

THE UNIVERSITY OF NOTTINGHAM  
SIR PETER MANSFIELD IMAGING CENTRE  
SCHOOL OF PHYSICS AND ASTRONOMY

---

---

**Developing Techniques for Quantitative Renal Magnetic Resonance Imaging**

---

*Author:*  
Alexander James DANIEL MSci

*Supervisor:*  
Prof. Susan FRANCIS

3<sup>rd</sup> February, 2021



Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy



## **Abstract**

Science will happen, but this bit can be read by muggles on ‘tinterweb.

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>vii</b>
<b>Abbreviations</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Imaging in the Clinic . . . . .	1
1.2 Clinical Motivation . . . . .	2
1.3 Thesis Overview . . . . .	5
1.4 References . . . . .	7
<b>2 Principles of Nuclear Magnetic Resonance Imaging</b>	<b>10</b>
2.1 Source of the NMR Signal . . . . .	11
2.1.1 Nuclear Spin . . . . .	11
2.1.2 Application of an External Magnetic Field . . . . .	12
2.1.3 Precession . . . . .	13
2.1.4 Resonance . . . . .	14
2.2 Relaxation and Contrast Mechanisms . . . . .	15
2.2.1 Longitudinal Relaxation ( $T_1$ ) . . . . .	15
2.2.2 Transverse Relaxation ( $T_2$ and $T_2^*$ ) . . . . .	17
2.2.3 Dipole-Dipole Interactions . . . . .	21
2.2.4 Diffusion Imaging . . . . .	23
2.2.5 Optimisation of Tissue Contrast . . . . .	24
2.3 Forming an Image . . . . .	25
2.3.1 Signal Localisation . . . . .	25
2.3.2 Image Acquisition Acceleration . . . . .	31

2.3.3	Image Acquisition Schemes . . . . .	33
2.4	Conclusion . . . . .	38
2.5	References . . . . .	39
<b>3</b>	<b>Assessment of Renal <math>T_2</math> Mapping Methods</b>	<b>40</b>
3.1	Introduction . . . . .	41
3.2	Methods . . . . .	42
3.2.1	Data Acquisition . . . . .	42
3.2.2	Post Processing . . . . .	48
3.2.3	Assessment of Data . . . . .	49
3.3	Results . . . . .	52
3.3.1	Sequence Accuracy . . . . .	52
3.3.2	Image Quality . . . . .	55
3.3.3	In-Vivo . . . . .	55
3.4	Discussion . . . . .	59
3.5	Conclusion . . . . .	59
3.6	Acknowledgements . . . . .	59
3.7	References . . . . .	60
<b>4</b>	<b>Quantitative Methods to Measure Renal Oxygenation</b>	<b>64</b>
4.1	Introduction . . . . .	65
4.2	Methods . . . . .	67
4.2.1	Susceptibility-Based Oximetry . . . . .	67
4.2.2	$T_2$ Relaxation Under Spin Tagging . . . . .	68
4.2.3	Inducing Changes in Oxygenation of Blood in the Renal Vein . . . . .	74
4.3	Results and Discussion . . . . .	76
4.3.1	Susceptibility-Based Oximetry . . . . .	76
4.3.2	$T_2$ Relaxation Under Spin Tagging . . . . .	78
4.4	Conclusions and Future Work . . . . .	85
4.5	Acknowledgements . . . . .	85
4.6	References . . . . .	86
<b>5</b>	<b>Automated Segmentation of Kidneys using Machine Learning</b>	<b>90</b>
5.1	Introduction . . . . .	91
5.2	Neural Networks for Image Segmentation . . . . .	94
5.2.1	Artificial Neural Networks . . . . .	94

