3D Unsupervised Kidney Graft Segmentation Based on Deep Learning and Multi-Sequence MRI

Leo Milecki, Sylvain Bodard, Jean-Michel Correas, Marc-Olivier Timsit, Maria Vakalopoulou

3D UNSUPERVISED KIDNEY GRAFT SEGMENTATION BASED ON DEEP LEARNING AND MULTI-SEQUENCE MRI

Leo Milecki[1], Sylvain Bodard[2], Jean-Michel Correas[2], Marc-Olivier Timsit[3], Maria Vakalopoulou[1]

[1] Mathématiques et Informatique pour la Complexité et les Systèmes, CentraleSupélec, Université Paris-Saclay, Inria Saclay, France.
[2] Department of Adult Radiology, Paris University and Necker University Hospital, France.
[3]Department of Urology, Paris University and Hôpital Européen Georges-Pompidou, France.

ABSTRACT

Image segmentation is one of the most popular problems in medical image analysis. Recently, with the success of deep neural networks, these powerful methods provide state of the art performance on various segmentation tasks. However, one of the main challenges relies on the high number of annotations that they need to be trained, which is crucial in medical applications. In this paper, we propose an unsupervised method based on deep learning for the segmentation of kidney grafts. Our method is composed of two different stages, the detection of the area of interest and the segmentation model that is able, through an iterative process, to provide accurate kidney draft segmentation without the need for annotations. The proposed framework works in the 3D space to explore all the available information and extract meaningful representations from Dynamic Contrast-Enhanced and T2 MRI sequences. Our method reports a dice of $89.8 \pm 3.1\%$, Hausdorff distance at percentile 95% of 5.8 ± 0.41mm and percentage of kidney volume difference of $5.9 \pm 5.7\%$ on a test dataset of 29 patients subject to a kidney transplant.

Index Terms— Unsupervised segmentation, kidney graft, multi-sequence MRI, Deep Learning, CNN

1. INTRODUCTION

Segmentation of organs or other structures such as vessels or cells in biomedical imaging is often an important first step or even a prerequisite to analyzing medical volumes. They are essential to describe and monitor shapes, volume characteristics [1] or further used in more complex pipelines, as in radiomics analysis [2] and precision medicine. In recent years, automatic segmentation in biomedical images has been subject to extensive search, with deep learning based techniques holding the state of the art performances in various applications and modalities, including Magnetic Resonance Imaging (MRI) [3]. However, one of the main limitations of common deep learning supervised approaches is the need for expensive and time-consuming annotations.

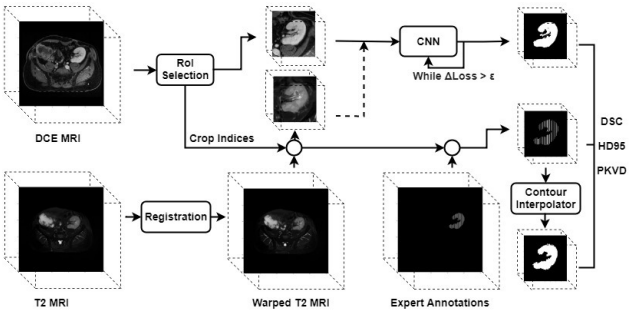

Fig. 1. Overview of the proposed method for the automatic segmentation of kidney graft using multimodal MR Imaging.

Nowadays, unsupervised approaches investigate methods that are able to generate models without the need for annotations. In particular, concerning unsupervised kidney segmentation, image-based methods have been proposed to obtain a segmentation through an image processing pipeline using tools such as thresholding [4] or morphological operators [5, 6]. Some other methods propose clustering frameworks based on K-means [7] or on wavelet-based clustering [8]. Deformable models were also designed in level-set methods [9] or exploiting active contours [10]. Advanced deformable boundaries methods are also explored in [11] with however the use of training samples for parts of their formulations.

This paper proposes an unsupervised method for the segmentation of kidney graft from Dynamic Contrast-Enhanced (DCE) and T2 sequences obtained 12 months after the transplantation. Our method consists of a two-step procedure that is summarised in Figure 1. First, we propose an unsupervised method to detect the kidney region using thresholding techniques and morphological operators. The second step corresponds to the pixel-wise segmentation of the kidney graft. Our method is

based on [12] extending the formulation to 3D to explore all the available information from the MRI modalities. Moreover, it proposes a multimodal formulation where multiple MRI sequences are integrated towards a robust kidney graft segmentation tool.

