CNN-based segmentation with a semi-supervised approach for automatic cortical sulci recognition

Léonie Borne LEONIE.BORNE@CEA.FR
Jean-François Mangin JEAN-FRANCOIS.MANGIN@CEA.FR
Denis Rivière DENIS.RIVIERE@CEA.FR
Neurospin, CEA Saclay, 91191 Gif-sur-Yvette, France

Editors: Under Review for MIDL 2019

Abstract

Despite the impressive results of deep learning models in computer vision, these techniques have difficulty achieving such high performance in medical imaging. Indeed, two challenges are inherent in this domain: the rarity of labelled images, while deep learning methods are known to be extremely data intensive, and the large size of images, generally in 3D, which considerably increases the need for computing power. To overcome these two challenges, we choose to use a simple CNN that tries to classify the central voxel of a 3D patch given to it as an input, while exploiting a large unlabelled database for pretraining. Thus, the use of patches limits the size of the neural network and the introduction of unlabelled images increases the amount of data used to feed the network. This semi-supervised approach is applied to the recognition of the cortical sulci: this problem is particularly challenging because it contains as many structures to be recognized as labelled subjects, i.e. only about sixty, and these structures are extremely variable. The results show a significant improvement compared to the BrainVISA model, the most used sulcus recognition toolbox.
Keywords: CNN, segmentation, semi-supervision, cortical sulci.

1. Introduction

1.1. Why automatic cortical sulci recognition?

The cortical surface is made up of many convolutions, called gyri, delimited by folds, called sulci. The main sulci provide a kind of road map delimiting functionally different regions. The shape of the sulci (length, depth, etc.) is used as biomarkers of developmental and neurodegenerative diseases. Despite the many tools available to visualize sulci in 3D, their labelling according to the anatomical nomenclature is long and fastidious. For each brain, it takes at least one hour for an expert to label all the sulci. However, because of the large variability of the folding pattern in the general population, inferring reproducible biomarkers requires the mining of thousands of brains. Hence the automation of sulcus recognition is essential.

Nevertheless, learning to label the cortical sulci is a complex challenge for several reasons. First of all, the sulci are **highly variable structures**, some of them are even present in only 30% of subjects. In addition, each brain contains more than **120 different sulci**, but few segmentation algorithms are designed to deal with such a large amount of struc-

© 2019 L. Borne, J.-F. Mangin & D. Rivière.

tures. Moreover, from one sulcus to another, the average size can vary from a simple to a hundredfold. Hence, **the voxelwise problem is particularly unbalanced**. Finally, the **number of manually labelled subjects currently available is very limited**.

Many methods have been proposed to take up this challenge. A large part of the algorithms rely on graph-based representations, allowing the representation of relative positions of the sulci as well as their location in a standardized space (Blida, 2014; Rivière et al., 2002; Royackkers et al., 1998; Shi et al., 2007; Vivodtzev et al., 2006; Yang and Kruggel, 2009). In order to improve the recognition robustness, other methods have sought to model inter-subject variability using several frameworks ranging from PCA to Bayesian approaches, including multi-atlas segmentation (Behnke et al., 2003; Borne et al., 2018; Fischl et al., 2004; Lohmann and von Cramon, 2000; Perrot et al., 2011). To the best of our knowledge, despite their current popularity, no CNN has yet been proposed for cortical sulci recognition.

1.2. How to manage large image sizes?

CNNs have proven their worth in 2D image analysis, but when it comes to adapting these networks to 3D, the increase in the number of parameters is such that the computational cost becomes a limiting or even blocking factor. For this first CNN-based approach dedicated to the cortical sulci recognition, we have chosen to adapt the method proposed in (Ciresan et al., 2012) to 3D segmentation. In this article, a CNN is trained to classify the central voxel of an input patch. Thus, thanks to this patch-based approach, the impact of the transition to 3D is limited, allowing, among other things, the use of a **3D patch and 3D convolutions**, which would have been extremely expensive otherwise. Moreover, since the sulci recognition is generally based on the patterns of neighbouring sulci, **this problem is particularly suitable for a local patch approach**. Finally, thanks to the BrainVISA tools (www.brainvisa.info), it is already possible to precompute a voxel-based representation of the unlabeled folds based on the skeleton of a negative mold of the brain. Thus, using this preprocessing, the sulcus recognition amounts to **labeling only the voxels belonging to a fold** and not the background, which represents a **considerable time saving**.

