Dynamic parametric MRI and deep learning: Unveiling renal pathophysiology through accurate kidney size quantification

Tobias Klein1,2 | Thomas Gładysz1 | Jason M. Millward1 | Kathleen Cantow3 | Luis Hummel3 | Erdmann Seeliger3 | Sonia Waiczies1 | Christoph Lippert2,4 | Thoralf Niendorf1,5

1Berlin Ultrahigh Field Facility (B.U.F.F.), Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
2Digital Health - Machine Learning Research Group, Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany
3Institute of Translational Physiology, Charité – Universitätsmedizin, Berlin, Germany
4Hasso Plattner Institute for Digital Health, Icahn School of Medicine at Mount Sinai, New York City, New York, USA
5Experimental and Clinical Research Center, a joint cooperation between the Charité Medical Faculty and the Max Delbrück Center for Molecular Medicine, Berlin, Germany

Correspondence
Thoralf Niendorf, Berlin Ultrahigh Field Facility (B.U.F.F.), Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Robert Rösse Strasse 10, 13125 Berlin, Germany. Email: thoralf.niendorf@mdc-berlin.de

Funding information
This study was funded in part (T.K., T.N.) by the iNAMES-MDC-Weizmann Helmholtz International Research School (HIRS) for Imaging and Data Science from the Nano to the MESo.

Abstract
Renal pathologies often manifest as alterations in kidney size, providing a valuable avenue for employing dynamic parametric MRI as a means to derive kidney size measurements for the diagnosis, treatment, and monitoring of renal disease. Furthermore, this approach holds significant potential in supporting MRI data-driven preclinical investigations into the intricate mechanisms underlying renal pathophysiology. The integration of deep learning algorithms is crucial in achieving rapid and precise segmentation of the kidney from temporally resolved parametric MRI, facilitating the use of kidney size as a meaningful (pre)clinical biomarker for renal disease. To explore this potential, we employed dynamic parametric T2 mapping of the kidney in rats in conjunction with a custom-tailored deep dilated U-Net (DDU-Net) architecture. The architecture was trained, validated, and tested on manually segmented ground truth kidney data, with benchmarking against an analytical segmentation model and a self-configuring no new U-Net. Subsequently, we applied our approach to in vivo longitudinal MRI data, incorporating interventions that emulate clinically relevant scenarios in rats. Our approach achieved high performance metrics, including a Dice coefficient of 0.98, coefficient of determination of 0.92, and a mean absolute percentage error of 1.1% compared with ground truth. The DDU-Net enabled automated and accurate quantification of acute changes in kidney size, such as aortic occlusion (−8% ± 1%), venous occlusion (5% ± 1%), furosemide administration (2% ± 1%), hypoxemia (−2% ± 1%), and contrast agent-induced acute kidney injury (11% ± 1%). This approach can potentially be instrumental for the development of dynamic parametric MRI-based tools for kidney disorders, offering unparalleled insights into renal pathophysiology.

Keywords
deep learning, kidney, kidney size, MRI, parametric mapping, segmentation

Abbreviations used: ABSM, automated bean-shape model; ADPKD, autosomal-dominant polycystic kidney disease; AKI, acute kidney injury; CKD, chronic kidney disease; CM, contrast medium; CO, cortex; CPU, central processing unit; CRF, contrast reduction factor; DDU-Net, deep dilated U-Net; deoxyHb, deoxygenated hemoglobin; DL, deep learning; EMA, European Medicines Agency; FDA, US Food and Drug Administration; GFR, glomerular filtration rate; GPU, graphics processing unit; i.v., intravenous; IM, inner medulla; KS, kidney size; LAGeSo, Landesamt für Gesundheit und Soziales; MAPE, mean absolute percentage error; ML, machine learning; nU-Net, no new U-Net; OM, outer medulla; PKD, polycystic kidney disease; pO2, partial oxygen pressure; R, Pearson correlation coefficient; R2, coefficient of determination; ReLU, rectified linear unit; USPIO, ultrasmall superparamagnetic iron oxide; ΔKS, change in kidney size.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. NMR in Biomedicine published by John Wiley & Sons Ltd.
With ever-increasing incidence, kidney diseases affect approximately 10% of the worldwide population, constituting a major public health burden. Despite the clear need, current options for the prevention and therapy of kidney disorders remain inadequate. A major reason for this is that the clinically available diagnostic and prognostic markers—such as serum creatinine-based metrics—only reveal kidney abnormalities at a late stage, and fail to detect early events in the pathological process. Several alternative blood- or urine-based diagnostic markers reflecting renal injury have been proposed. The modest performance of these markers has not yet delivered on their promise, and none has advanced to the point of use in clinical practice.

This major health issue requires noninvasive imaging to complement conventional approaches for the assessment of kidney diseases. Meeting this urgent need calls for synergistic approaches that include MRI. MRI allows for noninvasive imaging of kidney structure with whole organ coverage, and assessment of functional kidney features, with high spatial resolution. Many functional MRI methods have emerged that are sensitive to pathophysiological changes associated with renal hemodynamics, oxygenation, fibrosis, inflammation, and microstructure. MRI has the potential to provide quantitative imaging markers to detect various stages of pathophysiology, improve prediction and interception of disease progression, and evaluate treatment of renal disease.

