

Received March 9, 2021, accepted April 20, 2021, date of publication May 10, 2021, date of current version May 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3078430

Kidney Segmentation in Renal Magnetic Resonance Imaging - Current Status and Prospects

FRANK G. ZÖLLNER^{ID 1}, (Senior Member, IEEE), MAREK KOCIŃSKI^{ID 2,4},
LAURA HANSEN¹, ALENA-KATHRIN GOLLA^{ID 1}, AMIRA ŠERIFOVIĆ TRBALIĆ^{ID 3},
ARVID LUNDERVOLD², (Life Member, IEEE),
ANDRZEJ MATERKA^{ID 4}, (Life Senior Member, IEEE),
AND PETER ROGELJ^{ID 5}

¹Medical Faculty Mannheim, Mannheim Institute for Intelligent Systems in Medicine, Computer Assisted Clinical Medicine, Heidelberg University, 68167 Mannheim, Germany

²Department of Biomedicine, Mohn Medical Imaging and Visualization Centre, University of Bergen, NO-5020 Bergen, Norway

³Faculty of Electrical Engineering, University of Tuzla, 75000 Tuzla, Bosnia and Herzegovina

⁴Institute of Electronics, Łódź University of Technology, 90-924 Łódź, Poland

⁵Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, 6000 Koper, Slovenia

Corresponding author: Frank G. Zöllner (frank.zoellner@medma.uni-heidelberg.de)

This work was supported in part by the COST Action Magnetic Resonance Imaging Biomarkers for Chronic Kidney Disease (PARENCHIMA) through the European Cooperation in Science and Technology (COST) under Grant CA16103, in part by the Research Campus M2OLIE through by the German Federal Ministry of Education and Research (BMBF) within the Framework Forschungscampus: public-private partnership for Innovations under Grant 13GW0388A, in part by the Slovenian Research Agency (ARRS), Research Programme Metrology and Biometric Systems under Grant P2-0250, and in part by the Trond Mohn Research Foundation to A.L. and M.K., under Grant BFS2018TMT07.

ABSTRACT Magnetic resonance imaging has achieved an increasingly important role in the clinical work-up of renal diseases such chronic kidney disease (CKD). A large panel of parameters have been proposed to diagnose CKD among them total kidney volume (TKV) which recently qualified as biomarker. Volume estimation in renal MRI is based on image segmentation of the kidney and/or its compartments. Beyond volume estimation renal segmentation supports also the quantification of other MR based parameters such as perfusion or filtration. The aim of the present article is to discuss the recent existing literature on renal image segmentation techniques and show today's limitations of the proposed techniques that might hinder clinical translation. We also provide pointers to open source software related to renal image segmentation.

INDEX TERMS Renal MRI, image segmentation, deep learning.

I. INTRODUCTION

Magnetic resonance imaging (MRI) has achieved an increasingly important role in the clinical work-up of renal diseases [1]. Today a panel of parameters can be measured minimal-invasive that can play an important step for the diagnosis and monitoring of renal diseases. This comprises among others assessment of kidney volumes [2], [3], microstructure via diffusion weighted imaging [4], hemodynamic parameters by arterial spin labeling (ASL) [5], or dynamic contrast-enhanced (DCE-) MRI [6], [7] and eventually oxygenation by blood oxygen level dependant (BOLD) [8]. Total

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandra Bertoldo.

kidney volume (TKV) is the most accessed parameter in patients with autosomal dominant polycystic kidney disease (ADPKD). It has been shown that with disease progression the total volume of the kidney increases while kidney function declines [9]. TKV has been recently qualified as a biomarker by the Federal Drug Association (FDA) [10] for use in drug development in ADPKD. It is also the only MRI-based biomarker so far. Figure 1 shows an example of patients with ADPKD at different disease stages and therefore, increased load of cysts. The TKV is depicted by the green lines in the images. The segmentation of the kidney to derive the total kidney volume is delineated in for the left and right kidney respectively. Such segmentation can be either derived by manual annotation which is tedious and operator

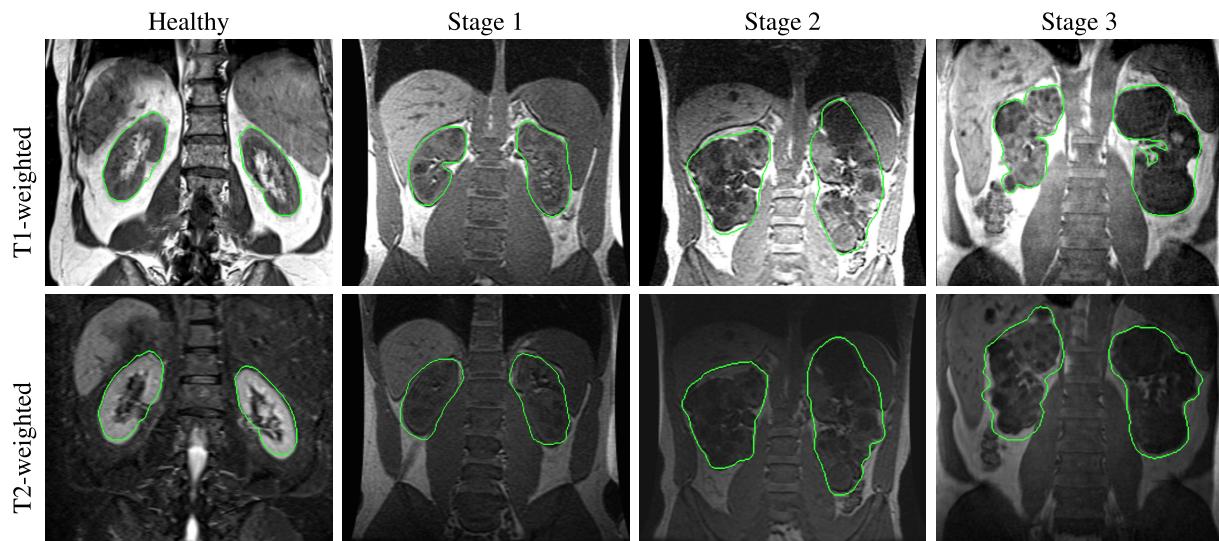


FIGURE 1. Example of MRI of patients with different stages of cyst load. The segmentation of the kidney is in green for the left and right kidney respectively. Repeating such segmentation for all slices of the data set enables the calculation of the TKV.

dependant, or by automated segmentation algorithms as shown in Figure 2.

Image segmentation is therefore an important step in the assessment of TKV but also can be used to derive kidney contours and its compartments i.e. the renal cortex and medulla [12], [13], and renal tumors [14] or cysts [15]–[17]. This drives further the automation of voxel based analysis of functional MRI techniques like perfusion, diffusion or BOLD to help improving diagnosis in kidney diseases like renal hydronephrosis, renal hypoplasia and chronic kidney disease (CKD) related to diabetes, cardiovascular disease, hypertension, and obesity. Furthermore, assessment of graft function in renal transplantation via volumetry and application in renal artery stenosis (RAS) have been reported. Image segmentation has also been reported in the treatment of renal tumors via cryoablation [18].

A challenge in renal image segmentation is similar signal intensities of the renal tissue to adjacent liver, spleen, vertebrae, and parts of the gastrointestinal tract. This problem increases in thin subjects where the fat surrounding the kidney is lacking [19]. Thus signal intensity based approaches like thresholding alone do not provide sufficient information for a robust segmentation. More complex approaches like shape analysis of the kidney is needed. This involves regularization to provide required robustness to noise, low contrast, heterogeneity, and highly variable shape of the organ. Thereby, also multi-modal approaches [15] seem promising incorporating complementary information e.g. from T1-and T2-weighted images.

The Working Group 2 of the COST (European Cooperation in Science and Technology) action PARENCHIMA (Magnetic Resonance Imaging Biomarkers for Chronic Kidney Disease) (<http://www.renalmri.org>) investigates renal data analysis algorithms including image registration [20] and segmentation to provide a core software library for a comprehensive and standardized approach to renal data analysis.

In 2012, Zöllner *et al.* published a review on image segmentation techniques for renal MRI summarizing existing literature at that time [2]. The survey of Torres *et al.* reviews beside MRI also methods applied to computer tomography and ultrasound [3]. The aim of the present article is to discuss the recent existing literature on renal image segmentation techniques and show today's limitations of the proposed techniques that might hinder clinical translation. We also provide pointers to open source software related to renal image segmentation.

II. METHODS

A. SEARCH STRATEGY AND SELECTION CRITERIA

We performed a computer assisted search on PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). We restricted the search to the following selection criteria:

- papers published between January 2013 and December 2020
- studies using magnetic resonance images
- study on human subjects
- describe image segmentation of the kidney
- published in English language

From this the following query was derived:

```
(“kidney”[MeSH Terms] OR “kidney”[All Fields]
OR “kidneys”[All Fields]) AND (“segmentation”
[All Fields] OR “segmentations”[All Fields]
OR “segmented”[All Fields] OR “segment-
ing”[All Fields]) AND (“magnetic resonance
imaging”[MeSH Terms] OR (“magnetic”[All
Fields] AND “resonance”[All Fields] AND
“imaging”[All Fields]) OR “magnetic resonance
imaging”[All Fields] OR “mri”[All Fields])
AND (“human”[All Fields] OR “humans”[MeSH
Terms]) AND (2013/01/01:2020/12/31[Date - Pub-
lication] AND “English”[Language])
```

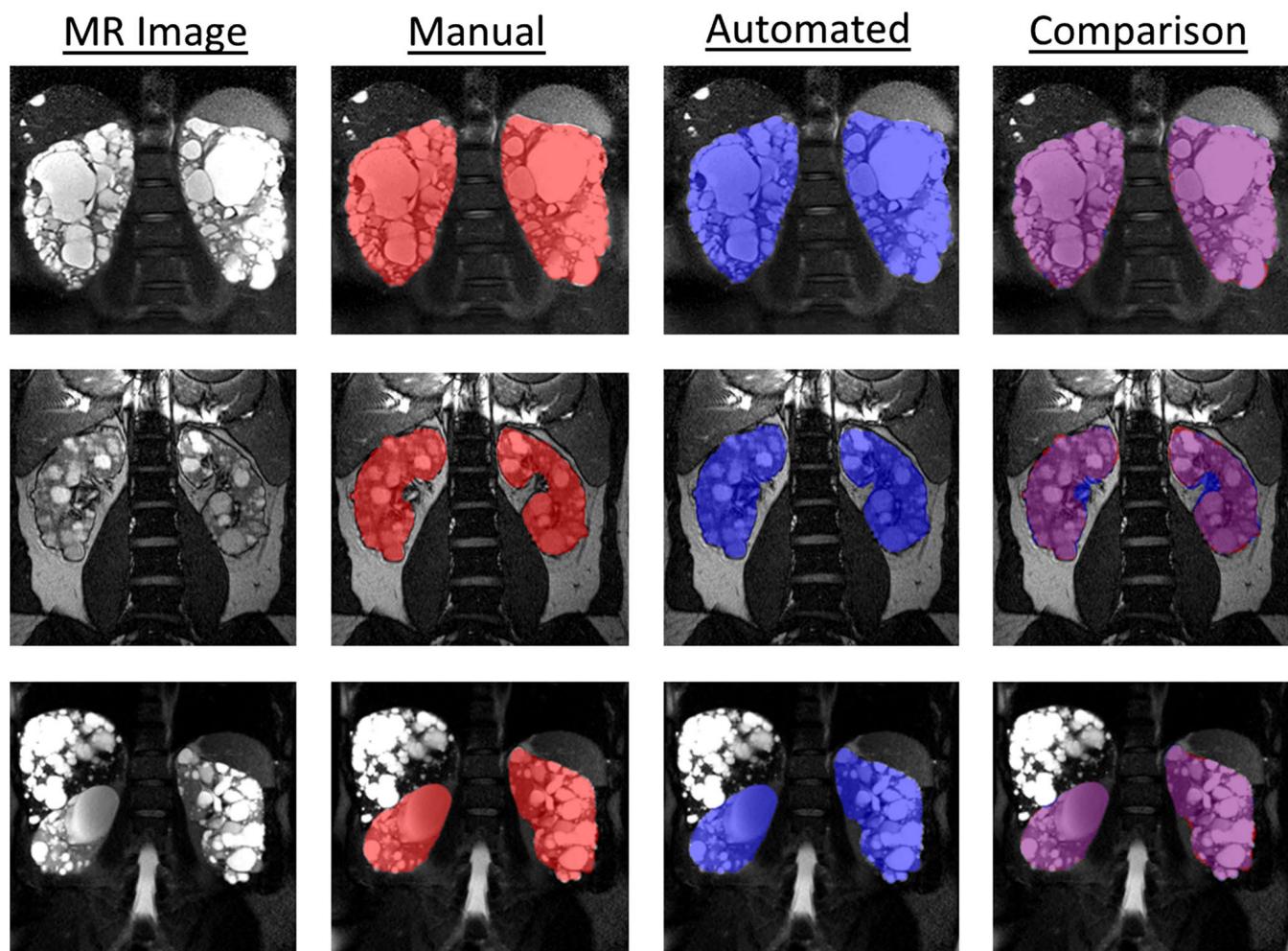


FIGURE 2. Example of TKV segmentation in ADPKD using a deep learning multi observer approach. Reproduced from Kline et al. [11].

The found papers were reviewed to exclude reviews, case reports and papers not focusing the kidney and segmentation such as studies on cardio-renal syndromes, clinical imaging studies and image acquisition developments.

B. SEARCH RESULTS

The PubMed search resulted in 110 papers. Out of these 42 papers passed our manual review of the automated search and were included in this review.

III. IMAGE SEGMENTATION TECHNIQUES

One of the approaches to classification of the segmentation methods stems from needed amount of user participation in the process of ROI delineation. Accordingly, the methods are manual, semi-automated (with limited user intervention) and automated (i. e. fully automated). The obvious trend, especially considering the increasing demand for fast and accurate extraction of medical diagnostic information from 3D and 4D datasets, is to focus on development of automated and semi-automated techniques. The most recent and high performing automated methods are deep learning segmentation approaches. However, their design and training phases

typically require large amounts of labeled data for which manual labeling methods are extensively used.

The amount of automation of the reviewed methods will be indicated in the following. Moreover, to better illustrate the principles, performance, and comparison of selected segmentation techniques experiments were carried out on a set of images downloaded from a free and openly available kidney MRI database.

A. MANUAL SEGMENTATION

Manual segmentation of the kidney is used in different contexts:

- to delineate kidney region boundaries in situations where automated tools are not incorporated in the clinical setting or in the kidney research environment,
- for high quality labeling of imaging data in the training phase of deep learning end-to-end workflow designs, and
- to obtain “ground truth” either from a single operator or by consensus within a panel of operators (typically experienced abdominal radiologists) in order to assess performance of (semi)automated segmentation

algorithms. The inference part of a deep learning approach is then regarded as a fully automated method.

The intrinsically laborious and time-consuming manual techniques are used as a starting point for semi-automated ones [21], [22]. Image editing software applications, for example Analyze in [16], [23]; MRIcron in [24]; Mango in [22], are used for tracing the kidney boundaries. Results of manual segmentation play the role of a reference (ground truth) data to the other methods, e.g. in [16].

Stereology data, produced by labeling the nodes of a sparse regular grid superimposed on the image by a trained operator, are used for renal cysts and parenchyma volumetric measurements [25]. Stereology is not a segmentation method, but it is much faster compared to planimetry [24]. Stereology was used in [23] to generate seed points for semi-automated kidney segmentation. The processing times were 45-90 min., 10-20 min., and 7s, respectively for planimetry, stereology, and automatic mathematical morphology-based segmentation.

Due to the subjective nature of the manual methods, the effects of inter-observer and intra-observer variability have to be quantified. A consensus planimetry tracing by a few experts should be considered as a reference for evaluation of automated techniques performance.

B. IMAGE-PROCESSING-BASED SEGMENTATION METHODS

The image segmentation techniques addressed in this section, loosely classified as “image-processing-based”, originate in a variety of different theoretical approaches to characterization and differentiation of the regions of interest (ROIs). The considered ROIs are the kidney organ as a whole (separated from other abdominal structures through the segmentation), and its internal structures – pelvis, medulla, cortex, and/or cysts. In the reviewed papers, the distinct properties of the regions (used to quantify voxels) are image intensity and, in the cases of DCE-MRI, time courses of it. One assumes that voxels belonging to a region share the selected property and that property takes different values in the other ROIs. Some dedicated, advanced algorithms are designed to detect and quantify glomeruli, based on their shape and convexity of the intensity spatial function [26].

Image-processing-based methods [27]–[31] applied to renal MRI segmentation in the reviewed papers are

- intensity thresholding [19], [22], [32]–[35],
- seeded region growing [16], [21], [33],
- watershed segmentation [23],
- intuitionistic fuzzy sets [35],
- graph cuts [36]–[38],
- mathematical morphology operations [29], [30],
- connected component analysis [16], [33].

Typically, these methods do not require much computation time. Some of them can be implemented on a parallel-processor machine. On the other hand, they usually do not produce accurate segmentation results, even after post-processing. In some cases, they are used as a preliminary

step for further, more accurate segmentation with more complex and computation-time demanding algorithms.

Intensity thresholding is perhaps the most intuitive segmentation technique. In the simplest case, for input gray-level images (e.g. T1-weighted MRI), the output image is a binary one of the same size as the input. If the intensity $\mathbf{I}(i,j,k)$ of a voxel at coordinates (i,j,k) is larger than a specified threshold t , the output image voxel $\mathbf{S}(i,j,k)$ is assigned one of two values, say 1. In the other case, it takes the other value:

$$\mathbf{S}(i,j,k) = \begin{cases} 1 & \text{if } \mathbf{I}(i,j,k) > t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

In thresholding, image intensity is the only property taken into account to distinguish voxels located inside objects from those in the background. No spatial information, such as the region shape or voxel neighbourhood, is considered. A post-processing step of the resulting binary image is then needed to obtain geometric representation of the organs and tissues of interest. Moreover, thresholding result depends on image random noise, always present in MRI and causing rough appearance of otherwise smooth object edges. This might increase segmentation errors, e.g. resulting in inaccurate estimation of kidney geometrical parameters. On top of that, due to image intensity/contrast nonuniformity, e.g. caused by spatially varying sensitivity of the MR scanner coils, single value of threshold t (so-called global threshold) is not appropriate. It should be locally adjusted to reflect spatially varying image properties, e.g. intensity profile around the edges. Optimum threshold selection is a classical topic in image processing; one of the most frequently used algorithms is the one proposed in [39], based on image regions histogram. For two voxel classes (foreground and background), the optimum Otsu's threshold maximizes the inter-class intensity variance.