1.3. How to compensate for the limited number of labelled images?

In medical imaging, manual labelling of an image is generally too long to obtain large training databases for segmentation, whereas it is essential to train a CNN. Considering segmentation as a classification of voxels, as explained above, the **number of training samples is already considerably increased**, but these samples are extremely redundant because of the overlap between adjacent patches.

However, it regularly happens that huge unlabelled databases are easily accessible and remain unused. In order to exploit this data, we propose to first train a CNN on the labelled database, then apply the obtained model to the massive unlabelled datebase that will be used for the pre-training of a new neural network that will then be finetuned on the labelled database. This **semi-supervised approach** makes it possible to better represent the variability of possible patterns and considerably increase the number of labelled images.

2. Method

The description of the model is organized in three steps: first, the extraction of the fold skeleton from the MRI, second, the classification of skeleton voxels using a simple CNN, third, the spatial regularization of the results, via the use of elementary folds or adjacent points.

2.1. Fold extraction

Thanks to the BrainVISA pipeline, already widely used for studying cortical anatomy, the folds are represented by a set of voxels corresponding to a skeleton of the cerebrospinal fluid filling the fold. This representation of the folds can be understood as a negative mold of the brain. BrainVISA is also providing a clustering of the skeleton voxels into elementary cortical folds, the building blocks of cortical morphology.

2.2. Voxel classification thanks to CNN

The approach proposed in (Ciresan et al., 2012) addresses a segmentation problem as a classification of each voxel based on its environment contained in a patch. In our case, as only voxels belonging to the fold skeleton need to be classified, it reduces the number of voxels to be identified by a factor of 10^3.

2.2.1. Patch design

As the data are particularly affected by the type of MRI sequence, the age of the subject or even the pathologies, it was chosen to represent only the skeleton of the folds into the patches, in order to normalize the data as much as possible. Thus, when a patch voxel belongs to the fold skeleton, it is assigned the value one, the other voxels have the value zero.

An additional normalization lies in the affine alignment of all the brains to the standard Talairach space using a common resolution of 2*2*2mm which is sufficient for our problem. We fixed the size of the patches a priori at 60*60*60mm because it seemed large enough to learn the variability of the sulci patterns and small enough to limit the calculation costs.

2.2.2. Network design

The neural network used is a 3D adaptation of the famous LeNet initially proposed in (LeCun et al., 1998), with additional ReLU regularization layers (after each convolutional layer and each linear layer). We have chosen to use 3D convolutional layers, despite the exponential increase in the number of parameters that this involved, because the nature of the problem leads us to believe that a 2D multi-view network as in (Su et al., 2015) would not be sufficient. Indeed, for a given sulcus, its depth, length, position, neighbourhood, etc. are all essential parameters for its recognition, whereas they are difficult to access in 2D.

At the output of the neural network, a score per sulcus label is obtained. An additional score is calculated for the label "unknown", also present in the training database, and for the ventricles that are not considered as sulci but are included in the BrainVISA skeleton.

2.2.3. Training design

First, the CNN described above is **trained on the manually labelled database**. Then, this model is used to **label 500 unlabelled subjects**, using regularization by elementary folds, as explained in the following section. A new CNN is then **pre-trained on the 500 automatically labelled subjects** and then **finetuned on the manually labelled database**.

During pretraining, at each epoch, 100 points are randomly selected by subject to participate in the training. The model obtained after 15 epochs is selected to proceed further with the training process.

During the first training step and the finetuning step, the same scheme is used: 10% of the subjects are selected for the validation set and the rest belongs to the training set. At each epoch, 1000 points are randomly selected by subject of the training set and the obtained model is tested on all voxels of the subjects of the validation set. After 15 epochs, the model of the epoch with the highest score on the validation set is selected.

Optimization is performed by a stochastic gradient descent with a momentum of 0.9 (Sutskever et al., 2013). The learning rate is initially set at 0.001 and then divided by 10 every 7 epochs. The loss function corresponds to the cross entropy loss which is particularly useful for unbalanced training sets.