Several renal pathologies are associated with changes in kidney size (KS), thus offering an opportunity for MRI-derived KS to be used as a parameter for the assessment of renal disease. In polycystic kidney disease (PKD), KS correlates with disease progression and the decline in glomerular filtration rate (GFR). KS has been approved as a prognostic marker for use in clinical trials of new therapies for autosomal-dominant PKD. Detecting KS reduction because of parenchymal atrophy, sclerosis, and fibrosis has been recognized as a marker to identify chronic kidney disease (CKD) and to determine its severity. Longitudinal MR-based KS monitoring is proposed as a key measure for several kidney disorders, including hyperfiltration in early diabetic nephropathy, renal transplants, renal artery stenosis, and vesicoureteral reflux.

Serial MRI in experimental models has demonstrated that KS changes can indicate pathophysiological developments. These include models of clinical conditions such as acute ureteral obstructions because of urolithiasis or during upper urinary tract endourologic procedures, administration of X-ray contrast media (CM) for cardiac procedures, obstructions of the renal vein during partial nephrectomy or because of renal cell carcinoma-derived thrombus formation, and clamping of the suprarenal aorta or renal artery during surgery, or the low arterial target pressure during cardiopulmonary bypass. MR-based KS assessment has also been used in experimental diabetes, mutant models mimicking PKD, and renal allografts. A recent preclinical study demonstrated that monitoring KS allows for physiological interpretation of MRI-based oxygenation changes in acute pathophysiologically relevant scenarios.

In vivo KS assessment requires a robust and fast segmentation of the kidney from MR images. Manual segmentation is time-consuming and prone to observer bias. This poses a major impediment for dynamic and longitudinal studies, and severely limits the potential for KS assessment in translational research. These constraints can be offset by analytical models using pre-established geometrical shapes like ellipsoids or hybrid level-set methods. These approaches are semiautomated and provide up to 70-fold improvement in segmentation speed compared with manual segmentation. Nevertheless, analytical model-based renal segmentation requires an expert observer to manually preselect measurements and/or landmarks to initialize the models.

Machine learning (ML)-based renal segmentation has gained momentum as a tool for decoding the links between KS as a mesoscopic marker, and data from histopathological, physiological, and functional measurements, with the goal of deciphering determinants and associations of renal disease. ML approaches provide a viable solution for deforming the kidney shape using a constrained statistical model-based algorithm trained upon a dataset. This approach requires minimal user interaction. Fully automated supervised and unsupervised ML algorithms have been explored for renal segmentation from MRI using convolutional neuronal network models. MRI studies using neural networks for renal segmentation reported processing times as good as 1–10 s per subject. Most of these deep learning (DL) approaches take advantage of U-Net variants. Recent application of DL-based kidney segmentation has focused on automation of renal cyst and kidney volume measurements in healthy subjects and patients with autosomal-dominant PKD and CKD. The feasibility and reliability of dynamic or longitudinal MRI-based KS monitoring using DL in acute pathophysiologically relevant scenarios, where changes may be more subtle than in autosomal-dominant PKD or CKD, has not yet been investigated.

Recognizing this opportunity, the current study examines the feasibility and reliability of dynamic parametric MRI-based automated KS assessment in rats using DL. Addressing this challenge, a custom-tailored deep dilated U-Net (DDU-Net) was developed and validated to facilitate quantification of acute changes in KS (AKS) in pathophysiologically relevant experimental setups mimicking realistic clinical scenarios. We demonstrate that our approach has the potential for establishing MRI-based diagnostic tools for various kidney disorders, and for gaining new insights into the mechanisms of renal pathophysiology.
2 | EXPERIMENTAL

2.1 | Animal preparation
Investigations were approved by the Landesamt für Gesundheit und Soziales (LAGeSo) of Berlin in accordance with German Animal Protection Law and EU Directive 2010/63/EU. Male Wistar rats (n = 52, aged 12–13 weeks, 270–300 g, Harlan-Winkelmann, Borchen, Germany) were anesthetized with urethane (1.2 g/kg; Sigma-Aldrich, Steinheim, Germany) during MRI examinations. Fourteen animals were surgically implanted with MR-safe remote-controlled inflatable occluders, applied around the suprarenal aorta and the left renal vein, as previously described.71–73 During MR scanning, respiration and body temperature were monitored continuously.

2.2 | MRI experiments
Experiments were performed on a 9.4-T animal MR system (Bruker Biospec 94/20; Bruker Biospin, Ettlingen, Germany). A linear birdcage radio-frequency (RF) volume resonator (inner diameter = 72 mm; Bruker Biospin) was employed for transmission, and a four-channel surface RF coil array customized for rats was used for signal reception.

Following T2-weighted pilot scans, selective shimming of magnetic field homogeneity on a volume of interest accommodating the left kidney was conducted. T2 mapping was performed with a respiratory-gated (Model 1025; SA Instruments, New York, NY, USA) multipin-echo technique: TR = 500 ms, number of echoes = 13, first TE = 6.4 ms, interecho time ΔTE = 6.4 ms, number of averages = 1, and t_echo = 58 s. A midcoronal oblique slice was acquired: in-plane spatial resolution = 226 × 445 μm², field of view = 38.2 × 50.3 mm², matrix size = 169 × 113 (zero-filled to 169 × 215), and slice thickness = 1.4 mm. Parametric maps of absolute T2 were calculated by pixel-wise monoeXponential fitting to the signal intensities of the T2-weighted images acquired at different echo times (in-house developed program; MATLAB, R2010a, MathWorks, Natick, MA, USA). T2 maps were clipped at T2 = 0 and 200 ms and rescaled to [0, 1].

