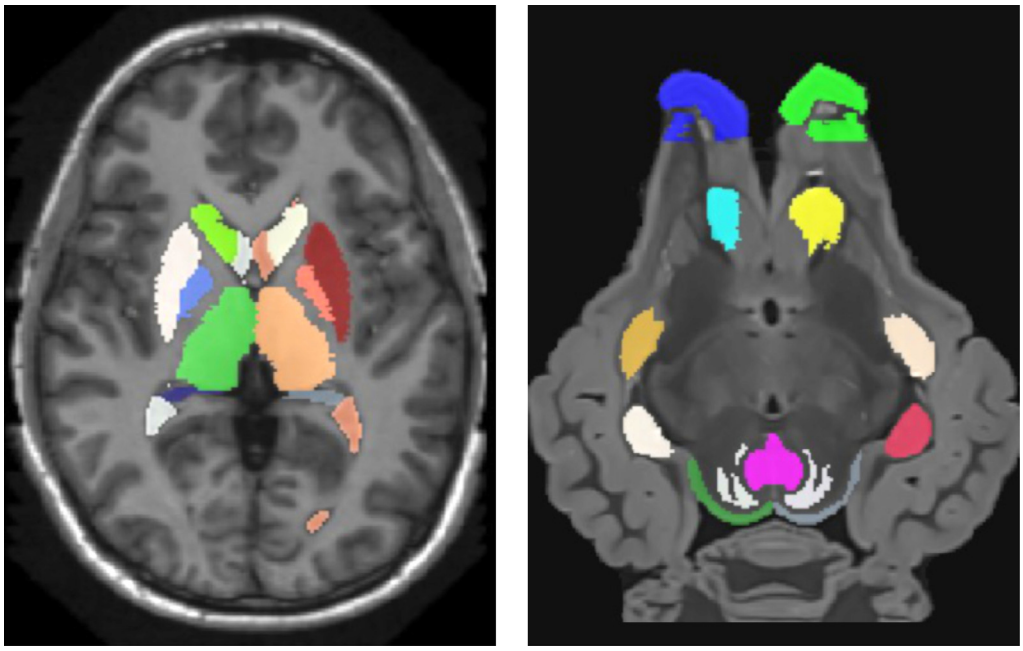

Image from the MICCAI'12 Image from the NeuroGeoEx

Fig. 5. Examples of images from the selected datasets.

Table 1
Average Dice ratio of the anatomical structures for different local classifiers (MICCAI'12 dataset).
(LIS_{GT} = all regions are well positioned with the GT, LIS_2 = the first 2 regions are well positioned according to the GT).

Method/Dice	Brain stem	Caud. Nuc.	Hippoc.	Ventricles	Pallidums	Putamens	Thalamus	Avg
LIS_{GT-CNN}	0.929	0.804	0.825	0.860	0.820	0.890	0.900	0.856
$LIS_{GT-HMRF}$	0.915	0.805	0.780	0.788	0.832	0.888	0.881	0.838
LIS_{GT-SV}	0.933	0.847	0.805	0.773	0.853	0.901	0.906	0.854
LIS_{2-CNN}	0.621	0.672	0.086	0.499	0.509	0.381	0.900	0.523
LIS_{2-HMRF}	0.901	0.738	0.694	0.730	0.732	0.853	0.881	0.811
LIS_{2-SV}	0.844	0.751	0.567	0.651	0.749	0.855	0.904	0.753

It is important to note that the choice of the local classifier is dependent on many parameters. The type of images and anatomical structures to study, the number of training images, the variability between subjects and the use of spatial relationships are important criteria which can drive this decision. Since certain of them are region dependent, it could be interesting to use a region specific classifier. This is not covered in this article but would be a good opportunity for the proposed framework.

4.3.2. Comparison with classical global methods

We compare the proposed local segmentation method with two other methods. *FreeSurfer* (Fischl, 2012) is a method that is commonly used to segment entire brains (*i.e.* subcortical as well as cortical structures). A human brain atlas is already integrated into this method so that training images are not needed for the segmentation of the test images. The *Freesurfer* process consists of a probabilistic atlas associated with Markov random fields. The joint label fusion (*JLF*) method described in Wang et al. (2013) is a multiatlas-based method that provides the best results during a challenge of multiatlas segmentation of brain images in MICCAI (Landman et al., 2012). *JLF* is based on the fusion of information computed on a patch with weighting minimizing the expected errors between the different atlases. *FreeSurfer* and *JLF* were applied to the MICCAI'12 dataset, but only *JLF* was applied to the NeuroGeoEx database.

Fig. 6 shows the Dice ratio for the human brain images. The segmentation quality of *FreeSurfer* is lower than the quality of the proposed approach; however, the *JLF* method provides better results. This result is verified for all of the 13 regions studied in this experiment. The difference between the proposed local method and *JLF* depends on the anatomical structures, and is large for ventricles (+ 10.8%) but smaller for putamen (+2.19%). The results of the proposed method with two well positioned regions (LIS_2) are also similar to the results with all regions well positioned (LIS_{GT}). Fig. 7 shows the quantitative results with the MRI test images. This image is relatively different from those in the training database, which consists mainly of elderly brains. This difference is characterized, among other factors, by ventricles with important sizes. *FreeSurfer* segmentation suffers from different errors, particularly on the putamen and pallidum, which are well managed in *JLF* and LIS segmentation. The thalamus and primarily the ventricle were better localized with *JLF* than with LIS. Elongated and thin regions, such as the ventricles, seem to be the most problematic segmentation with local atlases. Table 2 shows the global results of this experiment. The average Dice ratio of the proposed approach without using the spatial relationships (LIS_{GT}) is 7.7% greater than that with *FreeSurfer* and 5 % less than that with *JLF*.

Table 2
Average Dice ratio on the 13 anatomical structures studied (MICCAI'12 dataset).

Method	Dice ratio
LIS_1	0.689
LIS_2	0.811
LIS_{GT}	0.838
FreeSurfer	0.761
JLF	0.888

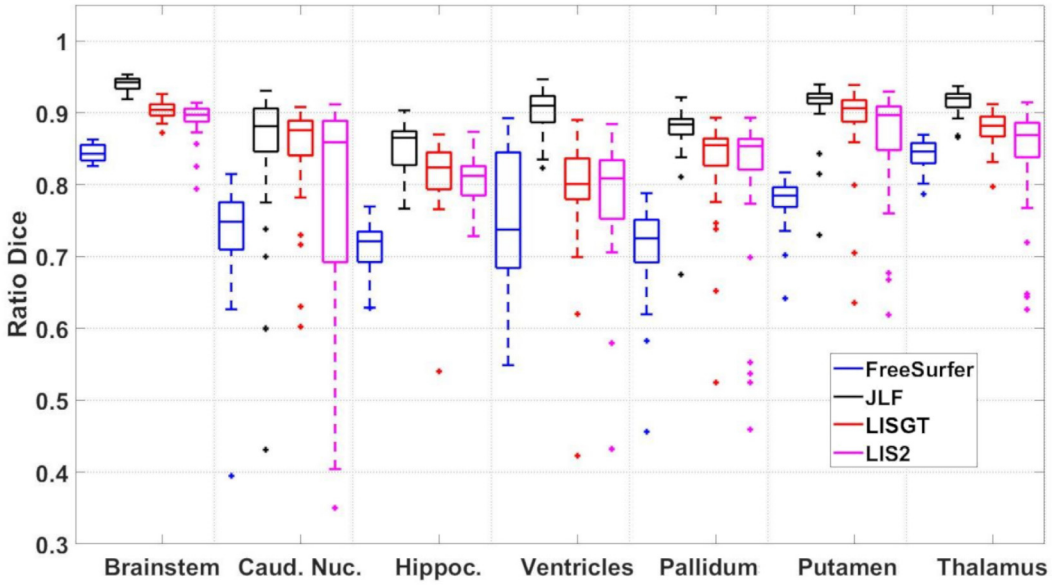

Fig. 6. Comparison of Dice score results among *FreeSurfer* (blue), *JLF* (black), LIS_{GT} (red) and LIS_2 (pink) on the MICCAI'12 dataset. The Dice ratios of bilateral structures have been averaged. Each box depicts the 25th and 75th percentiles and the central mark depicts the median. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

