Deep convolutional neural networks in application to kidney segmentation in the DCE-MR images

Artur Klepaczko[1], Eli Eikefjord[2], and Arvid Lundervold[2,3,4]

[1] Lodz University of Technology, Łódź, Poland, aklepaczko@p.lodz.pl
[2] Dept. Health and Functioning, Western Norway University of Applied Sciences
[3] Department of Biomedicine, University of Bergen, Bergen, Norway
[4] Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University Hospital, Bergen, Norway

Abstract. This paper evaluates three convolutional neural network architectures – U-Net, SegNet, and Fully Convolutional (FC) DenseNets – in application to kidney segmentation in the dynamic contrast-enhanced magnetic resonance images (DCE-MRI). We found U-Net to outperform the alternative solutions with the Jaccard coefficient equal to 94% against 93% and 91% for SegNet and FCDenseNets, respectively. As a next step, we propose to classify renal mask voxels into cortex, medulla, and pelvis based on temporal characteristics of signal intensity time courses. We evaluate our computational framework on a set of 20 DCE-MRI series by calculating image-derived glomerular filtration rates (GFR) – an indicator of renal tissue state. Then we compare our calculated GFR with the available ground-truth values measured in the iohexol clearance tests. The mean bias between the two measurements amounts to -7.4 ml/min/1.73m^2 which proves the reliability of the designed segmentation pipeline.

Keywords: Semantic segmentation · Convolutional neural networks · Dynamic Contrast-Enhanced MRI · Kidney perfusion modeling.

1 Introduction

1.1 Diagnostics of the kidney

Contrast-enhanced magnetic resonance imaging (CE-MRI) is one of the methods routinely used in clinics for the diagnosis of renal impairments. It allows visualization of the kidney lesions such as tumor, cysts or focal segmental glomerulosclerosis. Moreover, if image acquisition is performed in a dynamic sequence, resulting in a temporal series of CE-MRI volumes, it is possible to determine the glomerular filtration rate (GFR) based on image data. GFR can be affected by various renal diseases which lead to the loss of kidney filtration performance. Thus, quantification of renal perfusion provides an objective way for assessment of the potential of the kidney to restore its functional characteristics. In contrast to gold standard serum creatinine clearance test, image-derived GFR can

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

be calculated for a single kidney, thus allowing lateral differentiation of kidney diseases, while providing spatially-resolved information on tissue demage.

In principle, the dynamic contrast-enhanced (DCE) MRI examination produces a series of T1-weighted volumes acquired at multiple time steps in the time interval covering the passage of a contrast agent (CA) through the abdominal arterial system. While the CA bolus enters the capillary bed and then tubular network of the kidneys, it effectively increases the T1 relaxation time of the penetrated tissues and modifies the contrast in the image. The temporal dynamics of this image signal intensity reflects physiological conditions of kidney function and constitutes the basis for pharmacokinetic (PK) modeling.

There have been numerous PK models proposed in the literature. Among those specifically dedicated to the kidney, one should mention the 2-compartment separable and filtration models proposed in [17] and [19], respectively. In this paper, the latter model will be used to estimate GFR for the experimental data available in our study. There were also more complex models proposed, such as e.g. the six-compartment formulation in [12]. However, these approaches became less popular due to unstable behavior of optimization procedure while fitting models parameters to image data.

In any case, application of the PK models require prior segmentation of the kidney parenchyma into cortex and medulla. Moreover, the pelvis needs to be separated as it does not contribute to the renal filtration process. Only voxels which contain renal nephrons have to be take into account. Thus, precise segmentation of the kidney regions is a crucial step in automated analysis of the dynamic contrast-enhanced MRI.

1.2 Related work

The problem of kidney segmentation has been tackled by many authors. Frequently, voxels are classified based on their intensity time courses. For example in [23], voxels are clustered by the k-means algorithm. This approach was further developed in [13], where signal intensity time courses were preprocessed by the discrete wavelet transform. Image volumes were first manually cropped to cubic regions of interest (ROI) covering single kidneys. Then, the extracted wavelet coefficients of the ROI voxel time curves were submitted to the k-means clusterer. A problem which appears here is that there arises a need for the clustering algorithm to extract not only the three renal classes but also to separate them from the surrounding tissues. It is a frequent source of error since other tissues (liver, spleen and pancreas) exhibit signal dynamics similar to the kidney.

Therefore, a common strategy consists in firstly separating a whole kidney from other parts of an image. The delineated regions of interest should precisely fit kidney borders in order to get rid of all neighboring voxels. An example of such a solution are the area-under-the-curve maps (AUC) [3]. Those voxels in the DCE-MRI sequence which are penetrated by the tracer agent appear bright on AUC maps due to the largest area under their signal intensity time courses.

