

Transfer learning-based approach for automated kidney segmentation on multiparametric MRI sequences

Rohini Gaikar,^{a,*} Fatemeh Zabihollahy^b, Mohamed W. Elfaal,^c
Azar Azad,^d Nicola Schieda,^e and Eranga Ukwatta^{b,a}

^aUniversity of Guelph, School of Engineering, Biomedical Engineering, Guelph, Ontario, Canada

^bJohns Hopkins University School of Medicine, Department of Radiation Oncology and Molecular Radiation Sciences, Baltimore, Maryland, United States

^cUniversity of Alberta, Department of Radiology, Edmonton, Alberta, Canada

^dA.I. VALI Inc., Toronto, Ontario, Canada

^eUniversity of Ottawa, Department of Radiology, Ottawa, Ontario, Canada

Abstract

Purpose: Multiparametric magnetic resonance imaging (mp-MRI) is being investigated for kidney cancer because of better soft tissue contrast ability. The necessity of manual labels makes the development of supervised kidney segmentation algorithms challenging for each mp-MRI protocol. Here, we developed a transfer learning-based approach to improve kidney segmentation on a small dataset of five other mp-MRI sequences.

Approach: We proposed a fully automated two-dimensional (2D) attention U-Net model for kidney segmentation on T1 weighted-nephrographic phase contrast enhanced (CE)-MRI (T1W-NG) dataset ($N = 108$). The pretrained weights of T1W-NG kidney segmentation model transferred to five other distinct mp-MRI sequences model (T2W, T1W-in-phase (T1W-IP), T1W-out-of-phase (T1W-OP), T1W precontrast (T1W-PRE), and T1W-corticomedullary-CE (T1W-CM), $N = 50$) and fine-tuned by unfreezing the layers. The individual model performances were evaluated with and without transfer-learning fivefold cross-validation on average Dice similarity coefficient (DSC), absolute volume difference, Hausdorff distance (HD), and center-of-mass distance (CD) between algorithm generated and manually segmented kidneys.

Results: The developed 2D attention U-Net model for T1W-NG produced kidney segmentation DSC of $89.34 \pm 5.31\%$. Compared with randomly initialized weight models, the transfer learning-based models of five mp-MRI sequences showed average increase of 2.96% in DSC of kidney segmentation ($p = 0.001$ to 0.006). Specifically, the transfer-learning approach increased average DSC on T2W from 87.19% to 89.90%, T1W-IP from 83.64% to 85.42%, T1W-OP from 79.35% to 83.66%, T1W-PRE from 82.05% to 85.94%, and T1W-CM from 85.65% to 87.64%.

Conclusions: We demonstrate that a pretrained model for automated kidney segmentation of one mp-MRI sequence improved automated kidney segmentation on five other additional sequences.

© 2022 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.9.3.036001](https://doi.org/10.1117/1.JMI.9.3.036001)]

Keywords: multiparametric MRI; kidney segmentation; CNN models; transfer learning.

Paper 22015GR received Jan. 21, 2022; accepted for publication May 31, 2022; published online Jun. 16, 2022.

1 Introduction

It is estimated that 87,100 adults will be diagnosed with kidney cancer in Canada and the United States in 2022 and 15,870 will die from kidney and renal pelvis cancer.^{1,2} Due to extensive use of cross-sectional imaging, the majority of kidney cancers are incidentally discovered on abdominal imaging scans namely ultrasound (US) and computed tomography (CT) performed for other

*Address all correspondence to Rohini Gaikar, rgaikar@uoguelph.ca

reasons.^{3,4} Magnetic resonance imaging (MRI) is increasingly being investigated for kidney imaging due to its potential ability to differentiate malignant and benign renal masses. MRI offers advantages compared with CT relating to the ability to better enhancement, differentiate various soft tissues, and to some extent evaluate tissue function. These properties may improve renal mass characterization.^{5,6} Renal mass MRI requires a comprehensive multiparametric MRI (mp-MRI) protocol, which combines anatomic and functional pulse sequences e.g., T1-weighted (T1W), T2-weighted (T2W), T1W chemical shift imaging, dynamic contrast-enhanced (CE) imaging, and more recently diffusion weighted imaging.^{7–9}

Segmentation of the kidney is typically the first step in an image-based computer-aided diagnosis (CAD) pipeline for kidney cancer. Since CT is typically performed more frequently for renal mass evaluation than MRI, previous methods for kidney cancer detection were developed mainly for CT imaging. Heller et al.¹⁰ summarized results of highest ranked kidney and kidney tumor segmentation models of the KiTS19 challenge in contrast-enhanced CT imaging. Of the top five models, four of them^{11,12} used cascaded models segmenting the kidney boundaries and kidney cancer in two steps. Another coarse-to-fine segmentation framework by Yue et al.¹³ used CNN model with component analysis for abnormal detection and correction of kidney segmentations resulted from coarse model on KiTS19 test data. The CE-CT image study done by Zabihollahy et al.^{14,15} employed ensemble learning to localize renal mass and achieved classification as renal cell carcinoma versus benign renal masses using cascaded network of convolutional neural networks.

Recently, several methods have been reported for kidney segmentation on MR images,^{16–24} which are listed in Table 1. However, most of these methods are developed for a single MRI

Table 1 Summary of literature review for kidney segmentation on MRI images.

Research study	MRI sequence, #images	Method	#Test cases	Kidney DSC, volumetric difference (VD)
Traditional image processing algorithms				
Sandmair et al. ¹⁷	T2W, 24	Semiautomatic, unimodal thresholding on manually extracted kidney ROI	48 kidney ROIs	VD = 1.5 ml
Tirunagari et al. ¹⁸	DCE-MRI	Dynamic mode decomposition for renal mass quantification	120	DSC = 0.87
Obaidullah ¹⁹	Low contrast MRI, 230	Semiautomated, fractional-based energy minimization using active contour model	230	DSC = 0.93
Machine learning and deep learning algorithms				
Haghghi et al. ²⁰	DCE-MRI, 30	3D U-Net model (pediatric data)	6	DSC = 0.87
O'Reilly et al. ²¹	T2W-MRI, 13	<i>K</i> -means clustering followed by active (ADPKD ^a Data) contour snake algorithm on kidney ROIs	12	VD = 21.63%
Bevilacqua et al. ²²	T2W-TSE, 18	CNN model on kidney ROIs (ADPKD Data)	3	DSC = 0.85
O'Reilly et al. ²³	T2W-MRI, 132	Ninefold cross validation V-Net model (ADPKD Data)	14	DSC = 0.88
Artur Klepaczko and Eikefjord ²⁴	DCE-MRI, 20	Implementation of U-Net, SegNet and DenseNet models	20	DSC = 0.94(U-Net), 0.93 (SegNet), 0.91(DenseNet)
Agarwal et al. ²⁵	CE-MRI, 118	Cascaded approach of U-Net models for kidney localization and renal mass segmentation	23	DSC = 0.91, VD = 7.89%

^aADPKD, autosomal dominant polycystic kidney disease.

sequence and were tested on small dataset. Though deep learning-based methods have been applied successfully for a single sequence in renal MRI, they may fail to generalize to other MRI sequences, due to domain shifts in image features. Although domain shifts in kidney mp-MRI have not been investigated, but domain shifts in brain tumor MRI have been reported. Ghafoorian et al.²⁶ described a transfer learning-based approach for brain lesion segmentation in MRI sequences by applying pretrained weights of FLAIR MRI images to T1 weighted image model. Similarly, Wacker et al.²⁷ proposed a method for three-dimensional (3D) and two-dimensional (2D) U-Net brain tumor segmentation using transfer learning of pretrained weights on contraction path and training the upsampling path of U-Net on BraTS¹⁶ dataset.

