Evaluation of cortical segmentation pipelines on clinical neonatal MRI data

C. Tor-Díez[1], C.-H. Pham[1], H. Meunier[2], S. Faisan[3], I. Bloch[4], N. Bednarek[2,5], N. Passat[5] and F. Rousseau[1]

Abstract—**Magnetic Resonance Imaging (MRI) can provide 3D morphological information on brain structures. Such information is particularly relevant for carrying out morphometric brain analysis, especially in the newborn and in the case of prematurity. However, 3D neonatal MRI acquired in clinical environments are low-resolution, anisotropic images, making segmentation a challenging task. In this context, preprocessing techniques aim to increase the image resolution. Interpolation techniques were classically used; super-resolution (SR) techniques have recently appeared as an emerging alternative. In this paper, we evaluate the performance of different SR methods against the classical interpolation in the application of neonatal cortex segmentation. Additionally, we assess the robustness of different segmentation methods for each estimation of high resolution MRI input. Results are evaluated both qualitatively and quantitatively with neonatal clinical MRI.**

Index Terms—**Segmentation, super-resolution, MRI, neonatal brain, cortex.**

I. Introduction

Morphometry based on neonatal brain Magnetic Resonance Imaging (MRI) data has shown to be of great interest for brain development studies at early stages and for prediction of the outcome of prematurely born infants. However, neonatal MRI acquisition is often limited by time in order to avoid motion artefacts. Clinical data are thus characterized by slice thickness greater than 2 or 3 mm. This is a limiting factor for accurate brain morphometry studies such as brain folding and cortical maturation. In this context, it appears that the implementation of robust image processing pipelines is a key point to improve the understanding of brain development.

One of the first steps of such pipelines is the upsampling of the clinical low-resolution (LR) data to $1 \times 1 \times 1$ mm or $0.5 \times 0.5 \times 0.5$ mm. This upsampling step is usually performed using interpolation techniques, leading to blurry edges and loss of details. Single image Super-Resolution (SR) [1], which purpose is to estimate a high-resolution (HR) image from one LR image, constitutes a promising alternative approach [2], [3], [4], [5]. However, SR is a challenging inverse problem. In particular the estimation of texture and details remains difficult and SR techniques are not yet applied in routine on clinical data. Indeed, the impact of SR reconstruction on morphometry analysis needs to be accurately investigated.

The second major step is image segmentation. In the context of neonatal brain MRI analysis, segmentation is challenging due to low contrast-to-noise ratio, rapid change of size of brain structures, complex brightness modifications in structural MRI reflecting rapid white matter structuring through myelination, rapid change and large variability of anatomical shapes. Various approaches were investigated for neonatal brain data, such as mathematical morphology [6], deformable models [7], statistical classification [8] and multi-atlas methods [9], [10]. However, most of these techniques were only evaluated on HR images, typically with slice thickness lower than 1 mm. There is a clear need to evaluate state-of-the-art segmentation algorithms in actual clinical settings. NeoBrainS12 challenge was a first step toward such an evaluation [11].

The purpose of this paper is to assess the influence of various upsampling techniques on segmentation map estimation and the robustness of several segmentation pipelines, with a focus on cortical structures. In particular, our aim is to carry out such a study on real neonatal brain MRI data, acquired in a clinical context.

Section II presents the general approach of a LR neonatal cortex segmentation. We describe three upsampling approaches and four segmentation methods. Our experimental, comparative study is then proposed in Section III, where we evaluate both qualitatively and quantitatively the 12 combinations of SR and segmentation techniques. In Section IV, we propose a discussion of these results, leading to concluding remarks on future work in Section V.

[1]IMT Atlantique, LaTIM U1101 INSERM, UBL, Brest, France
[2]Service de médecine néonatale et réanimation pédiatrique, CHU de Reims, France
[3]ICube UMR 7357, Université de Strasbourg, CNRS, FMTS, Illkirch, France
[4]LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
[5]Université de Reims Champagne-Ardenne, CReSTIC, Reims, France

II. Low-resolution Neonatal Cortex Segmentation Pipeline

In this work, we consider multi-atlas segmentation pipelines for clinical LR images consisting of the following steps: image upsampling (Section II-A), bias correction using N4 [12], nonlinear registration of atlases on the input HR image using ANTs [13], intensity normalization using histogram matching techniques and image segmentation (Section II-B). In this study, we focus on two steps: HR estimation and image segmentation.