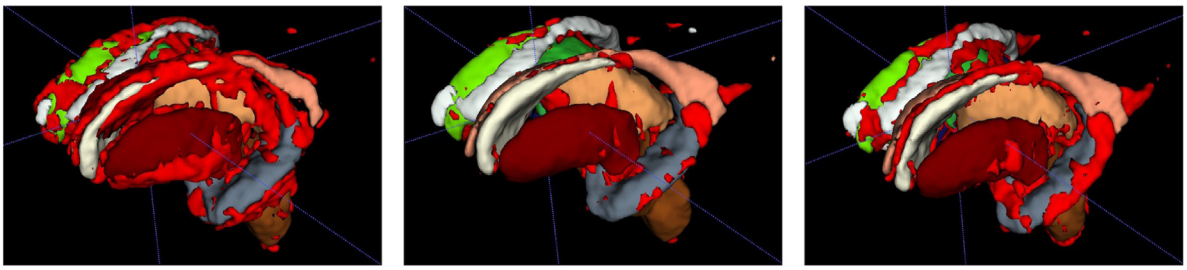

Fig. 7. Segmentation results of the MICCAI'12 data set with *FreeSurfer* (left), *JLF* (middle) and LIS (right). Errors with the ground truth are presented in red.

Fig. 8 and Table 3 show the results of the same experiments applied on the NeuroGeoEx data set. One of the primary differences between the two databases is the number of available images. Only 5 images are used during training with NeuroGeoEx. The Dice ratio of the *JLF* method is 1.7% greater than that of the proposed approach, which is less than that of human brain images. For example, the segmentation of the olfactory bulbs is better with LIS. The Dice score variability between the segmentations is also similar, which contrasts with the human brain images. This experiment tends to show that local atlas methods perform well even with a small amount of training images.

From a qualitative perspective, Fig. 9 shows the olfactory bulb segmentation provided by LIS_{GT} and *JLF*. When the local atlases are well positioned, the segmentation is more similar to the ground truth than with the *JLF* method. The external nature of this anatomical structure is not easily managed by global registration. Fig. 10 shows the segmentation of the colliculi regions. For small regions, the *JLF* method tends to provide smoother results than LIS_{GT}, as shown in its better segmentation results (*i.e.* superior colliculi in Fig. 10 (green)).

When evaluating a segmentation method, the precision of the results provided by comparison with subjective ground truth should not be the only criterion. Time complexity should also be considered. In this study, the computational times of the methods are markedly different. The other processes are applied to the entire image. For human brain images, 11 h are required to apply *FreeSurfer* to an image, and 6 h are required for JLF. With LIS, the segmentation time is dependent on the number of regions to be segmented. The computation time for 13 anatomical structures was 11 min. Also, user time was not considered but could be short in cases in which the user only positions the first two regions (*e.g.* the previous experiment). For sheep brain images, the computation time of the JLF method is shorter than that for human brain images due to the size of the training images but is still longer than that of the proposed method.

Table 3
Average Dice ratio of the 13 anatomical structures studied (NeuroGeoEx dataset).

Method	Dice ratio
LIS_{GT}	0.813
JLF	0.830

4.3.3. *Influence of the number of regions positioned with the ground truth*

Spatial relationships are usable when at least one region has already been segmented with manual positioning, but how many must be positioned before obtaining correct segmentation with the spatial relationships must be studied in more detail. This information is important because it shows the number of user interactions that is necessary to obtain satisfactory results. In our experiments, a *well-positioned* region is a region that is segmented by considering the ground truth (*i.e.* the labeled image) to position its bounding box. This ideal case can occur if the user or the spatial relationships do not include any mistakes. In the following experiment, the test database is segmented by positing different numbers of regions perfectly with the ground truth, while the other regions are positioned using the spatial relationships. The order of segmentation is fixed.

Fig. 11 shows the Dice ratio obtained for each situation. The segmentation with 13 regions well positioned depicts the case in which

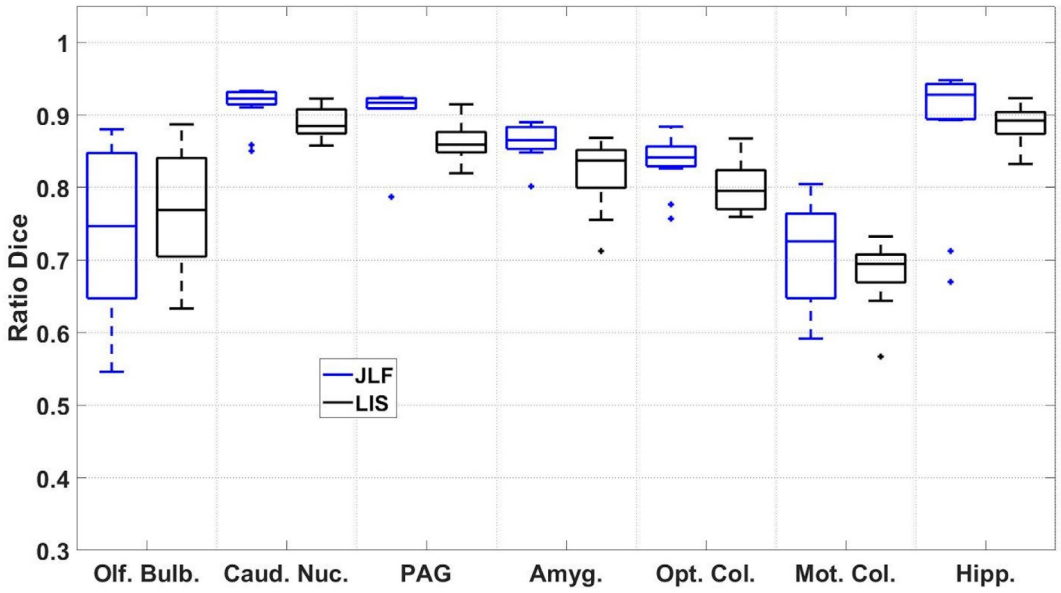

Fig. 8. Comparison of Dice score results among *JLF* and LIS_{GT} on the NeuroGeoEx dataset. The Dice ratio of bilateral structures has been average. Each box depicts the 25th and 75th percentiles and the central mark depicts the median.

Table 4
Overall comparison of SILA 3D with several methods and systems discussed in Section 2. (Comparisons between duration should be considered with caution as they come from different images and hardware configurations).

Method	JLF	Freesurfer	BIFSeg and DeepIGeos	DeepMedic	SILA 3D
Possible input data	2D/3D medical image	3D brain MRI image	3D MRI image	3D multi-channel MRI image (TBI, BRATS)	3D MRI/CT, *in vivo* or *ex vivo* brain/heart image
Required quantity of training data	Usually around 15 images, same as in our experiments	No learning step (static brain model already included)	>150 for 2D fetal and for 3D images (BRATS)	Dense training (10K patches) from at least 46 images	3 minimum for each desired ROI
Required level of interaction	None	None	All BB have to be manually positioned + manual or automatic fine tuning	None	1 BB positioning minimum + next BB repositioning if desired
Processing time	6 or 7 h for 13 regions	11 h for 13 regions	More than 65s for 1 region (tumor)	more than 65s for 1 region	11 min for 13 brain regions in a fully automatic mode - 20 to 54 min for a fully interactive segmentation of the 13 regions (90s to 250s in average for 1 region)[a]
Type of results	Full anatomical segmentation of 3D MRI image	Tissue and subcortical segmentation of human brain	Tumors segmentation	Specific lesions detection	Partial/intermediate/full segmentation of 3D MRI image

[a]Strongly user, image and region dependent.

every region has been well positioned without the use of any spatial relationships. When only one region is perfectly located, the Dice ratio was minimal, showing more incorrect and variable results, compared to the results achieved with more than two well positioned regions. However, results show that LIS_2 (and LIS_x with $x > 2$) produces a similar quality to LIS_{GT} (*i.e.* difference in the Dice ratio of less than 2.7%). These results are nearly equal to those produced by LIS_7. These results do depend on the order of segmentation and the type of anatomical structures studied but also highlight the efficiency of the proposed model in learning and using spatial relationships. This robustness is also due to the margin added around the bounding boxes compensating for small errors in the positioning.