Example 1: To illustrate the operation and performance of selected methods of image segmentation applied to kidney MRI, T1-weighted (T1w) in-phase and T2-weighted (T2w) volumes stored for healthy subject #1 in CHAOS database [40], were used. The T1w image consists of 35 5.5-mm-thick slices. Its in-plane pixel size is 1.895 mm × 1.895 mm. Applying the ITK-SNAP program [41] at mutual information setting, the T2w image was co-registered and resampled to the T1w volume. In brief, the rigid registration model was selected in the ITK-SNAP context menu. The T2w was set the “moving” image and mutual information served as the similarity metric. The “multiresolution schedule” comprised scales from 8x down to 1x. The moving image was finally resampled to the T1w mesh using linear interpolation. The corresponding expert-marked ground-truth kidney regions in the slices are available in the CHAOS database as PNG files. The right kidney image was selected in this example; its sections are visible in slices #11–#28. Figure 3 shows slices #16 for T1w, co-registered T2w, and the corresponding ground-truth binary objects. A MATLAB® (The MathWorks Inc., Natick, MA, USA) code was developed in-house for the task of analysis and

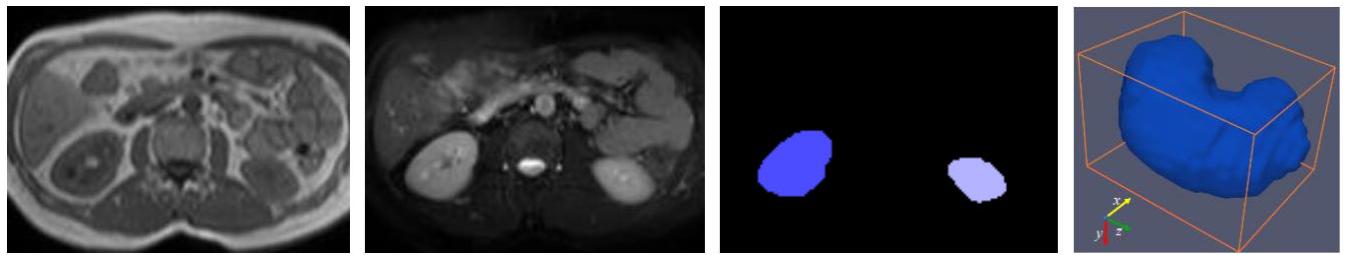


FIGURE 3. Example MRI slices #16 for subject #1 from CHAOS database: T1w in-phase (left); T2w registered to T1w (middle-left); expert annotation – right kidney in dark blue, left kidney in light blue (middle-right). Right: visualization of an expert-annotated right-kidney ground-truth (G-T) 3D region inside a cuboid volume of interest (VOI).

segmentation of the considered T1w and T2w images. It was made freely available in the supplemental files to this paper on GitHub (<https://github.com/MMIV-ML/KidneySegm>) and Code Ocean (<https://codeocean.com/capsule/b2a34704-867e-4545-b069-10699685a26b/>).

Altogether, there are 10,076 voxels representing the right kidney in the MRI of subject #1. The values of mean and standard deviation of those voxels intensity are, respectively, 307.1 ± 61.6 for T1w and 658.4 ± 129.0 for T2w. Although the coefficient of variation for both T1w and T2w inside the organ ROI seemed to be quite small (around 0.2), a closer inspection showed the actual range of kidney T1w intensity values overlapped much more with the values representing the surrounding organs, compared to the corresponding T2w case. One can find in fact, the histogram of T1w images is far from being bimodal, Fig. 4. There is no peak representing the kidney and a very well separated peak for its background, which would lead to a standard threshold placement in a histogram valley. In general, multilevel thresholding methods [42]–[44] better suit extraction of the whole-organ ROI. Among the two kinds of MR images, T2-weighted ones are more specific for the kidney ROI segmentation task (Fig. 4, right column). This is illustrated in Fig. 5 where results of T1w and T2w image thresholding within the ranges limited by respective mean \pm standard deviation values are compared (in the last two rows).

Apparently, it seems difficult to separate the kidney ROI from other objects in the T1w-thresholded image ($I > 245$ and $I < 370$). This is more easy in the case of binary images obtained via T2w thresholding ($I > 530$ and $I < 790$).

One can see in Fig. 6 that retaining the largest *connected components* (CC) in thresholded T2w slices allows better identification of the kidney organ silhouette. (Other conditions could include selection of the most-kidney-shaped CC objects or the ones whose centroids are closest to the likely localization of the organs estimated from anatomical atlas.) Still the shapes of the identified objects are different from the ground-truth ones. They can be made closer to the ground truth via postprocessing, e.g. through holes filling.

An alternative approach to thresholding is the *region-growing* segmentation. This technique needs seed points placed inside the kidney region, e.g. manually or taken from anatomical atlas. In this example, the seed points were computed as 3×3 pixel squares located at centroids of the

regions shown in the lower row of Fig. 6 (marked by red crosses).

To form the initial region for region-growing segmentation in a zero-valued image A, the seed pixels were assigned the value of 1. The mean m of this region in T2w image was then computed. Each 4-connected neighbor of each seed point was then added to the region, on the conditions it has not been the region member already and its intensity value was within the range $(m-d, m+d)$, where d is a hyperparameter of the algorithm. In this example, $d=0.2m$ for slices #11–#26 and $d=0.16m$ for slice #29. Each time a new pixel was added, the region mean value m was updated, until no 4-neighbor had satisfied the region inclusion criteria. The results are shown in the upper row in Fig. 7. The m value update allows the algorithm to adapt to T2w intensity variation within the organ, especially noticeable in the case of slice #16. The large opening visible in the upper part of this kidney section in Figs. 5 and 6 (caused by T2w local average value increasing from its lower left to upper right) is much reduced in Fig. 7.

Post processing of the results of region-growing segmentation, with the use of *mathematical morphology* operations, allows filling the holes in the regions and somewhat smoothing their boundaries, as shown in the middle row of Fig. 7. Visual comparison with the ground truth is illustrated in the bottom row of this figure. The corresponding quantitative metrics are given in Table 1 and discussed in Section IV.

The whole-organ kidney region obtained in result of T2w image segmentation can be used as a mask for further delineation of the kidney internal parts. Figure 8 illustrates an attempt to extraction of the renal cortex via T1w thresholding: the intensity I of all the T1w pixels within the kidney mask being $I <= 260$ and $I >= 450$ was turned down to zero.

Other segmentation techniques, especially those involving machine and deep learning can be successfully used to refine the initial segmentation resulting from application of the basic approaches illustrated in this example.

C. MODEL-BASED IMAGE SEGMENTATION

Imposing some reasonable geometrical constraints on the expected result of image segmentation may help in extracting the target region among other objects, such as those representing the abdominal organs that surround the kidney. The constraints can also improve robustness of the result, e.g. to image noise or to local average intensity variations.

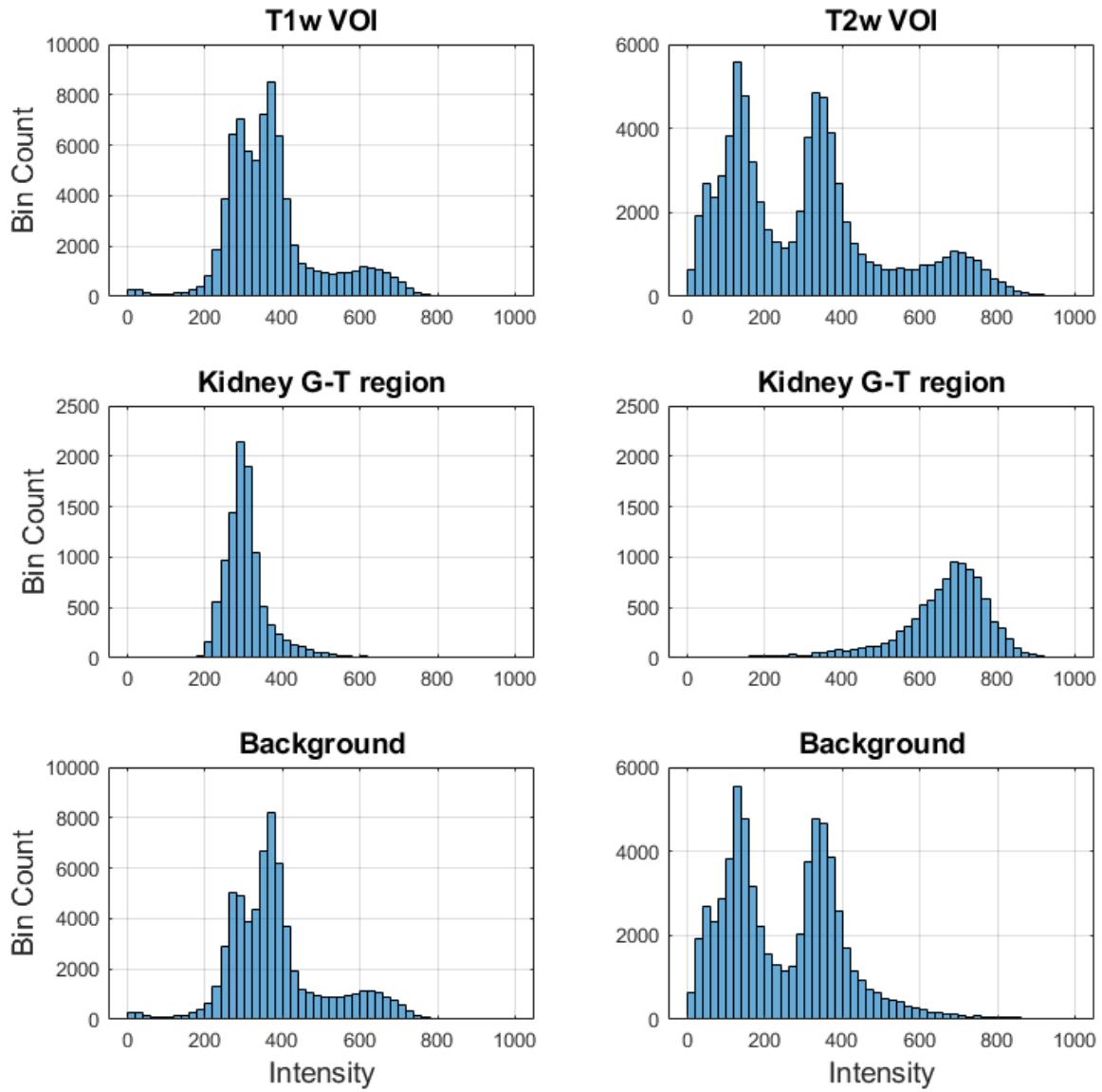


FIGURE 4. Histograms of T1w (left column) and T2w (right column) MR images for subject #1 in CHAOS database. Upper row: a cuboid $52 \times 56 \times 26$ -voxel volume of interest (VOI) comprising the right kidney; middle row: expert-annotated ground-truth (G-T) 3D region for the right kidney; lower row: background voxels located inside VOI and outside the G-T kidney region.

Parametric or non-parametric geometrical models are used in this category of segmentation techniques. Such models can be fitted to the image via their parameters optimization or, respectively, they are obtained as an iterative solution of a differential equation constrained by the image properties and expected kidney shape. The model-based methods used in the reviewed papers comprise

- level set (LS) segmentation [12], [45],
- active contours (snakes) [46],
- use of *a priori* knowledge about kidney shape and its localisation [45].

Example 2: Results of 3D image segmentation with the use of combined model-based, manual, and mathematical morphology techniques are illustrated in Fig. 9. The segmentation was implemented in 3 steps. First, in step 1,

double-threshold-based level-set technique was applied to T2w volume stored for subject #1 in CHAOS database. The ITK-SNAP package [41] was used for this task with thresholds defined by $I > 530$ and $I < 790$. The obtained binary volume was corrected in step 2 via manual removal of the elongated protrusions which represented blood vessels in this case. (Automatic cleaning could be applied as well, as those elongated objects are detectable with the use of scale-space Frangi filtering.) In step 3, the 3D holes present in the segmented object were filled out via morphological post-processing which involved the closing operation using a “spherical” structuring element of radius 3. One can notice in Fig. 9 that the subsequent steps make the segmented object closer to the ground truth. Quantitative similarity metrics emphasize specific shape differences between

| Image\Slice | 11 | 16 | 21 | 26 | 29 |
|--|----|----|----|----|----|
| T1w | | | | | |
| T2w | | | | | |
| Ground truth | | | | | |
| Thresholded T1w ($I > 245$)&($I < 370$) | | | | | |
| Thresholded T2w ($I > 530$)&($I < 790$) | | | | | |

FIGURE 5. Selected slices of subject #1 right kidney images. The bottom two rows show results of thresholding the intensity I of the T1w and T2w sections.

| Image\Slice | 11 | 16 | 21 | 26 | 29 |
|-------------------------------|----|----|----|----|----|
| Largest CC in thresholded T1w | | | | | |
| Largest CC in thresholded T2w | | | | | |

FIGURE 6. The largest connected components (CC) in the thresholded images of Fig. 5.

the ground-truth and the obtained 3D binary regions. Those topics are discussed in Section IV and illustrated quantitatively in Table 1. Apparent local average intensity variations in the T2w kidney image under consideration affected the overall accuracy of LS-based segmentation. These effects can be reduced by using an adaptive version of the level-set algorithm.

D. MACHINE LEARNING AND DEEP LEARNING APPROACHES

A fourth category of methods and different from the “classical” image-processing-based and model-based approaches,

is the application of machine learning (ML) techniques, for which the deep learning sub-class (DL) is generally the most prominent regarding segmentation speed and accuracy. This subsection introduces some terms and definitions related to ML and DL where these are applied to different medical image segmentation domains and tasks. Our aim is to familiarize the novice reader with some basic concepts that are important to ML/DL based kidney image segmentation methods. The inquisitive reader will find multitude of source materials related to machine and deep learning in medical applications e.g. [47]–[55]. Since DL models applied to MR kidney images are scarce, we refer to CT

| Image\Slice | 11 | 16 | 21 | 26 | 29 |
|--------------------------------------|----|----|----|----|----|
| Region growing in T2w | | | | | |
| Region growing in T2w, close-opening | | | | | |
| Ground-truth comparison | | | | | |

FIGURE 7. Upper row: segmentation of T2w images using region growing. Middle row: results of binary morphological close-opening with a 5-pixel octagon structuring element (a cross). Bottom row: comparison with the ground truth.

| Image\Slice | 11 | 16 | 21 | 26 | 29 |
|---|----|----|----|----|----|
| Thresholded T1w ($I > 260 \& I < 450$), within kidney | | | | | |

FIGURE 8. Renal cortex extraction through T1w image thresholding within the whole-kidney mask.

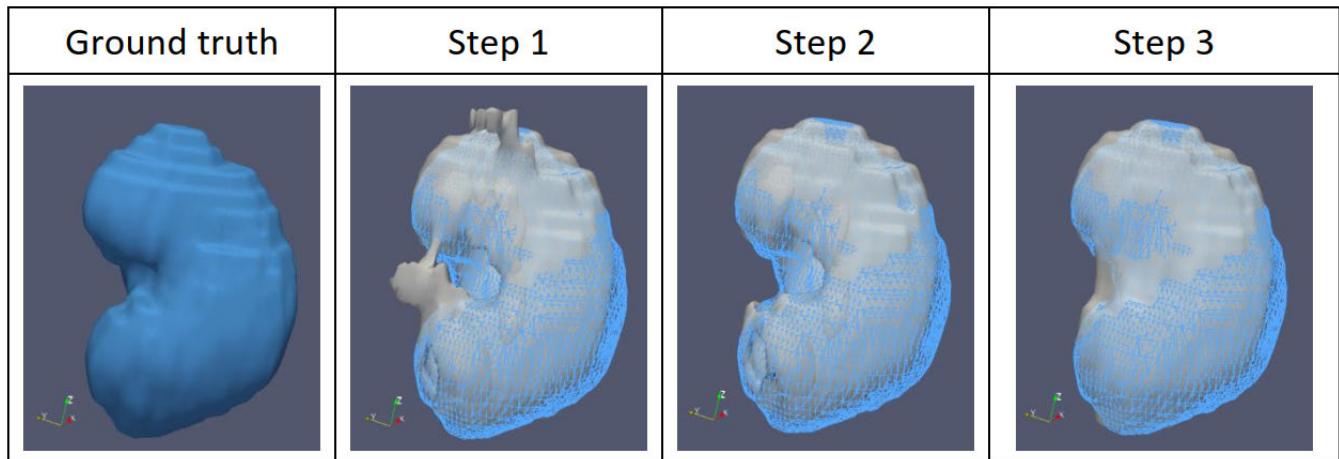


FIGURE 9. Example visualization of 3 segmentation steps of the right-kidney volume in subject #1 T2w MRI from CHAOS database. The blue wireframe model in 3 pictures on the right represents the ground-truth surface shown on the left. Step 1: Level-set segmentation; Step 2: Removing protrusions; Step 3: Applying morphological closing.

kidney image segmentation examples and challenges [56]–[59] as well.

Machine learning has been gaining enormous development for the last years in many aspects of medical image segmentation and medical image analysis in general. In contrast to

traditional programming, in which one must encode a set of rules, ML algorithms build a mathematical model based on provided sample data, known as *training data*. In other words, to make predictions or decisions such models can learn from data without being explicitly programmed. This is

why ML algorithms are often seen as a sub-field of *artificial intelligence* domain. The trained model is able to perform its task, e.g. to make diagnostic prediction in the future on the basis of unseen data, as a regular piece of software.

As it was mentioned in previous sections, it is not an easy task to prepare labels for large training data collections. As a consequence, the vast majority of the available data is unlabeled. That is the reason why ML approaches, among others utilized within biomedical image segmentation, are divided into two main categories: supervised and unsupervised learning. In *supervised learning* an algorithm learns from data, that are described by an expert – called labeled data. Such labels might be e.g. manually segmented MR kidney images. Supervised learning tasks are divided into two main areas:

- *classification*, that predicts a category, e.g. renal tissue or background
- *regression* which predicts continuous values, like quantified kidney volume.

In *unsupervised learning* the training data are unlabeled. It is said that the learning system tries to find patterns and relationships hidden in data without a “teacher”. In kidney image segmentation tasks, these methods are often a part of a collection of more advanced algorithms. Within unsupervised methods, one of the most popular algorithms of objects grouping is *clustering*. This method, when applied to kidney image segmentation, groups voxels that share common feature into clusters. In MR imaging, such a feature might be voxel intensity or texture of its neighbourhood. Combination of supervised learning with unsupervised one, where a labeled training dataset is used together with a much larger set of unlabeled data is also utilized, known as *semisupervised learning*. Semisupervised learning can be of great practical value since it contributes to increased model accuracy at reduced cost, as unlabeled images are much more easily available.

Most of ML methods used for kidney image segmentation are automatic, they do not need any user interaction [38], [45], [60]. However, both supervised and unsupervised methods are often supported by various image pre– or post–processing procedures e.g. manual, image processing–based or model–based (Sections III-A–III-C).