The coarse-to-fine segmentation strategy was also applied in [21], where the concept of Maximally Stable Temporal Volumes (MSTV) was introduced. The

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

MSTV features allow to recognize kidneys by detecting spatially homogeneous and temporally stable structures. Spatial homogeneity is defined in terms of image binarization performed with a large span of thresholds. Independently from a threshold value renal voxels remain bright and possess bright neighbors in all 3 directions of the 3D space. Temporal stability, in turn, is reflected in the fact that spatial homogeneity of kidney voxels is observed in adjacent time frames of the imaging sequence. Fine-grained segmentation is obtained by reducing voxels time courses to vectors of principal components, which are next partitioned by k-means to multiple clusters.

Similarly, in [22], the first stage of the segmentation procedure employs GrubCut algorithm to create renal masks. Fine-tuning is achieved by classifying voxels with a pre-trained random forest classifier. Voxels are characterized by their respective image intensities in selected time frames of the dynamic sequence as well as their location within the ROIs constructed in the first stage. Although both MSTV- and GrubCut-based contributions seem to produce satisfactory results, they are rather conceptually complex algorithms, unavailable in open-source software. As such, they cannot be easily adopted by the clinical community.

On the other hand, with the advent of deep learning (DL) methods, semantic segmentation networks offer an attractive computational methodology for the problem of automatic kidney delineation in MR images. For example, in [10] various network architectures, i.e. fully convolutional network [16], SegNet [1], U-Net [15], and DeepLabV3+ [2], have been tested for segmentation of prostate cancer in T2-weighted MRI. Anatomical MR images of polycystic kidneys were segmented by a custom convolutional neural network (CNN) in [11]. Another approach has been presented in [20], where deep learning was employed for direct inference of brain perfusion maps from a DCE-MRI sequence without explicitly fitting a PK model to measured signals. However, there has been only a moderate number of DL-based approaches targeting DCE-MRI of the kidney. As one of few exceptions, the study described in [5] presents a cascade of two CNN networks. The first network roughly localizes the left and right kidney in a 4D DCE-MR image, whereas the second one performs fine delineation of renal borders.

1.3 Current contribution

This paper presents a novel computational framework for automated segmentation of the kidney and its internal structures in the DCE-MR images using:

1. a convolutional neural network for delineation of the kidney parenchyma;
2. a classifier trained in supervised manner to partition parenchymal voxels into cortex, medulla and pelvis.

With regard to task 1, in order to find an optimal solution we have test three encoder-decoder CNN architectures: U-Net [15], SegNet [1] and 100-layers Fully Convolution DenseNets [9], referred to as *Tiramisu*. For the task 2, we have selected the support vector machine (SVM) classifier with the radial basis function kernel as it proved in our experiments to outperform other tested algorithms.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

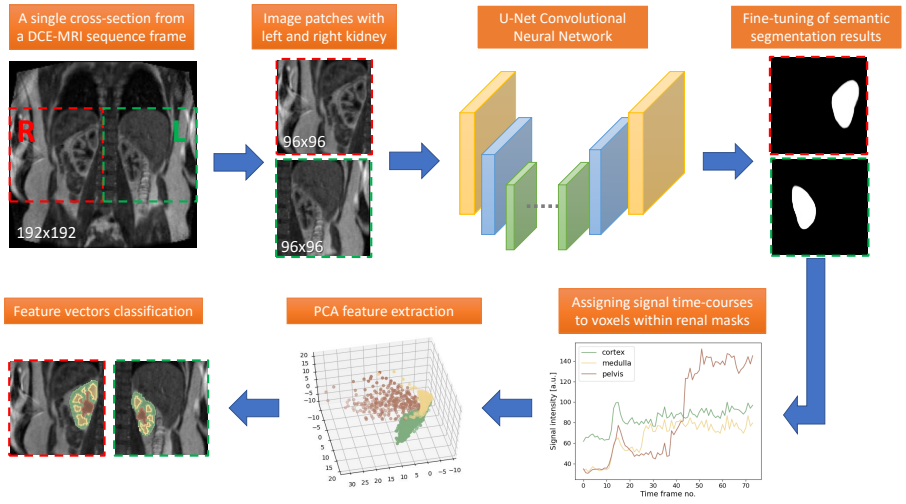

Fig. 1. Overview of the designed segmentation pipeline.