4.3.4. Influence of the size of training database

We argue that the proposed method can be used when the number of available labeled images in the training database is small. In the following experiment, the quality of the results is evaluated depending on the number of images used to construct the AGM during the learning step.

The first AGM is built only with the first image of the training dataset, the second with the first two images, etc. Figs. 12 and 13 show the quality for each different graph with LIS_{GT} and LIS_2, respectively. LIS_2 is chosen because the segmentation of LIS_1 is too variable to highlight an evolution depending on the data used to construct the AGM. With LIS_{GT}, which uses only 4 images for graph

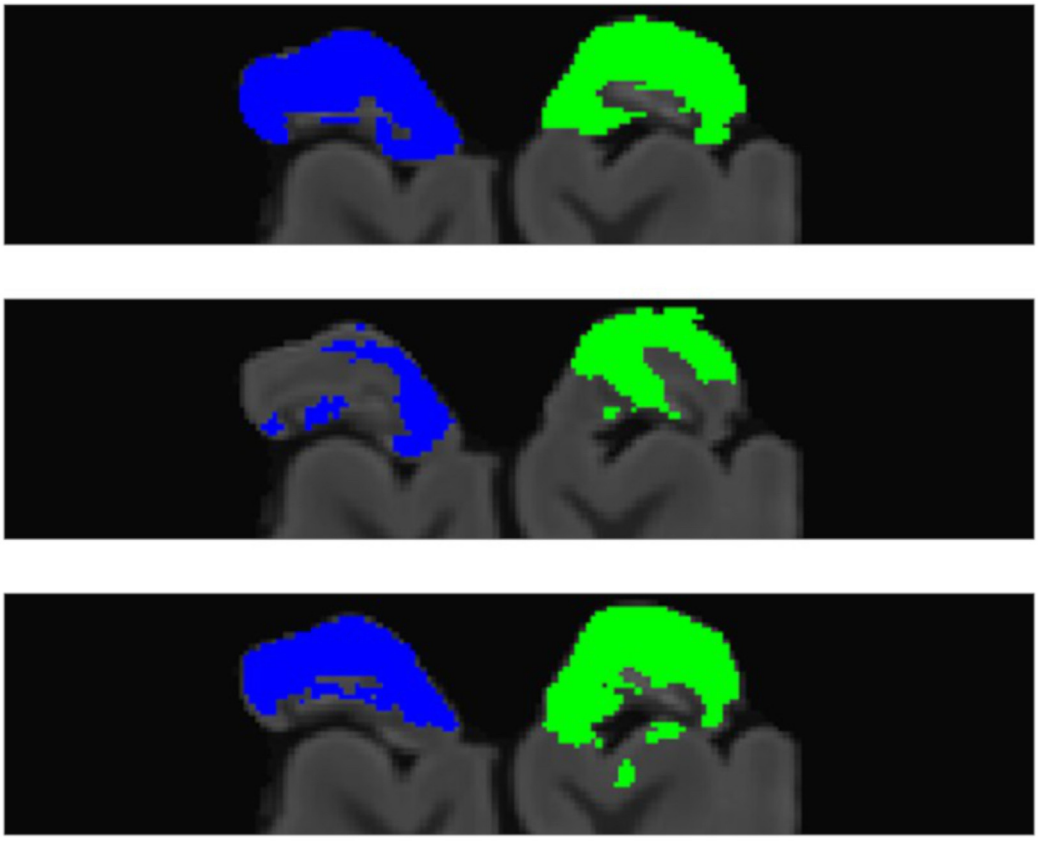

Fig. 9. Comparison of qualitative results of segmentation of the olfactory bulbs, ground truth (top), JLF (middle) and LIS_{GT} (bottom).

Fig. 10. Comparison of qualitative results of segmentation of the colliculi, ground truth (left), LIS_{GT} (middle) and JLF (right).

construction, the Dice ratio is near its maximum. A small number of images is required to attain satisfactory results, and adding more images to build the AGM does not improve the results. The construction of probabilistic information is perhaps not well adapted to a large number of images. This problem can also arise from the similarity between the training and testing images, with the first training images being more similar to the test images than the other. It would be interesting to test other sets of images to build the AGM. The same characteristic can be visualized in Fig. 13. The use of spatial relationships tends to make the process more variable, but the evolution of segmentation quality is similar to that with LIS_{GT}.

4.4. Impact of the user interaction

4.4.1. Segmentation order

Because the spatial relationships are relative between regions, the order of segmentation of the substructures impacts the final quality of the results. The quality of segmentation and the type of substructures already localized determine the quality of the information provided by the spatial relationships. In this experiment, we evaluate the influence of this order, and we validate the algorithm for automatic ordering proposed in Section 3.2.2. Experiments were performed using spatial relationship information. Only the first region is well positioned according to the ground truth. The other successive regions are positioned automatically, and the order is selected automatically using the heuristics described in Section 3.2.2.

The results of this experiment show that the medial structures of the brain are selected and segmented first. The average rank for the caudate nucleus, thalamus, pallidum and putamen was less than 7.1, while the average rank for the brainstem, ventricle and hippocampus was greater than 8.6 (among the 13 structures). The regions are also segmented hemisphere by hemisphere: all medial structures on one side are localized, followed by the medial structures of the other side. These coherence criteria are used to define an automatic order that tends to segment the medial region first, which is usually stable in terms of position and size. The external and variable regions are localized at the end of the process. This automatic order has been compared to several random or user orders and has provided better results: the segmentation quality provided is 6.3%, which is greater than the user order of 3.1% compared to a random order.

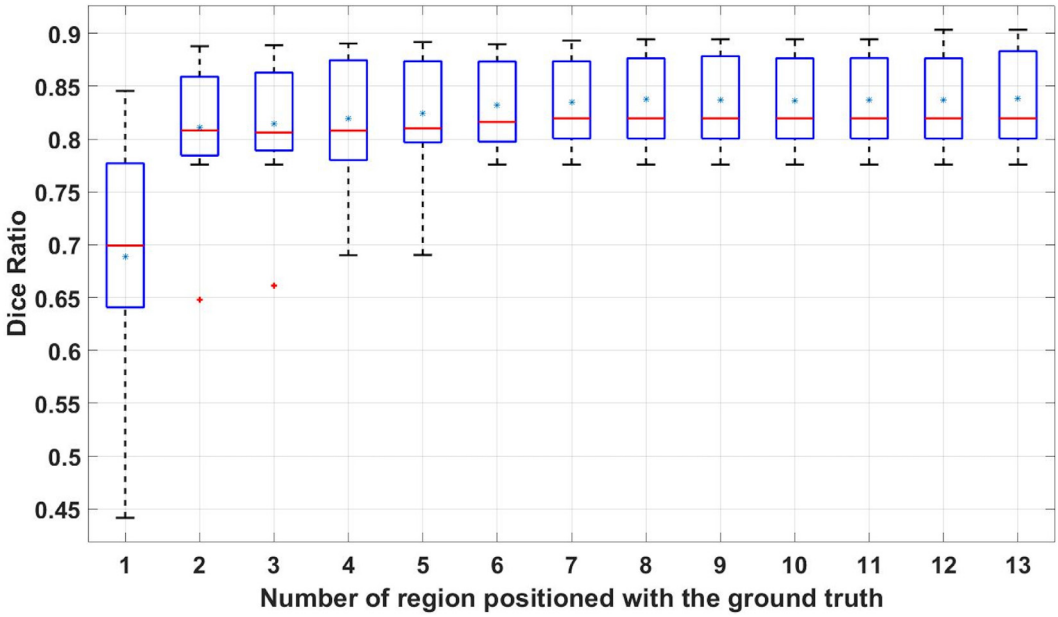

Fig. 11. Dice ratio of the segmentation with the MICCAI'12 dataset depending on the amount of regions positioned with the ground truth. Each box depicts the 25th and 75th percentiles and the central mark depicts the median.