The MR relaxation parameter T2 was monitored to examine renal blood oxygenation as a surrogate of renal tissue oxygenation. These data were also used for dynamic assessment of the renal size. Dynamic T2 mapping of the central coronal slice of the left kidney was performed in 43 rats under baseline conditions (without any intervention, at three subsequent time points), and in 30 of the 43 rats during or following the respective interventions and recovery periods, where applicable, according to the specific subgroup.

In addition, dynamic T2 mapping was performed in another nine rats before and after bolus injection of the X-ray CM iodixanol into the thoracic aorta.44 T2 maps were obtained from an analogous MSME protocol (TR = 550 ms, number of echoes = 7, first TE = 10 ms, and interecho time ΔTE = 10 ms).

2.3 | Longitudinal quantification of changes in KS upon pathophysiological interventions
Six pathophysiologically relevant interventions that alter renal tissue oxygenation were performed.71,72

Rats with vascular occluders (n = 14) underwent serial T2 mapping prior to, during, and following occlusion of the aorta (n = 12 successful). The occlusion was applied for 3.8 ± 0.3 min, then deflated, and the rats were allowed to recover for 7 min to ensure restoration of preocclusion hemodynamics and oxygenation. Time-of-flight MR angiography was performed immediately after inflation/deflation of the occluder to confirm occlusion/reperfusion of the vessels.71–73 After recovery, the same procedure was applied for renal venous occlusion (n = 12 successful) and for combined aortic-venous occlusion (n = 8 successful). Occlusion of the aorta results in decreased renal blood volume, occlusion of the renal vein induces an increase in renal blood volume, and simultaneous occlusion of both vessels does not affect renal blood volume.

A separate group of five rats, received an intravenous (i.v.) bolus of 5 mg/kg furosemide (Ratiopharm, Ulm, Germany) followed by a 0.2-mL saline chaser.44

In a separate group, 11 rats underwent T2 mapping during a control period of normoxia (inspiratory oxygen fraction [FiO2] of 21%) during 3.8 ± 0.1 min of hypoxemia (FiO2 = 10%), followed by 10 min of recovery at normoxia.

A final group of nine rats received an i.v. bolus of 1.5 mL of the X-ray CM iodixanol (320 mg/mL iodine, Visipaque; GE Healthcare Buchler, Braunschweig, Germany), followed by a 0.2-mL saline chaser. Administration of CM and furosemide are expected to have longer term effects, increasing the renal tubular volume fraction.

For the occlusion of the suprarenal aorta, occlusion of the renal vein, and the bolus administration of furosemide, interventions were repeated following i.v. administration of ultrasmall superparamagnetic iron oxide (USPIO); 2 mg Fe/kg, ferumoxytol (Feraheme; AMAG Pharmaceuticals, Lexington, USA).73 Ferumoxytol reduces the signal intensity and image contrast due to T2 shortening,73 but does not affect renal physiology at this dose.74
2.4 | Manual segmentation of the kidney by independent observers

Accuracy and precision of DDU-Net was evaluated versus the ground truth KS under baseline physiological conditions, based on manual segmentation of T2 maps by a reader with 1 year of experience in renal MRI analyses in rats. Following the manual segmentation procedure, a consensus segmentation involving the five expert observers was conducted to determine the ground truth. The manual segmentation prepared by the first observer was set as a starting point for the consensus segmentation. Following the manual segmentation procedure, a consensus involving the five expert observers was conducted to determine the ground truth. The manual segmentation prepared by the first observer was set as a starting point for the consensus segmentation. Subsequently, each of the 129 T2 maps was presented to all observers in ITK-SNAP together with a transparent overlay of the area determined as belonging to the kidney. Training was performed using two T2 maps as examples: one T2 map with very good image quality (classified as acceptable), and one T2 map with lower image quality (classified as marginal). From these training data, a consensus was reached regarding the inclusion/exclusion of voxels with apparent partial volume effects. For the remaining consensus reading, renal segmentation T2 maps were used. During the consensus reading, the observers remained blinded to the individual T2 map number, the animal identification, the numerical KS derived from the independent renal segmentation, and the number of voxels of the overlay. Upon presentation of each T2 map along with the corresponding transparent overlay, all observers agreed in real-time if it was necessary to add or delete (a) certain voxel(s) to/from the overlay, to improve the accuracy of the renal boundaries, whereupon a consensus on the total area assigned as renal tissue was reached. Changes were adopted in real-time using ITK-SNAP. A time limit of 90 s was set for the consensus renal segmentation of each T2 map. The ground truth comprised consensus readings of manual segmentations from a total of 129 T2 maps (n = 43 rats, with three baseline scans per rat).

2.5 | Data handling and data simulation

The 52 rats used in the study were split into five subgroups (Figure 1). In the first subgroup (n = 13), the training set, three baseline T2 maps were obtained per rat, and used for training and validation of DDU-Net. Another cohort (n = 30), the test set, was divided into three subgroups according to the interventions applied, as described above. For these animals, three baseline T2 maps were acquired prior to the intervention, followed by serial T2 maps obtained during the intervention and recovery period; the baseline T2 maps were used for training, validation, and testing of the neural network against the ground truth manual segmentation. The serial T2 maps regarding the intervention were used for the application of the network. For the remaining cohort (n = 9), the longitudinal data following administration of the X-ray CM were included. For the definition of the ground truth, all three baseline T2 maps of the 13 + 30 = 43 animals were anonymized. For these 43 rats, the longitudinal data following administration of the X-ray CM were included. For the definition of the ground truth, all three baseline T2 maps of the 13 + 30 = 43 animals were anonymized. For these 43 rats, the longitudinal data following administration of the X-ray CM were included.