One CNN is trained for each hemisphere.

2.3. Regularization

As the CNN performs voxelwise classification, it is essential to spatially regularize the results. For this purpose, the elementary folds provided by BrainVISAs pipeline can be used as sets of voxels that must have the same label, as done by the BrainVISA sulcus recognition model. This regularization is used to label the 500 unlabelled subjects. However, this division into elementary folds is sometimes incorrect, which can lead to large errors. This is why, for the second stage of training, which relies on a better sampling of the population, we propose to trust the voxelwise classification and limit the regularization to a local filtering based on a vote in each voxels neighborhood.

2.3.1. Regularization per elementary folds

Once the fold skeleton has been extracted, the BrainVISA pipeline proposes to cut it into elementary folds based on geometrical and topological constraints specific to the definition of sulci. Thus, for each elementary fold, the scores output by the CNN are averaged by label and the label with the highest score is retained. However, as mentioned above, the fragmentation into elementary folds sometimes presents inconsistencies and is particularly unstable to segmentation hazards.

2.3.2. Local regularization

In order to regularize without using elementary folds, we use the labels assigned to the neighboring voxels: the final label assigned to a given voxel corresponds to the one most present in its neighborhood. If two or more labels are equally present, the final label is chosen randomly among the most present.

3. Experiment

Error rates are assessed by leave-one-out cross validation: each time, one subject belongs to the test set and the rest to the training set. The training set is first used to train the first CNN which is then applied to 500 unlabelled subjects, then these 500 subjects are used to pre-train the CNN, which is fine-tuned using the training set.

3.1. Database

The training base is composed of 62 healthy brains selected from different heterogeneous databases and labelled with a model containing 63 sulci for the right hemisphere and 64 for the left. Most of the subjects are right-handed male persons, between 25 and 35 years old. The elementary folds of each brain were manually labelled according to the sulcus nomenclature following a long iterative process leading to achieve a consensus across a set of several experts of the cortex morphology. Each elementary fold is initially extracted thanks to the BrainVISA pipeline, however the fragmentation technique has shown several limits, therefore the elementary folds have been manually cut out when needed for this training database. Compared to (Perrot et al., 2011), the database is the same but four additional sulci were used, and a new iteration of the experts has been performed. The 500 unlabelled MRIs are taken from the publicly available database of the Human Connectome Project.

3.2. Error rates

As in (Perrot et al., 2011), two measures are used to compare the different models proposed above: *Elocal* at the sulcus scale and *ESI* at the subject scale.

For each subject, we use one manually labelled segmentation for training and ten unlabelled segmentations for error quantification (Figure 1). Using ten different segmentations for each sulcus highlights the weaknesses of the BrainVISA preprocessing since we can compute errors from the worst result, usually associated to an undersegmentation issue. To quantify errors, for each new segmentation, the manual labelling on the initial segmentation must be transferred to the new one. Because of the variability of the segmentations obtained and the sparcity of the fold skeleton, the simple superposition of images is insufficient. To remedy this, a Voronoi diagram of the training segmentation is used to label the voxels of any other segmentation. Note that the elementary folds are not used to transfer the labelling and that the true labelling is indeed on the voxel scale.

For each subject, from the ten segmentations, **the average (ESI_{mean} and $Elocal_{mean}$) and/or maximum (ESI_{max} and $Elocal_{max}$) errors are calculated**. Note that the training segmentation used for manual labelling is not used in the error calculation because it would bias our evaluation.