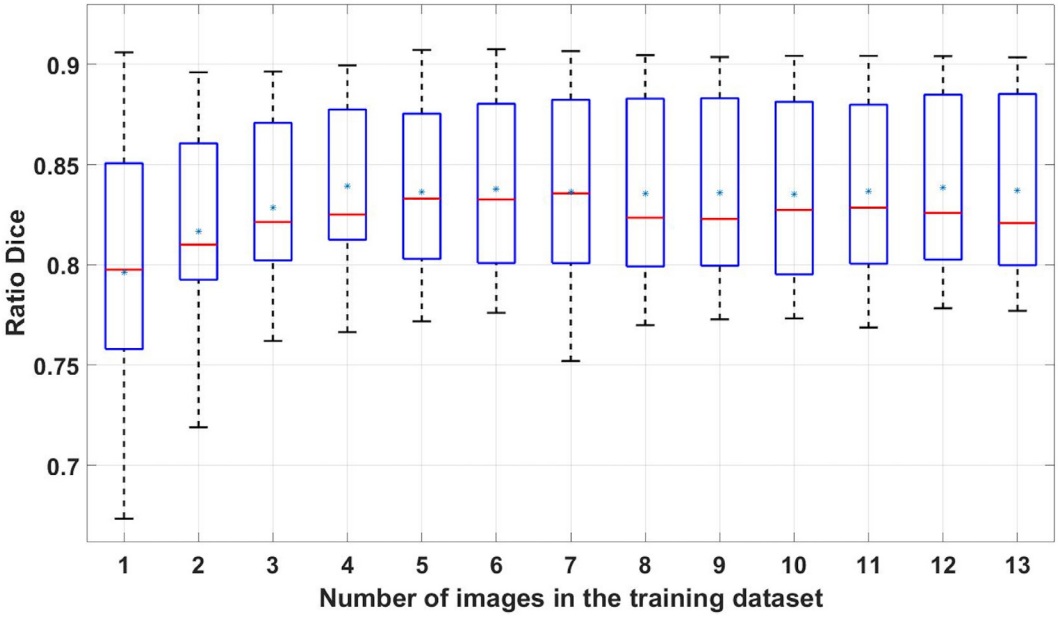

Fig. 12. Dice ratio of LIS_{GT} segmentation on the MICCAI'12 dataset depending on the amount of training images used to learn the graph. Each box depicts the 25th and 75th percentiles and the central mark depicts the median.

4.4.2. Manual and ground truth positioning comparison

In the previous section, the positioning of the bounding box was performed with the ground truth, modeling the ideal case in which the local bounding box is well positioned. In practice, the users must position the bounding box themselves or use the automatic positioning algorithm; thus, positioning will not be perfect. During these experiments, the quality of the manual positioning is evaluated compared to the ground truth, and sheep brain images are studied. The data collected during the user workshop is used to compute the different results (see Section 4.2.3). The Fig. 14 describes the average positioning error for each region in cases of automatic and manual positioning. Regions with important sizes, such as the caudate nucleus, or with visible edges, such as the amygdala, are well positioned by users; smaller structures, such as colliculi, are more difficult to localize. Automatic positioning is usually better than that performed by users, which can be explained by weak contrast and by some users not being experts in the anatomy of the sheep brain. However, the range of errors is similar to the ground truth positioning, even though this was the first time that users had applied this segmentation method.

4.4.3. Interactive improvement of the segmentation

Fig. 15 shows two examples of segmentation of sheep brain images to demonstrate the possible improvement of segmentation quality obtained with user repositioning. In these situations, the segmentation of an olfactory bulb and an amygdala shows that the ideal positioning (LIS_{GT}) leads to a markedly different segmentation result compared to the ground-truth segmentation: the size of the region is larger than it should be. In practice, when this situation occurs, the user can correct

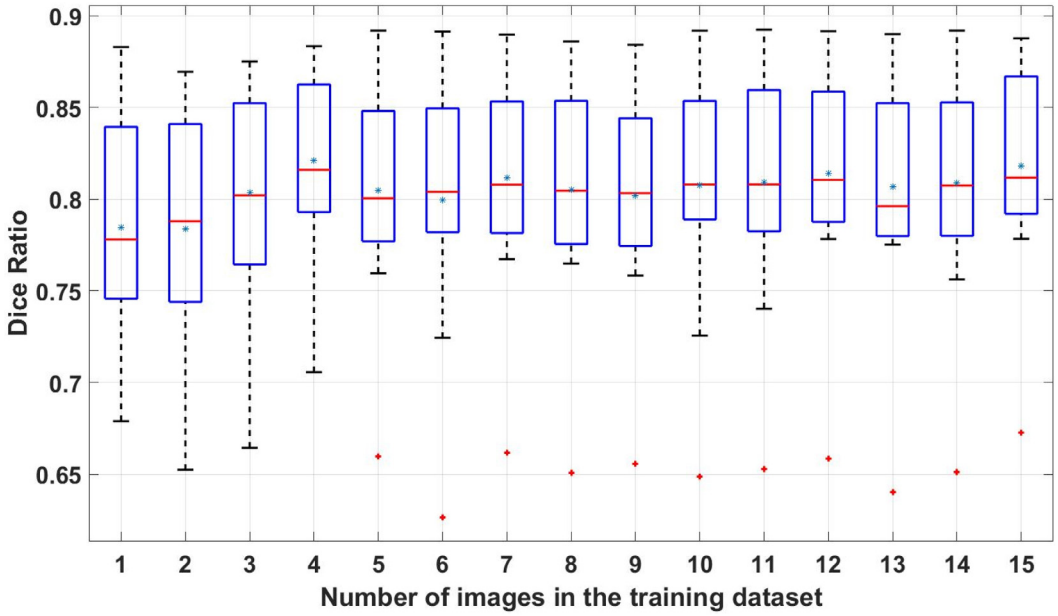

Fig. 13. Dice ratios of LIS_2 segmentation on the MICCAI'12 dataset depending on the amount of training images used to learn the graph of a priori knowledge. Each box depicts the 25th and 75th percentiles and the central mark depicts the median.

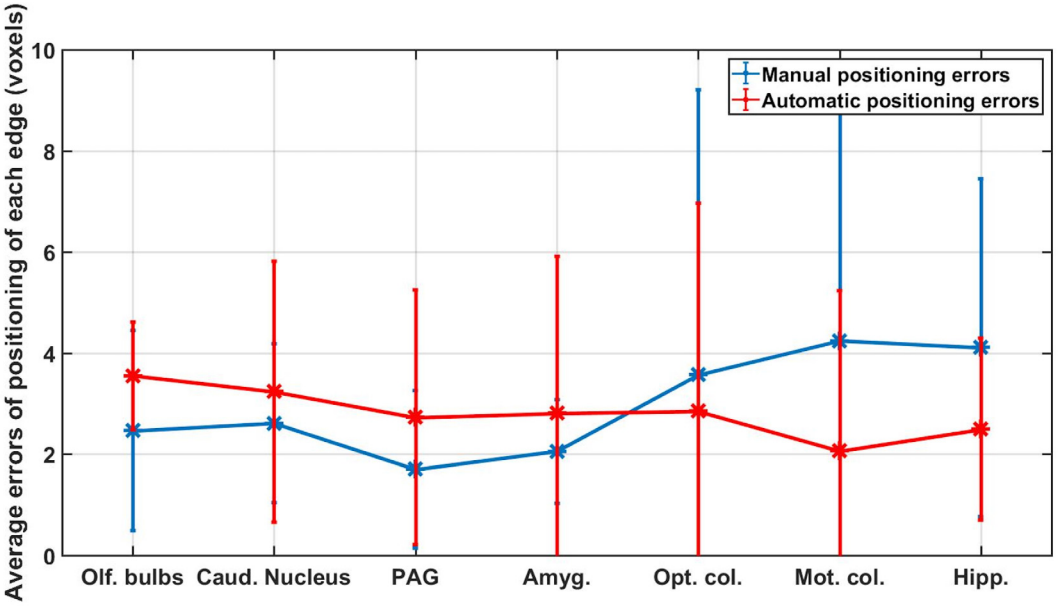

Fig. 14. Average errors of positioning on the NeuroGeoEx database from the user experiment. The automatic errors (red) and manual errors (blue) are compared. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and improve the result by modifying the bounding box. The user can cancel the segmentation and return to the previous state. Then, they can try positioning the region again to compensate for the errors. In these experiments, the bottom edges are positioned above where they are in the ideal case. Figs. 15(c) and 15(f) show the results that were more similar to the true segmentation. Different positioning implies the extraction of a different volume that can be more (or less) suitable for the co-registration process with the local atlas. Better registration drives the HMRF toward segmentation of better quality.