The proposed segmentation pipeline is visualized in Fig. 1. The initial coarse segmentation is accomplished by a neural network. This step is performed on subsequent two-dimensional cross-sections of a single volumetric image from the DCE-MRI sequence. This image corresponds to the frame of the highest signal enhancement in the cortex region, when the partitioning of the renal parenchyma into cortex and medulla is clearly visible.

Each cross-section is divided into left and right sides of 96-pixel width. On a given side, a centrally located image patch of 96-pixels height is selected. In this way, we ensure that left and right kidneys are processed separately. Prior to segmentation, we perform DCE image series matching in the time domain using B-splines deformable registration [8] to suppress motion artifacts. Thanks to image registration, kidney masks generated in one frame can be applied to all other frames of the dynamic series. Thus, renal voxels are prescribed feature vectors composed of MRI signal intensity values measured in subsequent time points. In order to obtain more general and compact characteristics of the signal dynamics, we extract feature aggregates using PCA transform. Eventually, a classifier trained to discriminate between temporal characteristics of cortex, medulla and pelvis regions assigns a voxel to an appropriate category.

In order to prove scalability of the designed framework, the proposed methodology was verified in the leave-one-subject-out manner based on the cohort of 10 healthy volunteers, each scanned twice with the DCE-MRI method. Hence, CNN network and classifier training was repeated 10 times, each time with one patient put apart. It was then possible, to objectively verify, how the system behaves in case of a subject not seen during the training phase. Between the scanning sessions, renal performance of each subject was evaluated using iohexol clearance

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

procedure to establish the ground-truth value of the glomerular filtration rate and enable validation of the image-derived GFR measurements.

2 Semantic segmentation of the kidney

2.1 U-Net

The U-Net convolutional neural network was originally developed for segmentation of neuronal structures in electron microscopic stacks and proved effective in numerous other biomedical applications. As said, the input to our model is a 2D grey-level image – a 96×96-pixel patch of a DCE-MRI volume cross-section.

The U-Net network contains two symmetric paths – a contractive and an expansion one. The goal of the contractive path is to encode image pixel intensity patterns. It is accomplished by convolution with a series of 3×3 filters of trainable weights. Filters outputs activate the main processing components of the network – the neurons called rectified linear units (ReLU). They allow for modeling non-linear relationship between image features and the output segmentation map. It is followed by the 2×2 max-pooling operation which down-samples the feature maps.

Contraction is repeated four times to extract image descriptors on various scale levels. Each level is formed by two convolutional layers followed by a batch normalization layer. The convolution and normalization layer pairs are separated by the dropout layer, which randomly sets 20% of the input nodes to 0. This mechanism, active only during the training phase, prevents the network from overfitting [18].

The output of the last down-sampling block is passed on to the expansion path. It is built up from the same number of up-sampling levels as the contractive part and its main task is to recover original spatial resolution. In this study, up-sampling is realized by transposed convolution. Decoding blocks are also composed of two pairs of convolutional and batch normalization layers. However, no dropout mechanism is inserted in-between. Moreover, the high-resolution feature maps extracted in the down-sampling path not only feed the subsequent encoding layers but they are also concatenated to the inputs of the decoding layers at the respective levels of the up-sampling path. These additional connections help the decoding blocks to restore kidney segments localization more precisely.

The output of the last up-sampling block is connected to a convolutional layer with a 1×1-size filters. It performs pixel-wise convolution of the filter kernel with a 64-element feature vector and then submits the result to an output activation function. In our design, a sigmoid activation is used since the final decision is binary – a pixel belongs to renal parenchyma or background.

2.2 SegNet

SegNet is another encoder-decoder architecture [1], whose main characteristic is the application of the VGG16 [14] topology as the encoder backbone. Specifically it uses its first 13 convolulational layers to extract image features. Also, the

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

method introduced in the decoder path to up-sample feature maps was different than the strategies used in fully convolutional networks. In this alternative approach, SegNet keeps record of the pixel indices selected by the max-pool operation and uses them in the corresponding decoder levels to perform non-linear up-sampling. Up-sampled maps are, in principle, zero-padded in positions not indicated by the memorized indices. Eventually, dense feature maps are created by convolution of up-sampled maps with trainable filter banks. Originally, SegNet, as majority of semantic segmentation neural networks, were developed for outdoor and indoor scene understanding, usually represented by the RGB image files. It conforms with the input of the VGG16 architecture pretrained for the color-coded images. Thus, its use for greyscale DCE-MR data requires duplication of the single image intensity channel to two other color channels, which is obviously a computational overhead.