A. HR Image Estimation

The resolution of clinical image is too low to carry out cortical morphometry studies. Upsampling (to isotropic resolution of $1 \times 1 \times 1$ mm or $0.5 \times 0.5 \times 0.5$ mm) appears then to be a key step in neonatal brain morphometry pipelines. In this work, we consider three different upsampling techniques: spline interpolation, non-local MRI upsampling [4], [5] and deep learning-based method [3].

Image interpolation is a widely used approach to compute isotropic data. However, interpolation models generally fail in accurately recovering fine details and textures in clinical LR data, leading mostly to blurred results.

Alternatively, the purpose of SR methods is to estimate finer reconstructions of HR images. Non-local MRI upsampling (NMU) [5] iteratively applies the non-local means algorithm to the upsampled LR image. An observation model using a predefined point spread function is used to relate LR observation and HR estimation. The non-local approach provides an adaptive regularization approach and does not need to be trained, thus reducing data resource requirements.

The third method considered in this work is a deep learning-based approach, namely Super-resolution residual-learning convolutional neuronal network (SRReCNN) [3]. SRReCNN is a supervised method based on the assumption that the restoration model can be learned from data. The problem is then expressed as the estimation of a restoration matrix which, combined to the spline interpolation of the LR input, provides its HR version. The restoration matrix is learned using a CNN with a set of HR images similar to the target image, in terms of modality, observation, etc.

B. Cortex Segmentation

Popular approaches for brain segmentation rely on atlas strategy. Such methods consist of three steps [14]: (1) registration of the atlas images onto the target image; (2) application of the induced transformations to atlas labels; and (3) fusion of the transformed labels. In this work, we consider four different atlas-based segmentation pipelines, described below.

In atlas-based approaches, the registration is carried out in two stages: first, affine and second, non-rigid registration, in order to first avoid local minima and then obtain accurate matching of fine structures. The obtained registration fields are then used for mapping atlas label images onto the target image. Once the registration step is performed, segmentation maps can be obtained by fusing the deformed atlases. A popular approach for label fusion is to average the deformed atlases and to apply the majority voting rule (MV) at voxel-scale. One of the limitations of MV is the strong dependency on registration accuracy.

To improve the robustness of MV, several fusion strategies have been developed. An iterative multi-atlas patch-based approach (IMAPA) [10] has been recently proposed to overcome this issue. It is an iterative non-local method, that computes atlas weights by minimizing a patch-based cost function. Moreover, this iterative approach takes advantage of current segmentation estimates as in cascading classifiers, to add regularization constraints on estimated segmentation maps. This approach has been shown to be very effective on HR neonatal brain MRI data [10].

The third approach evaluated in this study is called DrawEM [9]. In this method, the input MR image is first brain-extracted and corrected for field inhomogeneity. Then, atlases are registered to the target image and the atlas labels are propagated to the image. The propagated labels are averaged in a locally-weighted scheme and subdivided with the use of subject-specific tissue priors obtained with k-means clustering. An expectation-maximization scheme is used for the estimation of the segmentation map.

The last method is MANTiS (Morphologically Adaptive Neonatal Tissue Segmentation) [8], which is based on the unified segmentation approach [15] implemented in Statistical Parametric Mapping (SPM) software. MANTiS makes use of a combination of unified segmentation, template adaptation via morphological tools and topological filtering, to segment the neonatal brain into eight tissue classes.

III. EXPERIMENTS AND RESULTS

A. Data

1) Training datasets: Considering the HR image estimation step, SRReCNN is the only supervised approach that requires a set of training data. To learn the restoration matrix, we have used the 40 images provided by the Developing Human Connectome Project (dHCP)[1] [16]. The age of subjects varies between 37 and 44 weeks. Acquisitions were made with overlapping axial slices with voxel size $0.8 \times 0.8 \times 1.6$ mm every 0.8 mm. The resulting images have an isotropic resolution with voxel size $0.5 \times 0.5 \times 0.5$ mm. TR and TE are $12\,000$ and 156 ms, respectively.

The four considered segmentation approaches make use of atlases. For MV and IMAPA, we have used the dHCP data as the multi-atlas set. The open-source implementation of DrawEM uses the ALBERT dataset [17] and MANTiS is provided with a dedicated probabilistic template as a SPM toolbox.

