# AUTOMATIC RENAL SEGMENTATION IN DCE-MRI USING CONVOLUTIONAL NEURAL NETWORKS

**Marzieh Haghighi**[*,†], **Simon K. Warfield**[†], and **Sila Kurugol**[†]

[*]Electrical and Computer Engineering Department, Northeastern University, Boston, MA, 02115

[†]Radiology Department, Boston Childrens Hospital; and Harvard Medical School, Boston MA 02115

## Abstract

Kidney function evaluation using dynamic contrast-enhanced MRI (DCE-MRI) images could help in diagnosis and treatment of kidney diseases of children. Automatic segmentation of renal parenchyma is an important step in this process. In this paper, we propose a time and memory efficient fully automated segmentation method which achieves high segmentation accuracy with running time in the order of seconds in both normal kidneys and kidneys with hydronephrosis. The proposed method is based on a cascaded application of two 3D convolutional neural networks that employs spatial and temporal information at the same time in order to learn the tasks of localization and segmentation of kidneys, respectively. Segmentation performance is evaluated on both normal and abnormal kidneys with varying levels of hydronephrosis. We achieved a mean dice coefficient of 91.4 and 83.6 for normal and abnormal kidneys of pediatric patients, respectively.

## Keywords

DCE-MRI; CNN; Kidney segmentation; Fully-automated

## 1. INTRODUCTION

Hydronephrosis refers to the fluid-filled enlargement of the kidney as a result of obstruction in its output of urine. It is found in 2-7% of all maternal ultrasound scans and 10% of these children may have a significant urological problem. Delayed intervention in infants with severe hydronephrosis may lead to permanent loss of kidney function with the potential for lifelong complications associated with chronic renal insufficiencies. MRI can provide clinically important markers of kidney function such as glomerular filtration rate (GFR) without exposing patient to ionizing radiation. Dynamic contrast-enhanced (DCE) MRI signal can be analyzed using pharmacokinetic (PK) models, and MR based GFR can be computed for determining whether a patient with persistent hydronephrosis will be referred for surgery or will receive conservative treatment. The PK models use time intensity curves of the kidney parenchyma region to calculate both a single kidney GFR and a GFR map. Accurate segmentation of kidney parenchyma is an important step to compute a robust and reliable GFR measure. Manual segmentation can take several hours. An accurate and robust technique for automated segmentation of kidney parenchyma (i.e. cortex and medulla) will

reduce the burden on radiologist and accelerate the translation of MR based GFR technique into clinical practice.

DCE-MR image series include 3D volumetric images acquired at different time points after contrast injection. The time intensity curves for different organs have different shapes. This temporal information can be used to discriminate kidney parenchyma from the other abdominal organs in the segmentation process. Several automated segmentation techniques have been proposed for renal segmentation using this temporal information [1] and [2], however, for patients with diseased kidneys, these methods often fail. A software with user interface (CHOP-fMRU) is available for semi-automated segmentation and functional analysis of DCE-MR images [3], however, it requires several manual inputs from the user, such as drawing several initial boundary curves around the regions of interest. Recent studies has attempted to combine spatial and temporal information using a series of heuristic steps [4] ,[5], some of which might fail in patients with pathological kidney and these approaches have relatively longer running times. Unlike previous segmentation techniques, deep learning algorithms process new test data very fast with running times on the order of seconds. Another drawback of the previous methods is the usage of hand crafted features and thresholds which fail to perform well for patients with abnormal kidneys with enlarged pelvis and thinned parenchyma regions. In contrast, our proposed network learns a hierarchical representation of spatial and temporal features during the training process, which results in improved performance in both normal and abnormal kidneys.

In this work, we propose a fully automated segmentation framework based on U-Net [6] architecture which was initially developed for 2D microscopy image segmentation. U-Net has the ability to capture local and global information for the image segmentation task and is able to generalize well from a small set of training samples. Variations of this network have also shown successful results for volumetric medical image segmentation [7],[8],[9]. We propose a 3D segmentation algorithm based on the 3D version of this architecture for automatic renal parenchyma segmentation in DCE-MR images. We will incorporate temporal features as the channel information for each voxel. However, applying the 3D U-Net for renal segmentation task has multiple challenges. First, each subject has a very large sized 4D data along with a large sized network parameter file, all of which can not be fitted into limited GPU memory and will therefore result in slow training process. Moreover, the resulting segmentation usually needs a refining step such as dense conditional random fields [10],[9], using auto-context [11] or similar methods for reducing false positives which makes the algorithm more time inefficient. Thus, we propose to divide our problem into two sub-problems that can be solved more efficiently in terms of time and memory when separated out: First, we apply a modified 3D U-Net on low resolution and augmented data for localizing the right and left kidneys; and second, we will apply U-Net on each extracted kidney region from the previous step for segmentation. Each of these sub-problems can be solved more quickly and need less memory compared to the naive approach. Our total test time is < 5 seconds for each new patient.