2.3 Fully convolutional DenseNets

The Tiramisu network builds on the concept of Densely Connected Convolutional Networks, which occurred effective in multiple classification tasks [7]. In this approach, both the encoding and decoding paths contain the so-called dense blocks. Each block is composed of batch normalization layer, ReLU activation, 3×3 convolution and dropout (with proability = 0.2). The input to each layer is concatenated with its output to feed the next layer. Also, each layer output is concatenated to the final output of the dense block. In between of the dense blocks there are transition down and transition up units which perform either max-pooling or up-sampling. The latter operation is conducted by transposed convolution, similarily to the U-Net architecture. The number of layers within each dense block can be adjusted to the needs of a given application. In our experiements we used the same configuration as it was proposed in the original paper. Therefore, we used 4, 5, 7, 10 and 12 layers in the contractive path, and the same number of layers in the reverse order in the expansion part of the network. Together with the transition down and up layers, there were 103 layers in total.

2.4 Network training

In case of each network, trainable weights were initiated to random state by the method of He et al. [6]. The training process was conducted on image patches cropped from the DCE-MRI volumes, each containing a single, left or right kidney cross-section. As described above, 96×96-pixel image patches were extracted from volumes of the DCE sequence corresponding to the perfusion phase, i.e. time frames of the maximum signal contrast between cortex and medulla. In order to increase the number of training images, for each study we actually selected 3 such time frames – the one with maximum signal enhancement in the cortex region plus one preceding and one succeeding time frame. In each image volume, a single kidney is visible on 12 slices on average. It gives approximately 1440 training patches.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

Additionally, we enlarged the training set through data augmentation. This was accomplished by picking 10 different vertical positions of the image patch and by randomly mirroring it in the horizontal direction. While selecting patch positions, we made sure that it embraced sufficiently large portion of the image center containing significant fragment of the renal parenchyma (see Fig. 2). Overall, the number of images available for training reached the value of 13964. One-third of the training images were separated for the validation purposes.

As said above, we have trained 10 different CNN models, one for every patient. While building a model dedicated to a given subject, its corresponding image patches (irrespectively of the examination session) were removed from the training and validation sets and used only for testing. Weights of the network were updated using the stochastic gradient descent algorithm with the constant learning rate = 0.01 and momentum = 0.99. The loss function chosen to optimize was the binary cross-entropy criterion, defined as

$$\mathcal{H} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(p(y_i)\right) + (1-y_i)\log\left(1-p(y_i)\right) \tag{1}$$

where N is the number of voxels, y_i is the true voxel label, and $p(y_i)$ denotes the probability that an i-th voxel belongs to y_i category. Additionally, in order to

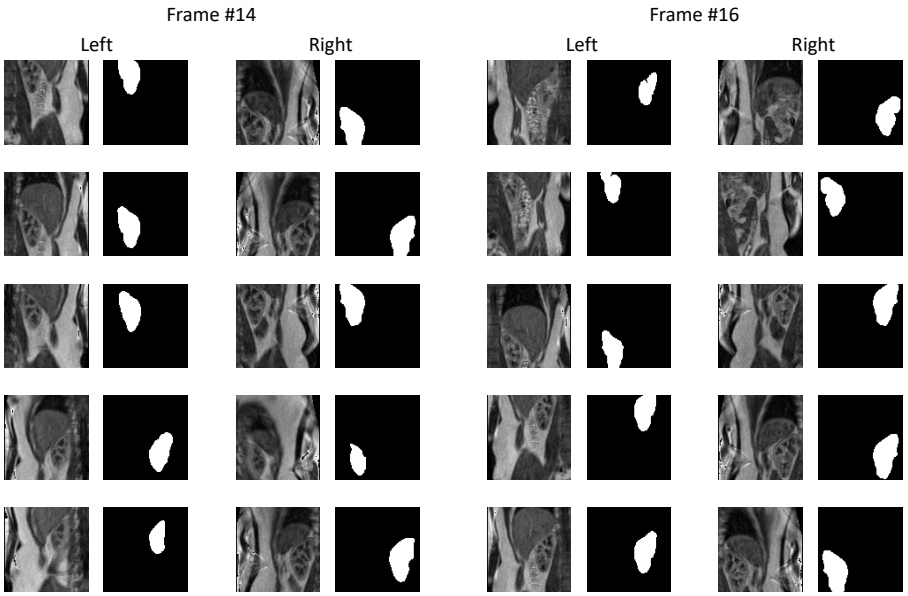

Fig. 2. Examples of training image patches extracted from left and right kidneys from two time frames of Subject 1. Data augmentation was realized by image flipping in horizontal direction and vertical shifting of patch location relative to image center.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

monitor the quality of kidney segmentation over training epochs, we calculated the Jaccard coefficient, hereafter designated as IoU (intersection-over-union)

$$IoU = \frac{\sum_{i=1}^{K} y_i \wedge y_i^{pred}}{\sum_{i=1}^{N} y_i \vee y_i^{pred}}, \tag{2}$$

where K designates the number of pixels in a processed slice and y_i^{pred} is the predicted pixel category. Here, categories are Boolean-valued and a pixel is labeled *True* if it belongs to the kidney, *False* otherwise. In the case of each subject, the optimization algorithm was run for 50 epochs. The stored model corresponded to the epoch with the minimum score on the loss function obtained for the validation data set.