5.2.2	Convolutional Neural Networks . . . . .	97
5.3	Methods . . . . .	99
5.3.1	MRI Data Acquisition . . . . .	99
5.3.2	Manual Segmentation . . . . .	100
5.3.3	Automated Segmentation Using a CNN Architecture . . . . .	100
5.3.4	Statistical Analysis . . . . .	103
5.4	Results . . . . .	104
5.4.1	Characteristics of the Training Cohort . . . . .	104
5.4.2	Reducing Acquisition Time . . . . .	105
5.4.3	Accuracy of Manual Segmentation . . . . .	106
5.4.4	Network Testing . . . . .	108
5.5	Discussion . . . . .	113
5.5.1	Evaluation of Methodology . . . . .	114
5.5.2	Future Directions . . . . .	116
5.6	Conclusions . . . . .	118
5.7	Acknowledgements . . . . .	118
5.8	References . . . . .	119
<b>6</b>	<b>Ex-Vivo Renal MRI</b>	<b>124</b>
6.1	Introduction . . . . .	125
6.2	Methods . . . . .	127
6.2.1	Sample Acquisition and Fixation . . . . .	127
6.2.2	$T_1$ Mapping . . . . .	127
6.2.3	$T_2$ Mapping . . . . .	129
6.2.4	$T_2^*$ Mapping . . . . .	130
6.2.5	Apparent Diffusion Coefficient . . . . .	131
6.2.6	Diffusion Tensor Imaging . . . . .	131
6.2.7	Layer Based Analysis . . . . .	133
6.3	Results and Discussion . . . . .	136
6.3.1	Fixation and Protocol Development . . . . .	136
6.3.2	Monitoring Changes in MR Parameters Post Fixation . . . . .	138
6.3.3	Comparing MR and Histological Measures in Aged Kidneys . . . . .	143
6.4	Conclusions and Future Work . . . . .	144
6.4.1	Protocol Validation on a Single Sample . . . . .	145
6.4.2	Ex-Vivo Sample Coil . . . . .	145

6.4.3	Human Organs . . . . .	146
6.5	Acknowledgements . . . . .	146
6.6	References . . . . .	147
<b>7</b>	<b>Conclusion</b>	<b>150</b>

# **Todo list**

■ Make sure these are in the correct order . . . . .	5
Figure: Flow phantom schematic . . . . .	43

## Chapter 5

# Automated Segmentation of Kidneys using Machine Learning

### Abstract

Total Kidney Volume (TKV) is an important measure in renal disease detection and monitoring. Here a fully automated method to segment the kidneys from  $T_2$ -weighted Magnetic Resonance Imaging (MRI) to calculate TKV of Healthy Control (HC) and Chronic Kidney Disease (CKD) patients is developed.

This automated method uses machine learning, specifically a 2D Convolutional Neural Network (CNN), to accurately segment the left and right kidneys from  $T_2$ -weighted MRI data. The dataset consisted of 30 HC subjects and 30 CKD patients. The model was trained on 50 manually defined HC and CKD kidney segmentations. It was subsequently evaluated on 50 test data sets, comprising data from five HCs and five CKD patients each scanned five times in a scan session to enable comparison of the precision of the CNN and manual segmentation of kidneys.

The unseen test data processed by the 2D CNN had a mean Dice score of  $0.93 \pm 0.01$ . The difference between manual and automatically computed TKV was  $1.2 \pm 16.2 \text{ mL}$  with a mean surface distance of  $0.65 \pm 0.21 \text{ mm}$ . The variance in TKV measurements from repeat acquisitions on the same subject was significantly lower using the automated method compared to manual segmentation of the kidneys.

The 2D CNN method provides fully automated segmentation of the left and right kidney and calculation of TKV in under ten seconds on a standard office computer, allowing high data throughput and is a freely available executable.

This work was presented as an aural presentation at the International Society of Magnetic Resonance in Medicine (ISMRM) 28th Annual Meeting (2020) [1].