4.4.4. Time

In Section 4.3.2, computational time did not consider user interactions. However, user time may be important. During the user experiments, the times required to segment the sheep brain images (with 13 anatomical structures) were measured. The segmentation time decreased when the user became accustomed to the segmentation software. The average segmentation time for a new user was 74 min for the first image and 54 min for the third image, and an expert was able to segment a 3D brain image in approximately 20 min (average time).

4.5. Generalizability of the method

The proposed method is an interactive segmentation method and can be applied to various types of problems and not only brain MRI images.

4.5.1. Intermodal segmentation

In previous segmentation, training and testing images from the same modality were acquired, but the proposed approach could also

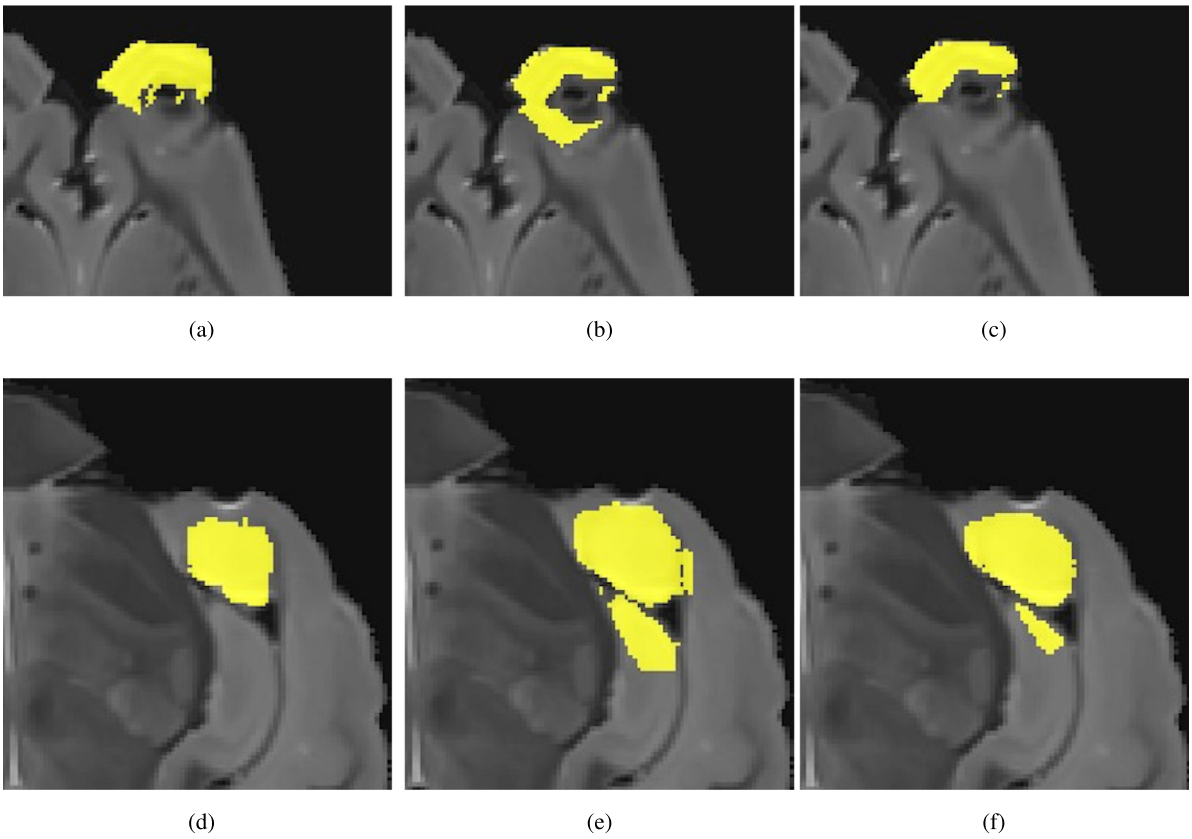

Fig. 15. 15(a) olfactory bulb ground truth, 15(b)LIS_{GT}, 15(c)LIS with manual positioning, 15(d) amygdala ground truth, 15(e)LIS_{GT} and 15(f)LIS with manual positioning.

be efficient in cases of intermodal segmentation. An AGM is learned with six images from the NeuroGeoEx database. This information was then used to segment T1 *in vivo* images of the sheep brains. In contrast to the *ex vivo* images, the brain is not extracted, and many signals are present. Seven anatomical structures are localized: left and right caudate nucleus; left and right hippocampus; left and right olfactory bulbs; and periaqueductal gray. The other regions are not visible on this image because the resolution ($0.5 \times 0.5 \times 0.5$) and contrast are less than with the NeuroGeoEx database ($0.3 \times 0.3 \times 0.3$). The intensities of the local atlas and the image to be segmented are different, so the metric used during the registration process is mutual information. The results presented are qualitative due to a lack of ground truth. Two segmentations are shown in Fig. 16; the first image 16(a) comes from an animal also present in the NeuroGeoEx database, and segmentation is performed manually (*i.e.* users interact when necessary); the second image 16(d) does not come from an animal present in the training database and is segmented automatically, except for the first region (a caudate nucleus), which is manually positioned. When the process is driven by an expert, segmentation is consistent. The caudate nucleus, hippocampus and PAG do not suffer from particular problems with the first images. The olfactory bulbs are more difficult and seem to be too small compared to what they should look like. When the bounding boxes are automatically set, the segmentation is more variable but still coherent. The caudate nucleus and the hippocampus are well localized. The PAG is distorted but still in a good position. However, the olfactory bulbs are shifted compared to the real position. The spatial relationships that link the medial regions and external regions of the brain are less accurate, which is also the case in intermodal situations. These qualitative results emphasize that the proposed method is usable for intermodal segmentation. The spatial relationships are still efficient for the medial regions of the brain, although the resolution and the context (*in vivo/ex vivo*) are different. Due to the local atlases, the presence of tissues around the brain in the *in vivo* images is not a problem.

4.5.2. Heart images

3D MRI and CT scans of the heart are processed with the proposed approach. The AGM is built from 20 MRI images of the heart and applied to 40 MRI images. This database comes from the Multi-Modality Entire Heart Segmentation Challenge (Zhuang et al., 2010; Zhuang & Shen, 2016). The same process is performed with the CT scan images, and the images are composed of 8 anatomical structures: left and right ventricle blood cavities, left and right atrium blood cavities, myocardium of the left ventricle, ascending aorta and pulmonary artery. Mutual information is used as a metric for the registration (training and testing) of the MRI images. The segmentation of the CT scan was performed automatically; only the left ventricle was positioned manually, and the other structures were positioned with spatial relationships. The segmentation of the MRI images is thus driven more by the user, and the user manually positions the bounding box if the quality of automatic positioning is poor. Two example results are shown in Fig. 17. The high resolution of the CT scan images allows satisfactory results to be obtained, even with automatic positioning. Overall, the shape of the anatomical structures conforms to the ground truth. A small amount of error comes from a non-smooth border between the regions, as shown in Fig. 17(b)). MRI images are more difficult to segment due to field inhomogeneity artifacts and their lower resolutions, which is why users interacted during segmentation in contrast to the CT scan images. The segmentation quality of the ventricles or the myocardium is satisfactory, but the segmentation of the arteries is more variable. The global Dice ratios on the CT scan images are 83.7 % and 81.7% for the MRI images. These qualities are marginally lower than those of the other methods that are based on deep learning. However, this last method considered only heart images in contrast to the proposed method. Readers can also refer to Zhuang et al. (2019) to see the initial results of proposed method during the Multi-Modality Whole Heart Segmentation challenge organized in conjunction with MICCAI 2017.