3 Renal voxels classification

3.1 Feature extraction

Differentiation of voxels representing particular renal compartments could be based on raw signal intensity time courses. We propose, however, to transform signal waveforms into the space of reduced dimensionality using principal component analysis (PCA). The purpose of this transform is not only to decrease the complexity of the resulting classification model but also to extract a more general characteristics of the kidney tissue, representative for various subjects. We presumed that the extracted PCA components should explain at least 90% of the original data set variance. Therefore, in the case of our experimental data (see Sect. 5.1), where each dynamic sequence consisted of 74 time frames, vectors of 74 temporal features (i.e. image signal intensities in subsequent time steps) were transformed into the space of 20 PCA feature aggregates.

3.2 Feature vectors classification

Assignment of renal voxels to cortex, medulla or pelvis is performed by a classifier trained in the supervised manner. In our approach, historical data serve as patterns for building appropriate decision rules, later applied to new studies. The training vectors were acquired from regions of interest manually annotated in the respective parenchymal locations. The annotations were made only in voxels whose membership was unambiguous (see Fig. 3a-b), thus letting a trained classifier to decide about the dominating tissue category in case of voxels partially filled with various compartments. The number of training vectors collected from the 20 available examinations exceeded the value of 60,000. This data set was partitioned into 10 folds, each containing data vectors from all but one subject, left apart for testing purposes. In a given fold, the class distribution was approximately as follows: cortex – 58%, medulla – 31%, pelvis – 11%. In order to give classifiers a chance to learn to discriminate categories with equal accuracy, in each training fold the subsets representing cortex and medulla were resampled

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

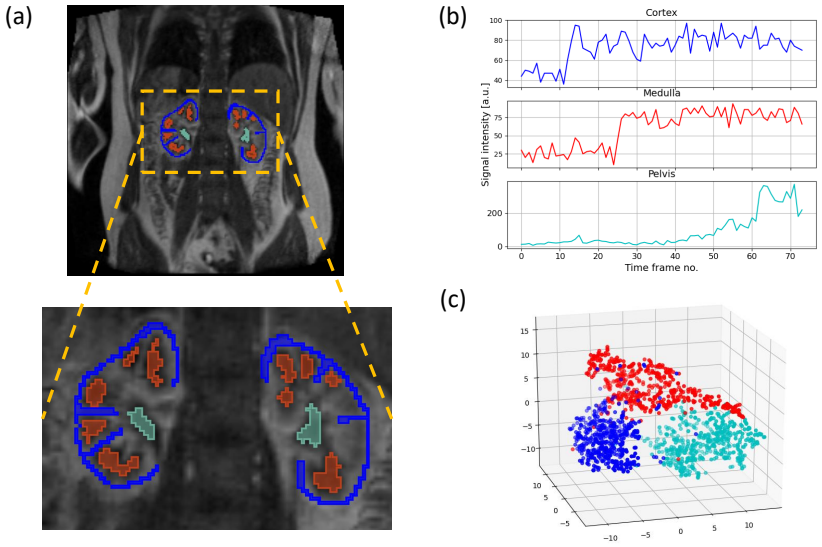

Fig. 3. Preparation of training data for supervised learning of classifiers: a) ROI placement in a DCE-MRI frame; b) signal time courses assigned to corresponding ROI voxels; c) 3-dimensional visualization of PCA feature vectors representing cortex (blue), medulla (red) and pelvis (magenta) ROIs. The visualization was obtained by transforming 20 PCA features using t-SNE method.

to match the size of the pelvis category. On average, the training set after resampling embraced over 16,000 vectors per fold. In a given training fold, data from both examination sessions were included. On the other hand, the testing folds contained from 600 to 4,800 vectors depending on the patient and examination session. Classifiers were evaluated using the balanced accuracy score calculated on the test sets. As previously noted, we used support vector machines with the radial basis function as the kernel to accomplish the classification task.