A very engrossing sub-field of ML that has seen dramatic developments, embraces methods known as Deep Learning (DL) approaches. Those methods are inspired by biological systems. The first attempt to build a simplified brain cell, that was named *neuron*, was performed by McCulloch and Pitts in 1943 [61]. Since then, an enormous increase of computational power took place. Now computers are capable to model many such neurons connected together to build an *artificial neural net* (ANN). Neurons arranged into many layers, with multitude of them in each layer, form a *deep neural network*. Each neuron is attributed with special features called *parameters* or *weights* whose values are modified during the learning process – to adjust the strengths of connections between neurons in the network. Thus the

input–output mapping implemented by the neural net is a mathematical function that is extremely flexible depending on its weights. The *universal approximation* theorem states that such a network can approximate any measurable function to any level of accuracy.

The increased capacity of the NN models was accompanied by substantial increase of the amount of available training data and the significant improvements of training algorithms efficiency. These resulted in a boost in prediction accuracy, surpassing that of humans in some applications.

The intended application of a DL model is to make predictions about new, unseen data. To ensure this, the available dataset is usually split into three collections:

- the *training set* seen during learning process,
- the *validation set* used only for evaluation,
- the *testing set* put aside and used for testing the trained model.

To achieve good learning results, many versions of the model are usually explored by changing its *hyperparameters*, like number of hidden layers, number of epochs, activation function type, learning rate, data augmentation strategy, and others. All of these actions aim to ensure the model to *generalize* well, i.e. learns general features from the training set such that the model makes good prediction on new, unseen data. When the model memorizes features from training data, instead of learning a general relation, *overfitting* takes place. To avoid such a situation, a *validation set* is employed, used to evaluate how well the model is doing on hold-out data during training. If the prediction error on the validation set starts increasing while the error over the training set still decreases, the network starts memorising the training data and the learning session should be stopped. However, when the model hyperparameters are repeatedly tuned, it sees both the training and validation sets many times. This might lead to overfitting the validation data also. That is the reason to introduce one more level of reserved data, which is the *test set*.

In the image processing domain, an especially effective model type is the *convolutional neural network* (CNN). This model is able to efficiently process spatial patterns with the use of one or more *convolutional layers*. Those layers consist of *kernels* (small-size spatial filters) that convolve with the image across it. Kernels contain the weights which are adjusted during training. CNNs have many applications in computer vision domain, among them is image segmentation. Thus, CNNs are applied to segment the whole kidney region or its subregions (cortex, medulla, pelvis) from images. Such a task, of linking each pixel/voxel within ROI/VOI to a class label it represents (e.g. cortex, medulla, pelvis), for this pixel, is called *semantic segmentation*.

A very handy technique in training a DL model is to take advantage of having another model which has already been trained on a distinct dataset to perform some tasks. Such a neural network is called a *pretrained* model. Selected weights of the pretrained network are copied to the new-application

model which learns more quickly, with more accuracy, possibly with less data and usually in shorter time. The use of the pretrained model for a task different than that it was originally trained for is known as *transfer learning*. An example of such approach is described in the following paragraph.

Example 3: An interesting approach to 3D kidney segmentation using CNNs, where a common stumbling block for supervised learning methods is the large number of labeled examples required for training, is to use *domain adaptation* and *transfer learning*. To address this approach Lunder-vold *et al.* [62] used segmentation results from automated brain hippocampus labeling to obtain fast semi-supervised segmentation of the kidneys from 4D DCE-MRI recordings. The “neuro → nephro” domain adaptation was based on transfer learning between the bean-shaped hippocampus in the temporal lobes of the brain as imaged in 3D T1-w MRI of the head and the kidney as appearing in the 4D DCE-MRI of the abdomen. After training a network (<https://github.com/deepmedic/deepmedic>) to produce accurate hippocampus segmentations (<https://surfer.nmr.mgh.harvard.edu>), they copied the weights to a CNN designed for segmenting kidneys, freezing the weights of the first few layers in this network during training. By a combination of transfer learning, dropout regularization, residual connections and semi-supervised learning through pseudo-labeling, they were able to train a three-dimensional CNN (Fig. 10) that could rather accurately [in terms of Dice coefficient values in the range 0.8-0.9, Eq. (4)] and very fast (in terms of seconds) segment both the left and right kidney, based on a small number of manually annotated training examples. For their experiments they were using a single standard NVIDIA GeForce 1080Ti GPU for training and executing the CNN model.

IV. EVALUATION OF IMAGE SEGMENTATION RESULTS

As it can be seen from previous sections, the extensive research has been done in creating many different image segmentation algorithms, but still there is no universal and ‘the best’ method yet. For many reasons, it is difficult to compare different segmentation algorithms, or to assess whether one of the algorithms is more suitable for a particular image, or part of the image, or set of images, or more generally, for a whole class of images, and also for a particular purpose. Evaluation of image segmentation is usually performed by a visual inspection, by comparison of the segmentation results with a ground truth or by computing some objective function value from the segmentation results. These evaluation approaches require an expert user interaction either for visual evaluation or to define the ground truth. The ground truth is usually obtained by manual delineation of the kidney or the kidney compartment borders by an expert user. But, manual delineation by an expert user from a large amount of MRI images generated in standard clinical routine, is a difficult and time consuming task and it is subject to intra- and inter-observer variability.

Therefore, to obtain a probabilistic estimate of the ground truth segmentation, some authors [24], [63] have used the simultaneous truth and performance level estimation (STAPLE) method. This method considers a collection of segmentations obtained by expert observers or an automated segmentation algorithm and computes a probabilistic estimate of the true segmentation and a measure of the performance level represented by each segmentation. However, in most cases, there are not enough expert observer segmentations available for using that method.

The most common way to quantitatively evaluate segmentation results is to calculate some metrics based on the overlap with the ground truth or on the spatial distance between corresponding region boundaries.

A. SPATIAL DISTANCE BASED METRICS

The primary spatial distance metric is the Hausdorff distance, which is defined as the maximum distance of a set to the nearest point in the other set. For image segmentation the sets A' and B' are defined as boundary pixels/voxels a and b of the segmented region (A) and the ground truth (B) respectively. The basic variant is a directed Hausdorff distance dHD and measures the maximum Euclidean distance for all boundary points of A' , to the closest boundary point of B' in millimeters [64].

$$dHD(A, B) = \max_{a \in A'} \min_{b \in B'} d(a, b) \quad (2)$$

The limitation here is that narrow excrescences of B do not affect the final result, as shown in Figure 11a. An improvement is the symmetric Hausdorff distance (mHD), defined as a maximum of two directed Hausdorff distances computed in opposite directions [65]. To evaluate an overall agreement of two regions and not only the largest error, an average directed Hausdorff distance ($adHD$) can be used, or more commonly its symmetric variant (aHD) also known as mean Hausdorff distance or average symmetric surface distance (ASSD), which is an average of the two directed average Hausdorff distances [65]–[67].

$$ASSD(A, B) = \frac{\sum_{a \in A'} \min d(a, B') + \sum_{b \in B'} \min d(b, A')}{|A'| + |B'|} \quad (3)$$

Another important distance measure is the 95% percentile Hausdorff distance, which is a value greater or equal to the distance for 95% of the total number of points and, thus, leaves out 5% of the boundary points with the highest distances. Usually, the symmetric variant (95pHD) is used, also called modified Hausdorff distance (MHD), which is again an average of 95% percentile Hausdorff distances in both directions [68]–[74].

Different variants of Hausdorff distance provide different and complementary information of the segmented region boundary agreement and some authors compute more than one variant [65].

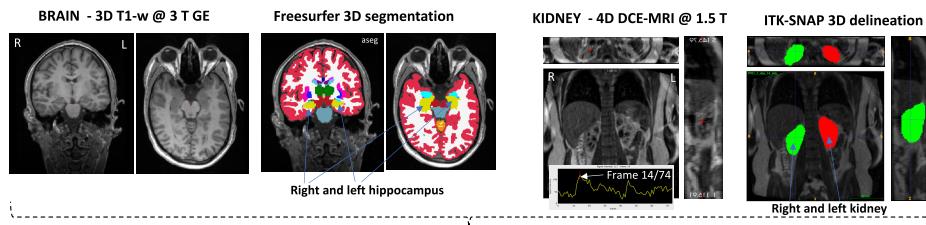
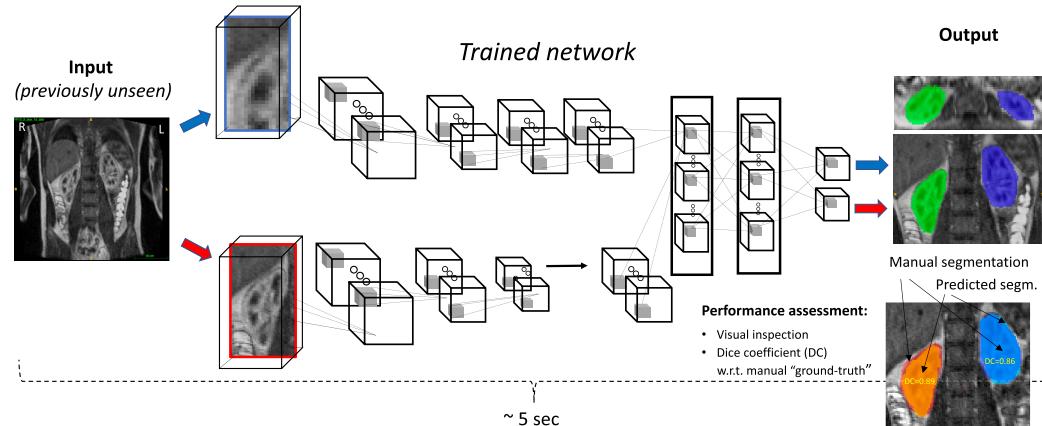
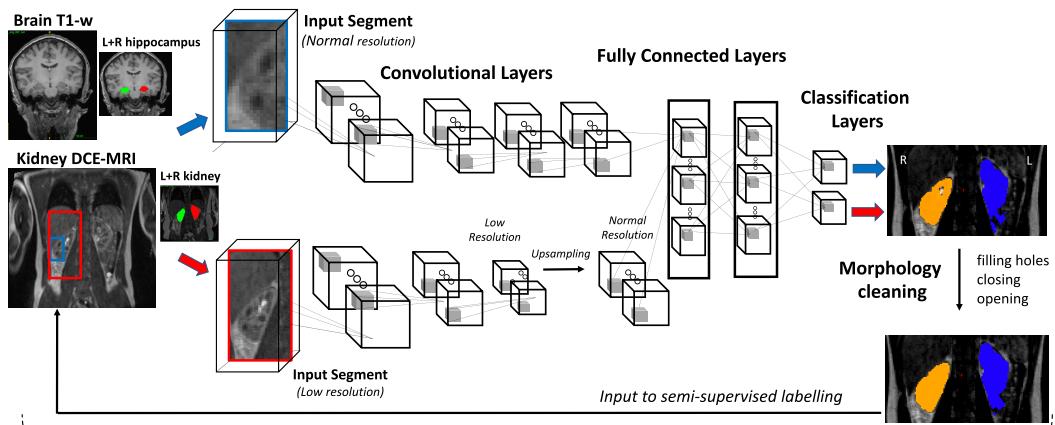
Labelling: $\sim 50\text{ hr}$ ($\sim 45\text{ hr}$ Freesurfer brain + $\sim 5\text{ hr}$ manual labelling kidney)Training:

FIGURE 10. Fast semi-supervised segmentation of the kidneys in 4D DCE-MRI using convolutional neural networks and transfer learning from brain hippocampus segmentation. Adapted from Lundervold et al. [62].

B. OVERLAP BASED METRICS

The most often used evaluation measure in medical image segmentation is a Dice similarity coefficient (*DSC*) also known as *F1* score. It measures the spatial overlap between the segmented region, *A*, and the reference region from the annotated image (the ground truth), *B*, by normalising the size of their intersection over the average of their sizes:

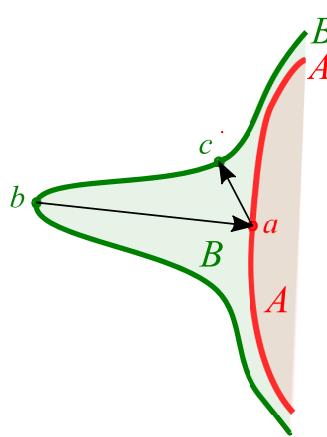
$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (4)$$

DSC values are in the range between 0 and 1, where it has value 0 if there is no overlap between the two segmentations and 1 if both segmentations are identical, see Figure 11b.

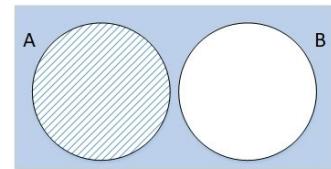
The overlap based metrics can be derived from true positive (TP), true negative (TN), false positive (FP) and false negative (FN) image region. TP represents an object region correctly labelled as the object and TN denotes the non-object region successfully labelled as such. FN is the object region falsely identified as the non-object region, while FP denotes the non-object region mislabelled as the object (Figure 11b). Using the definition of TP, FP, and FN, DSC can be written as

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (5)$$

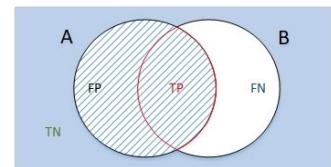
In [11], [17], [23], [32], [36], [45], [66], [68], [75]–[84], the accuracy of the segmentation methods is measured by using



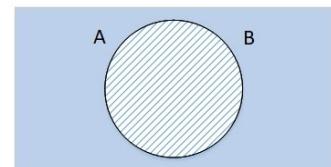
(a) Asymmetry of directed boundary distance measurement for Hausdorff distance (HD). The point a , belonging to A' , is the closest one to the point b which belongs to B' . On the other hand, c is the nearest point on B' to the point a lying on A' . Due to different distances between the points the directed HD is asymmetric and $dHD(A, B) \neq dHD(B, A)$.



No overlap: $DSC = 0$



Partial overlap: $0 < DSC < 1$



Complete overlap: $DSC = 1$

(b) Dice similarity coefficient.

FIGURE 11. Visualizations of evaluation metrics between segmentation A and ground truth B.

the mean and standard deviation of DSC , while in [24], [85] the median values of DSC have been calculated.

A similar measure is a Jaccard similarity coefficient (JSC) [11], [23], [32], [72], [86]–[90] which is defined as the intersection between two regions (segmented and ground truth) divided by their union. When computed in 3D it is also called volume overlap (VO) [24], [32].

$$JSC = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \quad (6)$$

DSC is closely associated with the JSC , and it is easy to convert the scores of one to the other. Therefore, using both of them as evaluation measures does not provide additional information.

$$DSC = \frac{2JSC}{JSC + 1} \quad (7)$$

Additional overlap based metrics have been used in [23], [72], [76], [88], [91], [92]. *Precision* also called the positive predictive value (PPV) indicates the correct portion of the segmented region. *Recall* also known as *sensitivity* or *true positive rate* (TPR) indicates the correctly segmented portion of the ground truth region. *Specificity* also known as *true negative rate* (TNR) represents the image fraction enclosed by the “true” boundary that was missed by the segmentation algorithm. Finally, the overlap error (OE) [32], does not differentiate between the type of error (FP or FN) and considers both.

$$PPV = \frac{|A \cap B|}{|A|} = \frac{TP}{TP + FP} \quad (8)$$

$$TPR = \frac{|A \cap B|}{|B|} = \frac{TP}{TP + FN} \quad (9)$$

$$TNR = \frac{TN}{TN + FP} \quad (10)$$

$$OE = \frac{|A \cup B| - |A \cap B|}{|B|} \quad (11)$$

PPV and TPR are joined in the DSC mentioned earlier:

$$DSC = \frac{2}{\frac{1}{PPV} + \frac{1}{TPR}} \quad (12)$$

C. OTHER EVALUATION CRITERIA

Another evaluation method is comparison of segmented and ground truth region volumes, which can be performed in absolute volume units (ml) or relatively, from the Total Kidney Volume (TKV). Volumetric estimation error (VEE) [84] is an absolute measure, but for more reliable indication of segmentation correctness, the overlap needs to be considered. In addition to previously described overlap methods volumetric measures include the volume error (VE) [32], [45], and volume overlap ratio (VOR) [69].

$$VEE = abs(|A| - |B|) \quad (13)$$

$$VE = abs\left(\frac{|A| - |B|}{|B|}\right) \quad (14)$$

$$VOR = \frac{|A \cap B|}{min(|A|, |B|)} \quad (15)$$

Even though in the literature used for this review, authors haven't used the probabilistic metrics, it is worth to mention

that such metrics are also an option for evaluation of the segmentation results. Examples of such metrics are the intraclass correlation (ICC), probabilistic distance, Cohen's kappa and area under ROC curve [93].

The last in our list but nevertheless the primary evaluation criterion is the visual inspection [36], [80]. It is generally used to quickly assess segmentation correctness with the disadvantage of being subjective, unsound and reliable only for expert observers.

D. SELECTION OF EVALUATION METRICS

By analysing the papers used for this review, one can notice that the authors are presenting their evaluation results using different evaluation metrics, but most of them prefer *DSC* as the main indicator of the segmentation algorithm accuracy.

DSC and other overlap based methods take into account only the correctly or incorrectly segmented image region without taking into account the respective spatial distribution. Thus, a segmentation result with a segmented region extending to a non-object part of the image will be considered similarly good or bad as a result with a segmented region propagating in another non-object part of the image, where the incorrectly segmented regions are in both cases of the same size. One advantage of the overlap based methods is that their values are in the range [0 1], so that makes the obtained results in the literature easily comparable. Specificity (*TPR*) and recall (*TNR*) are sensible to the size of the regions, and they are less suitable for evaluation of the renal MR image segmentation. It can be seen in Table 1 and Fig. 7, the slice 29 have *TPR* value 1, because the overlap region between the segmented region A and the ground truth B is equal to B. But, by visual inspection can be easily concluded that the slice was oversegmented, i.e. the surrounding tissue was included into the kidney region. So, oversegmentation is missed by *TPR* and *TNR*. Also, the visual inspection reveals that the kidney was undersegmented in the slices 11, 16 and 21, i.e. a part of the kidney was missed by the segmentation algorithm. The shape in the slice 26 is the simplest and most accurately segmented. On the other hand, *DSC* and *JSC* confirm the result of visual inspection.

PPV and *TNR* give high score to the undersegmented kidney regions, where *TNR* shows a little variability for all slices and all steps (Table 1). So, the experimental results presented in Table 1, Fig 7 and 9 give preference to Dice and Jaccard evaluation metrics over precision, *TPR* and *TNR*. But, taking into account the relation between *DSC* and *JSC*, only one of them should be used for the segmentation evaluation.