## 5.1 Introduction

Segmentation of the kidneys from Magnetic Resonance Imaging (MRI) is a time consuming aspect of many renal MRI studies [2–4]. Total Kidney Volume (TKV) gives insight into renal function and is therefore used as a measured parameter for a variety of renal pathologies. The use of TKV is an active area of ongoing research for Autosomal Dominant Polycystic Kidney Disease (ADPKD), which is characterised by an increase in TKV due to cyst formation. Disease progression can be monitored by recording TKV, with higher rates of TKV increase being associated with a more rapid decrease in renal function [5–7]. Measurements of TKV in Chronic Kidney Disease (CKD) subjects have shown a significant correlation with glomerular filtration rate [8], the primary measure of CKD severity [9], with more generally a decrease in TKV associated with a decrease in renal function [10]. When studying pathologies which commonly lead to a change in kidney function, total kidney perfusion is often measured, this metric relies on an accurate measurement of renal blood flow and kidney volume of each kidney, and allows investigators to ascertain if the blood flow is preserved as the organ changes in size or if tissue perfusion is impaired. In addition to TKV measurements, renal segmentation is an important first step for many other processing pipelines, be that for automated cortical-medullary segmentations or to carry out multiparametric mapping within only the kidney to reduce computation times.

The gold standards of kidney segmentation are manual Region Of Interest (ROI) boundary tracing [11] or stereology [12] by experienced and skilled experts, with blood vessels in the kidney and the hilum excluded. These manual processes are highly time consuming (taking approximately 15 – 30 minutes per subject [13–15] and can be biased by investigator judgement due to the similar signal intensities between the kidneys and surrounding organs, anatomical differences between subjects, cysts and image artefacts. Consequently, the resulting kidney ROIs produced are subject to intra- and inter-expert variability as a result of the varying expertise levels; experts may segment a specific image differently when performed more than once, or different experts may segment the same image differently. These factors mean that the development of a faster and ideally fully automated method of renal segmentation is highly desirable. However the same factors that make manual segmentation difficult can also limit fully automated methods, for example the signal intensity of the kidneys closely matches that of other abdominal structures such as the spleen.

A number of automated methods have been proposed with varied success [13]. Some simply assume the kidney is an ellipse and calculate the volume from measurements of the pole-to-pole distance [16, 17] or include a correction factor to reduce overestimations [18]. Unfortunately these techniques produce a large confidence interval and still require human intervention to define the pole-to-pole length, a process that can produce inconsistencies between readers and takes a reasonable amount of time ( $\approx 5$  min) [19]. Other semi-automated methods use classical image processing techniques such as thresholding [20], water-shedding [21], level sets [15, 22], and spatial prior probability mapping [23]. These methods can either be inaccurate, over-segmenting the kidneys, or include a number of parameters that need to be manually adjusted and are computationally intensive. Further, the fact that each technique is highly optimised for a specific dataset means that it needs to be re-written to be applied to different pathology, another time consuming and highly skilled process.

Machine learning methods have the potential to automatically detect different patterns from data given to a model which has been trained. Deep learning is a class of machine learning algorithms that can model high-level information in an image using several processing layers of transformations. This uses an architecture of multi-level linear and non-linear operations, described by layers, to learn complex functions that can represent high-level detail to map the input data to the output segmentations directly. As more data becomes available the algorithm can become more accurate and generalised, without a need to rewrite the underlying methods, thus making it a good choice for long term development.

In recent years, deep learning-based methods have been applied to the segmentation of medical images, especially successful has been the U-Net [24]. This modified fully Convolutional Neural Network (CNN) architecture uses a number of convolution, pooling and up-sampling layers to detect features in the input data at multiple resolutions. The convolution layers convolve a learnable kernel with the input data to generate spatial feature maps that are passed to subsequent layers in the network. By adjusting the kernels, the resulting feature maps can be optimised to detect the location of the kidneys. Pooling layers are used to down-sample the data and allow some convolution kernels to become tuned to approximate features, this also reduces the tendency of the network to overfit the training data. When the data has been fully down-sampled, up-sampling layers are used to increase the resolution of the feature maps back to that of the original data while more convolution layers also learn the precise location of the kidneys. Parameters are adjusted

## 5.1. Introduction

---

by comparing the output from the network to a known ground truth. CNN methods have been applied to segmentation in other areas of medical imaging [25–28], for example to prostate segmentation of MRI images [29], liver segmentation of x-ray Computed Tomography (CT) images [30] and segmentation of polycystic kidneys [31–33]. However, to date, these methods have not been successfully applied to CKD and healthy kidney segmentation from MR images.