2) Testing dataset: There are currently very few available ground-truth data of cortex segmentation maps in clinical neonatal brain MRI. The NeoBrainS12 study [11] provides three different image sets of preterm born infants, in order to perform reliable comparison of the performance of segmentation algorithms. In this study, we have used two subjects from 40 weeks axial set (as mentioned in [18], there are significant regions missing in 30 W coronal scans that would result in an invalid comparison). The size of voxels is $0.35 \times 0.35 \times 2$ mm. TR and TE are $6\,293$ ms and 120 ms, respectively.

To further assess the performances of segmentation pipelines on clinical LR images, manual delineation has been performed on two MR images acquired at the University Hospital of Reims as part of the ANR MAIA project. Acquisitions were made at the term age (between 38 and 42 weeks), with anisotropic resolution ($0.446 \times 0.446 \times 3$ mm). TR and TE are $3\,000$ ms and 200 ms, respectively.

[1]http://www.developingconnectome.org

6554

TABLE I

DICE SCORES FOR THE 12 COMBINATIONS OF UPSAMPLING (ROWS) AND SEGMENTATION METHODS (COLUMNS) ON THE 4 IMAGES OF THE MAIA
AND NEOBRAINS12 TESTING DATASET.

Subject	Upsampling	MV	MANTiS [8]	IMAPA [10]	DrawEM [9]
MAIA #1	Interpolation	43.43	70.52	63.94	71.98
	NMU [5]	44.51	71.90	65.51	73.84
	SRReCNN [3]	43.68	73.30	66.98	73.59
MAIA #2	Interpolation	41.93	71.18	64.43	70.27
	NMU [5]	42.42	71.82	64.97	72.08
	SRReCNN [3]	41.94	73.33	67.63	72.51
NeoBrainS12 #1	Interpolation	42.07	77.52	59.72	78.35
	NMU [5]	41.55	78.64	65.33	77.25
	SRReCNN [3]	41.26	77.26	65.42	78.39
NeoBrainS12 #2	Interpolation	44.70	80.05	62.02	81.96
	NMU [5]	44.49	80.74	66.17	82.14
	SRReCNN [3]	43.97	79.73	66.18	81.30

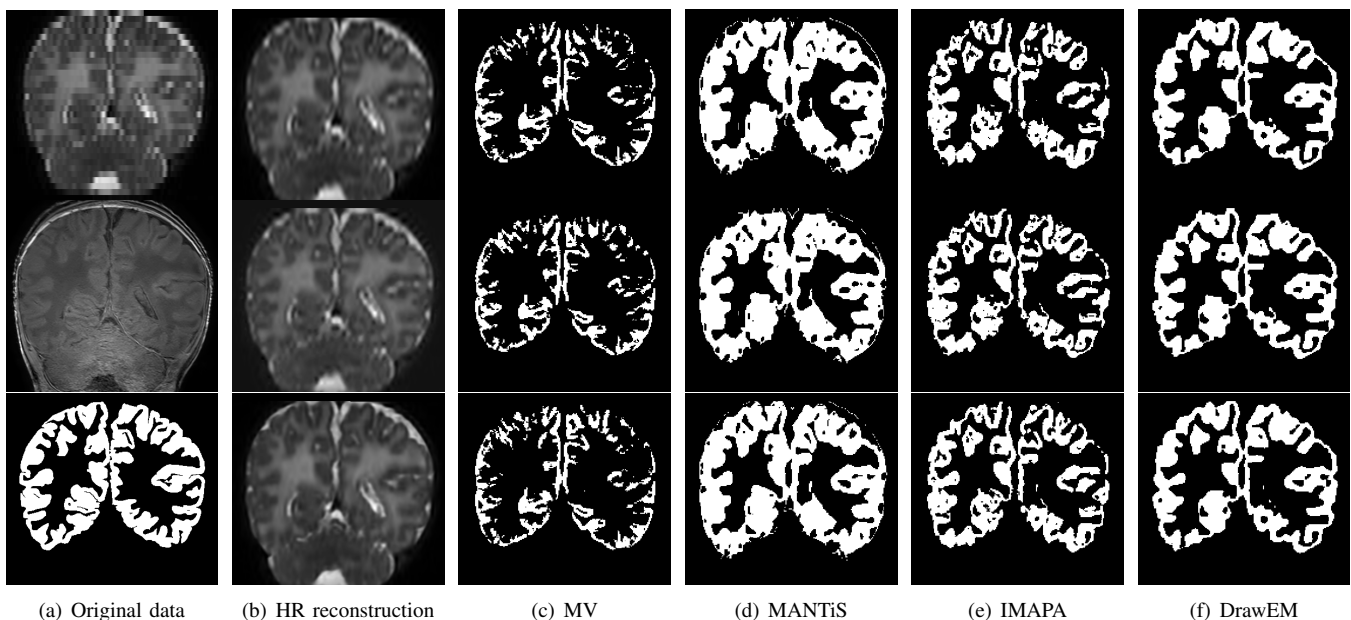| (a) Original data | (b) HR reconstruction | (c) MV | (d) MANTiS | (e) IMAPA | (f) DrawEM |