The distance based methods take into account the minimal distances between boundary points in the segmented region and the ground truth only. Among them, Hausdorff metric calculation is computationally expensive and extremely sensitive to outliers and holes within the segmented region, e.g. one can show for the segmented T2w slice #16 in Fig. 7 top row that the directional distances differ significantly:

$dHD(B, A) = 6.8$ mm and $dHD(A, B) = 20.8$ mm, while they take the same value for the slice #16 image shown in Fig. 7 middle row (Table 1). Similar Hausdorff metrics differences can be noticed in that table where large values of $dHD(A, B)$ correspond to external protrusions for Fig. 9 Step 1 and, respectively, internal holes for Fig. 9 Step 2. These metrics reflect the geometric mismatch of segmented boundaries, expressed in easily understandable distance units, usually millimetres.

Finally, in the authors opinion, the selection of the segmentation evaluation measure should depend on the purpose of the segmentation. If the purpose is to perform medical examination, the results are mostly affected by the overlap between the segmented and the correct volume and, thus, overlap based measures with *DSC* being the most accepted one are the most convenient choice. On the other hand, if the purpose includes treatment planning, the localization may become more important than the overlap and distance based measures, with the Haussdorff distance metrics as commonly accepted choice, provide better basis for risk estimation.

V. DATASETS AND DATABASES RELEVANT TO KIDNEY IMAGE SEGMENTATION

While for brain imaging databases of freely available datasets are provided [94] that allow for comparing software for automated segmentation of the brain from MR images [95] in renal MRI, however, such databases are not publicly available and therefore, the working group 2 of the EU COST action PARENCHIMA (<http://www.renalmri.org>) is working towards this. In the US, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) has compiled a large database of T1w and T2w MRI images of patients with ADPKD [96], [97] fueled from the CRISP consortium and its related studies [25], [98]. Besides dedicated renal imaging databases, cohort studies like the UK Biobank or the German National Cohort might be valuable resources to further foster renal image segmentation and its evaluation [99]. However, these resources are not Open Access, e.g. you have to apply for access to the data and for the UK Biobank you have to pay access fees.

VI. SOFTWARE AND GitHub PROJECTS RELEVANT TO KIDNEY IMAGE SEGMENTATION

As described in the aforementioned sections, the right choice of segmentation algorithm is a crucial task. Besides many algorithms and libraries – the building blocks for development of in-house programs, there are numerous freely available, ready to apply programs and GitHub repositories that can be used for initial experimentation at least. An extensive insight into different 2-D and 3-D MRI image segmentation techniques, for researchers who are new in this field or who test new methods with their data, provides [100]. When the number of MRI images is relatively small, ready to use solutions might be useful indeed. There exist several 3D segmentation software tools that enable test procedures or perform

TABLE 1. Exemplary evaluation metrics for segmented individual slices shown in the middle row of Fig. 7 (Slices) and for the segmented image volumes in Fig. 9 (Steps). Arrows indicate whether higher or lower metric values represent a better segmentation. Hausdorff metrics $dHD(B, A)$, $dHD(A, B)$ and ASSD are expressed in millimetres, other quantities are dimensionless.

| | Slices | | | | | Steps | | |
|------------------------|--------|-------|-------|-------|-------|-------|-------|-------|
| | 11 | 16 | 21 | 26 | 29 | 1 | 2 | 3 |
| $DSC \uparrow$ | 0.831 | 0.916 | 0.898 | 0.979 | 0.734 | 0.851 | 0.858 | 0.913 |
| $JSC \uparrow$ | 0.711 | 0.845 | 0.815 | 0.959 | 0.580 | 0.740 | 0.751 | 0.839 |
| $PPV \uparrow$ | 0.958 | 0.987 | 0.984 | 0.977 | 0.580 | 0.921 | 0.940 | 0.927 |
| $TPR \uparrow$ | 0.734 | 0.855 | 0.826 | 0.981 | 1.000 | 0.790 | 0.789 | 0.899 |
| $TNR \uparrow$ | 0.999 | 0.996 | 0.996 | 0.995 | 0.983 | 0.990 | 0.992 | 0.989 |
| $dHD(B, A) \downarrow$ | 5.36 | 6.83 | 8.04 | 3.79 | 5.36 | 6.94 | 8.04 | 9.47 |
| $dHD(A, B) \downarrow$ | 5.36 | 6.83 | 11.05 | 3.79 | 8.47 | 25.05 | 20.64 | 9.36 |
| $ASSD \downarrow$ | 1.96 | 2.14 | 2.43 | 0.51 | 2.54 | 2.74 | 2.53 | 1.50 |
| $VE \downarrow$ | 0.234 | 0.134 | 0.161 | 0.004 | 0.723 | 0.142 | 0.161 | 0.030 |

some part of complex algorithms, e.g. to prepare ground truth images [101], or to design new methods from a set of implemented universal segmentation routines (ITK [102], SimpleITK [103], VTK [104], SciPy [105], scikit-image [106]).

Challenges and open-source projects are continuously getting more and more popular. A few GitHub repositories on kidney segmentation are freely available. Yoruk [38], [107] aimed at evaluation of a fully automated renal segmentation technique for glomerular filtration rate (GFR) assessment in children. Yang et. al proposed a method for renal compartment segmentation in DCE-MRI images implemented in MATLAB [60], [108]. KiTS19 is a challenge of CT kidney tumor segmentation [59]. Its goal was to accelerate the development of reliable kidney and kidney tumor semantic segmentation methods. The organizers produced ground truth semantic segmentations for arterial phase abdominal CT scans of 300 unique kidney cancer patients who underwent partial or radical nephrectomy. MIScnn [109] is another general framework that allows fast building of medical image segmentation pipelines including: data I/O, preprocessing, data augmentation, patch-wise analysis, metrics, a library with state-of-the-art deep learning models and model utilization like training, prediction as well as fully automatic evaluation. It was applied to the KiTS19 segmentation task [110]. A semi-automatic segmentation tool called “rilcontour” [111] has been utilized by the MIROS [24] and the Sheffield TKV Tool [112]. It comprises semi-automated contouring combined with deep learning segmentation techniques.

VII. APPLICATIONS OF RENAL SEGMENTATION TECHNIQUES

In this section we will briefly describe renal segmentation approaches applied to renal diseases obtained by our literature search. The papers are grouped according to manual, image-based, model-based and machine learning methods similar to section III. An overview of the methods and its main characteristics like the used algorithms, number of datasets and type of data as well as the evaluation strategy and respective performance is given in Table 2. In all reported projects, the ground-truth reference was the result of image volumes segmentation by expert radiologists.

A. APPLICATIONS USING MANUAL KIDNEY SEGMENTATION

Compared to studies reported in 2013 [2], only a few papers reported manual delineation [113]–[115] to access TKV though the manual delineation is still considered as gold standard reference [116], [117]. In the study by Shi et al, the ellipsoid method (EL) vs. mid slice (MS) was compared with respect to the prognostic performance of the Mayo Clinic Imaging Classification (MCIC). They found a high intra class correlation (ICC) of TKV measurements in their 308 patient data set ($0.991, p < 0.001$) between EL vs. MS. However, in 5.5% of the cases a disagreement of TKV measurements $> 20\%$ was reported. A similar study was reported by Turco et al. [115]. Furthermore, a high degree of agreement of the individual MCIC risk classes was found but in nearly 14% of the cases EL based TKV measurements lead to a miss classification [113]. In the study of Lai et al. morphological imaging was supplemented with dynamic contrast enhanced perfusion imaging to access normal enhancing renal parenchima and fibrotic tissue in the kidney. The qualitative enhancement maps were used to guide the segmentation of the kidneys.

A comparison of renal cortex volume (RCV) segmented by manual delineation vs the EL method for estimating split renal function in DCE-MRI and CE-CT was investigated by Siedeck et al. [118]. For both imaging techniques the two volume estimation methods had no influence on the split renal function estimation.

Manual kidney segmentation using 3DSlicer tools was employed by Dwivedi et al. aiming at a 3D printed model of a renal cell cancer to allow for targeted tissue procurement and radiomics analysis of renal masses [119]. In this study single time points of a DCE-MRI scans of six patients maximizing the contrast between tumor and renal parenchima were selected for segmentation. The resulting label maps of the segmentations were transferred for to a 3D modelling system to produce STL style files for 3D printing. The mold was then used to slice the resected tumor specimen and to allow for a direct matching to the obtained MRI scans.

A similar work by Wake et al. was presented to create 3D printed renal tumor model that could be used in an augmented reality assisted surgery scenario [14]. An essential step is

the segmentation of the kidney and its tumor to be then transferred into STL files for printing. In this work, image segmentation is realized by manually placing ROIs and use subsequently image processing tools to further refine the segmentation [120]. This approach was evaluated on 20 patients with a single renal tumor using pre- and post-contrast axial T1w fat-saturated MR images. Image segmentation accuracy was not reported.

B. APPLICATIONS USING IMAGE-BASED APPROACHES

Mushrif *et al.* [35] proposed to apply intuitionistic fuzzy sets and rough sets theories for presegmented T2* kidney image analysis prior to its further multithreshold segmentation into medulla and cortex regions. The intuitionistic fuzzy-set image representation includes membership, non-membership and hesitance components for each voxel, to cope with the inherent heterogeneity within the tissue class and the overall intensity inhomogeneity present in the kidney MR images. The rough sets approach allows to map the imprecise histogram-based thresholding into an approximation space, limited by lower approximation (histogram) and upper approximation (histon). The histon is computed on the basis of fuzzy image representation. Finally, a roughness measure is defined combining the two approaches, being a better representation than histogram. Peaks and valleys of the roughness-intensity function are used to threshold the image intensity for medulla and pelvis segmentation. The study shows the potential of the method, whose performance needs to be further evaluated on larger datasets.

Warner at al. [23] studied a possibility of obtaining fully segmented polycystic kidney region from undersampled stereology grid data given as the *a priori* information. The planimetry (reference) and stereology images were prepared independently by two experts. The algorithm starts with morphological closing of the labeled stereology 2D image. The kidney region border together with the original 3D volume are passed to minimal-spanning-tree watershed algorithm. Due to noise, the resulting rough edges of left and right kidney are smoothed via thresholding of the image obtained after fuzzy-set processing with a spherical Gaussian membership function. The method is faster than manual planimetry segmentation and offers high accuracy.

The problem of renal region segmentation in very low contrast images is considered in [37]. The postulated method is a single-step process based on improved graph-cuts algorithm. The improvement stems from pixel/voxel connectivity component incorporated in graph-cuts formulation based on one of their three predefined different groups. In result, no shape prior and/or user intervention are needed. A high segmentation accuracy in terms of Dice coefficient is achieved as seen in Table 2.

Sandmair *et al.* [22] proposed and evaluated a simple semi-automatic technique of unimodal thresholding which requires approximate manual delineation which can be done by a non-expert with some training, using e. g. the Photoshop application. The Multi-image Analysis GUI Mango was used

in the study involving 24 T2-w images. Unimodal thresholding was performed within the delineated region, on the basis of Gaussian kernel probability density estimation. The maximum (or rarely maxima) of the density function are found and applied to geometric construction which defines the two threshold values – below and above the peak. A visual feedback is provided to the user for the thresholding result assessment.

In [21], [121], an attempt is made to develop a semi-automatic technique of polycystic kidney segmentation for renal volume estimation from T2-w images. Both coronal and axial views were considered and compared. An operator is expected to mark a single point within renal parenchyma region in the central slice of the volume. This is a starting point for region growing with similar or higher intensity level. The result of this segmentation step is refined using morphological operators with subsequent contour refinement using curvature motion in areas of negative curvature.

Woodard *et al.* [34] designed an automated method for kidney volume segmentation in DCE-MRI. Raw images were preprocessed to reduce signal inhomogeneity. Each kidney 3D image region was then registered to pre-contrast reference image. The unenhanced kidney image was thresholded to extract the organ contour, with some manual correction where needed. Templates for synchronized intensity time course were then identified for cortex, medulla and the collecting system using 29 out of 493 volumes. Signal intensity was normalized by linear rescaling. Each voxel of the kidney region was classified by comparison (sum of squared errors and correlation coefficient) with the collected templates. The estimated kidney parts volumes were well correlated with independently evaluated values of eGFR, urine-creatinine ratio (ACR) and risk factors for and complications of CKD. Based on that extensive study, it was concluded that DCE-MRI approach to automatic segmentation of kidney images can provide novel and unique information about the organ structure and function.

Yoruk *et al.* [38] developed an automatic technique for kidney and its parts segmentation in DCE-MRI. They applied the graph-cuts GrabCut algorithm implemented in OpenCV framework, originally designed for color image segmentation. Based on a uniquely-distinctive time-course appearance of the medulla voxels enhancement signal, a medulla-score was defined as a function of the signal values at selected moments of time. Then, the medulla-score map was thresholded with the Otsu's method to reveal the medulla clusters, further morphologically dilated to obtain a bounding box containing the whole kidney and some background. To achieve input data compatibility with GrabCut, the authors applied principal component analysis (PCA) to voxel enhancement signals inside the bounding box and mapped the 3 largest PCA coefficients of the signal expansion to RGB input channels of GrabCut. The GrabCut algorithm was applied to the renal bounding boxes to obtain the segmentation map of the renal parenchyma. After that, a random forest classifier

TABLE 2. Summary of applications of automated and semi-automated image processing-based methods to kidney MRI segmentation.

| Literature reference | Methods | Type of MRI | No. of images | Aim of segmentation | Linear regression | Dice | Volume error | Software/comput. time |
|--------------------------|--|--------------|---------------|--|--|---|--|---|
| Bae 2013 [16] | Manual segmentation, region growing, morphologic watershed, level-set | pre-T2-w | 20 | Renal cysts number and volume estimation | Number: 0.98, $R^2 = 0.98$, $P < 0.001$ Volume: 0.98, $R^2 = 0.98$, $P < 0.001$ | – | – | Analyze, ITK, C++ |
| Hanson 2013 [36] | Mumford-Shah model with local and non-local regularization, graph cuts | DCE | 3 | Kidney segmentation | – | – | – | MATLAB, 15 min |
| Khalifa 2013 [79], [122] | 2D Markov-Gibbs random field with level-set, three feature speed function | DCE | 50 subjects | whole kidney, cortex | – | Kidney: 0.97 ± 0.02 , Cortex: 0.90 ± 0.03 | – | – |
| Rudra 2013 [37] | Graph cuts, pixel connectivity | T1-w | 25 | Kidney segmentation from very low contrast images | – | 0.98 ± 0.005 | – | – |
| Bae 2014 [127] | iterative thresholding | T2-w | 241 | cyst segmentation and counting | average difference in the total numbers of cysts: -0.52 | – | – | ImageJ plugin |
| Chav 2014 [78] | renal shape model | T1-w | 10 | whole kidney segmentation | – | 1.9 ± 1.4 | – | MATLAB |
| Chiusano 2014 [125] | dictionary representation, clustering | DCE-MRI | 26 | whole kidney segmentation | – | 0.83 ± 0.09 | – | – |
| Hodneland 2014 [136] | normalized gradients as data term for registration; Mahalanobis distance between signal time courses to a training set for supervised segmentation | DCE-MRI | 10 | simultaneous registration and segmentation of the whole kidney | – | 0.719 | 36.4 ml | – |
| Warner 2014 [23] | Planimetry, stereology, morphologic closing, watershed, fuzzy-logic smoothing | T2-w | 12 | Polycystic kidney | – | 0.969 ± 0.007 | – | Analyze/45-90 min. 10-20 min. Python/7s |
| Will 2014 [32], [129] | Adapted thresholding, shape analysis | T1-w T2-w | 12 | Whole organ, cortex, medulla, pelvis | Kidney: $R^2 = 0.98$, $R^2 = 0.88$, $R^2 = 0.92$, $R^2 = 0.87$ Cortex: 0.001 , 0.003 , 0.003 , 0.004 Medulla: 0.808 , 0.726 , 0.726 , 0.719 ± 0.007 Pelvis: $7.03 \pm 5.56\%$ Pelvis: 0.004 , 0.004 , 0.004 , 0.004 | Kidney: $0.936 \pm 4.97 \pm 4.08\%$, $0.001 \pm 7.03 \pm 5.56\%$ Cortex: $0.001 \pm 7.03 \pm 5.56\%$ Medulla: $0.808 \pm 7.03 \pm 5.56\%$ Pelvis: $0.726 \pm 7.03 \pm 5.56\%$ | Kidney: $\pm 4.97 \pm 4.08\%$, 50 s for the whole MRI dataset | MATLAB/ whole MRI \pm dataset |

Continued on next page

TABLE 2. (Continued.) Summary of applications of automated and semi-automated image processing-based methods to kidney MRI segmentation.

| Lit. ref. | Methods | MRI | # img. | Aim | Lin. reg. | DSC | VE | SW/CT |
|---------------------------------|---|--|---------|--|-----------|---|-------------|---|
| Abdulahi [33] | Thresholding, region growing, fast level-set, connected component shape analysis | T1-w | 10 | Whole organ | – | – | – | MATLAB/ <u>Threshold:</u> ~4 s, <u>Level-set:</u> ~40 s |
| Gloer [12], [45] | prior shape masks, subject-specific & probability maps T1-w and level set DIXON segmentation (in-phase, out-phase, fat/water image) | T2-w | 35 | renal tissue volume-try (cortex, medulla) | – | 0.92 ± 0.022 (cortex: 0.821 ± 0.037, medulla: 0.749 ± 0.058 | 6.1 ± 5.5% | – |
| Goceri [137] | multi-layer perceptron | Spectral-Pre-saturation Inversion Recovery | – | whole kidney segmentation | – | – | – | – |
| Liu 2015 [138] | 1st-order shape and 5th-order Appearance Priors | DCE-MRI | 50 | whole kidney segmentation | – | 0.91±0.03 (1st order), 0.99±0.02 (5th order) | – | MATLAB and C++/ 125± 10 sec/ dataset (79 images) |
| Liu 2015 [18] | graph cut with adapted shape priors | T2-w | 13 | Iceball Segmentation – for renal tumor MRI-Guided Cryoablation | – | 0.92 ± 0.03 | – | C++ with a MATLAB interface, freely-available graph cut library (http://cbia.fi.muni.cz/cut-projects/graph-library.html / 20 s) |
| Shehata [68], [70], [92], [124] | Level sets | DWI | 65 | kidney segmentation | – | 0.95 ± 0.01 | 9.5 ± 2.2 % | – |
| Turco 2015 [21], [121] | Manual region growing, morphologic operations, curvature motion | seed, T2-w | 30 | Polycystic kidney – volume estimation | – | Left: 5.1 ± 4.0%, Right: 4.0 ± 4.7% | <2 min. | – |
| Woodart [34] | Thresholding, manual correction, voxel classification based on time-course templates | DCE | 493, 29 | Volume: kidney, cortex, medulla, perc. of fat, perc. of fibrosis | – | – | – | Image preparation: 15 min. |
| Yang 2015 [13], [60] | Maximally Stable Volume, PCA, clustering | DCE | 16 | kidney, cortex, medulla | – | 0.86-0.98 | – | – |