3.2.1. Elocal

Given a sulcus l,

$$Elocal_l = \frac{FP_l + FN_l}{FP_l + FN_l + TP_l} \tag{1}$$

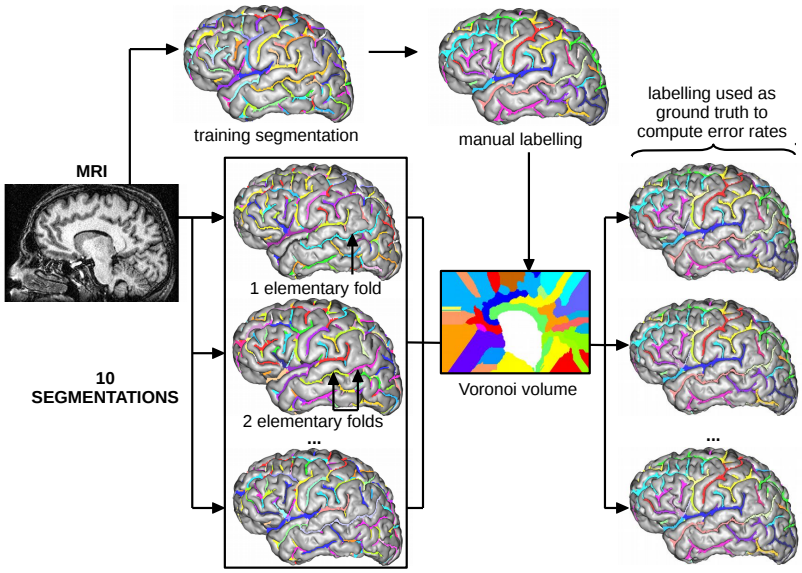

Figure 1: Strategy for measuring error rates.

with TP_l, FP_l and FN_l, respectively the number of true positive, false positive and false negative voxels for the sulcus l.

This error allows us to see local improvements that can be confused with brain-wide noise if only the proportion of false negatives had been taken into account.

3.2.2. ESI

Given a set of sulci L,

$$ESI = \sum_{l \in L} w_l * \frac{FP_l + FN_l}{FP_l + FN_l + 2 * TP_l} \tag{2}$$

with $w_l = \frac{s_l}{\sum s_l}$ and $s_l = FN_l + TP_l$, the sulcus l true size.

This error allows local errors to be synthesized in a single measurement. As explained in (Perrot et al., 2011), each component of the sum over labels differs on two points compared to $Elocal_l$. First, true positive measures count twice as false positive and negative ones, in order to remove errors shared by several labels, since each extra sulcal piece for a given label is a missing part for another label. Second, each component is weighted according to the sulcus true size so that each local component count as much as its size.

Compared to (Perrot et al., 2011), two labels are not included in the set of sulci ("unknown" and "ventricle"). Indeed, these two labels are not really considered as a sulcus label but correspond to others structures, not interesting for our purpose. Thus, the scores presented here for the BrainVISA method are worse than in (Perrot et al., 2011) for two reasons: on the one hand because the two labels removed considerably improved the results and on the other hand because of the consideration of undersegmentation errors during pre-processing thanks to the possibility to cut the elementary folds during manual labelling.

4. Results

4.1. Error at the subject level

In order to compare model performance at the subject level, we are interested in the average ESI_{mean} and ESI_{max}, that better reflect the robustness of the model to BrainVISA preprocessing errors (Table 1). By comparing the results with a matched T-test, the average ESI_{mean} and ESI_{max} are slightly (about 1%) but significantly improved thanks to the semi supervised strategy adopted ($p_{value} = 2.15e-26$ and $3.95e-27$). The same applies to the impact of the regularization technique used, which significantly reduces both error rates ($p_{value} = 8.11e-75$ and $1.26e-61$). Finally, the use of unlabelled data and regularization at the voxel scale results in a significantly better model than the BrainVISA model ($p_{value} = 9.47e-14$ and $3.09-26$), with more than 3% less ESI_{max} for the new model.

Table 1: $ESI_{mean/max}$ (2*standard deviation) in % per model

ESI	**Left (mean)**	**Right (mean)**	**Left (max)**	**Right (max)**
BrainVISA model	19.77 (5.63)	19.25 (6.26)	22.07 (5.91)	21.49 (6.89)
CNN	19.58 (4.32)	19.56 (4.55)	20.24 (4.58)	20.25 (4.65)
CNN + pretrain	18.24 (4.64)	18.20 (4.41)	18.91 (4.97)	18.85 (4.51)
CNN + pretrain + reg	**17.42 (4.78)**	**17.38 (4.48)**	**18.16 (5.06)**	**18.09 (4.53)**

4.2. Error at the sulcus level

In order to better understand the results, we have chosen to compare the average $Elocal_{max}$ for each sulcus of the final model (CNN + pretrain + reg) with those of the BrainVISA model (Figure 2). We observe that 64 sulci are significantly better recognized in the new model, compared to 5 sulci significantly worse. It is interesting to note that the vast majority of the better recognized sulci are the large sulci, which is particularly interesting in practice because they are the most studied in neuroanatomy.