Fig. 1. A visual summary of the different estimations of HR input and the corresponding segmentations for subject MAIA #1. (a) The original LR T2w and T1w images, and the reference segmentation. (b) HR reconstruction results. (c–f) Segmentation results with MV, MANTiS, IMAPA and DrawEM. (b–f) First row corresponds to interpolation, second to NMU and third to SRReCNN.

B. Evaluation

After the HR reconstruction step, all images have a common isotropic voxel size of $0.5 \times 0.5 \times 0.5$ mm. Importantly, this HR reconstruction does not have the same effect on MAIA and NeoBrainS12 data. Indeed, MAIA MR images have similar resolutions before / after the SR on two dimensions, whereas the third resolution is significantly increased (from 3 mm to 0.5 mm, i.e., a ×6 factor). By contrast, resolution of NeoBrainS12 MR images is slightly decreased in two dimensions, whereas the third resolution is increased (from 2 mm to 0.5 mm, i.e., a ×4 factor).

In order to quantitatively assess the segmentation results, they were compared to manual segmentations carried by experts, and used as reference.
MAIA: reference segmentations were obtained by manually

segmenting[2] T1-weighted (T1w) MR images, at a resolution $0.268 \times 0.268 \times 1.2$ mm. The estimated segmentation maps were then mapped onto the original T1w images by a rigid registration between HR T2w and T1w data.
NeoBrainS12: manual segmentations were carried out in the T2w space, at high resolution ($0.5 \times 0.5 \times 0.5$ mm) with the protocol presented in [11]. Then, these segmentation results were downsampled to reach the original LR T2w ($0.35 \times 0.35 \times 2$ mm).

It is worth mentioning that the MAIA reference maps were natively at a low resolution (1.2 mm) in the anisotropic dimension, whereas the low resolution in the same dimension

[2]A medical specialist in neonatology (H. Meunier) realized these manual segmentations. Brains were labelled into seven classes: cortical grey matter, unmyelinated / myelinated white matter, basal ganglia and thalami, brainstem, cerebellum, and ventricles. The cerebellum was delimited laterally and posteriorly by the cistern of the fossea and anteriorly by the brainstem.

was obtained by downsampling of a HR reference map, in the case of NeoBrainS12. In particular, these differences in the initial dimensions of the data and the distinct ways to express the LR reference segmentation for MAIA and NeoBrainS12 will shed light on slight variations observed in the results.

Parameters of the segmentation methods were set to their default values or following those proposed in the literature.

C. Results

Table I provides a quantitative assessment (Dice scores) of the results obtained with the proposed, combined up-samling and segmentation methods. Note that MV, MANTiS and DrawEM provide binary segmentations, by contrast to IMAPA that generates probabilistic segmentation maps. In order to compare all the results, the IMAPA outputs were thresholded at value 0.5. A more qualitative view is available in Figure 1 that illustrates results of these 12 combined methods on a coronal slice of one of the MAIA MR image.

IV. DISCUSSION

The Dice scores obtained with all methods are globally lower than those stated in the literature. This can be explained by the increased difficulty to process real images, acquired in a clinical context, with noise, artefacts, anisotropy effects (not fully corrected by oversampling) and a low contrast. This highlights the remaining gap to be filled for making current automated methods dedicated to neonatal brain MRI, fully efficient for actual clinical routine.

In this challenging context, MANTiS and DrawEM present comparable Dice scores (70–80), with a slight advantage for DrawEM. In particular, they overperform IMAPA and MV. However, this quantitative analysis has to be completed by qualitative elements exemplified in Figure 1. Indeed, one can observe that MANTiS and DrawEM oversegment the cortex, compared to IMAPA and MV, that undersegment this subpart. In other words, two classes of segmentations appear, with dual and antagonist specificity / sensitivity properties.