Based on the ground truth, 12 rats of the test set with baseline T2 maps showing the highest signal-to-noise ratio (SNR) were selected to simulate T2 maps emulating the interventions. The SNR was defined as

$$\text{SNR} = \frac{0.9535}{\text{median(abs(diff(SROI)))}}.$$  

FIGURE 1 Overview of the data handling: the total number of 52 rats were split into five distinct subgroups (blue) as shown in the figure. While the baseline subgroup only was used during the training and the cross-validation (yellow) and the subgroup of the administration of the X-ray contrast media (CM) only during the application on the in vivo longitudinal MRI data (green), the remaining subgroups were used during all steps of the study, including the testing (orange) and the evaluation (red).
with \( \text{abs} \) referring to the absolute value function, \( \text{diff} \) calculating the differences between adjacent signal intensities, and \( S_{\text{ROI}} \) the signal intensities inside the region of interest (ROI).

A series of five copies randomly chosen from the three baseline T2 maps for each subject was created for all combinations of 10 different changes in KS ranging from \(-5\% \) to \(5\% \), 10 SNR levels of the T2 maps ranging from 2.5 to 50, and four tissue contrast reduction factors (CRFs) ranging from 1 to 0.25.

To generate different SNR levels, a nonlocal means filter was applied to ensure that the intended SNR was higher than the original SNR. For lower SNR, the image was manipulated as follows:

\[
S_{\text{simulated}} = \exp \left( \log(S_{\text{original}}) + X_{\text{normal}} \frac{0.9535}{\sqrt{\frac{\text{SNR}_{\text{simulated}}}{\text{SNR}_{\text{original}}}}} \right),
\]

Here, \( S_{\text{simulated}} \) and \( S_{\text{original}} \) refer to the simulated and original signal intensity, and \( X \) is a random variable with the standard normal distribution.

The contrast reduction was performed by applying the following formula:

\[
S_{\text{simulated}} = \text{CRF} \cdot (S_{\text{original}} - \mu_S) + \mu_S.
\]

Here, \( \mu_S \) denotes the mean of the original signal intensities.

These simulations resulted in 400 series of five simulated images for each of the five randomly chosen baseline maps, for each of the 12 rats: one baseline map, three maps right after an intervention exponentially recovering back to the baseline, and a final baseline map. This set of \( 5 \times 12 \times 400 = 24,000 \) synthetically generated T2 maps was used for benchmarking DDU-Net against the ABSM and nnU-Net.

For automatic DDU-Net based quantification of KS changes in in vivo longitudinal MRI data, the serial T2 maps of the rats that underwent the interventions were used \((n = 30 + 9)\).

### 2.6 Architecture of DDU-Net

Our DDU-Net incorporates the general structure of the original U-Net and consists of an encoder and a decoder part with four resolution levels each. While the decoder part remains the same, the two original convolutional layers in each level of the encoder are replaced by a sequence of...
four convolutional layers with growing dilation and residual connections. This results in an increased receptive field and an asymmetric network architecture with a larger encoding part.\textsuperscript{62,65}

A schematic representation of the network architecture is shown in Figure 3A. The numbers of filters are 16, 32, 64, and 128, and 256 in the bottleneck. Each level of the encoding part consists of two $3 \times 3$ convolutional layers without dilations, two $3 \times 3$ convolutional layers with dilations of 2 and 4, respectively, and a final downsampling $2 \times 2$ convolutional layer with stride of 2. All convolutional layers are followed by a leaky rectified linear unit (Leaky ReLU) with a slope of 0.1 and a batch normalization. The outputs of the first four convolutional layers are concatenated and function as the input of the final downsampling layer. The composition of an encoding block is illustrated in Figure 3B. The bottleneck of the network consists of two $3 \times 3$ convolutional layers (followed by Leaky ReLU and batch normalization) without dilations. For the decoding part, the original U-Net structure is maintained: each level starts with an upsampling $2 \times 2$ transposed convolutional layer with stride of 2, followed by two $3 \times 3$ convolutional layers without dilations. As before, all convolutional layers succeed a Leaky ReLU and a batch normalization. In each level, the concatenated layer outputs of the corresponding encoding stage are concatenated to the output of the transposed convolutional layer.

![Figure 3](https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/nbm.5075)

**FIGURE 3** Schematic representation of the deep dilated U-Net (DDU-Net) and its encoding blocks: (A) The general structure of the DDU-Net is illustrated with its encoding blocks (orange), decoding blocks (blue), and the bottleneck (gray). The red arrows represent the downsampling, the green arrows the upsampling. Violet arrows show the skip connection between encoding and decoding blocks. Numbers on the left represent the numbers of channels of each level of encoding/decoding. (B) A schematic illustration of the encoding blocks is shown with two regular $3 \times 3$ convolutional (conv) layers without dilation (blue), followed by two dilated conv layers (violet) with dilations of 2 and 4. The concatenation of the output of those four layers is shown in orange. The $2 \times 2$ downsampling conv layer is represented in red. Because of the extra depth and dilations, the receptive field size created by one encoding block (including downsampling) is 16 compared with 6 by the original encoding block of the U-Net. This corresponds to an expansion of more than 2.5 times.
and serve as the input of the first convolutional layer. The final layer of the decoder is a $1 \times 1$ transposed convolutional layer with a sigmoid activation function.

Because the network design presumes an input shape of $192 \times 224$ because of architectural reasons, the input was rescaled to $192 \times 224$ and the output rescaled back to $169 \times 215$.