5. Conclusion

In summary, this approach once again shows the power of CNNs compared to the methods developed so far. Compared to the BrainVISA model, the proposed model is slightly but significantly better, with 50% of the sulci significantly better recognized. But above all, the new method makes it possible to get rid of regularization by elementary folds while their cutting is not robust enough, which is one of the major defaults of the BrainVISA model.

In the future, many possibilities remain available to improve the model's performance. First of all, it would be necessary to test using more unlabelled images. Then, as in (Ciresan et al., 2012), several CNNs can be trained on different patch sizesin order to combine the results obtained by each network. In addition, the CNN used in this study is particularly simple, so it would be possible to test deeper architectures, such as AlexNet (Krizhevsky et al., 2012) adapted to 3D. Finally, U-Net (Ronneberger et al., 2015) have proven their worth in 2D and the use of patches to limit computational costs, as in (Beers et al., 2017), is a very promising approach to this problem.

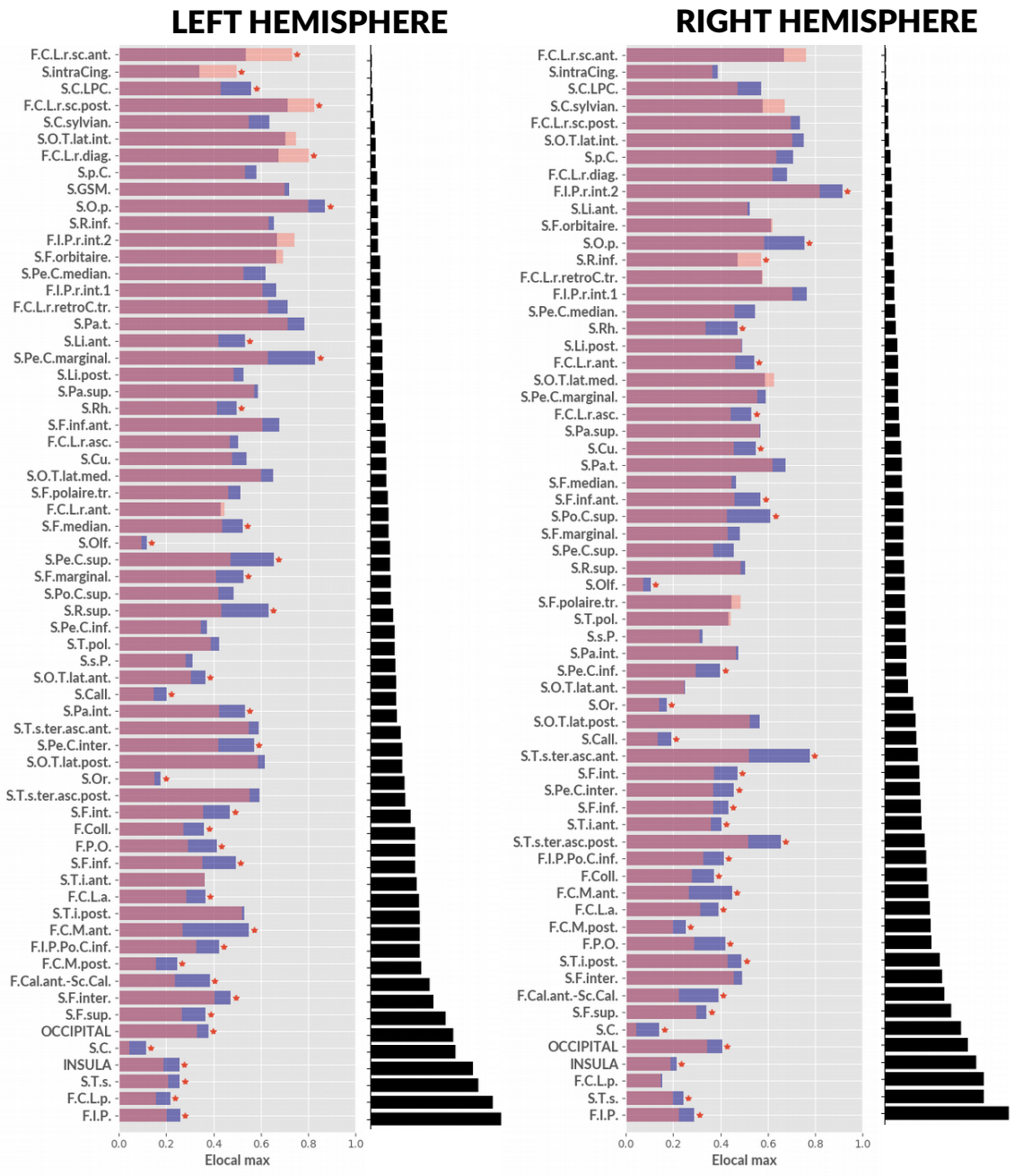

Figure 2: $Elocal_{max}$ **per sulcus.** The graph on the left and the graph on the right present the $Elocal_{max}$ for the sulci on the left hemisphere and on the right hemisphere, respectively. The BrainVISA model is represented in violet and the new model is represented in pink. The significative differences ($p_{value} < 0.05$) are marked with a star. Sulci are sorted from top to bottom from the smallest to the largest. The average sulci sizes, ranging from about 15 points to more than 2000 points per subject, are represented on the black graph.

References

Andrew Beers, Ken Chang, James Brown, Emmett Sartor, CP Mammen, Elizabeth Gerstner, Bruce Rosen, and Jayashree Kalpathy-Cramer. Sequential 3d u-nets for biologically-informed brain tumor segmentation. *arXiv preprint arXiv:1709.02967*, 2017.

Kirsten Judith Behnke, Maryam E Rettmann, Dzung L Pham, Dinggang Shen, Susan M Resnick, Christos Davatzikos, and Jerry L Prince. Automatic classification of sulcal regions of the human brain cortex using pattern recognition. In *Medical Imaging 2003: Image Processing*, volume 5032, pages 1499–1511. International Society for Optics and Photonics, 2003.