The SR upsampling tends to improve the Dice scores obtained by segmentation methods. However, these improvements are weak, compared to segmentation on interpolated images. They are also method-dependent. For instance, MV has similar results for the three upsampling methods, whereas IMAPA behaviour is clearly improved by SR. Such an improvement is also observed for MANTiS and DrawEM for the MAIA dataset, but not for NeoBrainS12.

V. CONCLUSION

The main conclusions of this study are the following. First, the current automated, neonate cortex segmentation methods can hardly provide satisfactory segmentation of real clinical low resolution MR images, even with the assistance of an oversampling preprocessing step. Secondly, the behaviour of these methods is not homogeneous, and two families emerge, that tend to oversegment and undersegment the cortex, respectively. Thirdly, oversampling with SR approaches tends to improve the overall quality of the segmented cortex, but this progress remains slight, compared to interpolation.

Future work emphasized by these results will consist in taking advantage of the complementary behaviours of the existing segmentation methods, in order to improve both specificity and sensitivity. To this end, we will consider segmentation fusion strategies. In addition, we will more deeply investigate why the current SR method does not allow us to significantly improve the overall quality of segmentation results on such reconstructed data.

REFERENCES

[1] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2008.

[2] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," in *IEEE ISBI*, 2017, pp. 197–200.

[3] C.-H. Pham, R. Fablet, and F. Rousseau, "Multi-scale brain MRI super-resolution using deep 3D convolutional networks," Research Report hal-01635455, 2017.

[4] F. Rousseau, "Brain hallucination," in *ECCV*, 2008, pp. 497–508.

[5] J. V. Manjón, P. Coupé, A. Buades, V. Fonov, L. Collins, and M. Robles, "Non-local MRI upsampling.," *Medical Image Analysis*, vol. 14, no. 6, pp. 784–92, 2010.

[6] L. Gui, R. Lisowski, T. Faundez, P. S. Hüppi, F. Lazeyras, and M. Kocher, "Morphology-driven automatic segmentation of MR images of the neonatal brain," *Medical Image Analysis*, vol. 16, no. 8, pp. 1565–1579, 2012.

[7] F. Leroy, J.-F. Mangin, F. Rousseau, H. Glasel, L. Hertz-Pannier, J. Dubois, and G. Dehaene-Lambertz, "Atlas-free surface reconstruction of the cortical grey-white interface in infants," *PloS One*, vol. 6, no. 11, pp. e27128, 2011.

[8] R.J. Beare et al., "Neonatal brain tissue classification with morphological adaptation and unified segmentation," *Frontiers in Neuroinformatics*, vol. 10, pp. 12, 2016.

[9] A. Makropoulos, I. S. Gousias, C. Ledig, P. Aljabar, A. Serag, J. V. Hajnal, A. D. Edwards, S. J. Counsell, and D. Rueckert, "Automatic whole brain MRI segmentation of the developing neonatal brain," *IEEE Transactions on Medical Imaging*, vol. 33, no. 9, pp. 1818–1831, 2014.

[10] C. Tor Díez, N. Passat, I. Bloch, S. Faisan, N. Bednarek, and F. Rousseau, "An iterative multi-atlas patch-based approach for cortex segmentation from neonatal MRI," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 73–82, 2018.

[11] I. Isgum et al., "Evaluation of automatic neonatal brain segmentation algorithms: The NeoBrainS12 challenge," *Medical Image Analysis*, vol. 20, no. 1, pp. 135–151, 2015.

[12] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[13] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[14] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, 2015.

[15] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.

[16] A. Makropoulos et al., "The Developing Human Connectome Project: A minimal processing pipeline for neonatal cortical surface reconstruction," *NeuroImage*, vol. 173, pp. 88–112, 2018.

[17] I. S. Gousias, A. D. Edwards, M. A. Rutherford, S. J. Counsell, J. V. Hajnal, D. Rueckert, and A. Hammers, "Magnetic resonance imaging of the newborn brain: Manual segmentation of labelled atlases in term-born and preterm infants," *NeuroImage*, vol. 62, no. 3, pp. 1499–1509, 2012.

[18] M. Liu, A. Kitsch, S. Miller, V. Chau, K. Poskitt, F. Rousseau, D. Shaw, and C. Studholme, "Patch-based augmentation of Expectation–Maximization for brain MRI tissue segmentation at arbitrary age after premature birth," *NeuroImage*, vol. 127, no. C, pp. 387–408, 2016.