4 Pharmacokinetic modeling

The 2-compartment filtration (2CFM) model assumes that signal measured in a given tissue voxel is a sum of contributions originating from intravascular (IV) and extracellular extravascular (EEV) spaces [19]. Furthermore, as in each PK model, the delivery of the gadolinium contrast agent through a feeding artery to the kidney, is encapsulated by the so-called arterial input function (AIF). Practically, AIF in case of the kidney studies, is the time-course of the contrast agent concentration in the abdominal aorta [4]. By convolving the AIF with a shifting and dispersion kernel one obtains tracer concentration in the IV compartment. Eventually, the time curve of the concentration in the EEV space is proportional

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

to the integral of the concentration in the IV compartment. The proportionality coefficient, denoted as K^{trans}, controls the rate of CA transfer from IV to EEV compartment. K^{trans} multiplied by volume of the organ leads to calculation of GFR.

Formally, the CA concentration in the tissue is given by

$$C_{tissue}(t) = K^{trans} \int_0^t C_p^{kid}(\tau) + v_p C_p^{kid}(t), \tag{3}$$

with

$$C_p^{kid} = C_p^{art} \otimes g(t) = \int_0^t C_p^{art}(t-\tau)\, g(\tau)\, d\tau, \tag{4}$$

where C_p^{art} denotes the arterial input function, v_p – plasma volume fraction, and C_p^{kid} – CA concentration in the blood plasma. The first term in (3) represents the CA concentration in the EEV space, whereas the second term covers the concentration in the IV space obtained by convolving arterial input function with the vascular impulse response function (VIRF), defined as

$$g(t) = \begin{cases} 0 & t < \Delta \\ \frac{1}{T_g} e^{-\frac{t-\Delta}{T_g}} & t \geq \Delta \end{cases}. \tag{5}$$

Variables T_g – the dispersion time constant, and Δ – the delay interval, together with the volume fraction v_p and transfer constant K^{trans} form the complete set of the 2CFM model parameters, which we find using the Levenberg-Marquardt non-linear least squares curve-fitting procedure.

5 Experiments

5.1 MRI data

Twenty DCE-MRI examinations were available for experiments. The datasets were collected for 10 healthy volunteers. Each subject was imaged twice, seven days apart (further, these examinations will be referred to as Session 1 and 2). The acquisition sequence used standard 3D FLASH spoiled gradient recalled echo technique with the following parameters: TR = 2.36 ms, TE = 0.8 ms, FA = 20°, in-plane resolution = 2.2 × 2.2 mm^2, slice thickness = 3 mm, acquisition matrix = 192 × 192, number of slices = 30. Prior to image acquisition, patients were administered 0.025 mmol/kg of GdDOTA at 3 mL/s flow rate. The contrast agent was injected intravenously. Then, 74 volumetric scans were gathered at 2.3 seconds time intervals. In order to validate the obtained estimates against ground truth values, volunteers underwent iohexol clearance tests. The measurement was carried out by administrating a dose of 5 mL of iohexol (300 mg I/mL; Omnipaque 300, GE Healthcare) and then by acquiring a venous blood sample after 4 hours. All volunteers gave their written informed consent for participation in the examinations. The study protocol, including its ethical aspects, was approved by the Institutional Review Board at the Haukeland University Hospital Bergen, Norway.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

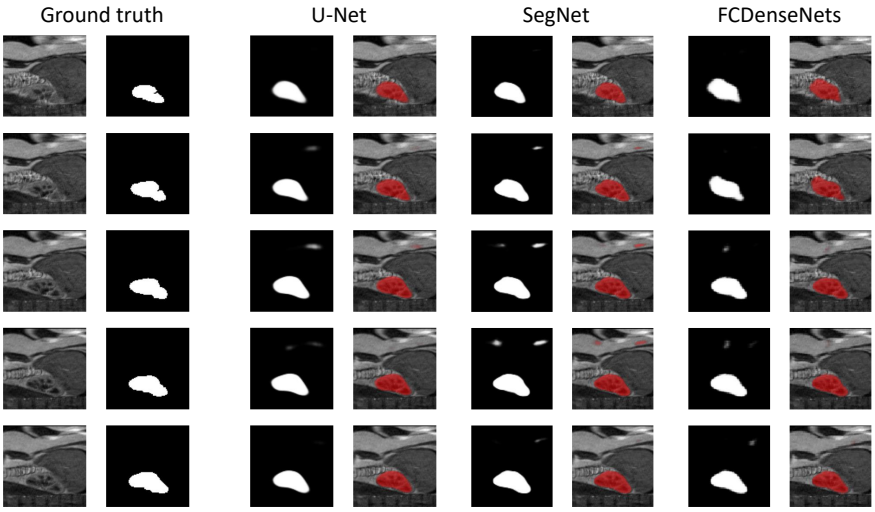

Fig. 4. Example outputs of the tested segmentation networks.