In this study, we describe a fully automated method for kidney segmentation using transfer learning for target domain on individual mp-MRI sequences using smaller datasets by utilizing pretrained weights from a source domain with consisting of a large, labeled dataset. To our knowledge, this study is the first to evaluate the utility of a transfer learning method for kidney segmentation on five individual mp-MRI sequences. We used a comparatively large dataset ($N = 108$) of T1W-NG MRI images as the source sequence and T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM as target MRI sequences evaluated with smaller datasets ($N = 50$) for transfer learning approach of kidney segmentation.

2 Materials and Methods

2.1 *mp-MRI Data Acquisition and Manual Segmentation*

This study was approved by the Ottawa Hospital Research Ethics Board (REB). Patient informed consent was waived by the REB and anonymized data were shared through a data sharing agreement between the Ottawa Hospital and the University of Guelph. We searched our Pathology and Picture Archiving and Communication Systems (PACS) database to create the study sample. The dataset consists of mp-MRI scans of patients with renal masses before percutaneous biopsy, partial or total nephrectomy without intervention of any chemotherapy or radiotherapy between January 2015, and December 2017.

The dataset consisted of 3D volumes of T1W-NG images for 108 patients and 3D volumes of each T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM for 50 patients. Segmentations of the kidneys were performed by an abdominal radiology fellow (BLINDED) on all sequences individually using ITKSnap. These images were acquired using heterogeneous MRI systems; Siemens, GE Healthcare, and Phillips with magnetic field strengths 1.5T and 3T at the Department of Radiology of the University of Ottawa and saved as anonymized images in digital imaging and communication in medicine format. The mp-MRI image acquisitions on GE Healthcare, Phillips, and Siemens were 32%, 3%, and 65%, respectively, in the dataset. The summary of the standard mp-MRI protocol used in clinical practice is provided in the [Appendix](#). The fellowship-trained abdominal radiologist manually segmented kidney boundaries in the dataset in a slice-by-slice manner. The kidney boundaries were established using ITKSnap version 3.2 and saved using NIFTI image format.

2.2 *Overview of Segmentation Method*

The developed algorithm for kidney segmentation on different mp-MRI sequences is implemented in two steps. First, the kidney segmentation was determined on T1W-NG images using DL-based attention U-Net model.²⁸ In the second step, the pretrained T1W-NG kidney segmentation model was then fine-tuned to delineate kidneys on T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM MRI sequences. We also implemented combined data training strategy by combining T1W-NG with each T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM data separately and multi-MRI input where training data of all six mp-MRI sequences used in training data.

2.3 *Automated Kidneys Segmentation in T1W-NG Sequence*

To study the transfer learning approach on different mp-MRI sequences, we identified relatively large dataset of T1W-NG ($N = 108$) as a source sequence as it was used for detection of renal

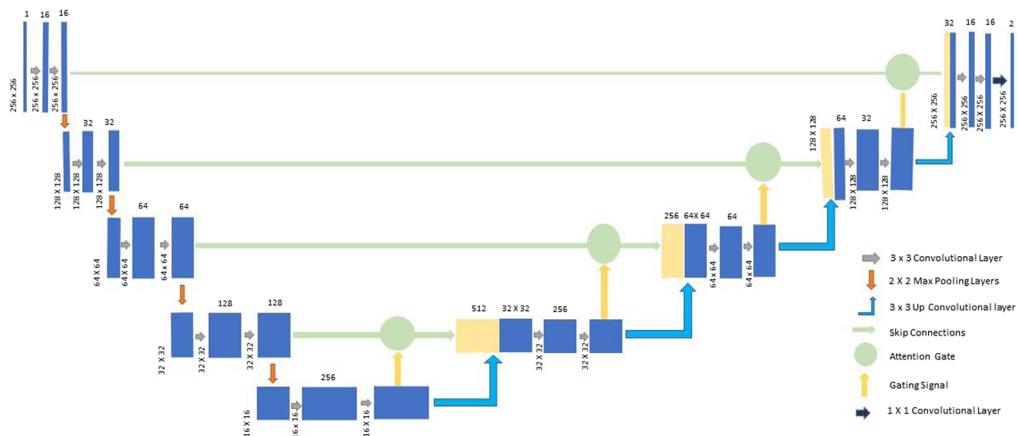


Fig. 1 Schematic of attention U-Net model designed for kidney segmentation on T1W-NG images.

masses in clinical studies. To delineate the kidney boundaries on 2D axial slices of T1W-NG, an attention U-net-based segmentation model was trained that suppresses the activation of irrelevant regions in kidney segmentation. Attention gates (AG)²⁸ are used in the expansion paths of U-Net to highlight salient features that are passed through the skip connections. The block diagram of implemented attention U-Net is shown in Fig. 1. The designed attention U-Net model is comprised of five encoder-decoder layers in which 16 filters of size 3×3 with the stride of 1×1 was used in the first layer of contraction path and were doubled in the subsequent layers. Each convolutional layer is followed by “ReLU” activation with “same” padding and “He Normal” as kernel initializer. Batch normalization²⁹ is employed as regularization technique to standardize the inputs to the layers after each minibatch of 16 training images to update the weights for faster convergence of the model.³⁰ To avoid the overfitting issue during the model training, dropout layers are implemented in last two layers of the contraction path.³¹ In the expansion path, an AG is employed in each layer as shown in Fig. 1. The final stage of the model used sigmoidal activation function on extracted features to generate a probability map that is the same size as an input image. To identify segmented kidneys, a threshold of 0.5 is applied on probability map such that the probability of 0.5 and above is a foreground kidney pixel and less than 0.5 is non-kidney background pixel. This model has total trainable parameters of 6,003,009.

The T1W-NG dataset contained 108 3D images acquired with different resolutions, minimum resolution of 192×256 and maximum resolution of 512×512 with number of axial slices from 20 to 144. To train the attention U-Net, test set of 20 MRI scans was prepared in random selection and remaining 88 MRI scans were randomly grouped into five different folds to implement fivefold cross validation model. Axial slices were extracted from 3D MRI scans and resized to same size of 256×256 . To improve the local contrast in the resized image, we used histogram equalization where number of bins equal to maximum intensity in the image. To achieve fractional intensity value between 0 and 1, the intensity normalization is done by maximum intensity of histogram equalized image. To augment the training image dataset, elastic deformation³² was used to extend the training dataset images. The elastic deformation was controlled by two values, elasticity coefficient σ as 512 and intensity of deformation controlled by α as 21. The binary cross entropy used as a loss function for binary kidney segmentation using ADAM optimizer with learning rate of 0.001 for training of attention U-Net on T1W-NG images. The test dataset contained 1838 axial slices from 20 patients to test model performance for kidney segmentation.

The target mp-MRI sequences' dataset contained T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM protocol acquired scans and corresponding manually annotated images of 50 patients; therefore, training and test datasets of 30 and 20 patients were prepared by random selection. To improve kidney segmentation on target mp-MRI sequences, we transferred features learned for kidney segmentation task using larger dataset of T1W-NG mp-MRI sequence to smaller dataset of five other distinct target mp-MRI sequences. A separate attention U-Net model was trained for each target mp-MRI sequence (T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM) which was similar in architecture to pretrained attention U-Net model of T1W-NG sequence. Each model

was initialized with the pretrained weights of attention U-Net model of T1W-NG kidney segmentation and fine-tuned at each layer by unfreezing all the layers during training process. The binary cross entropy used as a loss function and ADAM optimizer with learning rate of 0.001 for training of transfer learning-based attention U-Net model on individual target mp-MRI sequence model. These transfer learning-based models were tested on respective test data for kidney segmentation task. These model performances were compared with the similar models when trained with random initialization of weights.