## 2. DATASET

We use DCE-MR images of a total of 30 pediatric patients acquired at 3T for six minutes after injection of Gadavist using radial stack-of-stars 3D FLASH sequence (TR/TE/FA 3.56/1.39ms/12, 32 coronal slices, voxel size $1.25 \times 1.25 \times 3mm$). 18 of these cases were retrospectively collected from hospital database from patients with hydronephrosis who received MRI as part of their clinical protocol within the last 2 years. 12 of these patients were recruited under a protocol approved by the Institutional Review Board for subjects who receive contrast-enhanced MRI as part of their clinical protocol to undergo additional research imaging with DCE-MRI within the same imaging session. We optimized acquisition protocol to achieve a mean temporal resolution of 3.3 sec for the arterial phase (2 minutes) and 13 sec for the remaining phases (4 minutes). 4D dynamic image series were reconstructed offline from raw data using a compressed sensing algorithm to improve temporal resolution and image quality, effectively reducing the streaking artifacts [12]. The age range of our pediatric patient group was between 2 months to 17 years. The field of view varied in patients with hydronephrosis according to the clinical protocol.

## 3. METHODS

In this section, we describe a memory efficient renal segmentation framework which automatically segments kidneys given a 4D DCE-MRI as input. As described in Section 2, having a time dimension for each voxel creates a very large data tensor for each subject. On the other hand, U-Net architecture have shown to be successful in many MRI segmentation applications. However, the 3D version of this architecture is not memory efficient and requires many parameters to learn. Considering the nature of our 4D data, we need to reduce the data and network size in the context of fully convolutional network for the purpose of GPU processing. Our proposed algorithm divides the problem into two sub-problems of localization and segmentation which can be solved more efficiently in terms of time and memory when separated out. Figure 1 summarizes the steps and the input data size of each step. We have used the fact that localization doesn't need high resolution image in space dimension and have used downsampled version of the image for localization. High resolution image inside the bounding box will be segmented in the second step. The preprocessing and the details of the networks used at each step are described in Section 3.1 and 3.2.

### 3.1. Localization

The 3D CNN network used for localization task is based on the 3D U-Net. Temporal features are mapped to the channel information dimension of the network. U-net architecture learns the model with good generalization performance using a small number of training samples. However, considering our data variability described in Section 2, data augmentation is necessary for the network to learn and generalize from a small number of training samples. Given the large memory needed to load the training data and network parameters, augmentation choice is very limited. The approach we used here was to reduce the data size in all dimensions where redundant information is present so that the memory size will be reduced. This enabled us to use data augmentation in order to achieve an

improved model fitting. To this end, we resampled the data from $224 \times 224 \times 32$ to $64 \times 64 \times 64$ to have sufficient resolution required for localization. We also reduced the time dimension using principle component analysis (PCA) and kept the first 5 dimensions with the highest variance (i.e. explaining 98% of variance in the data). The 4D data of each subject was then augmented for various scales and fed into the localization network. Augmentation details are given in Section 4. The localization network, based on the U-net architecture, consists of a contracting and an expansive path. Each layer in contracting path contains two 3×3×3 convolution filters followed by a rectified linear unit (ReLu) and $2 \times 2 \times 2$ max pooling with strides of two for down-sampling. In the expansive path, each layer consists of a convolutional transpose of $2 \times 2 \times 2$ by strides of two in each dimension, followed by two $3 \times 3 \times 3$ convolutions each followed by a ReLu. Layers with equal resolution from contracting path are concatenated to their corresponding layers in the expansive path to add high-resolution features to the expansive path. Finally, a $1 \times 1 \times 1$ convolution reduces the number of output channels to the number of classes in the last layer. The input data to this network is a $64 \times 64 \times 64$ image with 5 channels. We used dropout layers after each maxpooling layer in the contracting path to reduce the chance of overfitting based on high resolution features in the first layers. Batch normalization was also used before the final $1 \times 1 \times 1$ convolution layer for having faster convergence and less overfitting. Input labels are forming three channels of foreground/background labels corresponding to each class.

The network discriminates between three classes, namely right kidney, left kidney and background. However, there is an imbalanced distribution of samples in the kidney classes compared to the background class. We used a weighted cross entropy loss [10] in order to compensate for this imbalance and achieve more accurate learning when training the network. Weighted cross entropy loss is given by

$$L_{cross-entropy} = -\frac{1}{N}\sum_{i=1}^{N} w_i^c \left[ \hat{p}_i \log p_i + (1 - \hat{p}_i) \log(1 - p_i) \right] \quad (1)$$

where $p_i$ is the probability of voxel $i$ belonging to the foreground in each output channel and $\hat{p}_i$ represents the true label in the corresponding input channel. We fix $w_i^c$ to be inversely proportional to the probability of voxel $i$ belonging to the foreground class. We used softmax with weighted cross-entropy loss for network output and true labels comparison. Cost minimization on 1000 epochs was performed using ADAM optimizer with learning rate of 0.0001. The training time for this network was approximately one hour on a workstation with an NVIDIA Quadro 5000 GPU.