Continued on next page

TABLE 2. (Continued.) Summary of applications of automated and semi-automated image processing-based methods to kidney MRI segmentation.

| Lit. ref. | Methods | MRI | # img. | Aim | Lin. reg. | DSC | VE | SW/CT |
|----------------------|---|----------------------------------|------------|---|---|--|---|---|
| Feng 2016 [132] | fuzzy k-means (FKM), k-harmonic means (KHM), reaction-diffusion evolution (RD-LSE) ,VR-TPDS | k-means image acquisition (FISP) | dynamic 40 | 3 images renal tumor patients | - | 0.74-0.97 | - | MATLAB scripts |
| Kim 2016 [17] | spatial prior ability map, propagated shape constraint, level sets | T2-w | 60 | total kidney volume – in ADPKD | - | 0.886 ± 0.08 | 885.46 569.7 ml | ± manual reference segmentation: Analyze 12.0 |
| Kline 2016 [24] | Planimetry, stereology, user-control active contour | T2-w | 40 | Polycystic kidney volume | - | 1.41 ± 4.78% | <u>MRIcron:</u> 45-90 min. <u>Analyze:</u> 30-60 min. <u>MIROS:</u> <5 min. | |
| Kline 2016 [87] | semi-automatic segmentation as in [24] on first point, two-way image registration of time points, refinement via geodesic active contours | T2-w | 40 | change detection in longitudinal studies of ADPKD | - | level accuracy: 0.99 0.79%; intraobserver variability: 0.77 ± 0.46%; interobserver variability: 1.34 ± 0.70% | of Python, Advanced Normalization Toolkit | |
| Rusinek 2016 [19] | Manual contouring, "blanketing", fat non-uniformity correction, unimodal thresholding, 3D morphology, active contours | T1-w, | 40 | Healthy and diabetes kidney volume estimation | Strongly linear, $R^2 = 0.98$ | - | User1: 6.8 ± 5.2%, User2: 1.5 min. per image, User3: 4.8 ± 4.7% RSS | |
| Sandmair 2016 [22] | Manual segmentation, unimodal thresholding | T2-w | 24 | Healthy kidney volume estimation | 0.98, $p < 2.2e - 16$ | - | - | Photoshop, Mango, Python |
| Khalifa 2017 [123] | extension of [79], [122] to 3D | | | | | | | |
| Kline 2017 [11] | 11 individual trained and parameterised T2-w deep learning segmentation networks, majority voting | T1-w/ T2-w | 2000 | total kidney volume – | - | 0.97 ± 0.01 | 0.68 ± 2.2% | Python, Keras, Theano |
| Lundervold 2017 [62] | CNN, transfer learning from brain segmentation | DCE-MRI | | whole kidney segmentation, compartments | - | 0.8-0.9 | - | 5s |
| Mushrif 2017 [35] | K-means clustering (KMC), intuitionistic fuzzy sets (IFS) representation, rough sets, multilevel thresholding | T2*-w | 1 | Medulla, cortex, – blood vessels, pelvis | Medulla KMC: 0.947, IFS: 0.956 Cortex KMC: 0.923, IFS: 0.958 | - | - | |

Continued on next page

TABLE 2. (Continued.) Summary of applications of automated and semi-automated image processing-based methods to kidney MRI segmentation.

| Lit. ref. | Methods | MRI | # img. | Aim | Lin. reg. | DSC | VE | SW/CT |
|-----------------------------|--|------------|---|--|--|-------------|--|--|
| Suess [130], [131] 2017 | thresholding, connected component analysis (semiautomatic method); ellipsoid formula | T2-w | 20 | whole kidney | – | – | semi automatic method: 0.3 ml, ellipsoid formular: 85.0 ml | Osirix plugin/ semi automatic: 220± 53 s, ellipsoid: 41 ± 11 s , manual: 408 ± 105 s |
| Barstugan [86] 2018 | K-means clustering to segment the abdominal region, active contour segmented the fat layer, adaptive thresholding, and image erosion for liver segmentation, region growing used to segment the adrenal tumor | T1-w/ T2-w | 113 | adrenal tumor | – | 0.50-0.82 | – | -/107.9 s |
| Bevilacqua [88], [135] 2018 | Cascading CNNs, bounding box detection followed by segmentation | T1-w/ T2-w | 57 scans from 4 patients, follow-up: 526 scans from 18 subjects | whole kidney – | – | 0.97 ± 0.01 | 0.68 ± 2.2% | Python, Keras, Theano |
| Haghghi [84] 2018 | CNNs for localisation and segmentation | DCE-MRI | total kidney volume | 30 | – | – | 12.9±2.4ml (healthy), 19.5±3.2ml (renal disease) | 3s |
| Lv 2018 [75] | modified Pyramidal Lucas-kanade registration | U-net, DWI | kidney segmentation to improve the registration accuracy | 10 | NRMS of DWI model after registration: 0.25± 0.06 (cortex), 0.22 ± 0.05 (medulla) | – | – | Python, Keras |
| Yoruk 2018 [38] | Time-resolved graph cuts for whole kidney, random forests for its parts | DCE | 26 | Whole organ, cortex, medulla, collecting system | Whole: 0.93 ± 0.012%, Cortex: 0.86 ± 0.006% | – | – | OpenCV, Python, C++/ 45 s per subject |
| Huang [139] 2019 | image enhancement and adaptive MRI thresholding, and labels of seed regions are assigned on a pixel-by-pixel basis; random walker model to segment renal compartments by using time-intensity signals and previous seed regions for training | DCE- MRI | 14 | fully automatic renal segmentation without using manually labeled data | – | 0.92 | – | – |

Continued on next page

TABLE 2. (Continued.) Summary of applications of automated and semi-automated image processing-based methods to kidney MRI segmentation.

| Lit. ref. | Methods | MRI | # img. | Aim | Lin. reg. | DSC | VE | SW/CT |
|----------------|---|-----------------------|----------------------------|--|---|------------------------------|-------------------|---|
| O'Reilly [46] | Decision classifier (kidney/non-kidney), contrast enhancement, k-means clustering, active contour | tree T1-w/ T2-w | 13 | Total polycystic kidney volume | – | – | 21.7% | Python, OpenCV, Scikit-Image, R Studio |
| Schnurr [140] | U-Net, additional skip connections, batch normalization | T1-w | – | sample mining – schemes and there influence on CNN kidney segmentation | <u>random</u> $0.885 \pm 0.046/$ 0.877 ± 0.053 $\mu \pm 3\text{Std}$ $0.879 \pm 0.053/$ 0.863 ± 0.076 <u>label</u> $0.835 \pm 0.115/$ 0.836 ± 0.130 | – | – | Python, TensorFlow 2.0 |
| Simms [112] | hybrid method | level set | T2-w | 65 | estimation of TKV – in ADPKD | – | $3.45 \pm 3.96\%$ | MATLAB 2016b (MathWorks), < 6 min |
| O'Reilly [134] | 3D FCN based on 3D U-Net and V-Net | T2-w | 135 scans from 55 subjects | whole kidney – | 0.787 0.060 | ± 187.5 ml 36.71 % | / | Keras, Tensorflow, Scikit-Learn, OpenCV, Python |

was trained on 10 000 renal voxels taken from manually selected ground truth images and used for voxel classification to obtain cortex, medulla and the collecting system clusters. Each voxel was represented by 7 scalar features – signal intensity at 6 moments of time from the bolus arrival and the depth of voxel in the renal tissue.

C. APPLICATIONS USING MODEL-BASED APPROACHES

In Kim *et al.* [17], kidney segmentation is reached by exploiting prior knowledge of spatial location of kidneys modeled as a spatial prior probability map (SPPM) and a propagated shape constraint (PSC). These were incorporated into a level set framework.

The ground-truth reference in [24] was constructed with the use of two planimetry segmentations and stereology-based [23] algorithm. The minimal interaction rapid organ segmentation algorithm MIROS was developed and applied to segmentation of 40 T2-w MR images. They were corrected for uniform intensity and co-registered prior to further processing. The MIROS interactive package requires the user to supply approximate boundaries of the kidney regions of interest in some slices. A mathematical-morphology modified geodesic active contour (GAC) is then used to approximate the original partial-differential equation to refine the contours. The narrow bands at the kidney region border, as found by GAC, is used for watershed edge detection on the Sobel-filtered image. This completes the semiautomatic MIROS procedure which typically requires less than 5 min of user interaction.

To detect changes in kidney volume over time e.g. to monitor the effect of a drug therapy, Kline *et al.* [87] used a two-way image registration in a longitudinal study. An initial semi-automatically segmentation as described before [24] is registered onto the follow-up data set. Then, geodesic active contours are used to refine the segmentation.

To evaluate renal graft rejection Khalifa *et al.* [79], [122] proposed an automated pipeline to analyse 2D DCE-MRI scans comprising kidney segmentation. Thereby, kidney segmentation is used to a) support contour based image registration of the dynamic scans to reduce motion in the signal intensity time curves and b) to derive the renal cortex and to extract the signal intensity time curves for pharmacodynamic modelling, i.e. to classify the functioning of the graft. To segment the kidney a joint Markov–Gibbs random field (MGRF) model is employed comprising a conventional level-set-based deformable model. The evolution of the contour by the level set is controlled by a stochastic speed function that accounts for three image features, namely first-order pixel-wise image intensities estimated using a linear combinations of discrete Gaussians (LCDG) intensity estimation model, a second-order pair-wise Potts–Markov–Gibbs random field (MGRF) spatial interaction model, and a weighted probabilistic shape prior. Their system has then been extended to evaluate also 3D DCE-MRI [123] and to incorporate also clinical information in the classification step [92].

Simms *et al.* [112] presented a tool for semiautomatic segmentation similar to the MIROS tool. In their study, the data was initially reprocessed removing image artefacts

such as intensity variations and motion and applying a level set segmentation which is initialized on the mid slice of the volume.

Shehata *et al.* proposed a level set approach [68] and evaluate its use in assessment of renal graft rejection from diffusion weighted imaging [70], [92], [124]. The employed segmentation strategy is similar to that reported by Khalifa *et al.* for DCE-MRI [123]. Here, the different b-values were treated as time domain compared to the DCE-MRI approach. A high performance of the approach was confirmed by the high DSC (0.95 ± 0.01), low 95-percentile modified Hausdorff distance (3.9 ± 0.76) mm, and low percentage kidney volume difference ($9.5 \pm 2.2\%$) relative to a manual segmentation by an MR expert on 65 DWI image data sets. In the study of Hollis *et al.* the kidney segmentations are used to statistically evaluate if the ADC which is calculated based on the segmentations, is a discriminator in graft rejection detection.

An approach utilizing renal image segmentation in renal tumor therapy, namely in MR guided cryo ablation was presented by Liu *et al.* [18]. The aim of the study was to segment the ice ball, i.e. the cryo ablated volume from time resolved T2W images to allow to monitor the ablation process during the intervention. For image segmentation a graph cut algorithm using a shape prior of the ice ball was used. The algorithm was tested retrospectively on data from 13 ablations of renal tumors. Image segmentation accuracy assessed by the Dice coefficient was 0.92 ± 0.03 . The computation time per time point was 20 s matching the cryo ablation procedure of 3- 5 minutes intervals.

To segment the kidney from DCE-MRI Chiusano *et al* proposed a dictionary learning approach [125]. Thereby, the signal time intensity curve of the whole data set is used to derive characteristic signal curves reflecting the kidney tissue forming the dictionary. Clustering similar to [126] is then used to discriminate between kidney and background. The technique was evaluated in a pediatric population with not specified renal dysfunction. In total 26 data sets were used and the segmentation accuracy (Dice) was 0.83 ± 0.09 for both kidneys.

D. APPLICATIONS USING IMAGE- AND MODEL-BASED TECHNIQUES

Some literature reviewed combine image- and model-based techniques to segment the kidney which do not match the two above subsections. Therefore, they are discussed in the following.

In [16], the kidney boundary was detected using the Analyze image editing software, and the cysts were marked and counted by two radiologists. The image was thresholded to estimate the area of the cysts in slices, and their volume by summing up the products of the areas and slice thickness. In the automatic approach, the image was first thresholded with ITK software, and voxels brighter than the background were grouped together using the k-means clustering algorithm, to form cysts candidate regions. The subsequent connected component analysis produced labels for individual

clusters. Then fast-marching level-set algorithm was applied to each cyst region, augmented by edges potential obtained from the initial segmentation via multiple filters. The outcome of shape-detection level-set processing was eroded in an iterative routine to identify the seed points for all cysts. After that, the morphological watershed algorithm produced the final segmentation. Another segmentation approach of the same group of authors [127] was presented in 2014 using an iterative thresholding method, first proposed by [128], in which the operator selects step by step thresholds to optimize the segmentation of the object, here the renal cyst. As post processing step, a flood fill operation was performed to close holes in the segmented images. The approach was evaluated on 240 patients data sets from the CRISP study [98].

Hanson and Lundervold [36] presented an extensive study on kidney DCE-MRI segmentation combining k-NN classification of pixels as belonging to non-overlapping foreground and background and Mumford-Shah piecewise constant image segmentation, regularized with both local (boundary size) and non-local (similarity in feature space) image information. Voxel intensity time courses (20 to 45 time frames) were included as multidimensional image features. The corresponding Mumford-Shah energy function in its discrete form was minimized using a modified graph-cuts algorithm. Flexibility and robustness of the proposed method applied to kidney DCE-MRI of 3 subjects is illustrated qualitatively in [36], showing the importance of non-local regularization and the superiority of Mahalanobis over Euclidean distance as a metric of feature separation.

O'Reilly *et al.* [46] used a decision tree classification and snake algorithm for TKV segmentation in polycystic kidney disease. In this two step approach first a decision tree was trained to roughly detect the kidneys while in a second step an active contour algorithm was used to segment the kidney outline.

Will *et al.* [32] aimed at automatic volumetry of the entire kidney and its internal parts – cortex, medulla and pelvis – from non-contrast-enhanced (low contrast) T1-w and T2-w MR images. The breath-hold acquired MRIs were co-registered with a rigid registration algorithm. Manual segmentation was performed for use as a ground-truth reference. The automated algorithms started with thresholding T2-w images, followed by some refinement using prior knowledge about the kidney shape and location. In the second step, both T1-w and T2-w images were again thresholded to distinguish between the internal structures (cortex, medulla and pelvis). The threshold values were determined and tested experimentally in several subjects and then held constant for all the data set. The time of segmentation was less than 5 s per subject. The accuracy of whole kidney segmentation is better than the accuracy of its partitioning into the three components (Table 2). The algorithm of Will *et al.* was further used in a study of quantifying renal and renal sinus fat volumes from DIXON MR scans [129]. Here, a total of 400 patients from cross sectional population study were included. Renal

and sinus and fat value was correlated to diabetes and prediabetes. Renal volumes estimated from the segmentation were 280.3 ± 64.7 ml for the normoglycemic individuals, 303.7 ± 67.4 ml for prediabetes and 320.6 ± 77.7 ml for diabetes, respectively.

Two automatic kidney segmentation techniques – Otsu thresholding and fast Chan-Vese level-set algorithm – are compared in [33] in terms of accuracy and speed. Then connected component analysis is performed followed by shape analysis of candidate binary regions with the use of compactness coefficient and Procrustes algorithm. It was found that the two algorithms feature similar performance while Otsu's thresholding is much faster than the level-set solution.

Forty non-contrast T1-weighted fat-suppressed MR images were used in [19] to evaluate the performance of their semi-automated “blanket” method for renal segmentation. The user is required to draw approximate contours of the kidney region on every tenth slice, to separate the organ from adjacent abdominal ones. Then the contour is filled with morphological operations, and the masks interpolated in the Z direction. Non-uniformity correction is performed within the bounding box of the kidney region. The program then finds a small seed volume for the kidney. Two values of intensity threshold are found based on the seed region intensities. A series of binary morphological operations are applied after range thresholding, to remove regions representing blood vessels and the collecting system. For comparison, the robust statistical segmenter (RSS) package available in 3D Slicer, based of variational framework is run to drive evolution of a closed surface through minimization of two-term energy functional. Those terms comprise information about similarity of statistical features of the region to the seed, and the area of the surface. As a matter of fact, the RSS method failed in 45% of cases. The authors conclude that the blanket method offers high speed, high accuracy, has the potential for clinical implementation, and is competitive with the 3D Slicer RSS method.

Thresholding combined with connected component analysis referred to as semiautomatic approach was compared to the ellipsoid formula by Suess *et al.* [130]. The semi automatic approach, tested on non contrast enhanced T2w MRI in 20 healthy volunteers, reached a high accuracy compared to the manual estimated ground truth. The ellipsoid formula had a higher deviation from the reference volumes estimated but segmentation could be obtained in roughly one fifth of the computation time (41 ± 11 s). The approach was then applied to renal arterial spin labeling (ASL) to estimate renal cortical and medullary perfusion [131].

Feng *et al.* [132] compared five different segmentation algorithms to be employed in tracking the kidneys in Magnetic Resonance Image-Guided Radiotherapy of renal tumors. The global thresholding method was chosen as a representative thresholding method. For unsupervised learning method, the fuzzy k-means (FKM) and k-harmonic means (KHM) were chosen. For partial differential equation-based deformable models, we selected a

modified reaction-diffusion level-set evolution (RD-LSE) method. For the atlas-based method, the ViewRay treatment planning and delivery system (VR-TPDS) was used. To compare all the gray-scale images within the same level, all the images tested were linearly transformed to have 256 gray levels. For the final contouring of the objective organ, necessary morphological image processing steps were used for postprocessing. For initialisation of the clustering a test run was performed to identify the cluster group (FKM and KHM), the thresholding, and the appropriate level set (RD-LSE). The mean target registration error (TRE) values of thresholding and RD-LSE for tracking the kidney were 1.76 mm and 1.75 mm. The mean TRE values of FKM and KHM are very similar: 0.92 vs. 0.96 mm for kidney. The VR-TPDS method had TRE values of 0.69 mm for the kidney.

The work of Barstugan *et al.* aims at segmenting adrenal tumors from MR image [86]. The authors combine various techniques like K-means clustering, active contours, adaptive thresholding, image erosion, and region growing to segment the liver, the kidney and eventually the adrenal tumor from the abdominal T1w and T2w scans. In total, 113 MRI scans from a not detailed number of patients is used in this work. The authors evaluated their technique using a bunch of metrics like sensitivity, specificity, accuracy, precision, Dice coefficient, Jaccard index and the structural similarity index (SSIM) to also differentiate subtypes of adrenal tumors. Segmentations could be computed in about two minutes at a Dice of 0.05 - 0.82 depending of the type of adrenal tumor.