### 2.7 Training, data augmentation, and postprocessing

DDU-Net was pretrained using the kidney labels and MRI data from the CHAOS dataset. For the actual training, the subjects of the test set ($n = 30$) were randomly split into three distinct test subsets of size $n_{\text{test}} = 10$. For each subset, a 5-fold cross-validation of DDU-Net was performed based on the two remaining subsets ($n = 20$) and the subgroup that was only used for training and validation ($n = 13$). This resulted in training sets of sizes $n_{\text{train}} = 26$ and $n_{\text{train}} = 27$ and validation sets of sizes $n_{\text{validation}} = 7$ and $n_{\text{validation}} = 6$. To achieve a more robust prediction and a lower generalization error, the five models of each cross-validation were combined into a 5-fold DDU-Net ensemble: the output logits before the final sigmoid activation were averaged, passed through the sigmoid activation function and rounded to zero and one, resulting in binary segmentation masks. For subjects not included in the training or validation (rats administered X-ray CM), networks of all cross-validations were combined into a 15-fold ensemble. Each single network was trained with a stochastic gradient descent optimizer with an initial learning rate of $10^{-2}$ and a momentum of 0.8, and a loss function combining the binary focal loss and the Dice loss. The learning rate was reduced by a factor of 0.5 to a minimal learning rate of $10^{-8}$ in case the validation loss did not improve for 50 epochs. Training was performed for a maximum number of 700 epochs, switching to stochastic weight averaging after 500 epochs to improve the generalization of the model. The schedule used for stochastic weight averaging was the cyclic cosine annealing from the initial learning rate to $10^{-8}$ with a duration of 25 epochs.

Each epoch, every training map underwent a data-augmentation procedure including random affine transformations ($x$- and $y$-zoom, $x$- and $y$-shear, rotation, shift), elastic transformation, randomized manipulations of the signal intensities (brightness, contrast, and gamma adjustment), and a random Gaussian noise.

After prediction by the network, the output masks were cleaned automatically and only the largest segmentation region was kept. Smaller distinct regions, often representing parts of the other kidney, were removed from the mask.

Preprocessing, network design, training, postprocessing, and evaluation were programmed in Python (PyTorch 1.13).

### 2.8 Validation of DDU-Net

For validation, DDU-Net was benchmarked against nnU-Net and the ABSM. To benchmark DDU-Net against a state-of-the-art DL approach, we trained nnU-Net based on the exact same distinct testing subsets. During the training, nnU-Net performs a 5-fold cross-validation by default for several network configurations and chooses the best model automatically. For the cross-validation, the exact same splits were used as for DDU-Net. The resulting trained nnU-Net was applied on the baseline test set, the simulated dataset, and the intervention data. As for DDU-Net, all predicted masks were cleaned as described above. For the baseline maps of the test sets, the Dice scores between the manual segmentations and the predicted masks by DDU-Net and nnU-Net were calculated. Furthermore, the $R^2$ of the renal area obtained from manual segmentations and DDU-Net/nnU-Net was determined and linear regression was performed for all three baseline time points. To meet the assumption of normality, two rats with unusually small KS of less than 180 mm² were identified as outliers and excluded from this analysis. Relative residuals were plotted and the mean absolute percentage errors (MAPEs) were determined.

To further benchmark DDU-Net, we used an established analytical ABSM. Although the manual segmentation and both DL-based predictions include the papilla while excluding the incoming and outgoing vessels, the ABSM is a geometrical model of a skewed ellipse with a straight line (border) drawn that connects the outermost points of the upper and lower medial curvatures of the kidney (Figure 2D). Therefore, the three methods can only be compared by evaluating their prediction of relative changes in KS.

For the validation of DDU-Net, synthetic data were generated as described in section 2.5. KS was determined by all three methods and the relative size of the peak of the intervention was compared with the calculated baseline. By considering only the relative KS changes upon an intervention, all three methods could be compared, regardless of the differing segmentation policies of the DL approaches and the ABSM. Based on this relative KS, DDU-Net, nnU-Net, and ABSM were compared by calculating the respective MAPEs against the ground truth change.

### 2.9 Statistical analysis

For the in vivo study of the intervention-induced changes in ΔKS, KS assessed in the $T_2$ mapping before the respective intervention was used as the control KS and ΔKS was calculated for the subsequent $T_2$ maps. $T_2$ values for the renal cortex (CO), outer medulla (OM), and inner medulla
(IM) were determined using baseline T2 maps before the respective interventions and the T2 map during or right after the respective intervention. The placement of ROIs was conducted using a standardized semiautomatic approach, following a previously described method. This method involved positioning the ROIs (five for the CO and OM, and three for the IM) in a way that excludes the transitional regions between renal layers, ensuring that partial volume effects are avoided. Because of the T2 shortage induced by ferumoxytol, T2 values were not analyzed after the administration of USPIO.

Relative changes in KS over time and changes in T2 were analyzed using the nonparametric repeated-measures Friedman test, followed by Dunn's post-hoc test with the Benjamini–Hochberg correction for multiple comparisons. p values less than 0.05 were considered to be statistically significant. All statistical analysis was performed in R v. 3.6.3 using the libraries “rstatix” and “dunn.test”.