Alegria Blida. Ontology driven graph matching approach for automatic labeling brain cortical sulci. *IT4OD*, page 162, 2014.

Léonie Borne, Jean-François Mangin, and Denis Rivière. A patch-based segmentation approach with high level representation of the data for cortical sulci recognition. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 114–121. Springer, 2018.

Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.

Bruce Fischl, André Van Der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, et al. Automatically parcellating the human cerebral cortex. *Cerebral cortex*, 14(1):11–22, 2004.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Gabriele Lohmann and D Yves von Cramon. Automatic labelling of the human cortical surface using sulcal basins. *Medical image analysis*, 4(3):179–188, 2000.

M. Perrot, D. Rivière, and J.-F. Mangin. Cortical sulci recognition and spatial normalization. *Medical Image Analysis*, 15(4):529–550, 2011.

D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin, and J. Régis. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical Image Analysis*, 6(2):77–92, 2002.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Nicolas Royackkers, Michel Desvignes, and Marinette Revenu. Une méthode générale de reconnaissance de courbres 3d: application à l'identification de sillons corticaux en imagerie par résonance magnétique. *Traitement du Signal*, 15(5):365–379, 1998.

Yonggang Shi, Zhuowen Tu, Allan L Reiss, Rebecca A Dutton, Agatha D Lee, Albert M Galaburda, Ivo Dinov, Paul M Thompson, and Arthur W Toga. Joint sulci detection using graphical models and boosted priors. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 98–109. Springer, 2007.

Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.

Fabien Vivodtzev, Lars Linsen, Bernd Hamann, Kenneth I Joy, and Bruno A Olshausen. Brain mapping using topology graphs obtained by surface segmentation. In *Scientific Visualization: The Visual Extraction of Knowledge from Data*, pages 35–48. Springer, 2006.

Faguo Yang and Frithjof Kruggel. A graph matching approach for labeling brain sulci using location, orientation, and shape. *Neurocomputing*, 73(1-3):179–190, 2009.