5.2 Results

Figure 4 shows example outputs of the tested semantic segmentation networks for one of the participating subjects along with the ground-truth annotation masks. In the selected image samples, it can be observed that SegNet in many cases produces false positive regions around actual renal tissue. Apparently, however, as also shown in Table 1, across all participating subjects, it was Tiramisu network which failed to precisely delineate parenchymal borders.

In the next stage, parenchymal voxels were classified into separate renal compartments. The results of classification for the test sets are presented in Table 2. All presented scores are mean values over 10 subjects. The SVM exhibits the balanced accuracy equal to 96% and also gains high rate of true positive detections, as well as it seems to be relatively robust against false predictions. Using our algorithm, we achieved the mean Jaccard coefficient for the cortex class in the left kidney equal to 93.2%. In case of the other regions the IoU equated approximately 91%, except for the pelvis class in the left kidney where it dropped to

Table 1. Mean Jaccard coefficients over all subjects and MR sessions.

	IoU	
CNN architecture	Left	Right
U-Net	0.941	0.940
SegNet	0.932	0.927
FCDenseNets	0.912	0.908

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

Table 2. Classification metrics averaged over the testing sets – subjects 1-10, both examination sessions.

Balanced accuracy	Cortex		Medulla		Pelvis	
	Recall	Precision	Recall	Precision	Recall	Precision
0.956	0.951	0.970	0.954	0.941	0.962	0.919

90.1%. The quality of fine segmentation can be visually confirmed by analyzing examples of the kidney decomposition into regions shown in Fig. 5.

Validity of the results was further verified by using the obtained segmentation masks in the process of GFR assessment. The mean signals in the cortex regions were fitted to the 2CFM pharmacokinetic model. Then, reproducibility of image-based GFR estimates were compared against iohexol-derived measurements using the Bland-Altman method. The mean difference μ_d for the MR examination Session 1 was equal to -7.4 ml/min/1.73 m^2. In the case of Session 2, the agreement with the reference method was found weaker (-12.9 versus -14.1 ml/min/1.73 m^2).

6 Conclusions

To conclude, in this paper we presented a computational framework for supporting quantitative assessment of kidney perfusion by providing an automated way of kidney parenchyma segmentation. We compared three CNN architectures for semantic segmentation. The obtained results demonstrated superior performance of the classic U-Net network over SegNet and FCDenseNets structures. Morever, we showed that classification of voxels belonging to the kidney masks automatically found by our designed U-Net network leads to reliable quantification of

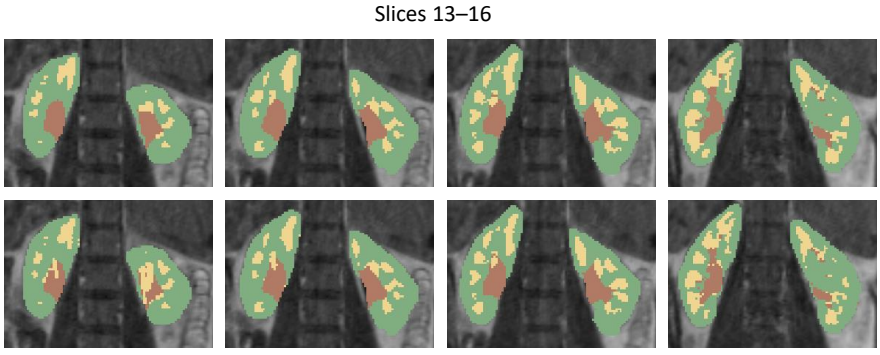

Fig. 5. Comparison of segmentation results obtained by the proposed method (bottom row) with ground truth annotations (top).