Dice similarity coefficient (DSC), absolute volume difference (AVD) in percentage with respect to true volume of kidneys, Hausdorff distance (HD), and center-of-mass distance difference (CD) were used to quantitatively evaluate the quality of the network-generated kidney segmentation compared to the radiologists' manual segmentation. The kidney segmentation algorithm was implemented in Python using Keras library, on the top of TensorFlow³³ open-source machine learning library. All models were trained and evaluated on GPU-accelerated high-performance computers of Shared Hierarchical Academic Research Computing Network (SHARCNET), which is a partner organization of Compute Canada national research computing platform.

2.4 Other State-of-the-Art Methods on T1W-NG Sequence

U-Net³⁴ is a widely used model for biomedical image segmentation. In U-Net model, the captured features on the encoder paths are concatenated with the outputs of transposed convolution in decoder path to maintain precise localization with respect to the entire image. The coarse-to-fine segmentation approach using cascade of 2 U-Net models evaluated for kidney segmentation.¹³ The U-Net model architecture is modified in the U-Net++³⁵ architecture by connecting encoder and decoder paths through nested skip connection at each level to overcome the semantic gaps in the features. Another variation in U-Net model is at the convolutional layer using dilated convolutions³⁶ for multi-scale contextual information without any compromise on resolution. This is also known as Atrous convolution which enlarges the field of view of filters without increasing the number of computational parameters. Sometimes, deep learning models encounter vanishing gradients problem,³⁷ therefore we used residual learning in the U-Net architecture, i.e., residual U-Net³⁸ with identity mappings which allows error to travel through a short path in the model.

Other than U-Net-based models, we evaluated the performance of SegNet³⁹ model for kidney segmentation on T1W-NG sequence images. The SegNet model is encoder-decoder type architecture in which pooling indices are transferred to expansion path instead of transferring the whole feature set such as in U-Net, resulting in less memory utilization. In another method, we tested 50 layers deep ResNet50 model⁴⁰ as a backbone for U-Net model for kidney segmentation on T1W-NG images trained from randomly initialized weights. This ResNet50-Unet model implemented using segmentation models library available for implementation in Keras. All state-of-the-art models were trained using fivefold cross validation method on T1W-NG sequence.

3 Results

3.1 Methods for Kidney Segmentation on the Source Sequence (T1W-NG)

The attention U-Net trained on T1W-NG images evaluated on respective 20 test cases containing 1838 axial slices. Algorithm-generated segmentations and their corresponding manual segmentation on axial slices for nine different patients and from different regions of kidneys are shown in Fig. 2. The row 1 in Fig. 2 shows slices from apex and bottom regions of kidney where kidneys are small in appearance and either one of the kidneys is visible instead of two. In Fig. 2, the row 2 shows segmentations on axial slices where both kidneys are visible but in different proportion. The row 3 in Fig. 2 shows axial slices from the middle part of kidneys where both kidneys are prominently visible. All axial slices show algorithm generated kidney segmentations in red contour follows manual kidney segmentation shown in yellow contours. The axial slice in third

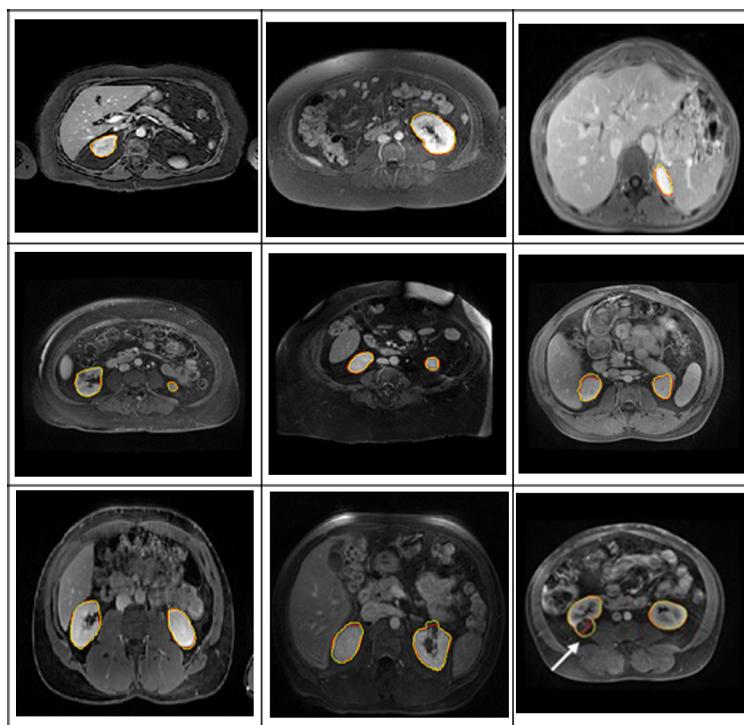


Fig. 2 Sample kidney segmentation using attention U-Net model on T1W-NG axial test slices from different patients, where yellow contours are ground truths and red contours are model predicted kidney segmentations.

column of third row has protruding renal mass present in the right kidney marked with white arrow. The model predicted renal mass as an integral part of kidney and segmented it including renal mass. The model produced kidney boundaries that closely match manual segmentations of the kidneys with a DSC of $89.34 \pm 5.31\%$ (mean \pm SD) and absolute kidney volume difference of $8.42 \pm 10.84\%$ (mean \pm SD). The HD between algorithm and manually generated kidney boundaries was 3.22 ± 2.17 mm and CD 1.11 ± 2.27 mm.

The seven different models designed for kidney segmentation on source sequence T1W-NG model were tested on 20 patients in test data. These model performances were compared with the proposed attention U-Net model based on kidney segmentation metrics DSC, AVD, HD, and CD as shown in Table 2. In these seven models, U-Net, dilated U-Net, and SegNet were performed

Table 2 Summary of fivefold cross validation model results on 20 T1W-NG test images in different models (mean \pm SD).

Model	DSC (%)	AVD (%)	HD (mm)	CD (mm)
U-Net	87.29 ± 10.2	9.32 ± 12.03	3.95 ± 2.26	1.87 ± 3.31
Cascade U-Net	86.64 ± 7.75	11.29 ± 12.43	5.62 ± 3.29	3.18 ± 2.11
U-Net++	88.03 ± 9.4	8.71 ± 11.37	3.52 ± 2.12	2.15 ± 3.61
Dilated U-Net	87.94 ± 11.43	9.27 ± 11.61	3.74 ± 2.33	1.64 ± 2.89
Residual U-Net	85.65 ± 12.87	7.16 ± 14.76	4.22 ± 3.04	6.86 ± 9.40
SegNet	87.26 ± 9.52	9.31 ± 12.85	3.89 ± 2.48	1.31 ± 2.70
ResNet50	88.62 ± 10.42	9.83 ± 11.65	3.62 ± 2.24	1.14 ± 2.39
Attention U-Net (proposed)	89.34 ± 5.31	8.42 ± 10.84	3.22 ± 2.17	1.11 ± 2.27

approximately similar in terms of DSC and AVD in kidney segmentation. The residual U-Net model was poor performing model with lowest Dice score of 85.65% as compared to all other models by generating false positive regions when kidney was absent in the scan. The ResNet50 model performed well on middle region of kidneys, but it did not eliminate false predictions. The attention U-Net model showed accurately segmented kidney contours matching with the manually segmented ground truth kidneys with DSC of 89.34% and AVD of 8.42%. It also reported minimum HD and average CD of 3.22 and 1.11 mm, respectively, between algorithm generated and manually segmented kidneys as compared with the other state-of-the-art models. Therefore, attention U-Net was proposed as best performing model for kidney segmentation on T1W-NG sequence images. The results of fivefold cross validation models tested on 20 patients' images trained on state-of-the-art models are shown in Table 2.