Here a single 2D U-Net model CNN is used for the segmentation of the kidneys in both Healthy Control (HC) participants and CKD patients using  $T_2$ -weighted MR images. Automatically generated kidney masks are compared with manual masks defined by experts and assessed for similarity using multiple voxel and surface based metrics and total segmented volume. A subset of subjects were scanned multiple times to assess the repeatability of the segmentations.

## 5.2 Neural Networks for Image Segmentation

### 5.2.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) aim to solve computational problems using a similar methodology to their biological namesake. Input data is passed through a series of connected nodes or neurons, each of which can have multiple input and output connections from and to other neurons mimicking synapses. At each neuron, a weighted sum of the input values is calculated before being passed onto the next hidden layer of neurons. The final layer of neurons is connected to the output layer which will give an estimation of the desired property, be that a number e.g. probability someone will like a television program, an image e.g. the probability that a pixel in an image is a road sign, or a sample in a time series e.g. audio in voice synthesis. More concisely, an ANN can be used to map a non-linear set of input data to an output dimension.

A very basic example could use the mass and colour of an animal to guess if it is a dog or a cat, Figure 5.1. The connections between neurons are initialised with random weights.

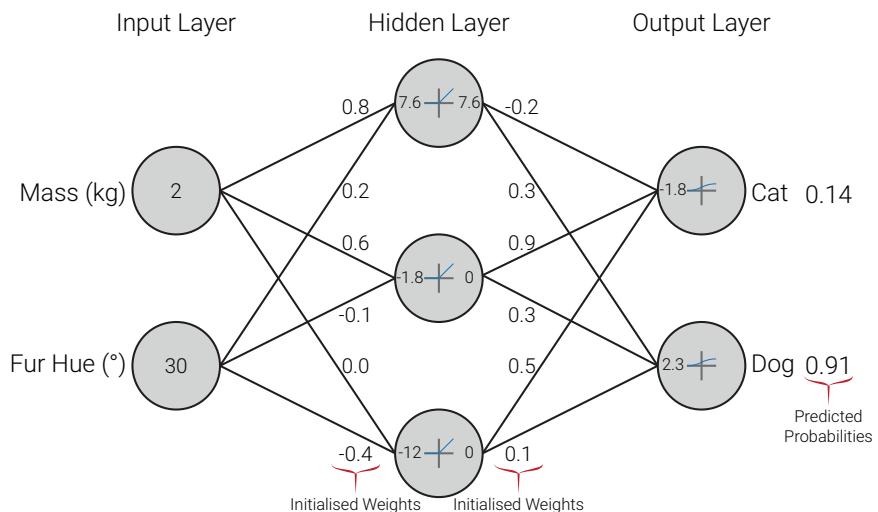


Figure 5.1: The ANN initialised with random numbers trying to predict the species of a small ginger cat.

At each neuron, a weighted sum of its inputs is taken, then an activation function applied, here a Rectified Linear Unit (ReLU), Figure 5.2a, for the hidden neurons and sig-

moid, 5.2b, for the output neurons. These activation functions allow the network to act non-linearly and are modelling the action potential of biological neurons. The ReLU function represents a higher rate of firing for signals above zero; as it is impossible for a biological neuron to reduce its firing rate below zero, the ReLU outputs zero when the input signal is negative. The sigmoid function maps all values between zero and one, and therefore ensures the network outputs a probability at the output nodes.