In Yang *et al.*, the whole kidney and the renal compartments (cortex, medulla) are segmented from 26 DCE-MRI datasets [13], [60]. A three step approach is performed segmenting the whole kidney via Maximally Stable Temporal Volume (MSTV) which is based on thresholding the time series and calculating spatial-temporal correlations of the kidney and PCA-kmeans clustering for compartment segmentation.

E. APPLICATIONS USING DEEP LEARNING BASED SEGMENTATION

Deep learning based image segmentation approaches applied to the kidney emerge. O'Reilley *et al.* implemented a 3D fully-convolutional network (FCN) to estimate TKV. The network was adapted from [133] and trained on 155 datasets of about 50 patients [134]. In the work of Bevilacqua *et al.* two approaches are combined [88]. First, a CNN is employed that detects automatically a ROI spanning each of the kidneys (R-CNN). Then, a classification CNN is used to perform a semantic segmentation, here, a classifying voxels into kidney and non kidney. In this initial work only a relatively small training set of 57 images from 4 patients were used for training and testing the method while in the follow up paper a larger cohort was used [135].

Kline *et al.* proposed an “Artificial Multi-observer Network” comprising 11 individual trained and parameterized deep learning segmentation networks [11]. The developed convolutional neural network (CNN) architecture is based on

semantic segmentation approach. They search for a range of networks that were shallower and deeper (in terms of layers), thinner and wider (in terms of number and size of kernel), as well as different activation functions (ReLU, tanh). Based on the best performing network, eleven separate networks were trained in order to create the artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys (see Fig. 2). For training, about 2000 data sets derived from the TEMPO study [141] were used and labels were generated using a semi-automated segmentation procedure as outlined before [24].

Deep learning relies on a reasonable large number of samples to allow for a robust training of the networks. In medical imaging often such amount of data is missing, therefore, to use domain adaptation or transfer learning might be a solution to overcome such limitations. In renal segmentation, Lundervold *et al.* [62] used segmentation results from automated brain hippocampus labeling to obtain fast semi-supervised segmentation of the kidneys from 4D DCE-MRI (Fig. 10).

Haghghi *et al.* used a 3D U-NET to automatically segments kidneys from 4D DCE-MRI [84]. In their implementation a focus was to build a memory efficient framework to allow for using a 3D U-NET and also a large data tensor. Therefore, the segmentation task was divided in first a localisation step (i.e. finding the kidney in the MR images) and then actual segmentation task. The network was trained on 24 subjects while 6 subjects were used for testing including healthy subjects and subjects with renal diseases. Error in volume estimated to a ground truth reference was 12.9 ± 2.4 ml for healthy subjects and 19.5 ± 3.2 ml for subjects with renal disease, respectively.

Furthermore, there are also approaches optimizing existing architectures and the training of the neural networks. A genetic algorithm for optimizing the CNN architecture, modeling the number of encoders, the structure of each encoder and the final fully-connected layers is proposed by Brunetti *et al.* [91]. Their algorithm is designed for detecting images containing the kidney and subsequently segment them classifying each pixel. All images containing the kidney were split into left and right side and used as input dataset for the segmentation procedure. These two-step classification strategy allows obtaining the final segmentation of images representing kidneys affected by ADPKD.

Schnurr *et al.* investigated different sample mining schemes and there influence on CNN kidney segmentation in T1w MRI [140]. These comprise a) a random sampling, i.e. patches are randomly selected from each slice; b) “ $\mu \pm 3\text{Std}$ ”, i.e. patch centers are constrained to the average label position \pm three times the standard deviation and eventually c) label based, i.e. patch centers are constrained to label positions. A U-Net architecture with additional residual connections and batch normalization were employed. The networks were trained using the weighted cross entropy loss and 3-fold cross validation using the Visceral Anatomy 3 challenge and the Visceral Silver Corpus data set [142].

VIII. DISCUSSION

In the recent years, efforts to position [1] and standardize renal imaging [143] to enable researchers to create evidence of its clinical value have been done. However, most of this important recent work focus on imaging techniques [4], [5], [8], [144] and the standardised acquisition of morphological and functional data and its reporting [145]–[148]. A crucial point in this is the analysis of the data obtained by imaging.

Thereby, kidney image segmentation is an important step when estimating renal function as reflected by the recent literature reviewed here. Most studies in this review employ kidney segmentation in ADPKD to estimate the TKV which already qualified as a biomarker by the FDA, recently. However, other parameters like MIR-based GFR, hemodynamic parameters like blood flow or perfusion, diffusion or oxygenation are yet not established as renal biomarkers and further research and especially large scale studies are needed. Renal image segmentation can hereby play an important role as it enables a robust, reader independent and automated way to derive not only kidney volume but also segmentation of the kidney compartments which are essential also to extract and map the renal biomarkers obtained from the multiparametric imaging data (see Table 2).

While manual delineation is nowadays mainly used as reference method for the evaluation of new kidney segmentation approaches, semi- and automatic approaches are still developed to improve segmentation accuracy. These approaches could be divided into image-based and model-based techniques (see section III). Regarding the recent literature presented here, however, the majority of approaches use a combination of algorithms from the above mentioned categories. This is probably due to the fact that the renal image data (cf. Fig. 1) is difficult to process (low SNR, different amount of cyst or tumor load, adjacent organs like liver and spleen, motion) and a single technique cannot alone sufficiently segment the kidneys. Also, there is a variety of imaging contrast employed in renal imaging that warrants a robust segmentation approach. The values of DSC reported in the reviewed papers range between 0.5 and 0.98 depending on the specific application. VE mean value and standard deviation are typically a few percent each.

With the rise of deep learning in medical imaging and subsequent promising results, also these techniques emerge in the field of renal image segmentation. The benefit of deep learning approaches is that they learn their given task, i.e. here, segmenting the kidneys. An explicit modeling of the segmentation task like in the image- and model-based approaches is not needed. Considering results presented in the literature it is clear that applications of the deep learning methods for fully automated MRI kidney segmentation are encouraging, but there is a room for further improvement. According to results provided by the authors, we cannot conclude that current deep learning solutions outperform traditional image processing kidney segmentation algorithms. There are several issues that need to be addressed in order

to improve the network performance, such as: the lack of the sufficiently large MRI kidney datasets; generalization ability of the network response in case the source of data (data acquisition device) changes; training with imbalanced data. There are also ideas of combining both deep learning and image processing techniques [88].

Moreover, in general there is still a lack of a comparison of methods with respect to segmentation performance. Already in their review in 2013, Zöllner *et al.* pointed out that such comparisons are highly valuable [2]. This is key aspect if such methods should be employed e.g. to automatically estimate a biomarker like TKV, or if included in larger renal data analysis pipelines to estimate their contribution to the overall performance of such systems [149]. Most methods are evaluated on own curated data that makes a comparison across publications difficult. The performance of the algorithms clearly depends on the data and several groups differentiate results according to patient health status [13], [60], [65]. Even the way how results are reported is very fragmented as seen in Table 2. For instance, not all studies report a DSC nor a VE. Thus, the results provided by the authors in the papers cannot be directly compared, even within the same segmentation task.

In CT imaging abdominal organs segmentation challenges [142] have been proposed which might be a resource for allowing such comparisons across methods based on the same curated data set and a fixed evaluation strategy. The CHAOS challenge [40] also provided MRIs however, the aim was a general organ segmentation. To the best of the authors knowledge, a dedicated challenge for renal image segmentation yet is not proposed. The only challenge on renal image data is the KiTS2019 challenge [150] on renal tumor segmentation, however, from CT images. For renal imaging such a challenge might be warranted to allow for method comparisons.

Furthermore, in brain imaging a set of well recognized open source image processing packages (SPM, Freesurfer, FSL) have emerged which are widely used in that field. Such an evolution of renal image segmentation methods would be warranted to set up a complete and automated renal image data analysis pipeline. Such system can provide the necessary infrastructure to drive renal image biomarker discovery and clinical value of renal imaging. A set of tools and algorithm reported in this work (see section VI) have been published open source and might be the seed of the aforementioned evolution towards commonly recognized tools and processing pipelines.

In this review, we focused on approaches applied to MR image datasets. However, renal image segmentation is also successfully performed on CT image data including deep learning approaches [56], [57], [67] which might be explored for/ or transferred to renal MRI.

IX. CONCLUSION

In conclusion, renal image segmentation is a valuable tool to further automate the analysis of renal MRI and thus might foster reader independent data analysis in e.g. multi-center

studies. Thereby, it might be possible establishing renal imaging biomarkers more easily. Most recent works combine image based and model based approaches to allow a robust segmentation undermining the non-trivial analysis of renal MRI data. Probably deep learning based approaches might be a solution to further boost results and generalization. Apart from the segmentation techniques, a broader comparison and evaluation of methods as demonstrated is needed. Therefore, curated publically available datasets are important and warranted.

REFERENCES

- [1] N. M. Selby *et al.*, "Magnetic resonance imaging biomarkers for chronic kidney disease: A position paper from the European cooperation in science and technology action PARENCHIMA," *Nephrol. Dial. Transplantation*, vol. 33, no. 2, pp. 4–14, Sep. 2018.
- [2] F. G. Zöllner, E. Svarstad, A. Z. Munthe-Kaas, L. R. Schad, A. Lundervold, and J. and Roervik, "Assessment of kidney volumes from MRI: Acquisition and segmentation techniques," *Amer. J. Roentgenol.*, vol. 199, pp. 1060–1069, Nov. 2012.
- [3] H. R. Torres, S. Queirós, P. Morais, B. Oliveira, J. C. Fonseca, and J. L. Vilaça, "Kidney segmentation in ultrasound, magnetic resonance and computed tomography images: A systematic review," *Comput. Methods Programs Biomed.*, vol. 157, pp. 49–67, Apr. 2018.
- [4] A. Caroli, M. Schneider, I. Friedli, A. Ljimani, S. De Seigneux, P. Boor, L. Gullapudi, I. Kazmi, I. A. Mendichovszky, M. Notohamiprodjo, N. M. Selby, H. C. Thoeny, N. Grenier, and J.-P. Vallée, "Diffusion-weighted magnetic resonance imaging to assess diffuse renal pathology: A systematic review and statement paper," *Nephrol. Dial. Transplantation*, vol. 33, no. 2, pp. ii29–ii40, Sep. 2018.
- [5] A. Odudu, F. Nery, A. A. Harteveld, R. G. Evans, D. Pendse, C. E. Buchanan, S. T. Francis, and M. A. Fernández-Seara, "Arterial spin labelling MRI to measure renal perfusion: A systematic review and statement paper," *Nephrol. Dial. Transplantation*, vol. 33, no. 2, pp. ii15–ii21, Sep. 2018.
- [6] R. A. Jones, J. R. Votaw, K. Salman, P. Sharma, C. Lurie, B. Kalb, and D. R. Martín, "Magnetic resonance imaging evaluation of renal structure and function related to disease: Technical review of image acquisition, postprocessing, and mathematical modeling steps," *J. Magn. Reson. Imag.*, vol. 33, pp. 1270–1283, Jun. 2011.
- [7] F. G. Zöllner, T. Gaa, F. Zimmer, M. M. Ong, P. Riffel, D. Hausmann, S. O. Schoenberg, and M. Weis, "Quantitative perfusion imaging in magnetic resonance imaging," *Der Radiol.*, vol. 56, pp. 113–123, Feb. 2016.
- [8] M. Pruijm, I. A. Mendichovszky, P. Liss, P. Van der Niepen, S. C. Textor, L. O. Lerman, C. T. P. Krediet, A. Caroli, M. Burnier, and P. V. Prasad, "Renal blood oxygenation level-dependent magnetic resonance imaging to measure renal tissue oxygenation: A statement paper and systematic review," *Nephrol. Dial. Transplant.*, vol. 33, pp. ii22–ii28, Sep. 2018.
- [9] A. C. M. Ong, O. Devyust, B. Knebelmann, and G. Walz, "Autosomal dominant polycystic kidney disease: The changing face of clinical management," *Lancet*, vol. 385, pp. 1993–2002, May 2015.
- [10] Center for Drug Evaluation and Research. (2016). *Qualification of Biomarker Total Kidney Volume in, Studies for Treatment of Autosomal Dominant Polycystic Kidney Disease Draft Guidance for Industry*. [Online]. Available: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/qualification-biomarker-total-kidney-volume-studies-treatment-autosomal-dominant-polycystic-kidney>
- [11] T. L. Kline, P. Korfiatis, M. E. Edwards, J. D. Blais, F. S. Czerwiec, P. C. Harris, B. F. King, V. E. Torres, and B. J. Erickson, "Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys," *J. Digit. Imag.*, vol. 30, no. 4, pp. 442–448, Aug. 2017.
- [12] O. Glöger, K. Tönnes, B. Mensel, and H. Völzke, "Fully automatized renal parenchyma volumetry using a support vector machine based recognition system for subject-specific probability map generation in native MR volume data," *Phys. Med. Biol.*, vol. 60, no. 22, pp. 8675–8693, Nov. 2015.
- [13] X. Yang, H. Le Minh, K.-T. Cheng, K. H. Sung, and W. Liu, "Renal compartment segmentation in DCE-MRI images," *Med. Image Anal.*, vol. 32, pp. 269–280, Aug. 2016.

- [14] N. Wake, J. S. Wysock, M. A. Bjurlin, H. Chandarana, and W. C. Huang, “‘Pin the tumor on the kidney’: An evaluation of how surgeons translate CT and MRI data to 3D models,” *Urology*, vol. 131, pp. 255–261, Sep. 2019.
- [15] T. L. Kline, M. E. Edwards, I. Garg, M. V. Irazabal, P. Korfiatis, P. C. Harris, B. F. King, V. E. Torres, S. K. Venkatesh, and B. J. Erickson, “Quantitative MRI of kidneys in renal disease,” *Abdominal Radiol.*, vol. 43, pp. 629–638, Mar. 2018.
- [16] K. Bae, B. Park, H. Sun, J. Wang, C. Tao, A. B. Chapman, V. E. Torres, J. J. Grantham, M. Mrug, W. M. Bennett, M. F. Flessner, D. P. Landsittel, and K. T. Bae, “Segmentation of individual renal cysts from mr images in patients with autosomal dominant polycystic kidney disease,” *Clin. J. Amer. Soc. Nephrol.*, vol. 8, pp. 1089–1097, Jul. 2013.
- [17] Y. Kim, Y. Ge, C. Tao, J. Zhu, A. B. Chapman, V. E. Torres, A. S. L. Yu, M. Mrug, W. M. Bennett, M. F. Flessner, D. P. Landsittel, and K. T. Bae, “Automated segmentation of kidneys from MR images in patients with autosomal dominant polycystic kidney disease,” *Clin. J. Amer. Soc. Nephrol.*, vol. 11, no. 4, pp. 576–584, Apr. 2016.
- [18] X. Liu, K. Tuncali, W. M. Wells, and G. P. Zientara, “Automatic iceball segmentation with adapted shape priors for MRI-guided cryoablation,” *J. Magn. Reson. Imag.*, vol. 41, no. 2, pp. 517–524, Feb. 2015.
- [19] H. Rusinek, J. C. Lim, N. Wake, J.-M. Seah, E. Botterill, S. Farquharson, A. Mikheev, and R. P. Lim, “A semi-automated ‘blanket’ method for renal segmentation from non-contrast T1-weighted MR images,” *Magn. Reson. Mater. Phys.*, vol. 29, pp. 197–206, Apr. 2016.
- [20] F. G. Zöllner, A. Šerifović-Trbali, G. Kabelitz, M. Kociäski, A. Materka, and P. Rogelj, “Image registration in dynamic renal MRI—Current status and prospects,” *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 33, no. 1, pp. 33–48, Feb. 2020.
- [21] D. Turco, S. Severi, R. Mignani, R. Magistroni, and C. Corsi, “Geometry-independent assessment of renal volume in polycystic kidney disease from magnetic resonance imaging,” in *Proc. 37th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 3081–3084.
- [22] 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 Res. Notes*, vol. 9, no. 1, p. 489, Nov. 2016.
- [23] J. D. Warner, M. V. Irazabal, G. Krishnamurthi, B. F. King, V. E. Torres, and B. J. Erickson, “Supervised segmentation of polycystic kidneys: A new application for stereology data,” *J. Digit. Imag.*, vol. 27, no. 4, pp. 514–519, Aug. 2014.
- [24] T. L. Kline, M. E. Edwards, P. Korfiatis, Z. Akkus, V. E. Torres, and B. J. Erickson, “Semiautomated segmentation of polycystic kidneys in T2-weighted MR images,” *Amer. J. Roentgenol.*, vol. 207, no. 3, pp. 605–613, Sep. 2016.
- [25] K. T. Bae, P. K. Commean, B. S. Brunsden, D. A. Baumgarten, J. F. King, L. H. Wetzel, P. J. Kenney, A. B. Chapman, V. E. Torres, J. J. Grantham, L. M. Guay-Woodford, C. Tao, J. P. Miller, C. M. Meyers, and W. M. Bennett, “Segmentation and volumetric measurement of renal cysts and parenchyma from MR images of polycystic kidneys using multi-spectral analysis method,” in *Medical Imaging: Image Processing*, J. M. Reinhardt and J. P. W. Pluim, Eds. Bellingham, WA, USA: SPIE, 2008, Art. no. 691449.
- [26] M. Zhang, T. Wu, S. C. Beeman, L. Cullen-McEwen, J. F. Bertram, J. R. Charlton, E. Baldeolmar, and K. M. Bennett, “Efficient small blob detection based on local convexity, intensity and shape information,” *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1127–1137, Apr. 2016.
- [27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New York, NY, USA: Pearson, 2018.
- [28] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. Upper Saddle River NJ, USA: Prentice-Hall, 2004.
- [29] E. Davies, *Computer and Machine Vision: Theory, Algorithms, Practicalities*. Waltham, MA, USA: Elsevier, 2012.
- [30] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer, 2004.
- [31] C. Rother, V. Kolmogorov, and A. Blake, “‘GrabCut’ interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.
- [32] S. Will, P. Martirosian, C. Wärslin, 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,” *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 27, no. 5, pp. 445–454, Oct. 2014.
- [33] W. A. Abdulahi and J. R. Tapamo, “Fast Chan-Vese without edges and connected component analysis for kidney segmentation in MRI images,” in *Proc. AFRICON*, Sep. 2015, pp. 1–5.
- [34] T. Woodard, S. Sigurdsson, J. D. Gotal, A. A. Torjesen, L. A. Inker, T. Aspelund, G. Eiriksdottir, V. Gudnason, T. B. Harris, L. J. Launer, A. S. Levey, and G. F. Mitchell, “Segmental kidney vol. measured, by dynamic contrast-enhanced magnetic resonance imaging and their association with CKD in older people,” *Am. J. Kidney Dis.*, vol. 65, pp. 41–48, Jan. 2015.
- [35] S. Mushrif, A. Morales, C. Sica, Q. X. Yang, S. Eskin, and L. Sinowa, “A novel intuitionistic fuzzy set approach for segmentation of kidney MR images,” in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2016, pp. 1–6.
- [36] E. A. Hanson and A. Lundervold, “Local/non-local regularized image segmentation using graph-cuts: Application to dynamic and multispectral MRI,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 8, no. 6, pp. 1073–1084, Nov. 2013.
- [37] A. K. Rudra, A. S. Chowdhury, A. Elnakib, F. Khalifa, A. Soliman, G. Beache, and A. El-Baz, “Kidney segmentation using graph cuts and pixel connectivity,” *Pattern Recognit. Lett.*, vol. 34, no. 13, pp. 1470–1475, Oct. 2013.
- [38] U. Yourk, B. A. Hargreaves, and S. S. Vasanaawala, “Automatic renal segmentation for mr urography using 3d-grabcut and random forests,” *Magn. Reson. Med.*, vol. 79, pp. 1696–1707, Mar. 2018.
- [39] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [40] A. E. Kavur et al., “CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation,” *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101950.
- [41] P. A. Yushkevich and G. Gerig, “ITK-SNAP: An interactive medical image segmentation tool to meet the need for expert-guided segmentation of complex medical images,” *IEEE Pulse*, vol. 8, no. 4, pp. 54–57, Jul. 2017.
- [42] P. L. Rosin, “Unimodal thresholding,” *Pattern Recognit.*, vol. 34, no. 11, pp. 2083–2096, Nov. 2001.
- [43] S. Arora, J. Acharya, A. Verma, and P. K. Panigrahi, “Multilevel thresholding for image segmentation through a fast statistical recursive algorithm,” *Pattern Recognit. Lett.*, vol. 29, no. 2, pp. 119–125, Jan. 2008.
- [44] A. Ben Ishak, “A two-dimensional multilevel thresholding method for image segmentation,” *Appl. Soft Comput.*, vol. 52, pp. 306–322, Mar. 2017.
- [45] O. Gloer, K. Tonnes, R. Laqua, and H. Volzke, “Fully automated renal tissue volumetry in MR volume data using prior-shape-based segmentation in subject-specific probability maps,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2338–2351, Oct. 2015.
- [46] J. A. O'Reilly, S. Tanpradit, T. Puttasakul, M. Sangworalasil, T. Matsuura, P. Wibulpolprasert, and K. Chousangsuntorn, “Automatic segmentation of polycystic kidneys from magnetic resonance images using decision tree classification and snake algorithm,” in *Proc. 12th Biomed. Eng. Int. Conf. (BMEICON)*, Nov. 2019, pp. 1–5.
- [47] A. S. Lundervold and A. Lundervold, “An overview of deep learning in medical imaging focusing on MRI,” *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019.
- [48] J. Bernal, K. Kushibar, D. S. Asfaw, S. Valverde, A. Oliver, R. Martí, and X. Lladó, “Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review,” *Artif. Intell. Med.*, vol. 95, pp. 64–81, Apr. 2019.
- [49] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [50] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. Way, E. Ferrero, P. Agapow, M. Zietz, N. M. Hoffmann, and W. Xie, “Opportunities and obstacles for deep learning in biology and medicine,” *J. Royal Soc. Interface*, vol. 15, no. 141, 2018, Art. no. 20170387.
- [51] F. Piccialli, V. D. Somma, F. Giampaolo, S. Cuomo, and G. Fortino, “A survey on deep learning in medicine: Why, how and when?” *Inf. Fusion*, vol. 66, pp. 111–137, Feb. 2021.
- [52] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, May 2015.
- [53] X. Zhao and X. M. Zhao, “Deep learning of brain magnetic resonance images: A brief review,” *Methods*, Sep. 2020, doi: [10.1016/j.jymeth.2020.09.007](https://doi.org/10.1016/j.jymeth.2020.09.007).