Figure 3 shows the comparison of kidney segmentation results on T1W-NG images by deep learning models U-Net, cascade U-Net, U-Net++, dilated U-Net, residual U-Net, SegNet, ResNet50, and attention U-Net. Each column indicates axial slices of different patient from different regions of kidneys. The apex and bottom regions of kidneys where kidney contours are small, all models except attention U-Net predict false positives which deviates from ground truth segmentations as shown in first and last columns of Fig. 3. The middle sections of the kidneys where kidney contours are bigger as compared with kidney apex and bottom regions, each deep learning model was performing well as shown in column two and three.

The comparison of 3D views of kidney segmentation on T1W-NG test images using different state-of-the-art models is shown in Fig. 4. The 2D coronal and sagittal views of kidney segmentation predicted on Attention U-Net model for sample T1W-NG test images are shown in Fig. 5. It shows that attention U-Net model accurately segments kidneys and do not generate false predictions for the test image having only one kidney in the scan.

3.2 Kidney Segmentation on Target mp-MRIs Using Transfer Learning Approach

The architecturally similar attention U-Net model used on T1W-NG sequence was designed for kidney segmentation on other target mp-MRI sequences. The pretrained T1W-NG attention U-Net weights were initialized to each target mp-MRI model and trained on respective datasets of T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM mp-MRIs. To improve the kidney segmentation in T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM mp-MRI sequences, the layers were unfrozen for fine tuning of the target sequence model. The comparison of kidney segmentation metrics with and without transfer learning models for different mp-MRI sequences was shown in Table 3.

The sample kidney segmentations on target mp-MRI sequences without and with transfer learning approaches along with corresponding manual segmentations are shown in Fig. 6 with one of the axial slice and complete 3D view of segmented kidneys. T2W and T1W-IP images in first and second rows in Fig. 6 show that without transfer learning approach model produced false segmentations in nearby regions and under segmented kidneys, respectively. On T1W-OP images, the predicted kidney segmentation with transfer learning approach followed the manual kidney segmented contour and did not show multiple contours like generated in without transfer learning model as shown in third row in Fig. 6. On T1W-PREimages, the transfer learning model produced accurate kidney contours as compared to without transfer learning-based model. The accurately segmented kidney boundaries in T1W-PRE following the manual ground truth labels can be seen in fourth row in Fig. 6. The transfer learning in T1W-CM outperformed without transfer learning approach as shown in last row in Fig. 6 by improving DSC for kidney segmentation. The transfer learning-approach showed improvement in mean standard deviation (SD) in target mp-MRI sequence for kidney segmentation. Each target mp-MRI sequence model with transfer learning approach showed significantly different results ($p < 0.05$) than the without transfer learning model results with average increase of DSC by 2.96%.

The transfer-learning approach increased average DSC on T2W from 87.19% to 89.90%, T1W-IP from 83.64% to 85.42%, T1W-OP from 79.35% to 83.66%, T1W-PRE from 82.05% to 85.94%, and T1W-CM from 85.65% to 87.64%. The corresponding AVD and CD errors were lowered in with transfer learning model prediction results as shown in Table 3.

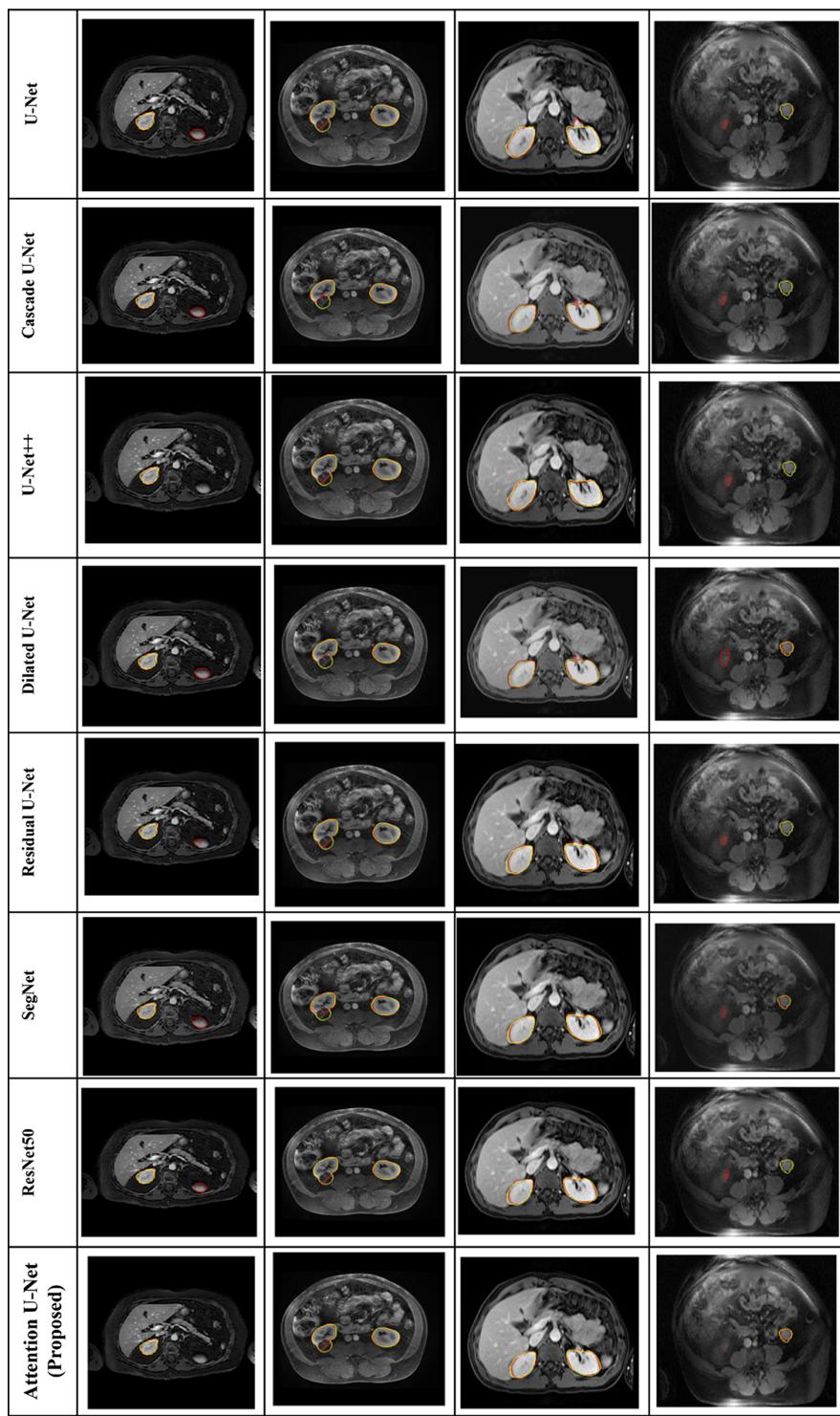


Fig. 3 Sample kidney segmentation on T1W-NG test images where each column is representing a different patient with yellow contours are ground truths and red contours are model predicted outputs of kidney segmentations. The first column shows one of the apex region slices where kidney becomes visible. The second and third columns represent kidney segmentation on middle slices and last column is one of the bottom region kidney slices.