3 | RESULTS

3.1 | DDU-Net evaluation on the ground truth obtained from manual segmentation of the kidney

For the baseline T2 maps of the test set, the mean KS predicted by DDU-Net was similar to the mean KS derived from the ground truth manual segmentation: 211 ± 1 mm² (mean ± standard error of the mean [SEM]). Figure 4A–C depict the linear regression plots for the KS predicted by DDU-Net versus the ground truth for each of the three baseline T2 maps. A mean coefficient of determination ($R^2$) = 0.92, a mean Pearson correlation coefficient ($r$) = 0.96, and an average MAPE = 1.1% were determined over the three baseline T2 maps. Figure 4D shows the distribution of relative residuals between both methods. DDU-Net yielded a median intrasubject variability of 0.3% compared with 0.8% for the ground truth and 0.6% for the previously published ABSTM. For the segmentations, DDU-Net reached an accuracy of more than 0.99, a recall of 0.98, a
precision of 0.98, a mean intersection-over-union of 0.96, and a Dice score of 0.98. Renal T2 mapping under baseline conditions revealed a median T2 of 47 ± 1 ms in the CO, 46 ± 2 ms in the OM, and 66 ± 5 ms in the IM.

On the ground truth baseline data, we further benchmarked DDU-Net against the self-configuring state-of-the-art nnU-Net. nnU-Net–predicted area achieved a mean $R^2 = 0.91$, a mean $r = 0.95$, an averaged MAPE = 1.2%, and a median intrasubject variability of 0.3%. For the predicted segmentations, nnU-Net reached an accuracy of more than 0.99, a recall of 0.98, a precision of 0.98, a mean intersection-over-union of 0.96, and a Dice score of 0.98.

### 3.2 Validation of DDU-Net against the ABSM and nnU-Net using simulated time series data with changes in renal size

Fitting the 24,000 simulated T2 maps took less than 4 min on a graphics processing unit (GPU) and less than 30 min on a central processing unit (CPU) for DDU-Net and nnU-Net, compared with approximately 100 h for the ABSM. This translates to a speed gain of ~1500 for the DL-based approaches versus the geometric model. Figure 5 depicts the MAPEs of the relative KS for all combinations of SNRs, CRFs, and changes in relative KS for all three methods. In 292 of the 400 scenarios (73%), DDU-Net outperformed the ABSM in terms of MAPE (one-sided binomial $p < 10^{-20}$). While failing for the lowest (nonrealistic) SNR and lowest contrast compared with the ABSM (Figure 5), for SNRs of more than 2.5 and CRFs of more than 0.25, DDU-Net was superior in 240 of the 270 scenarios (89%) ($p < 10^{-2}$), and achieved a median MAPE of 1.0% compared with 1.5% for the ABSM. DDU-Net was superior to nnU-Net in 205 of 400 scenarios (51%) ($p < 10^{-4}$). For realistic SNRs of more than 2.5 and CRFs of more than 0.25, DDU-Net outperformed nnU-Net in 155 of the 270 scenarios (57%) ($p < 10^{-2}$). The median MAPE achieved by nnU-Net was 1.1%, slightly higher than DDU-Net.

### 3.3 In vivo quantification of changes in renal size upon pathophysiologically relevant interventions

The various interventions the animals underwent altered the signal intensity and contrast of the respective T2 maps. KS quantified by the DDU-Net and by the nnU-Net showed similar longitudinal changes in renal size for the vascular occlusions, for the furosemide application and for...
hypoxemia. However, nnU-net KS assessment demonstrated higher SEM. For the low SNR data obtained upon administration of the X-ray CM, nnU-Net could not confirm the trend shown by DDU-Net. A closer examination of the nnU-Net segmentations revealed mis-segmented areas. Examples of suboptimal performance or failure of nnU-Net segmentation on the in vivo MRI data are shown in Figure 6. Examples of mis-segmentation include shifts of the segmented regions, extension of the boundary into adjacent tissue and organs, and inclusion of both kidneys or segmentation of the contralateral kidney by DDU-Net. Because of this limitation, the delta KS graphs obtained from nnU-Net are erroneous.

**FIGURE 6** Exemplary T2 maps and the corresponding prediction by deep dilated U-Net (DDU-Net) and no new U-Net (nnU-Net). The T2 maps illustrate cases in which nnU-Net failed to correctly segment the kidney, while DDU-Net performed effectively. (A) The whole mask predicted by nnU-Net was shifted and expanded. (B) An exemplary T2 map including both kidneys. While both networks were trained on the central slice of the left kidney, nnU-Net incorrectly segmented the right kidney even although the texture differs from the central slices. (C) The segmentation of nnU-Net extended irregularly far beyond the pelvic region, and in (D) The prediction was shifted and expanded in the pelvic region. These kinds of failure were not observed for DDU-Net.

**FIGURE 7** Relative kidney size (KS) over time for the occlusion of the suprarenal aorta. (A) Exemplary T2 maps are shown before (left) and during (center) the intervention, as well as after recovery (right). (B) Development of relative KS changes upon the occlusion of the suprarenal aorta before the administration of ultrasmall superparamagnetic iron oxide (USPIO). The blue line depicts the mean of each time point and the corresponding interval of the standard error of the mean (SEM) (n = 12 rats). The mean KS at baseline was 211 ± 4 mm². The duration of the occlusion is highlighted in green. (C) Analogous exemplary T2 maps of the same rat. (D) Time course of relative KS changes after administration of USPIO with a mean KS at baseline of 208 ± 4 mm². §: p < 0.001, #: p < 0.01.
Following aortic occlusion, DDU-Net–based segmentation revealed a significant reduction in KS of $-8\% \pm 1\%$ before (Figure 7A,B) and $-6\% \pm 0\%$ after USPIO (Figure 7C,D), accompanied by T$_2$ reductions of $22\% \pm 2\%$ in the CO, $27\% \pm 2\%$ in the OM, and $13\% \pm 3\%$ in the IM, highlighting the impact of altered blood flow on renal physiology.