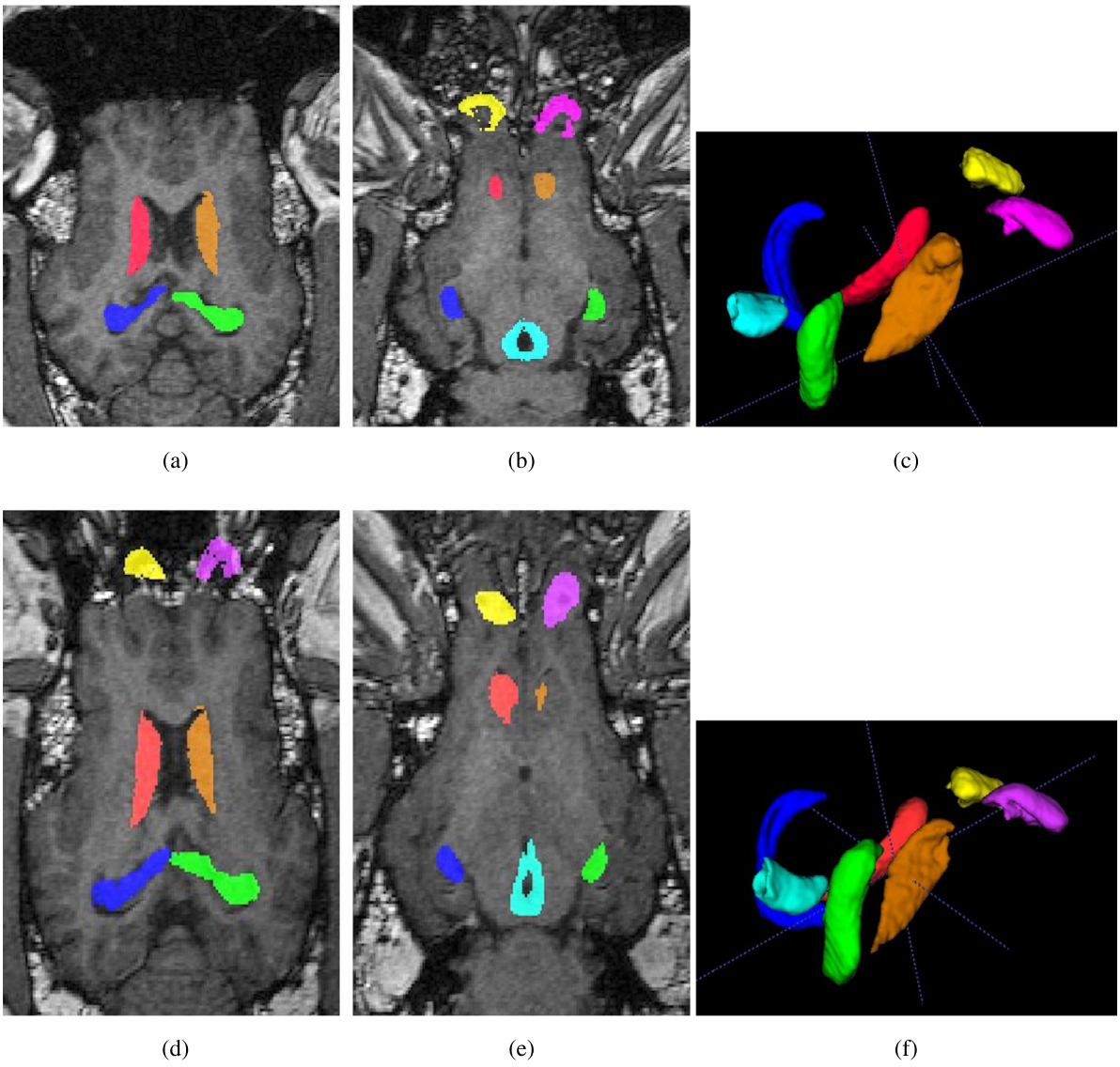

Fig. 16. Segmentation of *in vivo* sheep brain images, (a–c) LIS with a manual positioning, (d–e) LIS and an automatic positioning.

5. Discussion and conclusion

In this article, a new interactive machine learning method for 3D medical image segmentation was proposed. In contrast to the classical approach, in which anatomical structures are modeled inside an organ, the proposed method describes each subpart locally by learning a specific local model and storing it in an Anatomical Graph Model. The global aspect of the anatomical structure is learned by analyzing the relative spatial relationship and size between each possible pair of subparts. The edges of the Anatomical Graph Model are used to store this structural information. Visual information (shape and intensity) is then dissociated from the structural information (relative position and size information), but all information is stored in a unified structure (the Anatomical Graph Model). Each local model can be used independently from the others to enable partial segmentation. The segmentation of an image becomes an incremental process allowing user interaction before and after each local segmentation.

We also proposed an algorithm to determine the best order for the local segmentations automatically based on the coherence between spatial relationships. This order provides a good alternative to user interaction and allows the system to be used in nearly automatically if desired. These orders of segmentation have been shown to be efficient compared to user or yo random order of segmentation.

The experiments presented in this study enhanced the small training databases that are required to use the proposed method efficiently. Only 4 images were sufficient to reach a level of segmentation quality that equal that of other systems evaluated on the MICCAI'12 database. This information is critical in cases of segmentation of unusual data sets (*e.g.* organs or species).

Incremental and local segmentation allows for partial and rapid segmentation of images compared to classical global methods. Unlike such methods, SILA 3D does not require a global registration step. The computation time of segmentation of a few anatomical structures is significantly lower than that of a classical atlas-based method. The LIS method can use various types of local decision processes during the voxel classification step. The primary criteria that should be considered are (1) keeping the system sufficiently fast so that it can be used interactively, and (2) requiring a small number of training images. For instance, a graph-cut or more efficient multiatlas fusion method could also be used to classify the voxels inside a bounding box.

The proposed method could be used with organs that have a large number of different structures, as well as when users do not have to always segment the same regions. However, the proposed method is, for example, not well adapted to the segmentation of structures with a bounding box that is the same size as the organ (*e.g.* the white matter