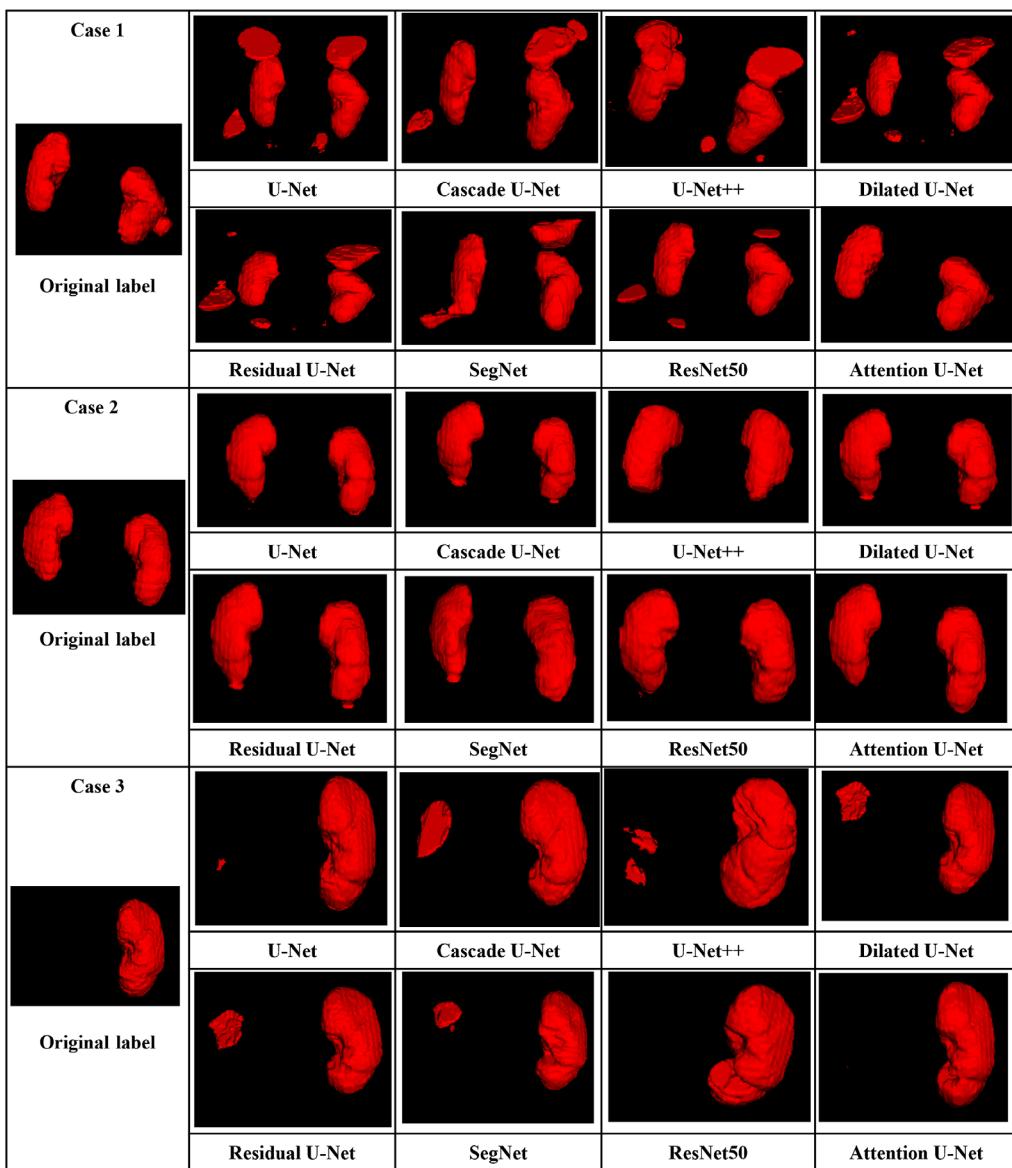


Fig. 4 Sample three cases showing 3D view of kidney segmentation on T1W-NG test images where original label (manual kidney segmentation) is compared with U-Net, cascade U-Net, U-Net++, dilated U-Net, residual U-Net, SegNet, ResNet50, and proposed attention U-Net model predicted kidney segmentation.

The coronal and sagittal views of kidney segmentations on different target mp-MRI sequences using with and without transfer learning approach is shown in Fig. 7.

3.3 Comparison with Combined Training Strategy

To study mp-MRI sequences behavior when combined with T1W-NG, we prepared combined datasets of T1W-NG and T2W, T1W-NG and T1W-IP, T1W-NG and T1W-OP, T1W-NG and T1W-PRE, and T1W-NG and T1W-CM. The proposed attention U-Net model similar in structure as discussed in Sec. 2.3 was trained on each combined dataset and evaluated its performances on respective test cases. The DSC for kidney segmentation used to know the generalization of model on two datasets.

When any one of the target sequences is combined with T1W-NG sequence, the attention U-Net model trained on combined dataset performed differently on test data from two different sequences. The comparison of kidney DSC values of T1W-NG and combined target mp-MRI

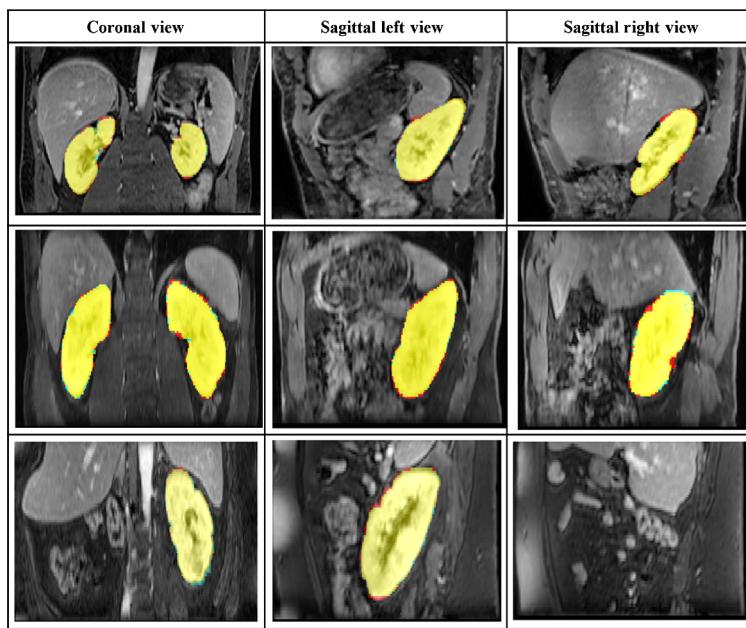


Fig. 5 Sample 2D coronal and sagittal views of kidney segmentation on T1W-NG test images on attention U-Net model where manual segmentation is in red, model prediction is in cyan, and overlap between manual segmentation and model predictions in yellow.

Table 3 Comparison of kidney segmentation DSC metric in different target mp-MRI sequences on 20 test cases using transfer learning of pretrained T1W-NG Attention U-Net model (mean \pm SD). Bold indicates the best performance.

mp-MRI sequences	Without transfer learning			With transfer learning		
	DSC (%)	AVD (%)	CD (mm)	DSC (%)	AVD (%)	CD (mm)
T2W	87.19 ± 8.4	11.63 ± 15.69	2.29 ± 1.49	89.96 ± 5.0 ($p=0.037$)	9.89 ± 12.59	1.11 ± 1.24
T1W-IP	83.64 ± 1.7	14.58 ± 13.16	1.61 ± 2.06	85.42 ± 6.0 ($p=0.043$)	9.83 ± 6.41	1.30 ± 1.33
T1W-OP	79.35 ± 9.7	23.49 ± 25.10	2.65 ± 3.89	83.66 ± 7.5 ($p=0.003$)	15.42 ± 13.35	1.57 ± 2.88
T1W-PRE	82.05 ± 6.7	15.88 ± 5.30	1.77 ± 1.02	85.97 ± 4.6 ($p=0.001$)	13.38 ± 7.20	1.13 ± 0.83
T1W-CM	85.65 ± 4.0	14.05 ± 4.87	1.07 ± 1.08	87.64 ± 3.5 ($p=0.006$)	6.14 ± 4.21	0.91 ± 0.84

sequence is shown in Table 4. The NA in table indicates that corresponding test data is not applicable on the corresponding model. The combined dataset models of T1W-NG and T2W produced kidney DSC of 88.80% and 88.87% on T1W-NG and T2W test data, respectively. This showed that during training the domain difference between T1W-NG and T2W is nullified by generalizing the model on T1W-NG and T2W MRI sequences. The combined dataset model of T1W-NG and T1W-IP did not generalize on both MRI sequences by producing DSC of 89.91% and 81.07% on T1W-NG and T1W-IP test dataset. Similarly, the combined training experiment of T1W-NG and T1W-OP sequences showed low DSC of 83.37% on T1W-NG and 81.35% on T1W-OP. In T1W-NG and T1W-PRE combined dataset, the trained model produced kidney DSC of 90.06% on T1W-NG and 85.14% on T1W-PRE images. In combined training model of T1W-NG and T1W-CM, the model recorded DSC of 91.51% on T1W-NG test data and lower performance by 3.30% on T1W-CM test data showing that the domain difference is minimized in combined training.