2. METHODS

2.1. Selection of the region of interest

The first step of the proposed method is the automatic detection of possible areas of interest. This step is essential to reduce the search space and boost the performance of the unsupervised segmentation technique eliminating the number of false positives. The detection of the region of interest is based on a relatively simple pipeline taking full advantage of the contrast-enhanced MRI. In particular, a thresholding operation is first applied into the entire DCE volume highlighting the highest intensities of the volume corresponding to regions that depict the graft kidney and blood vessels. In the pixels' volume histogram, the threshold is obtained as the abscissa of maximum intensities. After the volume thresholding, morphological operations are applied to remove artifacts and fill holes in the detected regions. The morphological step consists of a binary closing followed by a binary opening. For both operations, a sphere of radius 5 and 4 have been chosen as a structural element, respectively. Principal component analysis has been used in the thresholded volume. A boundary box indicating the highest volume component was selected as the region of interest and further processed. The same area is also indicated in the T2 sequence. Areas of interest ranged for $N_x \times N_y \times N_z \in [125, 173] \times [118, 199] \times [49, 75]$.

2.2. Unsupervised Segmentation Model

The proposed unsupervised segmentation model then processes the boundary boxes of various sizes containing the region of interest. Our method is based on differentiable feature clustering [12]. A simple straight-forward 3D convolutional neural network (CNN) is used to extract features and optimize two loss functions without any need for ground truth annotations. The first loss is designed to refine and extract relevant features, and the second to favor smoothness on the obtained labels.

Let us denote $X \in \mathbb{R}^{2\times N_x\times N_y\times N_z}$ the regions of interest extracted from the T2 and DCE sequences. Our unsupervised method is based on a 3D feature extractor CNN denoted as $h(.,\theta)$, where $\theta \in \Theta$ defines its trainable parameters. The p-dimentional feature vector obtained by the CNN is denoted as $Y = h(X,\theta) \in \mathbb{R}^{p\times N_x\times N_y\times N_z}$. From this feature vector we can propose a mapping strategy defined by a function g to obtain class labels for each of the voxels of the volume $C = g(Y) \in \mathbb{Z}^{N_x\times N_y\times N_z}$. Two different loss functions are defined for the optimization of the method,

$$L_{sim}(Y,C) = \sum_{i=1}^{N_x\times N_y\times N_z} \sum_{j=1}^{p} -\delta(j-c_i)\ln(y_{j,i}) \quad (1)$$

where $Y = \{y_{j,i}\}_{j\in[1,p],i\in[1,N_x\times N_y\times N_z]}$ and $C = \{c_{j,i}\}_{j\in[1,p],i\in[1,N_x\times N_y\times N_z]}$ and δ is the kronecker function. This loss correspond to the cross entropy between the network output and class labels used as intermediate targets that are obtained by an argmax operation.

$$L_{cont}(Y) = \sum_{k=1}^{N_x-1}\sum_{l=1}^{N_y-1}\sum_{m=1}^{N_z-1} \|y_{k+1,l,m} - y_{k,l,m}\|_1 + \|y_{k,l+1,m} - y_{k,l,m}\|_1 + \|y_{k,l,m+1} - y_{k,l,m}\|_1 \quad (2)$$

where $y_{k,l,m}$ corresponds to the p-dimensional feature vector of the voxel at (k,l,m). This loss aims to enforce spatial smoothness on the obtained segmentation masks.

Finally, the overall optimization is achieved by,

$$L = L_{sim}(Y,C) + \lambda L_{cont}(Y) \quad (3)$$

where λ is a weight that balances the contribution of each component to the final optimization.

The network parameters are optimized iteratively until the absolute difference of loss between two successive iterations is lower than a predefined value ϵ. Using this criterion, the number of classes obtained can be superior to 2 as the non-kidney elements can rarely be assigned in one class. Using the final predicted clustering, the kidney class can easily be obtained by selecting the first connected component class.

3. IMPLEMENTATION DETAILS

In this study, we first applied bias field correction to all the MRI images as a preprocessing step. The number of iterations was set to 2 and the number of fitting levels to 8, enabling to enhance peaks in volumes' histograms and, more particularly, contrast-enhanced regions. Moreover, even if the two sequences (DCE and T2) correspond to the same follow-up exam, they were acquired a few minutes apart, introducing misalignments. To address this issue and project properly one sequence to the other, we used an unsupervised deformable registration method [13]. The DCE modality has been used as target, and the T2 has been warped to it, using a variety of similarity metrics, including normalized cross-correlation, mean square error, and mutual information.

Our 3D CNN architecture consists of N convolutional blocks with p channels. Each block consists of a 3D convolution layer with a kernel size of $3\times3\times3$ and stride 1,

followed by a ReLU activation and a 3D batch normalization layer. The two different sequences were concatenated channel-wise and were given into the network. The choice to include as stacked channels the two sequences – early fusion – has been made and tested over middle (classification level fusion) or late fusion (decision level fusion) strategies [14].