- [54] I. Ulku and E. Akagunduz, "A survey on deep learning-based architectures for semantic segmentation on 2D images," 2021, *arXiv:1912.10230*. [Online]. Available: <https://arxiv.org/abs/1912.10230>
- [55] E. Goceri, "Challenges and recent solutions for image segmentation in the era of deep learning," in *Proc. 9th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2019, pp. 1–6.
- [56] K. Sharma, C. Rupprecht, A. Caroli, M. C. Aparicio, A. Remuzzi, M. Baust, and N. Navab, "Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease," *Sci. Rep.*, vol. 7, May 2017, Art. no. 2049.
- [57] V. Couteaux, S. Si-Mohamed, R. Renard-Penna, O. Nempong, T. Lefevre, A. Popoff, G. Pizaine, N. Villain, I. Bloch, J. Behr, M.-F. Bellin, C. Roy, O. Rouvière, S. Montagne, N. Lassau, and L. Boussel, "Kidney cortex segmentation in 2D CT with U-Nets ensemble aggregation," *Diagnostic Interventional Imag.*, vol. 100, no. 4, pp. 211–217, Apr. 2019.
- [58] F. Turk, M. Luy, and N. Barisci, "Comparison of Unet3D models for kidney tumor segmentation," *Preprints*, 2020, Art. no. 2020010314, doi: [10.20944/preprints202001.0314.v1](https://doi.org/10.20944/preprints202001.0314.v1).
- [59] K. C. Homepage. (Oct. 2020). *Kits19 Challenge Homepage*. [Online]. Available: <https://kits19.grand-challenge.org>
- [60] X. Yang, H. Le Minh, T. Cheng, K. H. Sung, and W. Liu, "Automatic segmentation of renal compartments in DCE-MRI images," in *Medical Image Computing and Computer-Assisted Intervention*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, eds. Cham, Switzerland: Springer, 2015, pp. 3–11.
- [61] W. McCulloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, pp. 127–147, Dec. 1943.
- [62] A. S. Lundervold, J. Rorvik, and A. Lundervold, "Fast semi-supervised segmentation of the kidneys in DCE-MRI using convolutional neural networks and transfer learning," in *Proc. 2nd Int. Sci. Symp.*, Berlin, Germany, 2017, pp. 79–81.
- [63] G. Villa, S. Ringgaard, I. Hermann, R. Noble, P. Brambilla, D. S. Khatir, F. G. Zöllner, S. T. Francis, N. M. Selby, A. Remuzzi, and A. Caroli, "Phase-contrast magnetic resonance imaging to assess renal perfusion: A systematic review and statement paper," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 33, no. 1, pp. 3–21, Feb. 2020.
- [64] F. Liang, P. Qian, K.-H. Su, A. Baydoun, A. Leisser, S. Van Hedent, J.-W. Kuo, K. Zhao, P. Parikh, Y. Lu, B. J. Traughber, and R. F. Muzic, "Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: An intelligent, multi-level fusion approach," *Artif. Intell. Med.*, vol. 90, pp. 34–41, Aug. 2018.
- [65] K. Heryan, A. Skalski, J. Jakubowski, T. Drewniak, and J. Gajda, "Automatic extraction of the pelvicalyceal system for preoperative planning of minimally invasive procedures," *Metrol. Meas. Syst.*, vol. 24, no. 1, pp. 3–18, Mar. 2017.
- [66] Y. Fu, T. R. Mazur, X. Wu, S. Liu, X. Chang, Y. Lu, H. H. Li, H. Kim, M. C. Roach, L. Henke, and D. Yang, "A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy," *Med. Phys.*, vol. 45, no. 11, pp. 5129–5137, Nov. 2018.
- [67] D. Xiang, U. Bagci, C. Jin, F. Shi, W. Zhu, J. Yao, M. Sonka, and X. Chen, "CorteXpert: A model-based method for automatic renal cortex segmentation," *Med. Image Anal.*, vol. 42, pp. 257–273, Dec. 2017.
- [68] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M. A. El-Ghar, A. C. Dwyer, R. Ouseph, and A. El-Baz, "A novel framework for automatic segmentation of kidney from DW-MRI," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 951–954.
- [69] M. Shehata, F. Khalifa, A. Soliman, M. Ghazal, F. Taher, M. A. El-Ghar, A. C. Dwyer, G. Gimel'farb, R. S. Keynton, and A. El-Baz, "Computer-aided diagnostic system for early detection of acute renal transplant rejection using diffusion-weighted MRI," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 539–552, Feb. 2019.
- [70] M. Shehata, M. Ghazal, G. Beach, M. A. El-Ghar, A. Dwyer, H. Hajjdiab, A. Khalil, and A. El-Baz, "Role of integrating diffusion mr image-markers with clinical-biomarkers for early assessment of renal transplants," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 146–150.
- [71] M. Shehata, A. Mahmoud, A. Soliman, F. Khalifa, M. Ghazal, M. A. 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, Oct. 2018, Art. no. e0200082.
- [72] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: A deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018.
- [73] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M. A. El-Ghar, A. C. Dwyer, R. Ouseph, and A. El-Baz, "A level set-based framework for 3D kidney segmentation from diffusion MR images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4441–4445.
- [74] M. Shehata, F. Khalifa, E. Hollis, A. Soliman, E. Hosseini-Asl, M. A. El-Ghar, M. El-Baz, A. C. Dwyer, A. El-Baz, and R. Keynton, "A new non-invasive approach for early classification of renal rejection types using diffusion-weighted MRI," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 136–140.
- [75] J. Lv, W. Huang, J. Zhang, and X. Wang, "Performance of U-net based pyramidal lucas-kanade registration on free-breathing multi-b-value diffusion MRI of the kidney," *Brit. J. Radiol.*, Mar. 2018, Art. no. 20170813.
- [76] L. Gui, C. Li, and X. Yang, "Medical image segmentation based on level set and isoperimetric constraint," *Phys. Medica*, vol. 42, pp. 162–173, Oct. 2017.
- [77] I. Aganj, M. G. Harisinghani, R. Weissleder, and B. Fischl, "Unsupervised medical image segmentation based on the local center of mass," *Sci. Rep.*, vol. 8, no. 1, Aug. 2018, Art. no. 13012.
- [78] R. Chav, T. Cresson, G. Chartrand, C. Kauffmann, G. Soulez, and J. A. de Guise, "Kidney segmentation from a single prior shape in MRI," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 818–821.
- [79] F. Khalifa, G. M. Beach, M. A. El-Ghar, T. El-Diasty, G. Gimel'farb, M. Kong, and A. El-Baz, "Dynamic contrast-enhanced MRI-based early detection of acute renal transplant rejection," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1910–1927, Oct. 2013.
- [80] Y. Xu, F. Gao, T. Wu, K. M. Bennett, J. R. Charlton, and S. Sarkar, "U-net with optimal thresholding for small blob detection in medical images," in *Proc. IEEE 15th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2019, pp. 1761–1767.
- [81] F. Liu, C.-A. Cuenod, I. Thomassin-Naggara, S. Chemouny, and Y. Rozenholc, "Hierarchical segmentation using equivalence test (HiSET): Application to DCE image sequences," *Med. Image Anal.*, vol. 51, pp. 125–143, Jan. 2019.
- [82] C. E. Cardenas, J. Yang, B. M. Anderson, L. E. Court, and K. B. Brock, "Advances in auto-segmentation," *Semin. Radiat. Oncol.*, vol. 29, pp. 185–197, Jul. 2019.
- [83] P. Jackson, N. Hardcastle, N. Dawe, T. Kron, M. S. Hofman, and R. J. Hicks, "Deep learning renal segmentation for fully automated radiation dose estimation in unsealed source therapy," *Frontiers Oncol.*, vol. 8, pp. 1–7, Jun. 2018.
- [84] M. Haghghi, S. K. Warfield, and S. Kurugol, "Automatic renal segmentation in DCE-MRI using convolutional neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1534–1537.
- [85] M. F. Bobo, S. Bao, Y. Huo, Y. Yao, J. Virostko, A. J. Plassard, I. Lyu, A. Assad, R. G. Abramson, M. A. Hilmes, and B. A. Landman, "Fully convolutional neural networks improve abdominal organ segmentation," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 10574, Mar. 2018, Art. no. 105742V.
- [86] M. Barstüaan, R. Ceylan, S. Asoglu, H. Cebeci, and M. Koplay, "Adrenal tumor segmentation method for MR images," *Comput. Methods Programs Biomed.*, vol. 164, pp. 87–100, Oct. 2018.
- [87] T. L. Kline, P. Korfiantis, M. E. Edwards, J. D. Warner, M. V. Izrazabal, B. F. King, V. E. Torres, and B. J. Erickson, "Automatic total kidney vol. measurement, on follow-up magnetic resonance images to facilitate monitoring of autosomal dominant polycystic kidney disease progression," *Nephrol. Dial. Transplantation*, vol. 31, pp. 241–248, Feb. 2016.
- [88] V. Bevilacqua, A. Brunetti, G. D. Cascarano, F. Palmieri, A. Guerrero, and M. Moschetta, "A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on magnetic resonance images," in *Intelligent Computing Theories and Application (Lecture Notes in Computer Science)*, vol. 10955, D. S. Huang, K. H. Jo, and X. L. Zhang, Eds. Cham, Switzerland: Springer, 2018, doi: [10.1007/978-3-319-95933-7_73](https://doi.org/10.1007/978-3-319-95933-7_73).
- [89] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
- [90] T. L. Kline, "Segmenting new image acquisitions without labels," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 330–333.
- [91] A. Brunetti, G. D. Cascarano, I. De Feudis, M. Moschetta, L. Gesualdo, and V. Bevilacqua, "Detection and segmentation of kidneys from magnetic resonance images in patients with autosomal dominant polycystic kidney disease," in *Intelligent Computing Theories and Application*, D.-S. Huang, K.-H. Jo, and Z.-K. Huang, Eds. Cham, Switzerland: Springer, 2019, pp. 639–650.