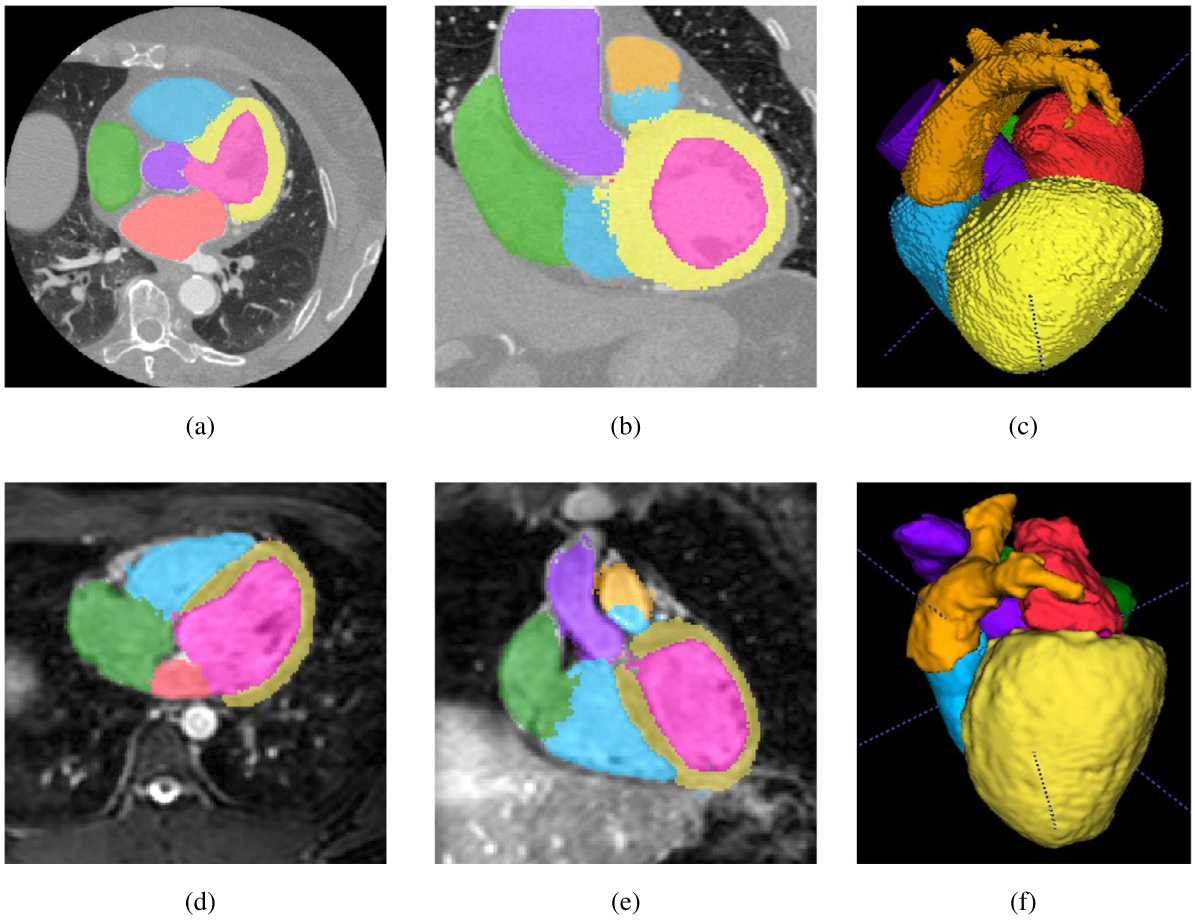

Fig. 17. Segmentation of CT scan heart images (a–c). Segmentation of MRI images (d–f).

region of the brain). In this situation, classical atlases are more suitable for segmenting many anatomical structures at once.

The proposed method has also been shown to be generalizable and efficient in various types of applications. The proposed method is directly applicable to different organs (heart) or different imaging modalities (T1, T2, CT scan, etc.), although it was first designed to segment of MRI brain images. The local modeling and relative nature of spatial relationships provide robust results, and relative distances can also be efficient when the resolution or global size of the anatomical structures is different. Thus, segmentation of the growing neonatal brain should be explored with this approach. One of the other advantages of the proposed model is that it stores information from different modalities; thus, it could be possible to model information from MRI and CT scan images in only one graph.

We also experimented with the impact of interactivity during a user experiment. The quality of the results and interaction were satisfactory but variable. First, the user must understand the anatomical structures of the imaged organ. Among all users, experts of brain anatomy or medical images produced significantly better results. User feedback about the quality of the segmentation and positioning was positive, and users mentioned the rapidity of the proposed segmentation method. However, the method's primary drawback was linked to human–computer interaction, which is an important parameter of the proposed process. Better visualization of the segmentation results and more ergonomic positioning should significantly improve users' experiences. Finally, the user experiment highlighted the ownership of the software and the process; after segmentation a few images, users became more comfortable with the segmentation tools.

SILA 3D thus achieves similar or higher accuracies with fewer user interactions in less time than traditional interactive segmentation methods. The proposed method has been integrated into an operational framework that is available online and has been used in several studies and different kinds of projects. It is probably true that this framework cannot compete with a finely tuned method dedicated to a specific dataset/population or a classification task where huge annotated data are available to learn a deep learning architecture fully automatically, but this situation is not so common in real life applications.

SILA 3D is rather dedicated to applications where very few annotated data are available or when there is a high variability inside the images to be segmented (global registration will become unusable). Local approach and interaction are then mandatory to obtain a correct result in such conditions. As demonstrated in the experiments section, SILA 3D is able to work with very few annotated images (3 is enough), does not require a full brain registration, allows the user to easily visualize and improve intermediate results (incrementality), and also tries to reduce the number of interactions by including a semi-automatic positioning of bounding boxes. Until today, these features have led to the use of SILA 3D for the following real-world applications: (i) in neuroethology for the exploration of new ROIs for which biologists and neuroanatomists do not have digital atlases; (ii) for Deep Brain Stimulation in order to segment specific parts of the brain in very high resolution images; (iii) faster creation of initial atlases of specific populations (lifetime, analysis of specific substructures, . . .) or of initial annotated datasets used thereafter to train CNN.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding and acknowledgments

This work was mainly supported by the Région Centre Val de Loire in France during NeuroGeo (2015–2017, n° 00091714) and Neuro2co (2017–2020, n° 00117257) projects. We thank Hans Adriansen for the *in vivo* MRI acquisitions performed at (CIRE-INRAE) and Cyril Poupon for the *ex vivo* MRI acquisitions performed at (NeuroSpin-CEA). These images are included in the sheep brain image datasets used for the experiments. We are also grateful to all the participants to the SILA 3D user workshop and especially to Ophélie Menant for her huge help regarding the data annotation.

References

Acuna, D., Ling, H., Fidler, S., & Kar, A. (2018). Efficient interactive annotation of segmentation datasets with polygon-RNN++. In *Conf. on computer vision and pattern recognition* (pp. 859–868). http://dx.doi.org/10.1109/CVPR.2018.00096.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851. http://dx.doi.org/10.1016/j.neuroimage.2005.02.018.

Bai, W., Shi, W., O'Regan, D. P., Tong, T., Wang, H., Jamil-Copley, S., Peters, N. S., & Rueckert, D. (2013). A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac MR images. *IEEE Transactions on Medical Imaging*, *32*(7), 1302–1315. http://dx.doi.org/10.1109/TMI.2013.2256922.

Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: A review. *Image and Vision Computing*, *23*(2), 89–110.

Boykov, Y. Y., & Jolly, M. (2001). Interactive graph cuts for Optimal Boundary Region segmentation of objects in N-D images. In *Proceedings eighth IEEE international conf. on computer vision, vol. 1* (pp. 105–112 vol.1). http://dx.doi.org/10.1109/ICCV.2001.937505.

Castrejón, L., Kundu, K., Urtasun, R., & Fidler, S. (2017). Annotating object instances with a polygon-RNN. In *2017 IEEE conf. on computer vision and pattern recognition* (pp. 4485–4493). http://dx.doi.org/10.1109/CVPR.2017.477.

Çiçek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from sparse annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), *Lecture notes in computer science*, *Medical image computing and computer-assisted intervention* (pp. 424–432). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

Chen, H., Qi, X., Yu, L., & Heng, P. (2016). DCAN: deep contour-aware networks for accurate gland segmentation. In *2016 IEEE Conf. on computer vision and pattern recognition* (pp. 2487–2496). IEEE Computer Society, http://dx.doi.org/10.1109/CVPR.2016.273.

Criminisi, A., Sharp, T., & Blake, A. (2008). GeoS: Geodesic image segmentation. In *Proceedings of ECCV* (pp. 99–112).

de Brébisson, A., & Montana, G. (2015). Deep neural networks for anatomical brain segmentation. In *2015 IEEE conf. on computer vision and pattern recognition workshops* (pp. 20–28). http://dx.doi.org/10.1109/CVPRW.2015.7301312.

Dou, Q., Yu, L., Jin, Y., Yang, X., Qin, J., Heng, P.-A., & Chen, H. (2017). 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*, *41*, 40–54. http://dx.doi.org/10.1016/j.media.2017.05.001.

Ella, A., & Keller, M. (2015). Construction of an MRI 3D high resolution sheep brain template. *Magnetic Resonance Imaging*, *33*(10), 1329–1337. http://dx.doi.org/10.1016/j.mri.2015.09.001.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J., Pieper, S., & Kikinis, R. (2012). 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, *9*, 1323–1341, URL https://www.slicer.org/.

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. http://dx.doi.org/10.1016/j.neuroimage.2012.01.021.

Galisot, G., Brouard, T., Chaillou, E., & Ramel, J. Y. (2019). A comparative study on voxel classification methods for atlas based segmentation of brain structures from 3D MRI images. In *14th International joint conf. on computer vision, imaging and computer graphics theory and applications, VISIGRAPP 2019, Vol. 4* (pp. 341–350).

Iglesias, J. E., Sabuncu, M. R., & Van Leemput, K. (2013). A probabilistic, non-parametric framework for inter-modality label fusion. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Eds.), *Medical image computing and computer-assisted intervention* (pp. 576–583). Springer Berlin Heidelberg.

Klein, A., Mensh, B., Ghosh, S., Tourville, J., & Hirsch, J. (2005). Mindboggle: Automated brain labeling with multiple atlases. *BMC Medical Imaging*, *5*(1), 7. http://dx.doi.org/10.1186/1471-2342-5-7.