Concerning the other parameters of our network, we set $\lambda = 5$ and $\epsilon = 0.005$. Our CNN model was trained using classic stochastic gradient descent with a momentum set to 0.9 and learning rate to 0.01. The overall optimization of a single volume needed approximately 1 minute on a GeForce GTX 1080 TI GPU (11.2 Go).

4. DATASET

The dataset used in this study corresponds to a total of 32 patients subject to a kidney transplant. All the MRIs were obtained 12 months after the transplantation. Our dataset consists of 2 different MRI sequences, namely T2 and DCE. DCE volumes being $512 \times 512 \times [64 - 88]$ included spacing ranging from $0.78 \times 0.78 \times 1.9$ to $0.94 \times 0.94 \times 2.5$ and T2 volumes being $512 \times 512 \times [25 - 36]$ from $0.70 \times 0.70 \times 5.0$ to $0.98 \times 0.98 \times 6.0$. For the optimization of the network's hyperparameters, 3 patients were used, while all the rest (29 patients) were used for evaluation of the proposed method. The 3 patients were selected as a sample of low, medium, and high segmentation difficulty volumes from preliminary studies. Pixel-wise kidney annotation was performed by a medical expert (a board-certified radiologist), by tracing the kidney boundaries using the DCE MRI in sagittal view. The annotation was available in every 2 or 3 slices, resulting in 945 annotated slices. We obtained the 3D annotation by interpolating the obtained manual annotations per subject to assess the method's performance.

5. EXPERIMENTAL RESULTS

We evaluated the graft kidney's segmentation accuracy using five metrics commonly used to assess biomedical segmentation tasks. First, the mean dice coefficient (DSC) metric is used to indicate the similarity between the manual annotation and prediction. Secondly, we used the mean of Hausdorff distance at percentile 95% (HD95) to measure how far two surfaces are from each other. Finally, the mean percentage of kidney volume difference (PKVD), precision, and recall between our predictions and the annotations were used to evaluate different methods' performance.

Starting with the evaluation of the proposed method, an ablation study benchmarking the influence of the parameter λ and the 3D CNN architecture has been performed for both 1 sequence (DCE) and 2 sequences (DCE & T2) models (Figure 2). Starting with the influence of λ, we used grid search and benchmark values in {0.01, 0.1, 1, 5, 10, 25, 50, 100}. The superiority of the 2 sequence model seems to be indicated with the best performances to be found in [1,25]. Moreover, for the 1 sequence model, the variations are less monotone, and the selection of an optimum value more difficult.

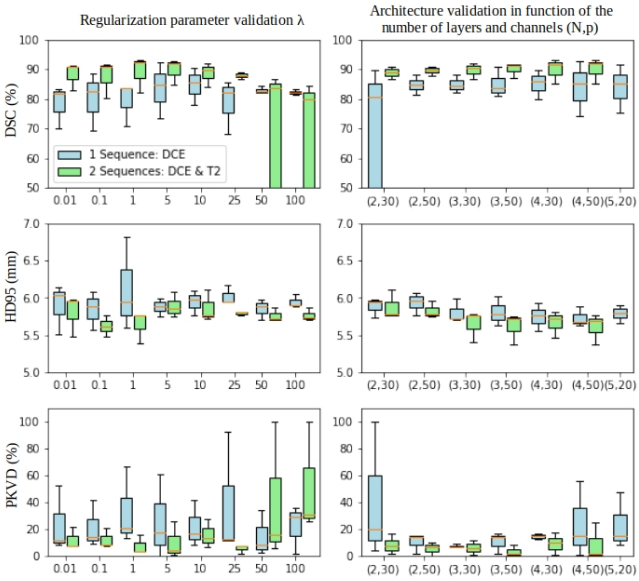

Fig. 2. Evaluation of the different components of the proposed method using different evaluation metrics on the validation set. The first column indicates the influence of λ and the second column the influence of the used architecture for different number of channels and layers.

Moreover, we analyzed the impact of the network's architecture in depth (number of convolutional blocks) and width (number of channels in each block) to extract relevant features. We evaluated these parameters (N convolutional blocks, p number of channel per block) in the set of {(2,30), (2,50), (3,30), (3,50), (4,30), (4,50), (5,20)}. The specific parameters seem to not really influence each model's performance, with again the 2 sequences one reporting more stable performance for all the experiments. Overall, the (2,50) configuration seems to report more stable performance, with less variance, on the validation set. This indicates that a relatively simple architecture is enough to address our problem.