- [92] M. Shehata, M. Ghazal, F. Khalifa, M. A. El-Ghar, A. Khalil, A. C. Dwyer, A. El-giziri, M. El-Melegy, and A. El-Baz, "A novel CAD system for detecting acute rejection of renal allografts based on integrating imaging-markers and laboratory biomarkers," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–6.
- [93] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, Dec. 2015.
- [94] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies (OASIS): Longitudinal MRI data in nondemented and demented older adults," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2677–2684, 2010.
- [95] I. Fellhaber, F. G. Zöllner, J. Schröder, C. Degen, L. Kong, M. Essig, P. A. Thomann, and L. R. Schad, "Comparison of automated brain segmentation using a brain phantom and patients with early Alzheimer's dementia or mild cognitive impairment," *Psychiatry Res., Neuroimag.*, vol. 233, no. 3, pp. 299–305, Sep. 2015.
- [96] P. Cooley, C. Scheper, Q. Ying, C. Turner, S. Cantor, and H. Ray, "The NIDDK central repository using legacy data & samples to address new questions," *Diabetes*, vol. 56, p. 290, Oct. 2007.
- [97] C. F. Turner *et al.*, "The NIDDK central repository at 8 years-ambition, revision, use and impact," *Database*, vol. 2011, Sep. 2011, Art. no. bar043.
- [98] J. J. Grantham, L. T. Cook, L. H. Wetzel, M. A. Cadnapaphornchai, and K. T. Bae, "Evidence of extraordinary growth in the progressive enlargement of renal cysts," *Clin. J. Amer. Soc. Nephrol.*, vol. 5, no. 5, pp. 889–896, May 2010.
- [99] T. Langner, A. Östling, L. Maldonis, A. Karlsson, D. Olmo, D. Lindgren, A. Wallin, L. Lundin, R. Strand, H. Ahlström, and J. Kullberg, "Kidney segmentation in neck-to-knee body MRI of 40,000 UK Biobank participants," *Sci. Rep.*, vol. 10, Dec. 2020, Art. no. 20963.
- [100] S. Shirly and K. Ramesh, "Review on 2D and 3D MRI image segmentation techniques," *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 15, no. 2, pp. 150–160, Jan. 2019.
- [101] A. Virzi, C. O. Muller, J.-B. Marret, E. Mille, L. Berteloot, D. Grévent, N. Boddaert, P. Gori, S. Sarnacki, and I. Bloch, "Comprehensive review of 3D segmentation software tools for MRI usable for pelvic surgery planning," *J. Digit. Imag.*, vol. 33, no. 1, pp. 99–110, Feb. 2020.
- [102] GitHub. (Oct. 2020). *The Insight Toolkit*. [Online]. Available: <https://itk.org>
- [103] GitHub. (Oct. 2020). *SimpleITK*. [Online]. Available: <https://simpleitk.org>
- [104] GitHub. (Oct. 2020). *The Visualization Toolkit*. [Online]. Available: <https://vtk.org>
- [105] GitHub. (Oct. 2020). *Scipy*. [Online]. Available: <https://scipy.org>
- [106] GitHub. (Oct. 2020). *Scikit-Image*. [Online]. Available: <https://scikit-image.org>
- [107] GitHub. (Oct. 2020). *Renal-Segmentation*. [Online]. Available: <https://github.com/umityoruk/renal-segmentation>
- [108] GitHub. (Oct. 2020). *Kidney-Compartment-Segmentation-From-MRR-Images*. [Online]. Available: <https://github.com/xinyang-hust/kidney-compartment-segmentation-from-MRR-images>
- [109] GitHub. (Oct. 2020). *Medical Image Segmentation With Convolutional Neural Networks*. [Online]. Available: <https://github.com/frankkramer-lab/MIScnn>
- [110] GitHub. (Oct. 2020). *Kits19 Participation With misCNN*. [Online]. Available: <https://github.com/muellero/kits19.MIScnn>
- [111] K. A. Philbrick, A. D. Weston, Z. Akkus, T. L. Kline, P. Korfiatis, T. Sakinis, P. Kostandy, A. Boonrod, A. Zeinoddini, N. Takahashi, and B. J. Erickson, "RIL-contour: A medical imaging dataset annotation tool for and with deep learning," *J. Digit. Imag.*, vol. 32, no. 4, pp. 571–581, Aug. 2019.
- [112] R. J. Simms, T. Doshi, P. Metherall, D. Ryan, P. Wright, N. Gruel, M. D. A. van Gastel, R. T. Gansevoort, W. Tindale, and A. C. M. Ong, "A rapid high-performance semi-automated tool to measure total kidney volume from MRI in autosomal dominant polycystic kidney disease," *Eur. Radiol.*, vol. 29, no. 8, pp. 4188–4197, Aug. 2019.
- [113] B. Shi, P. Akbari, M. Pourafkari, I.-A. Iliuta, E. Guiard, C. F. Quist, X. Song, D. Hillier, K. Khalili, and Y. Pei, "Prognostic performance of kidney volume measurement for polycystic kidney disease: A comparative study of ellipsoid vs. Manual segmentation," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 10996.
- [114] S. Lai, D. Mastroluca, C. Letizia, L. Petramala, A. M. Perrotta, A. DiGaeta, L. Ferrigno, M. Ciccarello, A. R. D'Angelo, and V. Panebianco, "Magnetic resonance imaging 3T and total fibrotic volume in autosomal dominant polycystic kidney disease," *Internal Med. J.*, vol. 48, no. 12, pp. 1505–1513, Dec. 2018.
- [115] D. Turco, M. Busutti, R. Mignani, R. Magistroni, and C. Corsi, "Comparison of total kidney volume quantification methods in autosomal dominant polycystic disease for a comprehensive disease assessment," *Amer. J. Nephrol.*, vol. 45, no. 5, pp. 373–379, 2017.
- [116] E. Banach-Ambroziak, M. Jankowska, M. Grzywińska, J. Pieákowska, and E. Szurowska, "MRI-derived markers for predicting a decline in renal function in patients with autosomal dominant polycystic kidney disease," *Polish J. Radiol.*, vol. 84, pp. 289–294, Oct. 2019.
- [117] R. Magistroni, C. Corsi, T. Martí, and R. Torra, "A review of the imaging techniques for measuring kidney and cyst volume in establishing autosomal dominant polycystic kidney disease progression," *Amer. J. Nephrol.*, vol. 48, no. 1, pp. 67–78, 2018.
- [118] F. Siedek, S. Haneder, J. Dörner, J. N. Morelli, S.-H. Chon, D. Maintz, and C. Houbois, "Estimation of split renal function using different volumetric methods: Inter-and intraindividual comparison between MRI and CT," *Abdominal Radiol.*, vol. 44, no. 4, pp. 1481–1492, Apr. 2019.
- [119] D. K. Dwivedi, Y. Chatzinoff, Y. Zhang, Q. Yuan, M. Fulkerson, R. Chopra, J. Brugarolas, J. A. Cadeddu, P. Kapur, and I. Pedrosa, "Development of a patient-specific tumor mold using magnetic resonance imaging and 3-dimensional printing technology for targeted tissue procurement and radiomics analysis of renal masses," *Urology*, vol. 112, pp. 209–214, Feb. 2018.
- [120] N. Wake, H. Chandarana, W. C. Huang, S. S. Taneja, and A. B. Rosenkrantz, "Application of anatomically accurate, patient-specific 3D printed models from MRI data in urological oncology," *Clin. Radiol.*, vol. 71, no. 6, pp. 610–614, Jun. 2016.
- [121] D. Turco, S. Severi, R. Mignani, V. Aiello, R. Magistroni, and C. Corsi, "Reliability of total renal volume computation, in polycystic kidney disease from magnetic resonance imaging," *Acad. Radiol.*, vol. 22, pp. 1376–1384, Nov. 2015.
- [122] F. Khalifa, M. Abou El-Ghar, B. Abdollahi, H. B. Frieboes, T. El-Diasty, and A. El-Baz, "A comprehensive non-invasive framework for automated evaluation of acute renal transplant rejection using DCE-MRI," *NMR Biomed.*, vol. 26, pp. 1460–1470, Nov. 2013.
- [123] F. Khalifa, M. Shehata, A. Soliman, M. Abou El-Ghar, T. El-Diasty, A. C. Dwyer, M. El-Melegy, G. Gimel'farb, R. Keynton, and A. El-Baz, "A generalized MRI-based CAD system for functional assessment of renal transplant," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 758–761.
- [124] E. Hollis, M. Shehata, M. Abou El-Ghar, M. Ghazal, T. El-Diasty, M. Merchant, A. E. Switala, and A. El-Baz, "Statistical analysis of ADCs and clinical biomarkers in detecting acute renal transplant rejection," *Brit. J. Radiol.*, vol. 90, no. 1080, Dec. 2017, Art. no. 20170125.
- [125] G. Chiusano, A. Staglianá, C. Basso, and A. Verri, "Unsupervised tissue segmentation from dynamic contrast-enhanced magnetic resonance imaging," *Artif. Intell. Med.*, vol. 61, no. 1, pp. 53–61, May 2014.
- [126] F. G. Zöllner, R. Sance, P. Rogelj, M. J. Ledesma-Carbajo, J. Rørvik, A. Santos, and A. Lundervold, "Assessment of 3D DCE-MRI of the kidneys using non-rigid image registration and segmentation of voxel time courses," *Comput. Med. Imag. Graph.*, vol. 33, no. 3, pp. 171–181, Apr. 2009.
- [127] K. T. Bae, H. Sun, J. G. Lee, K. Bae, J. Wang, C. Tao, A. B. Chapman, V. E. Torres, J. J. Grantham, M. Mrug, W. M. Bennett, M. F. Flessner, and D. P. Lansittel, "Novel methodology to evaluate renal cysts in polycystic kidney disease," *Amer. J. Nephrol.*, vol. 39, no. 3, pp. 210–217, 2014.
- [128] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 8, pp. 630–632, Oct. 1978.
- [129] M. Notohamiprodjo, M. Goepfert, S. Will, R. Lorbeer, F. Schick, W. Rathmann, P. Martirosian, A. Peters, K. Müller-Peltzer, A. Helck, S. Rospleszcz, and F. Bamberg, "Renal and renal sinus fat vol. as, quantified by magnetic resonance imaging in subjects with prediabetes, diabetes, and normal glucose tolerance," *PLoS ONE*, vol. 15, Oct. 2020, Art. no. e0216635.
- [130] H. Seuss, R. Janka, M. Prämmer, A. Cavallaro, R. Hammon, R. Theis, M. Sandmair, K. Amann, T. Bäuerle, M. Uder, and M. Hammon, "Development and evaluation of a semi-automated segmentation tool and a modified ellipsoid formula for volumetric analysis of the kidney in non-contrast T2-weighted MR images," *J. Digit. Imag.*, vol. 30, no. 2, pp. 244–254, Apr. 2017.
- [131] M. Hammon, R. Janka, C. Siegl, H. Seuss, R. Gross, P. Martirosian, R. E. Schmieder, M. Uder, and I. Kistner, "Reproducibility of kidney perfusion measurements with arterial spin labeling at 1.5 tesla MRI combined with semiautomatic segmentation for differential cortical and medullary assessment," *Medicine*, vol. 95, no. 11, p. e3083, 2016.

- [132] Y. Feng, I. Kawrakow, J. Olsen, P. J. Parikh, C. Noel, O. Wooten, D. Du, S. Mutic, and Y. Hu, "A comparative study of automatic image segmentation algorithms for target tracking in MR-IGRT," *J. Appl. Clin. Med. Phys.*, vol. 17, no. 2, pp. 441–460, Mar. 2016.
- [133] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 424–432.
- [134] J. A. O'Reilly, S. Tanpradit, T. Puttasakul, M. Sangworasil, T. Matsuura, K. Chousangsunthorn, and P. Wibulpolprasert, "Automatic segmentation of polycystic kidneys from magnetic resonance images using a three-dimensional fully-convolutional network," in *Proc. RSU Int. Rese. Conf.*, 2020, pp. 43–50.
- [135] V. Bevilacqua, A. Brunetti, G. D. Cascarano, A. Guerriero, F. Pesce, M. Moschetta, and L. Gesualdo, "A comparison between two semantic deep learning frameworks for the autosomal dominant polycystic kidney disease segmentation based on magnetic resonance images," *BMC Med. Informat. Decis. Making*, vol. 19, no. S9, pp. 1–12, Dec. 2019.
- [136] E. Hodneland, E. A. Hanson, A. Lundervold, J. Modersitzki, E. Eikefjord, and A. Z. Munthe-Kaas, "Segmentation-driven image registration-application to 4D DCE-MRI recordings of the moving kidneys," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2392–2404, May 2014.
- [137] N. Goceri and E. Goceri, "A neural network based kidney segmentation from MR images," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 1195–1198.
- [138] N. Liu, A. Soliman, G. Gimel'farb, and A. El-Baz, "Segmenting kidney DCE-MRI using 1st-order shape and 5th-order appearance priors," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), vol. 9349, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, doi: [10.1007/978-3-319-24553-9_10](https://doi.org/10.1007/978-3-319-24553-9_10).
- [139] W. Huang, H. Li, R. Wang, X. Zhang, X. Wang, and J. Zhang, "A self-supervised strategy for fully automatic segmentation of renal dynamic contrast-enhanced magnetic resonance images," *Med. Phys.*, vol. 46, no. 10, pp. 4417–4430, Oct. 2019.
- [140] A.-K. Schnurr, C. Drees, L. R. Schad, and F. G. Zöllner, "Comparing sample mining schemes for cnn kidney segmentation in t1w mri," in *Proc. 3rd Int. Conf. Funct. Renal Imag.*, Nottingham, U.K., Oct. 2019, pp. 1–5.
- [141] V. E. Torres, E. Meijer, K. T. Bae, A. B. Chapman, O. Devuyst, R. T. Gansevoort, J. J. Grantham, E. Higashihara, R. D. Perrone, H. B. Krasa, J. J. Ouyang, and F. S. Czerwiec, "Rationale and design of the TEMPO (Tolvaptan efficacy and safety in management of autosomal dominant polycystic kidney disease and its Outcomes) 3–4 study," *Amer. J. Kidney Diseases*, vol. 57, no. 5, pp. 692–699, May 2011.
- [142] O. Jimenez-del-Toro et al., "Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCELAR anatomy benchmarks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2459–2475, Nov. 2016.
- [143] I. Mendichovszky et al., "Technical recommendations for clinical translation of renal MRI: A consensus project of the cooperation in science and technology action PARENCHIMA," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 33, no. 1, pp. 131–140, Feb. 2020.
- [144] M. Wolf, A. de Boer, K. Sharma, P. Boor, T. Leiner, G. Sunder-Plassmann, E. Moser, A. Caroli, and N. P. Jerome, "Magnetic resonance imaging T1-and T2-mapping to assess renal structure and function: A systematic review and statement paper," *Nephrol. Dial. Transplantation*, vol. 33, no. 2, pp. ii41–ii50, Sep. 2018.
- [145] H. Ahn, "Consensus-based technical recommendations for clinical translation of renal phase contrast MRI," *J. Magn. Reson. Imag.*, pp. 163–176, Dec. 2020.
- [146] F. Nery et al., "Consensus-based technical recommendations for clinical translation of renal ASL MRI," *Magma*, vol. 33, pp. 141–161, Feb. 2020.
- [147] O. Bane et al., "Consensus-based technical recommendations for clinical translation of renal BOLD MRI," *Magma*, vol. 33, pp. 199–215, Feb. 2020.
- [148] H. Ahn, "Consensus-based technical recommendations for clinical translation of renal phase contrast MRI," *J. Magn. Reson. Imag.*, vol. 42, pp. 60–68, Oct. 2017.
- [149] E. Hanson, E. Eikefjord, J. Rørvik, E. Andersen, A. Lundervold, and E. Hodneland, "Workflow sensitivity of post-processing methods in renal DCE-MRI," *Magn. Reson. Imag.*, vol. 42, pp. 60–68, Oct. 2017.
- [150] N. Heller et al., "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101821.



FRANK G. ZÖLLNER (Senior Member, IEEE) was born in Bielefeld, Germany, in 1976. He received the Diploma and Ph.D. degrees (Dr.-Ing.) in computer science from the University of Bielefeld, Germany, in 2001 and 2004, respectively, the Ph.D. degree from the Bioinformatics Graduate Program (Graduiertenkolleg Bioinformatik), in 2004, and the Venia Legendi degree in medical physics from Heidelberg University, Germany, in 2014.

He joined the Applied Computer Science Group, in 2001. From 2004 to 2006, he worked as a Postdoctoral Researcher with the Applied Computer Science Group. From 2006 to 2007, he was a Researcher with the Section of Radiology, Institute for Surgical Sciences, Haukeland University Hospital, and the Neuroinformatics and Image Analysis Group, Section for Physiology, Department of Biomedicine, University of Bergen, Norway. He joined the Chair of Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, in 2008. Since 2014, he has been the Vice Chair of the Chair of Computer Assisted Clinical Medicine. He became an Adjunct Professor of medical physics with Heidelberg University, in 2017. He was appointed as the Head of the Preclinical MRI Corefacility, Medical Faculty Mannheim, Heidelberg University, in 2021. His research interests include pattern recognition, image processing, and imaging techniques. He is interested in applying computational methods from pattern recognition, image analysis to the fields of molecular imaging and medical image analysis.

Dr. Zöllner is a member of the IEEE EMB Society and the International Society for Magnetic Resonance in Medicine (ISMRM).



MAREK KOĆIŃSKI was born in Łódź, Poland, in 1977. He received the M.Sc. degree in electronics and telecommunications and the Ph.D. degree in computer science from the Lodz University of Technology (TUL), Poland, in 2003 and 2009, respectively.

Since 2009, he has been an Assistant Professor with the Institute of Electronics, TUL. He is currently employed as a Postdoctoral Researcher with the Mohn Medical Imaging and Visualization Centre (MMIV), Department of Radiology, Haukeland University Hospital, Bergen, Norway, and the Department of Biomedicine, University of Bergen, Norway. His research interests include biomedical image processing, 3D visualization, 2D/3D texture analysis, quantitative analysis and modeling of DCE MRI images, and computer modeling. He is interested in application of machine learning and deep learning in medical image analysis.



LAURA HANSEN was born in Datteln, Germany, in 1996. She has been studying medicine, since 2015, and passed the Second State Examination Test in medicine from the Medical Faculty Mannheim, Heidelberg University, Baden-Württemberg, Germany, in 2020.

From 2018 to 2020, she worked as a Surgical Assistant with the Department of Orthopedic and Trauma Surgery, University Medical Center Mannheim, Mannheim Medical Faculty, Heidelberg University. Her current research interest includes volume determination of kidneys with autosomal dominant polycystic kidney disease (ADPKD) using AI techniques on the basis of MRI images.



ALENA-KATHRIN GOLLA was born in Hamburg, Germany, in 1991. She received the B.S. and M.S. degrees in computational visualistics from Otto von Guericke University Magdeburg, Germany, in 2017.

She joined the Chair of Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, Germany, as a Research Assistant, in 2017. Her research interests include image processing and deep learning. Her work is focused on medical image segmentation, convolutional neural networks, and learning from limited data.



ANDRZEJ MATERKA (Life Senior Member, IEEE) was born in Łęczyca, Poland, in 1949. He received the M.Sc. degree in radio engineering from the Warsaw University of Technology, in 1972, the Ph.D. (Dr.Eng.) degree from the Technical University of Łódź (TUL), Poland, in 1979, and the D.Sc. (Habilitation) degree in electronics from the Technical University of Wrocław, in 1986.

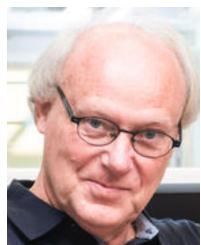
From 1972 to 1974, he was an Engineer with the Radio and TV Broadcasting Stations, Łódź. Since 1974, he has been with the Institute of Electronics, TUL, being its Director, from 1995 to 2015. From 1980 to 1982, he was a Monbusho Scholar with the Research Institute of Electronics, Shizuoka University, Hamamatsu, Japan. From 1992 to 1994, he was a Senior Lecturer with the Department of Electrical and Computer Systems Engineering, Monash University, Melbourne (Caulfield), VIC, Australia. For two terms, from 2002 to 2008, he was the Dean of the Faculty of Electrical and Electronic Engineering, TUL, where he is currently a Professor. He supervised 21 Ph.D. candidates. He published over 250 scientific articles and six books/monographs. His research interests include microwave circuit design, medical electronics, digital signal and image analysis, pattern recognition, artificial neural networks, secure database information systems design, electronic smart cards applications, human-computer interaction including brain-computer interfaces and aids for the visually impaired, as well as geometric modeling of blood vessel trees based on 3D images.

Prof. Materka was the Chairman of Software and Statistics international working group in COST B11 European Program Quantitative Texture Analysis on Magnetic Resonance Images from 1998 to 2003, and the Deputy Chairman of COST B21 Physiological Modelling of MR Image Formation action from 2004 to 2008. In result, a popular software package MaZda for quantitative medical image texture analysis has been developed under his guidance. He was a Co-Founder of the European Campus Card Association, where he served as the Vice-President, from 2004 to 2006, and the President, from 2009 to 2011. From 2017 to 2020, he was elected a member of the Polish State Committee for Scientific Titles and Degrees. He received the title of Professor of technical sciences (electronics, computer engineering) from the President of Poland, in 1996.



AMIRA ŠERIFOVIĆ TRBALIĆ was born in Tuzla, Bosnia and Herzegovina, in 1978. She received the Diploma, M.Sc., and Ph.D. degrees from the Faculty of Electrical Engineering, University of Tuzla, Bosnia and Herzegovina, in 2002, 2006, and 2011, respectively.

She is currently appointed as an Associate Professor with the Faculty of Electrical Engineering, University of Tuzla. Her research interests include image processing and analysis, pattern recognition, and intelligent systems. She is interested in medical image processing and analysis.



ARVID LUNDERVOLD (Life Member, IEEE) received the B.Sc. degree in mathematics and the M.D. degree from the University of Oslo, Norway, in 1982, and the Ph.D. degree in multispectral MR image analysis from the University of Bergen, Norway. He is currently a Professor of medical information technology and physiology with the Department of Biomedicine, University of Bergen, and the Head of the Neuroinformatics and Image Analysis Laboratory, Neural Networks and Microcircuits Research Group. He is also the Co-Leader of the Mohn Medical Imaging and Visualization Centre, Machine Learning Group, Department of Radiology, Haukeland University Hospital, Bergen, and an Adjunct Professor II of computational radiography with the Department of Health and Functioning, Western Norway University of Applied Sciences. His research interests include computational medicine, image processing, pattern recognition and machine learning, functional MR imaging, image registration, quantification and visualization, and mathematical modeling. He is a member of the IEEE Computer Society, the Norwegian Medical Association, and the American Mathematical Society.



PETER ROGELJ was born in Ljubljana, Slovenia, in 1975. He received the Ph.D. degree in electrical engineering from the University of Ljubljana, Slovenia, in 2003. He worked in the industry as a software developer, project manager, and product manager in companies in the fields of telecommunications and software engineering. He was also involved in research projects related to processing and analyzing medical image data. In 2010, he employed at the University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies (UP FAMNIT), as an Assistant Professor. From 2015 to 2020, he was a coordinator of the B.S. study Programme Bioinformatics at UP FAMNIT. His research interests include image and signal processing and analysis, computer vision, and intelligent systems. His current research activities remain in the fields of medical image processing and EEG data analysis. In 2012, he won first prize at a Business Plan Competition for an open information system for medical collaboration.