Similarly, during occlusion of the left renal vein, DDU-Net segmentation detected a significant increase in KS of $5\% \pm 1\%$ before (Figure 8A,B) and $6\% \pm 1\%$ after USPIO (Figure 8C,D), along with pronounced T$_2$ reductions of $33\% \pm 2\%$ in the CO, $39\% \pm 2\%$ in the OM, and $32\% \pm 6\%$ in the IM, reflecting the interplay between venous outflow and KS dynamics, as well as tissue oxygenation.

Additionally, simultaneous aortic-venous occlusion showed no significant change in KS, but immediate declamping revealed a significant decrease in KS of $-3\% \pm 0\%$ (Figure 9A,B), corresponding to T$_2$ reductions of $23\% \pm 1\%$ in the CO, $28\% \pm 1\%$ in the OM, and $20\% \pm 3\%$ in the IM, further emphasizing the acute response of the kidneys to vascular adjustments.

Furthermore, furosemide administration induced a significant KS increase of $2\% \pm 1\%$ both before (Figure 10A,B) and following USPIO (Figure 10C,D), accompanied by T$_2$ increases of $11\% \pm 2\%$ in the CO and $19\% \pm 2\%$ in the OM, indicating the functional impact of diuretics on renal fluid dynamics. The KS remained significantly elevated throughout the observation period without USPIO (Figure 10A,B), while even with reduced contrast, a significant furosemide-induced KS increase was detected at one timepoint, providing insights into the persistence of its effects (Figure 10C,D).

In the context of hypoxemia, DDU-Net detected a significant decrease in KS of $-2\% \pm 1\%$ (Figure 11A,B), with corresponding T$_2$ reductions of $18\% \pm 3\%$ in the CO, $15\% \pm 3\%$ in the OM, and $8\% \pm 3\%$ in the IM, revealing the response of the kidneys to altered oxygenation levels. Notably, following the administration of the X-ray contrast agent, DDU-Net detected a significant increase in KS of up to $11\% \pm 1\%$ (Figure 11C,D), while T$_2$ increased by $14\% \pm 3\%$ in the CO and by $17\% \pm 5\%$ in the OM, and decreased by $17\% \pm 7\%$ in the IM, illustrating the physiological impact of contrast-induced acute kidney injury (AKI) and its effects on tissue perfusion and oxygenation throughout the 60-min observation period.

These findings demonstrate the comprehensive physiological insights provided by combining DDU-Net–based segmentation and T$_2$ mapping, enabling a deeper understanding of renal pathophysiology under diverse experimental conditions.

**FIGURE 8** Relative kidney size (KS) over time for the renal venous occlusion. (A) Exemplary T$_2$ maps acquired before (left) and during (center) the intervention, as well as during recovery (right). (B) Development of relative KS changes upon the occlusion of the left renal vein before the administration of ultrasmall superparamagnetic iron oxide (USPIO). The blue line depicts the mean of each time point and the corresponding interval of the standard error of the mean (SEM) ($n = 12$ rats). The mean KS at baseline was $210 \pm 4$ mm$^2$. The intervention is highlighted in green. (C) The analogous exemplary T$_2$ maps of the same rat, and (D) Time course of relative KS changes after administration of USPIO with a mean KS at baseline of $212 \pm 4$ mm$^2$. §: $p < 0.001$; *: $p < 0.05$. 
4 | DISCUSSION

This work demonstrates the feasibility of automated quantification of acute changes in KS of the rat kidney using segmentation of parametric MR images with a DDU-Net. The feasibility and reliability of DDU-Net–based longitudinal KS monitoring is demonstrated on in vivo T2 mapping data, with acute pathophysiological interventions performed directly while the animal was inside the MRI scanner. Despite the limited number of subjects involved in the training of the network, data augmentation, ensemble training, and stochastic weight averaging permitted successful optimization of the neural network. When benchmarked against the ground truth, our DDU-Net achieved a highly accurate segmentation of the kidney (Dice = 0.98) and quantification of KS ($R^2 = 0.92$). With a MAPE of 1.1%, DDU-Net facilitated accurate and precise prediction of KS, and outperformed the state-of-the-art nnU-Net. By reproducing the observed KS changes before and after the administration of USPIO, DDU-Net demonstrated resilience towards severely reduced signal intensities, SNRs and image contrasts. The results from the subgroup that received the X-ray CM reinforce the ability of DDU-Net to overcome a data shift resulting from a different image acquisition protocol.

To date, the application of our DDU-Net has been focused on retrospective assessment of changes in KS during acute pathophysiological scenarios. Future steps include prospective and real-time application of DDU-Net, integrating our DDU-Net library directly into the image reconstruction and postprocessing pipeline of the MR scanner. While the CPU computation time was $\sim 70$ ms per segmentation, GPU implementation improved the computation time to 10 ms. These computation times align with the temporal resolution of preclinical renal T2 mapping, in the range of 10–90 s per T2 map; thus, DDU-Net is suitable for on-the-fly assessment of KS. While it might be conceptually appealing to extend the DDU-Net approach to three-dimensional (3D) datasets, whole kidney coverage T2 mapping comes at the cost of increased acquisition time, which is a temporal resolution constraint for dynamic studies. In the preclinical experimental context, the temporal resolution required for longitudinal