Konstantinos, K., Ledig, C., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., Glocker, B., & Newcombe, V. F. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, *36*, 61–78. http://dx.doi.org/10.1016/j.media.2016.10.004.

Koohbanani, N. A., Neda Zamani Tajadin, M. J., & Rajpoot, N. (2020). NuClick: A deep learning framework for interactive segmentation of microscopic images. *Medical Image Analysis*, *65*, Article 101771.

Kovacevic, N., Henderson, J., Chan, E., Lifshitz, N., Bishop, J., Evans, A., Henkelman, R., & Chen, X. (2005). A three-dimensional MRI atlas of the mouse brain with estimates of the average and variability. *Cerebral Cortex*, *15*(5), 639–645. http://dx.doi.org/10.1093/cercor/bhh165.

Landman, B. A., Warfield, S. K., Hammers, A., Akhondi-Asl, A., Asman, A. J., Ribbens, A., Lucas, B., Avants, B. B., Ledig, C., Ma, D., Rueckert, D., Vandermeulen, D., Maes, F., Holmes, H., Wang, H., Wang, J., Doshi, J., Kornegay, J., Hajnal, J. V., ... Heckemann, R. A. (2012). In B. A. Landman, & S. K. Warfield (Eds.), *MICCAI 2012 grand challenge and workshop on multi-atlas labeling*: *Technical report*, CreateSpace Independent Publishing Platform.

Leprince, Y., Schmitt, B., Chaillou, E., Destrieux, C., Barantin, L., Vignaud, A., Rivière, D., & Poupon, C. (2015). Optimization of sample preparation for MRI of formaldehyde-fixed brains. In *23rd annual meeting of ISMRM* (p. 1). International Society for Magnetic Resonance in Medicine.

Li, W., Wang, G., Ourselin, S., Cardoso, M. J., Vercauteren, T., & Fidon, L. (2017). On the compactness, efficiency, and representation of 3D convolutional networks: Brain parcellation as a pretext task. In *Information processing in medical imaging* (pp. 348–360). Springer International Publishing.

Lin, D., Dai, J., Jia, J., He, K., & Sun, J. (2016). ScribbleSup: scribble-supervised convolutional networks for semantic segmentation. In *2016 IEEE conf. on computer vision and pattern recognition* (pp. 3159–3167). IEEE, http://dx.doi.org/10.1109/CVPR.2016.344.

Ling, H., Kar, A., Chen, W., Fidler, S., & Gao, J. (2019). Fast interactive object annotation with curve-gcn. In *Proceedings of the IEEE/CVF conf. on computer vision and pattern recognition* (pp. 5257–5266).

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. http://dx.doi.org/10.1016/j.media.2017.07.005.

Matsakis, P., & Wendling, L. (1999). A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(7), 634–643. http://dx.doi.org/10.1109/34.777374.

McKinley, R., Wepfer, R., Gundersen, T., Wagner, F., Chan, A., Wiest, R., & Reyes, M. (2016). Nabla-net: A deep dag-like convolutional architecture for biomedical image segmentation. *BrainLes@MICCAI*, 119–128.

Mehtal, R., Majumdar, A., & Sivaswamy, J. (2017). BrainSegNet: A convolutional neural network architecture for automated segmentation of human brain structures. *Journal of Medical Imaging*, *4*(2), 1–11. http://dx.doi.org/10.1117/1.JMI.4.2.024003.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth international conf. on 3D vision* (pp. 565–571).

Moeskops, P., Wolterink, J. M., Gilhuijs, K. G. A., Leiner, T., Viergever, M. A., Išgum, I., & van der Velden, B. H. M. (2016). Deep learning for multi-task medical image segmentation in multiple modalities. In *Medical image computing and computer-assisted intervention* (pp. 478–486). Springer International Publishing.

Nitzsche, B., Frey, S., Collins, L. D., Seeger, J., Lobsien, D., Dreyer, A., Kirsten, H., Stoffel, M. H., Fonov, V. S., & Boltze, J. (2015). A stereotaxic, population-averaged T1w ovine brain atlas including cerebral morphology and tissue volumes. *Frontiers in Neuroanatomy*, *9*, 69. http://dx.doi.org/10.3389/fnana.2015.00069.

Nyul, L. G., Udupa, J. K., & Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, *19*(2), 143–150. http://dx.doi.org/10.1109/42.836373.

Pohl, K. M., Fisher, J., Grimson, W., L., E., Kikinis, R., & Wells, W. M. (2006). A Bayesian model for joint segmentation and registration. *NeuroImage*, *31*(1), 228–239. http://dx.doi.org/10.1016/j.neuroimage.2005.11.044.

Poon, M., Hamarneh, G., & Abugharbieh, R. (2008). Efficient interactive 3D livewire segmentation of complex objects with arbitrary topology. *Computerized Medical Imaging and Graphics*, *32*(8), 639–650. http://dx.doi.org/10.1016/j.compmedimag.2008.07.004.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention* (pp. 234–241). Springer International Publishing.

Scherrer, B., Forbes, F., Garbay, C., & Dojat, M. (2009). Distributed local MRF models for tissue and structure brain segmentation. *IEEE Transactions on Medical Imaging*, *28*(8), 1278–1295. http://dx.doi.org/10.1109/TMI.2009.2014459.

Voronin, P., Vetrov, D., & Ismailov, K. (2013). An approach to segmentation of mouse brain images via intermodal registration. *Pattern Recognition and Image Analysis*, *23*(2), 335–339. http://dx.doi.org/10.1134/S105466181302017X.

Wang, X., & Keller, J. M. (1999). Human-based spatial relationship generalization through neural/fuzzy approaches. *Fuzzy Sets and Systems*, *101*(1), 5–20. http://dx.doi.org/10.1016/S0165-0114(97)00035-3.

Wang, G., Li, W., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S., Vercauteren, T., & Zuluaga, M. A. (2018). Interactive medical

image segmentation using deep learning with image-specific fine tuning. *IEEE Transactions on Medical Imaging*, *37*(7), 1562–1573. http://dx.doi.org/10.1109/TMI.2018.2791721.

Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., & Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(3), 611–623. http://dx.doi.org/10.1109/TPAMI.2012.143.

Wang, G., Zuluaga, M. A., Pratt, R., Aertsen, M., Doel, T., Klusmann, M., David, A. L., Deprest, J., Vercauteren, T., & Ourselin, S. (2016). Slic-seg: A minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views. *Medical Image Analysis*, *34*, 137–147. http://dx.doi.org/10.1016/j.media.2016.04.009.

Wang, G., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S., Vercauteren, T., & Li, W. (2019). DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(7), 1559–1572.

Yeo, B. T., Sabuncu, M. R., Desikan, R., Fischl, B., & Golland, P. (2008). Effects of registration regularization and atlas sharpness on segmentation accuracy. *Medical Image Analysis*, *12*(5), 603–615.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, *31*(3), 1116–1128. http://dx.doi.org/10.1016/j.neuroimage.2006.01.015.

Zhao, L., & Jia, K. (2016). Multiscale CNNs for brain tumor segmentation and diagnosis. *Computational and Mathematical Methods in Medicine*, *2016*, 1–7.

Zhao, F., & Xie, X. (2013). An overview of interactive medical image segmentation. *Annals of the BMVA*, *2013*(7), 1–22.

Zhuang, X., Li, L., Payer, C., Stern, D., Urschler, M., Heinrich, M. P., Oster, J., Wang, C., Smedby, O., Bian, C., Yang, X., Heng, P., Mortazi, A., Bagci, U., Yang, G., Sun, C., Galisot, G., Ramel, J., & Yang, G. (2019). Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. *Medical Image Analysis*, *58*, http://dx.doi.org/10.1016/j.media.2019.101537.

Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J., & Ourselin, S. (2010). A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Transactions on Medical Imaging*, *29*(9), 1612–1625. http://dx.doi.org/10.1109/TMI.2010.2047112.

Zhuang, X., & Shen, J. (2016). Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Medical Image Analysis*, *31*, 77–87. http://dx.doi.org/10.1016/j.media.2016.02.006.