The comparison of kidney DSC values with respect to three models such as combined training, without transfer learning, and with transfer learning on different mp-MRI sequences is shown in

Target mp-MRI sequence	Original image		Without transfer learning approach		With transfer learning approach	
	Sample axial slice	3D view	Sample axial slice	3D view	Sample axial slice	3D View
T2W						
T1W-IP						
T1W-OP						
T1W-PRE						
T1W-CM						

Fig. 6 Sample axial and 3D kidney segmentation view on target mp-MRI sequence images using without and with transfer learning approaches; where in axial slices yellow contours are ground truth kidney labels and red contours are model predictions.

Target mp-MRI sequence	Without transfer learning approach			With transfer learning approach		
	Coronal view	Sagittal left view	Sagittal right view	Coronal view	Sagittal left view	Sagittal right view
T2W						
T1W-IP						
T1W-OP						
T1W-PRE						
T1W-CM						

Fig. 7 Sample 2D coronal and sagittal views of kidney segmentation on five different mp-MRI sequences test images using without and with transfer learning approaches where manual segmentation is in red, model prediction is in cyan, and overlap between manual segmentation and model predictions is in yellow.

Table 5. The corresponding sample kidney segmentation images on T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM mp-MRI sequences using these three models are shown in Fig. 8.

In another multisquences MRI input training strategy, where the training data pertaining to all the six sequences were used in training attention, U-Net model did not produce a generalized

Table 4 Attention U-Net model performance based on average DSE metric in combined training dataset.

Combined dataset	T1W-NG	T2W	T1W-IP	T1W-OP	T1W-PRE	T1W-CM
T1W-NG + T2W	88.80 ± 7.30	88.87 ± 5.69	NA	NA	NA	NA
T1W-NG + T1W-IP	89.91 ± 6.42	NA	81.07 ± 7.65	NA	NA	NA
T1W-NG + T1W-OP	90.16 ± 7.19	NA	NA	81.35 ± 8.52	NA	NA
T1W-NG + T1W-PRE	90.34 ± 6.16	NA	NA	NA	85.14 ± 6.08	NA
T1W-NG + T1W-CM	91.51 ± 5.57	NA	NA	NA	NA	88.21 ± 5.03

NA, not applicable.

Table 5 Summary of attention U-Net model performance in terms of kidney segmentation DSC with 3 different training strategies (mean ± SD). Bold indicates the best performance.

mp-MRI sequences	Combined dataset training (T1W-NG + one mp-MRI sequence)	Without transfer learning	With transfer learning
T2W	88.87 ± 5.6	87.19 ± 8.4	89.96 ± 5.0
T1W-IP	81.07 ± 7.65	83.64 ± 1.7	85.42 ± 6.0
T1W-OP	81.35 ± 8.52	79.35 ± 9.7	83.66 ± 7.5
T1W-PRE	85.14 ± 6.08	82.05 ± 6.7	85.97 ± 4.6
T1W-CM	88.21 ± 5.0	85.65 ± 4.0	87.64 ± 3.5

model for respective test data. This model achieved lowest DSC on T2W test data as compared to other training strategies. The model was biased toward T1W sequence protocols achieving the DSC such as transfer learning strategy. The comparison of average kidney segmentation DSC on different test data sequences is shown in Table 6.

4 Discussion

The experiments conducted to study the kidney segmentation on different mp-MRI sequences proved that though training set images were acquired on three heterogeneous MRI systems, the model generalized well on heterogeneous systems' scanned images. Compared to a previous study done on MRI sequences, our work is distinctive as each patient scan have renal masses. This study is the first to our knowledge that studied transfer learning on five other distinct mp-MRI sequences for automated kidney segmentation. We studied fivefold cross-validation for kidney segmentation on T1W-NG images using eight different deep learning models which are discussed in Sec. 2.4. The attention U-Net was the most accurate segmentation model on T1W-NG images by generating accurate kidney segmentation with DSC of 89.34% as compared with other state-of-the-art models. The time measured to predict the kidney segmentation of single image on proposed attention U-Net model was ~0.72 s using GPU platform.

When attention U-Net model trained with combined dataset of T1W-NG and T2W data yielded kidney DSC of 88.80% and 88.87% on T1W-NG and T2W test data, respectively, showing generalized model behavior for T1W-NG and T2W sequence images, but transfer learning of pretrained T1W-NG improved DSE in T2W target sequence by 2.77% giving average DSC 89.96% than scratch trained model on T2W images, which is greater as compared to previous studies^{20–23} related to T2W dataset.

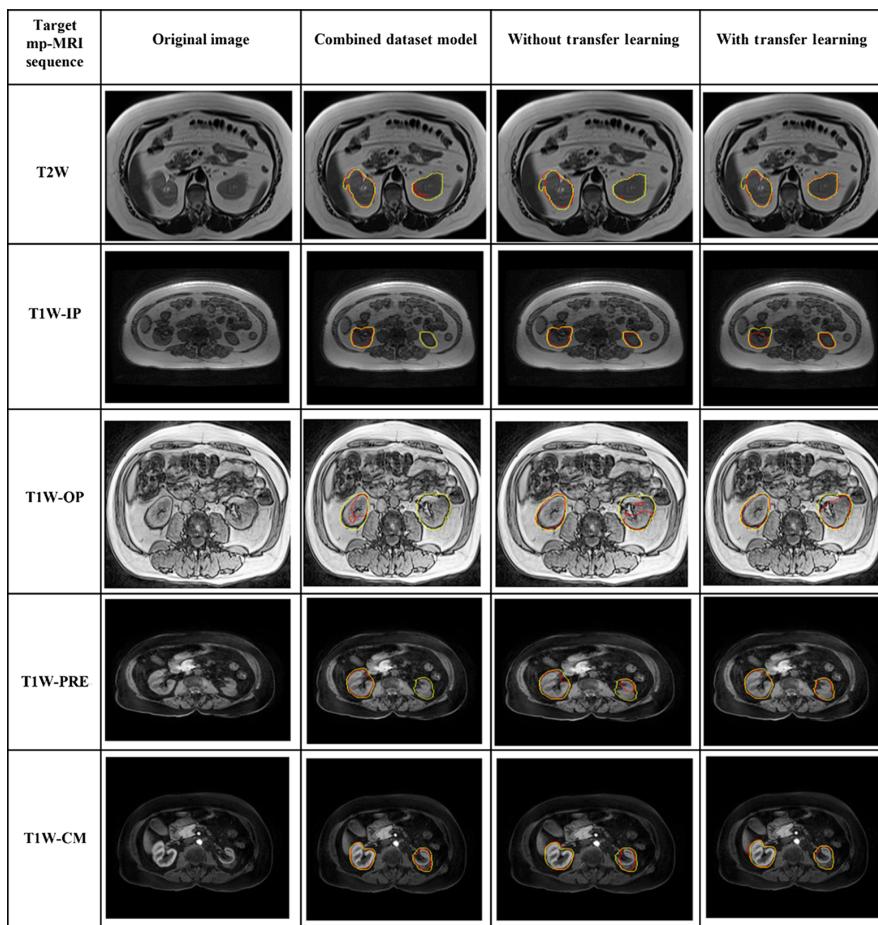


Fig. 8 Sample kidney segmentation on axial slices is shown using combined dataset model predictions in column 2, without transfer learning approach model predictions in column 3, and with transfer learning approach model in column 4. Each row represents a separate test case where yellow contours are ground truth kidney labels and red contours are model predictions.