### 3.2. Segmentation

We trained the second network, which performs the segmentation task, using the bounding boxes of the manually labeled kidneys in the training set. The kidneys were cropped and then fed into the second network for training. We resampled all cropped kidneys to a common spatial dimension of $64 \times 64 \times 64$. We also interpolated and resampled the time intensity curves of each subject to a common temporal resolution and a common maximum

acquisition time of 5 minutes. Fifty samples from 5 minutes acquisition were interpolated to ensure keeping the maximum variance of time intensity curves for different classes using minimum number of samples. The segmentation network used in this framework is the same as the localization network with the exception that the drop out layers were removed. The input data to this network is a $64 \times 64 \times 64$ image with 50 channels and input labels are two channels of foreground/background labels corresponding to each kidney/non-kidney class. We again used softmax along with weighted cross entropy loss to compare network output and true segmentation labels. Cost minimization on 500 epochs was performed using ADAM optimizer with learning rate of 0.0001. The training time for this network was approximately one hour on a workstation with an NVIDIA Quadro 5000 GPU.

## 4. EXPERIMENTAL RESULTS

We trained the networks in the proposed framework for automated segmentation of normal and abnormal kidneys, using the data of 24 subjects (9 with normal and 15 with abnormal kidneys). We used precision, recall, dice coefficient (DSC) or F1-score and volumetric estimation error (VEE) for evaluating the algorithm segmentation performance. F1-score, which is the harmonic average of precision and recall, reports the accuracy of the overlap between the predicted and true manual segmentation. We also report the performance of the model, trained using 24 subjects, and tested on 6 previously unseen subjects (3 patients with normal and 3 patients with pathological kidneys) that were not included in the training process. As explained in section 3, we train each of the localization and the segmentation networks independently using the training data and the manual segmentation masks. Segmentation results are shown in Figure 2 for one normal and one abnormal kidney example from the test set. Middle figure in each row is showing the result of bounding box detection. Predicted output consisted of three classes; right kidney, left kidney and background. After extracting three classes from initial segmentation masks and forming the bounding boxes, each class was scaled to $64 \times 64 \times 64$ volumes and the original time dimension was resampled, interpolated and added to the data as the channel information. Finally, the segmentation classifies each voxel in the high resolution image into kidney or non-kidney class. Third figure in each row is showing the result of segmentation and re-positioning each kidney back into the detected bounding box. The resulted average performance measures for final unseen test cases are reported in Table. 1. Mean F1-scores for three patients with normal and three with abnormal kidneys were 91.4 and 83.6 respectively.

## 5. CONCLUSION

In this work we proposed a time and memory efficient fully-automated framework for segmentation of renal parenchyma using DCE-MRI data. The proposed learning based framework consists of two cascaded CNNs for localization and segmentation of kidneys. The proposed fully automated algorithm performed well in both normal and abnormal kidneys.

## ACKNOWLEDGEMENTS

## 7. REFERENCES

[1]. Zöllner Frank G, Sance Rosario, Rogelj Peter, Ledesma-Carbayo María J, Rørvik Jarle, Santos Andrés, and Lundervold Arvid, "Assessment of 3d dce-mri of the kidneys using non-rigid image registration and segmentation of voxel time courses," Computerized Medical Imaging and Graphics, vol. 33, no. 3, pp. 171–181, 2009. [PubMed: 19135861]

[2]. Chevaillier Béatrice, Ponvianne Yannick, Collette Jean-Luc, Mandry Damien, Claudon Michel, and Pietquin Olivier, "Functional semi-automated segmentation of renal dce-mri sequences," in Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008, pp. 525–528.

[3]. Khrichenko Dmitry and Darge Kassa, "Functional analysis in mr urographymade simple," Pediatric radiology, vol. 40, no. 2, pp. 182–199, 2010. [PubMed: 20012602]

[4]. Yoruk Umit, Hargreaves Brian A, and Vasanawala Shreyas S, "Automatic renal segmentation for mr urography using 3d-grabcut and random forests," Magnetic Resonance in Medicine, 2017.

[5]. Yang Xin, Le Minh Hung, Tim Cheng Kwang-Ting, Sung Kyung Hyun, and Liu Wenyu, "Renal compartment segmentation in dce-mri images," Medical image analysis, vol. 32, pp. 269–280, 2016. [PubMed: 27236222]

[6]. Ronneberger Olaf, Fischer Philipp, and Brox Thomas, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2015, pp. 234–241.