To highlight our method's soundness, we compare it with other commonly used unsupervised methods in the literature and its 2D implementation [12]. In particular, our results were evaluated against the kidney graft segmentations obtained by Otsu thresholding [15] and watershed 3D [6] methods that are commonly used for unsupervised segmentation in a variety of studies. The obtained results are summarised in Table ,1 presenting

Method \ Metric	DSC (%)		HD95 (mm)		PKVD (%)		Precision (%)		Recall (%)	
Statistic	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Otsu Thresholding	71.04	10.97	6.55	0.65	84.71	57.34	57.07	13.03	97.53	0.91
Watershed	84.06	23.30	5.75	0.42	16.35	23.99	89.25	23.91	80.05	23.18
Proposed 2D (DCE only)	69.09	20.34	7.22	1.67	51.30	65.00	74.57	25.77	69.79	20.48
Proposed (DCE only)	86.48	4.01	5.92	0.42	11.08	10.13	83.93	8.32	89.97	3.67
Proposed 2D (DCE & T2)	75.75	10.80	6.39	1.11	32.81	41.56	81.65	14.24	81.65	14.66
Proposed (DCE & T2)	89.81	3.17	5.82	0.41	5.88	5.74	91.00	5.18	88.20	5.82

Table 1. Quantitative evaluation of the proposed method and other methods reporting: dice coefficient (DSC), Hausdorff distance at percentile 95% (HD95), percentage of kidney volume difference (PKVD), precision and recall.

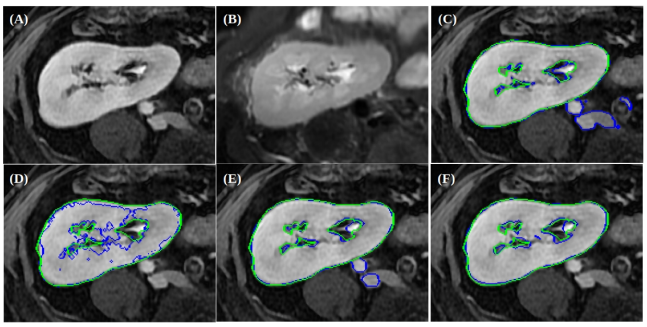

Fig. 3. One testing subject depicting one slice in axial view and the segmentation results as overlay for each of the evaluated methods (in blue) and the annotation from the expert (in green): (A) DCE, (B) Warped T2, (C) Otsu thresholding, (D) watershed based method, (E) Proposed (DCE), (F) Proposed (DCE & T2).

the mean and standard deviation (std) of each score on the whole test set. Even if additional postprocessing methods are presented in the literature [16, 6] in our study, we did not perform any additional postprocessing in any of the exploit methods. The lowest performance in terms of all the metrics is reported by the Otsu thresholding, which reaches a mean precision of $57.07 \pm 3.03\%$, highlighting the high number of false positives. Watershed seems to perform better, reporting a mean DSC of 84.06% with, however, a high standard deviation. The proposed method performs better in all the cases. Using only the DCE MRI as input, we obtain better (in terms of mean values) and more stable (in terms of std) performance than the previous two methods. The performance of the proposed method is further increased with the integration of the T2 sequence. The 2 sequence model reports 3% higher DSC than the model that uses only one sequence. The method's superiority is also indicated by the higher numbers of mean precision that increases from $83.93 \pm 8.32\%$ for the DCE to $91.00 \pm 5.18\%$ for the DCE & T2 models.Even if the two models report similar mean HD95, the DCE & T2 model's impact is further indicated by the significantly lower mean PKVD metric. Finally, our experiments indicated that the 2D [12] implementation is not performing well, highlighting the need for 3D formulations for this problem.

In Figure 3, we present some qualitative analysis of the obtained segmentation masks. One can observe that the accurate detection of the kidney between structures – essentially between the kidney and its irrigating vessels – is very challenging for unsupervised methods. However, the proposed formulation is able to separate the kidney graft from the different structures and provide an accurate segmentation.

6. CONCLUSION

In this paper, we proposed a 3D deep learning-based unsupervised kidney segmentation framework using multimodal MRI. Our framework consists of two losses, one that optimizes the cross-entropy between the predicted labels and the labels of clusters, while the second one is enforcing spatial continuity, producing smooth segmentation maps. Our 3D formulation works better than other commonly used unsupervised methods in the literature, including its 2D implementation [12]. Our experiments indicated that integrating more than one MRI sequence could significantly help remove detected structures close to the kidney graft, such as vessels irrigating the contrast agent. Our proposed method limitations include the number of hyperparameters that need to be defined, which are the same in all the deep learning based frameworks, and a lack of kidney prior information. Our quantitative results indicate that our tool could help clinical practice provide fast and robust monitoring of the kidney graft. In the future, we would like to investigate the impact of additional loss functions that could integrate more shape specific information. Moreover, we plan to investigate our segmentation framework's adaptation to the analysis of follow-up exams for patients subject to transplantation towards potential rejection.