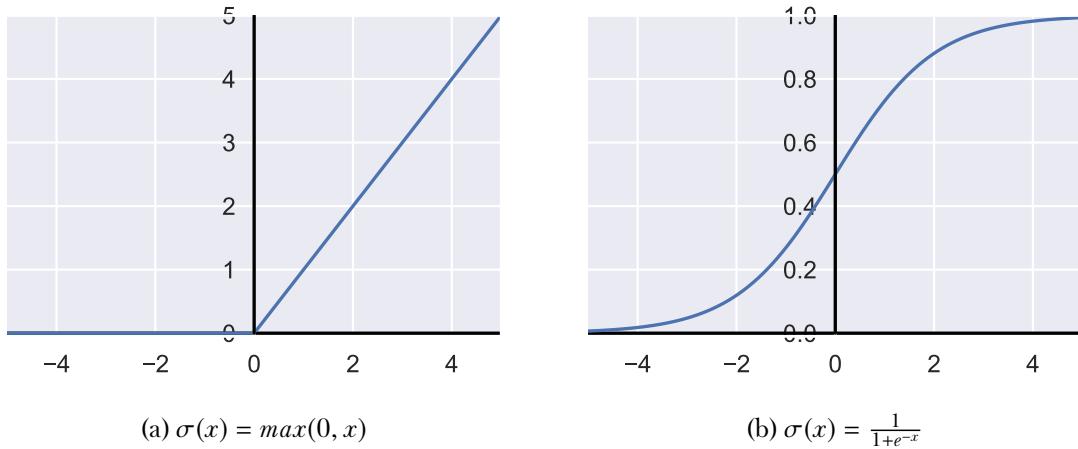


Figure 5.2: Activation functions.

As the weights were randomly initialised, the network has incorrectly predicted that the animal is a dog. By comparing the result output from the network to the known ground truth, the weights of the network can be adjusted in a process known as back propagation, Figure 5.3. Hyper-parameters such as learning rate and momentum control how much each weight is adjusted in response to the input data and subsequent result.

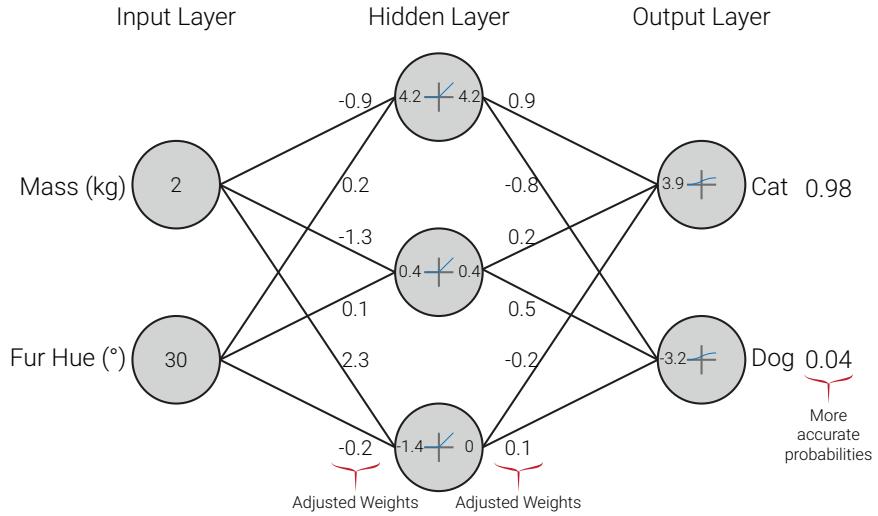


Figure 5.3: The weightings of each connection are adjusted so the output layer produces results closer to the ground truth.

When another animal is input to the network, here a smaller, darker coloured cat, the network now correctly predicts that it is a cat, 5.4. By repeating this procedure many times, comparing the result to ground truths and adjusting the weights, a process known as training, the network becomes more and more accurate. Once trained, the network can be used to infer the species of animals with no ground truth data.