[7]. Çiçek Özgün, Abdulkadir Ahmed, Lienkamp Soeren S, Brox Thomas, and Ronneberger Olaf, "3d u-net: learning dense volumetric segmentation from sparse annotation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 424–432.

[8]. Milletari Fausto, Navab Nassir, and Ahmadi Seyed-Ahmad, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016, pp. 565–571.

[9]. Kamnitsas Konstantinos, Ledig Christian, Newcombe Virginia FJ, Simpson Joanna P, Kane Andrew D, Menon David K, Rueckert Daniel, and Glocker Ben, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," Medical image analysis, vol. 36, pp. 61–78, 2017. [PubMed: 27865153]

[10]. Christ Patrick Ferdinand et al., "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 415–423.

[11]. Mohseni Salehi Seyed Sadegh, Erdogmus Deniz, and Gholipour Ali, "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging," IEEE Transactions on Medical Imaging, 2017.

[12]. Feng Li, Grimm Robert, Tobias Block Kai, Chandarana Hersh, Kim Sungheon, Xu Jian, Axel Leon, Sodickson Daniel K, and Otazo Ricardo, "Golden-angle radial sparse parallel mri: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric mri," Magnetic resonance in medicine, vol. 72, no. 3, pp. 707–717, 2014. [PubMed: 24142845]
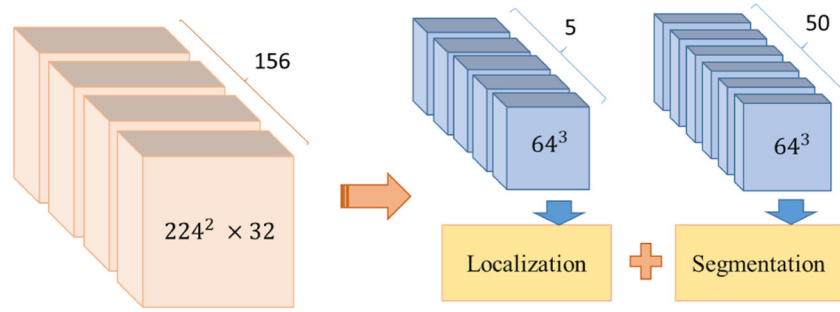
**Fig. 1.**
Automated segmentation steps: Original image segmentation problem is divided to: 1) localization based on low resolution data. 2) segmentation of each localized kidney.
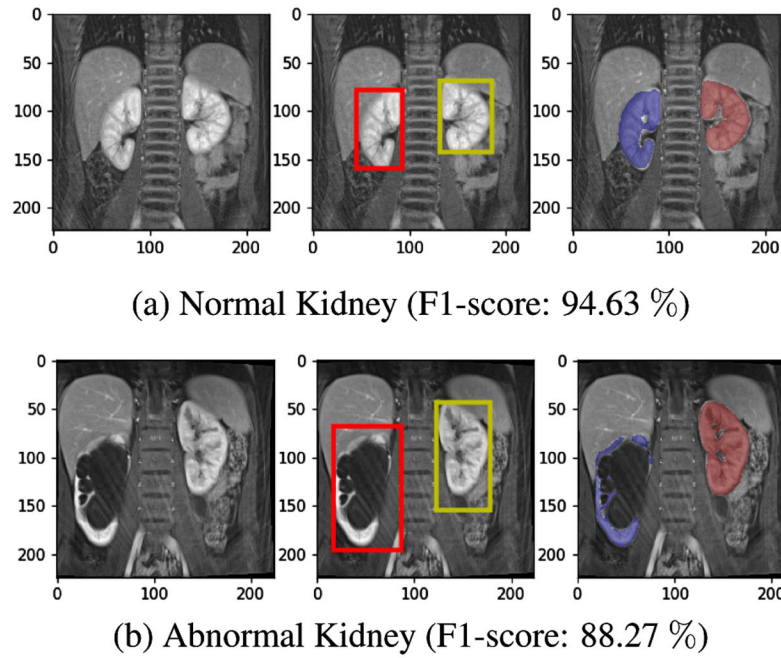
(a) Normal Kidney (F1-score: 94.63 %)

(b) Abnormal Kidney (F1-score: 88.27 %)

**Fig. 2.**
Kidney segmentation results:1) Localization using simplified 3D U-Net, 2) Segmentation of the cropped area using 3D U-Net

**Table 1.**

Renal segmentation performance.

| Session | Test (mean ± sd) | |
|---|---|---|
| Kidney | Normal | Abnormal |
| Precision | 94.8±0.05 | 90.6±0.03 |
| F1-score | 91.4±0.03 | 83.7±0.03 |
| VEE | 12.9±2.4 mL | 19.5±3.2mL |
| Time | ~3s | |