7. COMPLIANCE WITH ETHICAL STANDARDS

Our study was approved by our Institutional Review Board, which waived the need for patients' consent. Data cohort corresponds to study reference ID-RCB: 2012-A01070-43 and ClinicalTrials.gov identifier: NCT02201537. All the data used in this study were anonymized.

8. REFERENCES

[1] J.H. Lee, J.W. Won, and C.-K. Oh, "Impact of the ratio of graft kidney volume to recipient body surface area on graft function after live donor kidney transplantation," Clinical transplantation, vol. 25, pp. E647–55, September 2011.

[2] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A.G. Morganti, and M. Bellomi, "Radiomics: the facts and the challenges of image analysis," European radiology experimental, vol. 2, no. 1, pp. 1–8, 2018.

[3] A.S. Lundervold, and A. Lundervold, "An overview of deep learning in medical imaging focusing on mri," Zeitschrift für Medizinische Physik, vol. 29, no. 2, pp. 102 – 127, 2019, Special Issue: Deep Learning in Medical Physics.

[4] M. Sandmair, M. Hammon, H. Seuss, R. Theis, M. Uder, and R. Janka, "Semiautomatic segmentation of the kidney in magnetic resonance images using unimodal thresholding," BMC Research Notes, vol. 9, December 2016.

[5] H. Li, N. Bao, X. Xu, Y. Zhang, S. Jin, Y. Jin, and H. Sun, "A renal vascular compartment segmentation method based on dynamic contrast-enhanced images," Technology and health care : official journal of the European Society for Engineering and Medicine, vol. 24, April 2016.

[6] C. Yu, and Y. Li, "A watershed method for mr renography segmentation," 2012 International Conference on Biomedical Engineering and Biotechnology, pp. 700–703, 2012.

[7] X. Yang, H. Le Minh, K.-T. Tim Cheng, K.H. Sung, and W. Liu, "Renal compartment segmentation in DCE-MRI images," Med Image Anal, vol. 32, pp. 269–280, August 2016.

[8] S. Li, F.G. Zöllner, A.D. Merrem, Y. Peng, J. Roervik, A. Lundervold, and L.R. Schad, "Wavelet-based segmentation of renal compartments in DCE-MRI of human kidney: initial results in patients and healthy volunteers," Comput Med Imaging Graph, vol. 36, no. 2, pp. 108–118, March 2012.

[9] W.A. Abdulahi, and J.R. Tapamo, "Fast chan-vese without edges and connected component analysis for kidney segmentation in mri images," in AFRICON 2015, 2015, pp. 1–5.

[10] C.-Y. Yu, W.-S. Zhang, Y.-Y. Yu, and Y. Li, "A novel active contour model for image segmentation using distance regularization term," Computers & Mathematics with Applications, vol. 65, no. 11, pp. 1746 – 1759, 2013.

[11] M. Shehata, A. Mahmoud, A. Soliman, F. Khalifa, M. Ghazal, M. Abou-El-Ghar, M. El-Melegy, and A. El-Baz, "3d kidney segmentation from abdominal diffusion mri using an appearance-guided deformable boundary," PLOS ONE, vol. 13, pp. e0200082, July 2018.

[12] W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," IEEE Transactions on Image Processing, vol. 29, pp. 8055–8068, 2020.

[13] E. Ferrante, P.K. Dokania, R. Marini, and N. Paragios, "Deformable registration through learning of context-specific metric aggregation," in International Workshop on Machine Learning in Medical Imaging. Springer, 2017, pp. 256–265.

[14] L. Fang, X. Wang, and L. Wang, "Multi-modal medical image segmentation based on vector-valued active contour models," Inform. Sci., vol. 513, pp. 504–518, 2020.

[15] H. Bindu, and K.S. Prasad, "An efficient medical image segmentation using conventional otsu method," International Journal of Advanced Science and Technology, vol. 38, no. 1, pp. 67–74, 2012.

[16] S. Will, P. Martirosian, C. Würslin, and F. Schick, "Automated segmentation and volumetric analysis of renal cortex, medulla, and pelvis based on non-contrast-enhanced t1- and t2-weighted mr images," Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 27, pp. 445–454, 2014.