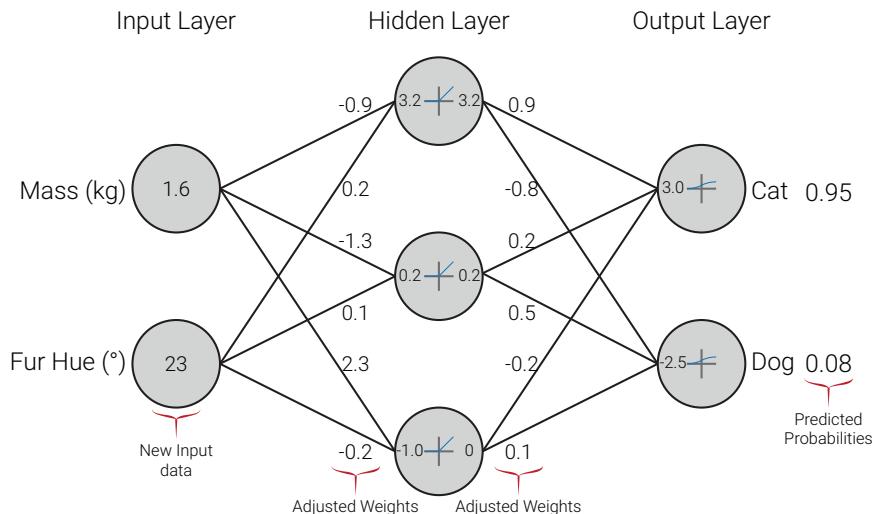


Figure 5.4: New data is presented to the network in the form of a smaller, darker coloured cat and the process of adjusting weights is repeated until more accurate results are produced.

The above example is highly simplified, real ANNs will have many more input nodes and hidden layers. In the case of imaging data, the input layer will simply be a node for each pixel in the image.

### 5.2.2 Convolutional Neural Networks

It was found that ANN segmentation performance was increased if additional features were input to the network. These features could be colour rather than greyscale date, different MRI contrasts e.g. fat/water images or artificially generated features. In Figure 5.5 a selection of artificially generated features are presented, some of these highlight the kidneys from the surrounding tissue e.g. the intensity range adjustment, 5.5b, and edge detector, 5.5f while others are better at providing contrast between the cyst in the right kidney and the renal tissue e.g. the intensity inversion, 5.5c and sharpening filter, 5.5e.