Table 6 Summary of attention U-Net model performance in terms of kidney segmentation DSC when six sequences in training strategies (mean \pm SD). Bold indicates the best performance.

Test data sequence	T2W	T1W-NG	T1W-IP	T1W-OP	T1W-PRE	T1W-CM
DSC	80.31 ± 14.12	87.24 ± 4.29	85.43 ± 4.7	84.02 ± 6.41	85.45 ± 3.77	86.18 ± 4.18

Compared with previous studies of automated kidney segmentation, this study is unique as it explored the transfer learning strategy of using one mp-MRI sequence to improve the kidney segmentation on five other distinct mp-MRI sequences such as T2W, T1W-IP, T1W-OP, T1-PRE, and T1W-CM. The combined dataset T1W-NG and T1W-CM trained model was generalized for T1W-NG and T1W-CM mp-MRI sequences producing DSC of 91.21% and 88.21% kidney segmentation dice on T1W-NG and T1W-CM test data, respectively. The transfer learning model of T1W-CM images also showed increase in DSC by 1.99%. The combined dataset model of T1W-NG and T1W-IP produced kidney DSC of 89.91% and 81.07%, respectively, which indicates that generalization of kidney segmentation did not perform well on T1W-NG and T1W-IP images. The same observation was made when models were trained with combined dataset of T1W-NG and T1W-OP. The T1W-NG and T1W-PRE combined training performed well on both sequences by producing DSC of 90.06% on T1W-NG and 85.14% on T1W-PRE. But transfer learning approach in respective model training improved kidney segmentation

of T1W-OP and T1W-PRE target mp-MRI sequences by 4.31% and 3.92%, respectively. Though transfer learning improved kidney segmentation performance on T1W-OP model but combined training of T1W-NG and T1W-OP produced lowest DSC of 83.37% for kidney segmentation on T1W-NG images as compared with the other combined training dataset models. This proved differences in acquisition protocols prohibited generalization of T1W-NG and T1W-OP mp-MRI sequences.

When training data from all the six sequences combined to train the model with multi-MRI input data, the model achieved lowest kidney segmentation DSC of 80.31% on T2W test data than other T1W protocols. The T1W-IP, T1W-OP, T1W-PRE, and T1W-CM test data DSC was approximately equal to transfer learning strategy due to six times increased training data of T1W protocols in multi-MRI input training strategy. Thus, this study proved that transfer learning from another mp-MRI sequence improves the kidney segmentation results in small size of labeled dataset. In future studies, there is scope to explore unsupervised segmentation techniques on mp-MRI sequences to overcome domain shift problems.

5 Conclusion

In this study, we described the transfer learning approach to improve fully automated kidney segmentation on small mp-MRI dataset sequences using attention U-Net deep learning model. Due to the difference in acquisition protocols in mp-MRI sequences, a single model will not generalize on all mp-MRI sequence images, but the knowledge gained by model during T1W-NG training when transferred on small size target dataset improved the kidney segmentation results. Combined datasets training of model did not help to improve DSC for kidney segmentation. Thus, this study showed that transfer learning in mp-MRI sequences improves automated kidney segmentation results.

6 Appendix: Summary of the Standard Multiparametric-MRI Protocols Used in Clinical Practice for Image Acquisition

The 3D mp-MRI scans were acquired using imaging at 1.5T and 3T. The summary of acquisition protocol is provided in Table 7.

Table 7 Summary of standard mp-MRI protocol used in image acquisition.

	Dual echo T1W GRE			T2W TSE/FSE		Volume interpolated T1W 3D GRE ^b		Diffusion weighted imaging ^c
Pulse sequence	3D GRE ^a			Single shot TSE/FSE	3T	1.5T	Single shot echo-planar imaging	
	2D GRE	3T	1.5T					
Physiology	Breath hold	Breath hold	Breath hold	Respiratory triggered	Breath hold	Breath Hold	Single shot echo-planar imaging	
				Breath hold				
Duration	21 s	16 s	20 s	3 to 4 min 22 s	20 s	21 s		
Fat suppression	N/A	N/A	N/A	N/A	Chemical or spectral inversion recovery	Spectral inversion recovery		
TE (IP/OP)^d; TR (ms)	(4.6/2.3); 160 to 180	(2.5/1.3); 5.5 and (2.2/1.1); 4.0	(4.6/2.3); 7.6	83 to 88; 1030	1.7 to 2.5; 4.0 to 4.5	1.4;4.3	60.8 to 74; 2075 to 4600	

Table 7 (Continued).

Pulse sequence	Dual echo T1W GRE			T2W TSE/FSE		Volume interpolated T1W 3D GRE ^b		Diffusion weighted imaging ^c
	3D GRE ^a			Single shot TSE/FSE			Single shot echo-planar imaging	
	2D GRE	3T	1.5T		3T	1.5T		
Flip angle (deg)	70	10 to 12	10	180	10 to 12	10 to 12	90	
Bandwidth (Hz)	260	700	313	450	325 to 460	488	250 to 1446	
Number of excitations	1	0.7 to 1	1	Half-Fourier	1	1	2	
Acceleration factor	2	2	1	1 2	2	2	2	
Matrix size	256/320 x 134/152	294 x 224	192 x 320	170 x 256	256 x 320	132 x 320	130 to 38; 96 to 75	
Field of view (cm)	25 x 35	25 x 35	25 x 35	25 x 35	25 x 35	25 x 35	40 to 380; 28 to 75	
Slice thickness (mm)	5 to 6	3 to 4	3 to 5	5	2.5 to 4	2.5 to 4	6	

^aImaging was performed on clinical 1.5 Tesla or 3 Tesla systems.^bVIBE (Siemens Healthcare), THRIVE (Philips Healthcare), LAVA (General Electric Healthcare).^cDiffusion weighted imaging performed with two *b* values (0 and 600 mm²/s) with ADC map automatically derived.^dIP = in phase, OP = opposed phase.

Disclosures

No conflict of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant awarded to Eran Ukwatta. Rohini Gaikar acknowledges the Ontario Graduate Scholarship (OGS) and MITACS accelerate internship grant awarded to her. We also acknowledge computing facilities provided by the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada.

References

1. R. L. Siegel et al., “Cancer statistics, 2022,” *CA Cancer J. Clin.* **72**(1), 7–33 (2022).
2. D. R. Brenner et al., “Projected estimates of cancer in Canada in 2022,” *CMAJ* **194**(17), E601–E607 (2022).
3. A. Volpe et al., “The natural history of incidentally detected small renal masses,” *Cancer* **100**(4), 738–745 (2004).
4. A. M. May et al., “Current trends in partial nephrectomy after guideline release: health disparity for small renal mass,” *Kidney Cancer* **3**(3), 183–188 (2019).
5. N. K. Ramamurthy et al., “Multiparametric MRI of solid renal masses: pearls and pitfalls,” *Clin. Radiol.* **70**(3), 304–316 (2015).
6. A. D. De Leon et al., “Role of virtual biopsy in the management of renal masses,” *Am. J. Roentgenol.* **212**(6), 1234–1243 (2019).