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

renal perfusion. These findings bring closer the clinical application of DCE-MR imaging as a routine method in kidney diagnostics. The designed segmentation method allows for increased objectivism of the image-derived perfusion parameters and also potentially faster diagnosis of renal impairments.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (2017). https://doi.org/10.1109/TPAMI.2016.2644615
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. pp. 833–851. Springer International Publishing (2018)
3. Choi, Y., Ahn, S., Lee, H.J., Chang, J., Kang, S.G., Kim, E., Kim, S., Lee, S.K.: The initial area under the curve derived from dynamic contrast-enhanced mri improves prognosis prediction in glioblastoma with unmethylated mgmt promoter. American Journal of Neuroradiology **38**(8), 1528–1535 (2017)
4. Cutajar, M., Mendichovszky, I., Tofts, P., Gordon, I.: The importance of aif roi selection in dce-mri renography: reproducibility and variability of renal perfusion and filtration. European Journal of Radiology **74**(3), e154–60 (2010)
5. Haghighi, M., Warfield, S.K., Kurugol, S.: Automatic renal segmentation in DCE-MRI using convolutional neural networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1534–1537 (2018). https://doi.org/10.1109/ISBI.2018.8363865
6. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1–11 (December 2015)
7. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243
8. Johnson, H.J., McCormick, M.M., Ibanez, L.: The ITK Software Guide: Design and Functionality. Kitware (2020)
9. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1175–1183 (2017). https://doi.org/10.1109/CVPRW.2017.156
10. Khan, Z., Yahya, N., Alsaih, K., Ali, S.S.A., Meriaudeau, F.: Evaluation of deep neural networks for semantic segmentation of prostate in T2W MRI. Sensors **20**(11), 3183 (June 2020). https://doi.org/10.3390/s20113183
11. Kline, T., Korfiatis, P., Edwards, M., Blais, J., Czerwiec, F., Harris, P., King, B., Torres, V., Erickson, B.: Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. Journal of Digital Imaging **30** (May 2017). https://doi.org/10.1007/s10278-017-9978-1
12. Lee, V.S., Rusinek, H., Bokacheva, L., Huang, A.J., Oesingmann, N., Chen, Q., Kaur, M., Prince, K., Song, T., Kramer, E.L., Leonard, E.F.: Renal function measurements from MR renography and a simplified multicompartmental model.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50

American Journal of Physiology-Renal Physiology **292**(5), F1548–F1559 (2007). https://doi.org/10.1152/ajprenal.00347.2006
13. Li, S., Zöllner, F., Merrem, A., Peng, Y., Roervik, J., Lundervold, A., Schad, L.: Wavelet-based segmentation of renal compartments in dce-mri of human kidney: Initial results in patients and healthy volunteers. Computerized Medical Imaging and Graphics **36**, 108–18 (June 2012). https://doi.org/10.1016/j.compmedimag.2011.06.005
14. Liu, S., Deng, W.: Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). pp. 730–734 (2015). https://doi.org/10.1109/ACPR.2015.7486599
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science. vol. 9351, pp. 234–241 (10 2015)
16. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 640–651 (2017). https://doi.org/10.1109/TPAMI.2016.2572683
17. Sourbron, S.P., Michaely, H.J., Reiser, M.F., Schoenberg, S.O.: MRI-measurement of perfusion and glomerular filtration in the human kidney with a separable compartment model. Investigative Radiology **43**(1), 40—48 (January 2008). https://doi.org/10.1097/rli.0b013e31815597c5
18. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research **15**(56), 1929–1958 (2014)
19. Tofts, P., Cutajar, M., Mendichovszky, I., Peters, A., Gordon, I.: Precise measurement of renal filtration and vascular parameters using a two-compartment model for dynamic contrast-enhanced MRI of the kidney gives realistic normal values. European Radiology **22**, 1320–30 (March 2012). https://doi.org/10.1007/s00330-012-2382-9
20. Ulas, C., Das, D., Thrippleton, M.J., Valdés Hernández, M.d.C., Armitage, P.A., Makin, S.D., Wardlaw, J.M., Menze, B.H.: Convolutional neural networks for direct inference of pharmacokinetic parameters: Application to stroke dynamic contrast-enhanced MRI. Frontiers in Neurology **9**, 1147 (2019). https://doi.org/10.3389/fneur.2018.01147
21. Yang, X., Le Minh, H., (Tim) Cheng, K.T., Sung, K.H., Liu, W.: Renal compartment segmentation in dce-mri images. Medical Image Analysis **32**(C), 269–280 (2016). https://doi.org/10.1016/j.media.2016.05.006
22. Yoruk, U., Hargreaves, B.A., Vasanawala, S.S.: Automatic renal segmentation for MR urography using 3D-GrabCut and random forests. Magnetic Resonance in Medicine **79**(3), 1696–1707 (2018). https://doi.org/https://doi.org/10.1002/mrm.26806
23. Zöllner, F., Sance, R., Rogelj, P., Ledesma-Carbayo, M., Rørvik, J., Santos, A., Lundervold, A.: Assessment of 3d dce-mri of the kidneys using non-rigid image registration and segmentation of voxel time courses. Computerized Medical Imaging and Graphics **33**, 171–81 (February 2009). https://doi.org/10.1016/j.compmedimag.2008.11.004

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_50