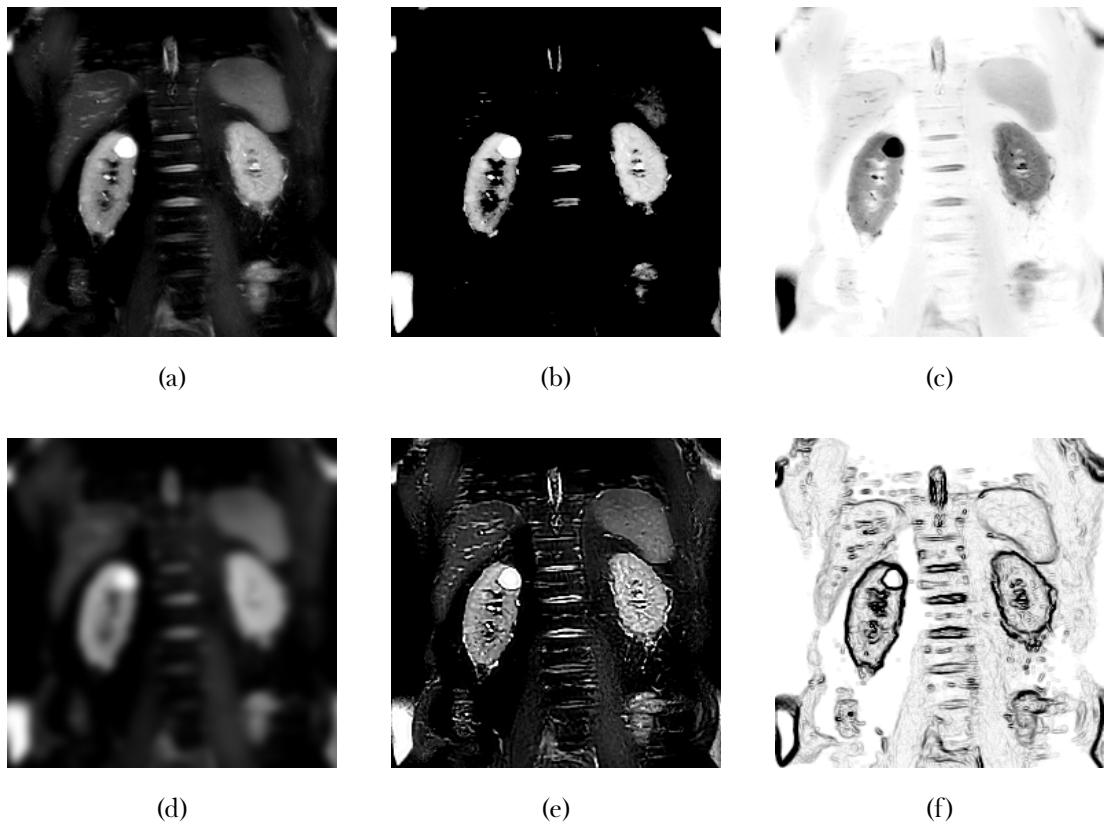


Figure 5.5: An example of the features that can be generated from a raw image (a). Implemented here are intensity range adjustments (b), intensity inversion (c), Gaussian blur (d), sharpening (e) and edge detection (f).

Many of these artificially generated features can be implemented as convolutional operations, these involve convolving a numerical kernel with every pixel in the image. By adjusting the values of each cell in the kernel, different features or filters can be produced. The control of these kernels can be handed over to similar optimisation processes to those used to adjust the weights of the connections between neurons. Over the training period, this enables the network to learn what features are useful for the task at hand and which are less useful rather than the network being given features that the programmer thinks will be helpful. Shallow layers of the network usually resemble features similar to those in 5.5 while deeper layers represent more complex and specific objects such as, using the cat and dog example above, pointy noses to distinguish dogs and triangular ears to distinguish cats. This architecture is known as a Convolutional Neural Network [34, 35].

To aid with feature extraction, the raw image is often downsampled by max-pooling layers, this enables different kernels to act on different scales of the image can help keep the network generalisable and avoid overfitting. The U-Net architecture [24] combines an arm with downsampling and feature extraction with an upsampling arm that returns the image to its initial dimensions making it especially useful for segmentations tasks.

## 5.3 Methods

The study was approved by the University of Nottingham Medical School Research Ethics Committee (H14082014 and E14032013), and East Midlands Research Ethics committee REC reference: 17/LO/2036 and 15/EM/0274.

### 5.3.1 MRI Data Acquisition

All kidney MRI scans were acquired on a 3T Philips Ingenia system (Philips Medical Systems, Best, The Netherlands) using a 2D  $T_2$ -weighted Half-Fourier Single-shot Turbo spin Echo (HASTE) sequence optimised to achieve the maximum contrast between the kidneys and surrounding tissue (Echo Time (TE) = 60 ms, Repetition Time (TR) = 1300 – 1800 ms, Sensitivity Encoding (SENSE) factor = 2.5, refocus angle 120°, bandwidth, 792 Hz, Field Of View (FOV) = 350 x 350 mm<sup>2</sup>, voxel size = 1.5 x 1.5 x 5 mm<sup>3</sup> and a slice gap of 0.5 mm with approximately 13 coronal slices, enough to image the entire kidney [36, 37], in a single 17 - 23 s breath hold.