7. N. E. Canvasser et al., “Diagnostic accuracy of multiparametric magnetic resonance imaging to identify clear cell renal cell carcinoma in cT1a renal masses,” *J. Urol.* **198**(4), 780–786 (2017).
8. R. A. Pooley, “AAPM/RSNA physics tutorial for residents fundamental physics of MR imaging,” *Radiographics* **25**, 1087–1099 (2005).
9. J. M. Willatt et al., “MR imaging in the characterization of small renal masses,” *Abdom. Imaging* **39**(4), 761–769 (2014).
10. 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.* **67**, 101821 (2021).
11. F. Isensee and K. H. Maier-Hein, “An attempt at beating the 3D U-Net,” arXiv1908.02182 (2019).
12. Y. Zhang et al., “Cascaded volumetric convolutional network for kidney tumor segmentation from CT volumes,” arXiv:1910.02235 (2019).
13. T. Yue et al., “Coarse-to-fine kidney segmentation framework incorporating with abnormal detection and correction,” arXiv:1908.11064 (2019).
14. F. Zabihollahy et al., “Automated classification of solid renal masses on contrast-enhanced computed tomography images using convolutional neural network with decision fusion,” *Eur. Radiol.* **30**(9), 5183–5190 (2020).
15. Z. Fatemeh et al., “Ensemble U-net-based method for fully automated detection and segmentation of renal masses on computed tomography images,” *Med. Phys.* **47**(9), 4032–4044 (2020).
16. B. H. Menze et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015).
17. M. Sandmair et al., “Semiautomatic segmentation of the kidney in magnetic resonance images using unimodal thresholding,” *BMC Res. Notes* **9**, 489 (2016).
18. S. Tirunagari et al., “Functional Segmentation through dynamic mode decomposition: automatic quantification of kidney function in DCE-MRI images,” arXiv1905.10218v1 (2019).
19. U. H. Obaidellah, “Kidney segmentation in MR images using active contour model driven by fractional-based energy minimization,” *Signal Image Video Process.* **14**(7), 1361–1368 (2020).
20. M. Haghghi et al., “Automatic renal segmentation in DCE-MRI using convolutional neural networks automatic renal segmentation in DCE-MRI using convolutional neural,” in *Proc. IEEE Int. Symp. Biomed. Imaging 2018*, 24 May 2018, pp. 1534–1537 (2018).
21. J. A. O'Reilly et al., “Automatic segmentation of polycystic kidneys from magnetic resonance images using decision tree classification and snake algorithm,” in *BMEiCON 2019 - 12th Biomed. Eng. Int. Conf.*, pp. 18–22 (2019).
22. V. Bevilacqua et al., “A comparison between two semantic deep learning frameworks for the autosomal dominant polycystic kidney disease segmentation based on magnetic resonance images,” *BMC Med. Inf. Decis. Making* **19**(Suppl. 9), 1–12 (2019).
23. J. A. O'Reilly et al., “Automatic segmentation of polycystic kidneys from magnetic resonance images using a three-dimensional fully-convolutional network,” in *RSU Int. Res. Conf.*, pp. 43–50 (2020).
24. A. L. Artur Klepaczko and E. Eikefjord, “Deep convolutional neural networks in application to kidney segmentation in the DCE-MR images,” *Lect. Notes Comput. Sci.* **12744**, 609–622 (2021).
25. A. Agarwal et al., “Deep learning-based ensemble method for fully automated detection of renal masses on magnetic resonance images,” submitted; in review (2022).
26. M. Ghafoorian et al., “Transfer learning for domain adaptation in MRI: application in brain lesion segmentation,” *Lect. Notes Comput. Sci.* **10435**, 516–524 (2017).
27. J. Wacker, M. Ladeira, and J. E. V. Nascimento, “Transfer learning for brain tumor segmentation,” arXiv:1912.12452 (2019).
28. O. Oktay et al., “Attention U-net: learning where to look for the pancreas,” in *Med. Imaging with Deep Learn. Conf.* (2018).

29. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Int. Conf. on Mach. Learn.*, pp. 448–456, PMLR (2015).
30. S. Santurkar, D. Tsipras, and A. Ilyas, "How does batch normalization help optimization?" <https://doi.org/10.48550/ARXIV.1805.11604> (2018).
31. G. Hinton et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
32. P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh Int. Conf. Doc. Anal. and Recognit. Proc.*, pp. 1–6 (2003).
33. M. Abadi et al., "TensorFlow: a system for large-scale machine learning this paper is included in the proceedings of the TensorFlow: a system for large-scale machine learning" in *OSDI'16: Proc. 12th USENIX Conf. Oper. Syst. Des. and Implement.* (2016).
34. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
35. Z. Zhou et al., "Unet++: a nested U-Net architecture for medical image segmentation," *Lect. Notes Comput. Sci.* **11045**, 3–11 (2018).
36. L. Chen et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV Proc.* (2018).
37. S. Hochreiter et al., "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Neural Networks*, J. F. Kolen and S. C. Kremer, Eds., pp. 237–243, Wiley (2003).
38. K. He et al., "Identity mappings in deep residual networks," *Lect. Notes Comput. Sci.* **9908**, 630–645 (2016).
39. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
40. K. He et al., "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).

Rohini Gaikar is currently pursuing her PhD under the guidance of Dr. Eranga Ukwatta on artificial intelligence (AI)-based computer aided methods for bio-medical image processing. She received her bachelor's degree in electronics from Savitribai Phule Pune University, India, in 2003. She graduated with Master of Engineering degree from Savitribai Phule Pune University, India, in 2011. She has worked as assistant professor at D. Y. Patil College of Engineering, Akurdi, Pune, India.

Fatemeh Zabihollahy is currently a postdoctoral research fellow working at Johns Hopkins University. Her research focuses on developing AI-based medical image analysis techniques for the detection, diagnosis, and staging of the cancers. She received her PhD in electrical and computer engineering in 2020 from Carleton University, Canada. She is the recipient of Carleton Medal for her outstanding graduate work at the PhD level.

Mohamed W. Elfaal is a diagnostic and body imaging fellow in the University of Alberta, Edmonton and former body imaging fellow in University of Ottawa, Canada. He was graduated from Ain Shames University Medical School (Cairo, Egypt) in November 2000 and was granted MSc in diagnostic radiology in 2008. He is a part of a research group interested in employing the recent AI techniques in diagnostic radiology.

Azar Azad received her doctorate degree in diagnostic medical laboratory sciences, followed by PhD in clinical biochemistry from Tehran University of Medical Sciences, in 1991. Her post-doctoral research focused on molecular pathway analysis in angiogenesis and apoptosis, at University of Toronto (1997). She worked at the University of Toronto, as co-director of Banting and Best Diabetes Core Lab, and as scientific staff at Mount Sinai Hospital, until 2017. She is an entrepreneur and founder of A.I. VALI Inc.

Nicola Schieda is an associate professor and the director of abdominal and pelvic MRI and prostate imaging at the University of Ottawa. He is actively involved in clinical research with

a focus on genitourinary imaging, adrenal, kidney, and prostate cancers. He is a member of the Radiological Society of North America, American Roentgen Ray Society, Society of Abdominal Radiology, and is the chair of the American College of Radiology Genitourinary Continuing Professional Improvement program.

Eranga Ukwatta received his PhD from Western University, Canada, in 2013. He was a multi-center postdoctoral fellow with Johns Hopkins University and University of Toronto. He is currently an assistant professor at the University of Guelph, Canada, and an adjunct professor in systems and computer engineering with Carleton University, Canada. He has more than 90 journal articles and conference proceedings. His research interests include medical image segmentation and registration, deep learning, and computational modeling.