![Image: Figure 9](https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/nbm.5075)
experiments with physiological interventions such as those used in the current study would be severely constrained by the longer acquisition times that would be needed to obtain 3D data.

nnU-Net has been previously applied on longitudinal MRI data, including assessment of lesion size and counts in patients with multiple sclerosis, and on cardiac chamber quantification. However, these cases generally do not include substantial reductions in image contrast or SNR. Our study illustrates the limitations of nnU-Net when dramatic changes do occur in the distribution of the data, signal intensities, or contrast of the ROI. We showed that nnU-Net failed to correctly segment the kidney in several cases with contrast changes, thereby compromising KS measurement. Conversely, our DDU-Net performed well in scenarios of reduced signal intensities, SNRs, and image contrasts. Furthermore, DDU-Net achieved a more precise determination of KS in terms of MAPE and $R^2$. While residual connections and dilations are possible options in the self-configuration of nnU-Net, the principles of our DDU-Net are not yet implemented in a self-configuring approach.

Several characteristics considered in our approach could be added to the portfolio of nnU-Net self-configuration: increased depth of the network, additional dilated convolutions, intra-level skip connections, concatenation of the layer output, and downsampling via strided convolution. Adding these configurations may reduce the number of channels needed in the network, and would decrease the memory needed for the model. In our case, the DDU-Net ensemble needed 135 MB, compared with 450 MB for nnU-Net, a reduction of 70%.

Fast, automated, and accurate kidney segmentation was facilitated by the DDU-Net architecture. This is of paramount relevance for monitoring KS to enable physiological interpretation of MRI-based assessments of renal oxygenation in acute pathophysiological scenarios. Renal tissue hypoperfusion and hypoxia are pivotal early events in the pathophysiology of AKI, and in progression to CKD. The MR relaxation parameters $T_2^*$, $T_2$ are surrogates of renal tissue oxygenation, because they are sensitive to the amount of deoxygenated hemoglobin (deoxyHb) per tissue volume. However, the relationship between renal tissue partial pressure of oxygen (pO$_2$) and $T_2^*$, $T_2$ is confounded by changes in hematocrit, the O$_2$ affinity of hemoglobin, and crucially by the blood and tubular volume fractions. Recognizing that events leading to acute renal hypoxia are often associated with changes in the blood and/or tubular volume fractions, and that these changes are mirrored by changes in KS, our recent preclinical study used simultaneous MRI-based measurements of changes in KS and $T_2^*$, $T_2$ for accurate physiological interpretation of acute renal oxygenation changes. If a decrease in $T_2^*$, $T_2$ is observed without any change in KS, this indicates that renal tissue oxygenation is reduced. However, if $T_2^*$, $T_2$ decrease and there is a corresponding decrease in KS, the reduction in tissue oxygenation should be regarded as more severe than if KS was unchanged. Conversely, if a similar decrease in $T_2^*$, $T_2$ is observed together with an increase in KS, the reduction in tissue oxygenation should be considered less severe than if KS remained unchanged. Thus, MRI-based measurement of renal oxygenation by $T_2^*$, $T_2$...
is highly dependent on monitoring accompanying changes in KS. To enable correct interpretation of $T_2^*$, $T_2$ derived from parametric MRI of renal oxygenation, accurate KS assessment is especially relevant, because a biophysical model can then be used to estimate changes in O$_2$ saturation of hemoglobin from changes in $T_2^*$ and KS.

The current study demonstrates that DDU-Net meets this need. As KS can be readily obtained from DDU-Net–based automated segmentation of $T_2$ maps without the need for additional MRI scans, this should always accompany assessment of MRI-derived oxygenation results.

The insights obtained from this work are not limited to experimental and preclinical studies. The basic principle of our approach can be adapted to the segmentation of human kidney MR images obtained from large-scale population MRI studies, like the German National Cohort, which include longitudinal MRI data with baseline and follow-up examinations. This wealth of data offers ample potential to train ML algorithms to establish real-time KS assessment in routine clinical practice. Leveraging MRI-based KS assessment from these large datasets would also support deciphering of the relationships between KS, molecular and biochemical markers, and data from physiological measurements, medical history and lifestyle data included in large cohort studies. These insights will advance our understanding of the determinants and associations of renal disease, and would promote extension of the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) guidelines on KS assessment, which are currently restricted to the use of KS as a biomarker for autosomal-dominant PKD. An expanded perspective would advance nephrology, radiology, and physiology, involving clinicians and patients, and provide a foundation for new insights into renal physiology and pathology.

ACKNOWLEDGMENTS

The authors thank A. Pohlmann, J. Periquito (Max-Delbrueck Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany), A. Anger, B. Flemming (Institute of Translational Physiology, Charité – Universitätsmedizin, Berlin, Germany), and Fabian Isensee (Helmholtz Imaging, German Cancer Research Center [DKFZ], Heidelberg, Germany) for technical and other support. The authors further thank the iNAMES - MDC-Weizmann Helmholtz International Research School (HIRS) for Imaging and Data Science from the NANo to the MESo. Open Access funding enabled and organized by Projekt DEAL.
CONFLICT OF INTEREST STATEMENT
The authors declare that they have no competing interests.

ORCID
Tobias Klein https://orcid.org/0000-0003-2236-8490

REFERENCES


How to cite this article: Klein T, Glaty T, Millward JM, et al. Dynamic parametric MRI and deep learning: Unveiling renal pathophysiology through accurate kidney size quantification. NMR in Biomedicine. 2024;37(4):e5075. doi:10.1002/nbom.5075