The dataset consisted of 60 subjects, 30 HC (10 female, 20 male) with a mean age of  $26 \pm 11$  (19–77) years and 30 CKD patients (6 female, 24 male) with a mean age of  $59 \pm 14$  (19–80) years and mean CKD Stage  $3.5 \pm 1.2$  (1-5). Ten of the subjects (5 HCs and 5 CKD patients) were scanned five times in the same scan session for use as test data. In each test data scan session, subjects were repositioned between each acquisition (removed from the scanner, asked to sit up and move on the bed), additionally the scanner operator attempted to vary the acquisition geometry between each scan while still acquiring full kidney coverage. These repeated test datasets allow the consistency of the networks ability to measure TKV to be assessed.

In total, 649 2D image slices from the 50 subjects in the training data and 650 2D image slices from the 10 subjects in the test data, were collected. A summary of the data collected is provided in Table 5.1 and Figure 5.9.

## 5.3. Methods

---

### 5.3.2 Manual Segmentation

The manual binary mask of the kidneys of each subject were generated by one of three observers (A, B and C who had been trained on kidney segmentation and had an average of 2 years of experience), with each observer segmenting data from both the training and testing datasets. Kidney boundaries were manually traced using freely available software (MRIcron [38]) and any area of non-renal parenchyma, such as the renal hilum and cysts, were excluded from the manual definition. Binary masks of the kidney were generated, and the volume of each kidney was computed from the product of the number of voxels in each kidney mask and the voxel volume. Separate kidney volume for the left and right kidneys was determined and summed to compute TKV. All measurements were performed by observers blinded for patient number and previous TKV measurements.

For the training phase, for each subject a manual mask was used from a single observer (randomised between observer A, B, or C). For the testing phase, all five scans from a given subject were segmented by a single reader with the ten subjects being segmented by a mix of the three readers i.e. the test data comprised of subjects segmented by all readers but the repeat scans of each subject were segmented by the same reader. For four HC subjects from the test dataset, manual masks were drawn by all three observers for all five repeat acquisitions to allow assessment of inter-observer variability in the manual masks. HCs were chosen for this analysis as they healthy kidneys have a more consistent morphology and thus will give a best-case measure of observer variability and provide a comparison of the automated method to the highest standard of manual segmentation.

### 5.3.3 Automated Segmentation Using a CNN Architecture

Voxel intensities were normalised between 0 and 255, where 0 was set to the mean voxel intensity minus 0.5 times the standard deviation of that slice and 255 was set to the mean voxel intensity plus four times the standard deviation of the volume. This empirically derived windowing led to a clear contrast between the kidneys and surrounding tissue while negating the effects of bulk signal changes between volumes. Each dataset volume was then split into 2D coronal slices and resampled to a matrix size of  $256 \times 256$ . Twenty percent of slices were reserved for validation during the network optimisation process, this validation data was used to monitor over-fitting and direct the optimisation process between

### 5.3. Methods

---

epochs. Once the data had been split into training and validation sets, the slice order was randomised within sets. Splitting the data before slice randomisation limited the possibility of slices from only one subject being split over both the training and validation datasets. During training, data augmentation was applied. At the start of each epoch, a batch of images and their corresponding masks was selected at random from the training data and a series of random shifts (up to 25 % of the image in both the horizontal and vertical direction), zooms (between 0.75 and 1.25 magnification), rotations (within a 20° range), and sheers (within a 5° range) were applied to the image/mask pair to produce different yet anatomically reasonable images. The weights of the network were then adjusted based on this augmented data before selecting a new batch of images for the next epoch. Augmenting the data reduces the tendency of a model to over-fit the training data and thus increases accuracy when the model is applied to unseen images.

The U-Net consists of two Fully Convolutional Neural Network-like structures that are cascaded in the form of an encoder-decoder (autoencoder) structure. The encoder is used for feature extraction and the decoder is used for feature mapping to the original input resolution. A summary of the network architecture is shown in Figure 5.6. The convolution layers use a set of small parameterised filters, referred to as kernels, to perform convolution operations to produce different feature maps of their input. Here each convolution and deconvolution layer uses a 3 x 3 kernel. Activation layers use a ReLU. Following convolution at each resolution, max pooling with a stride 2 is used on the encoding half of the network.

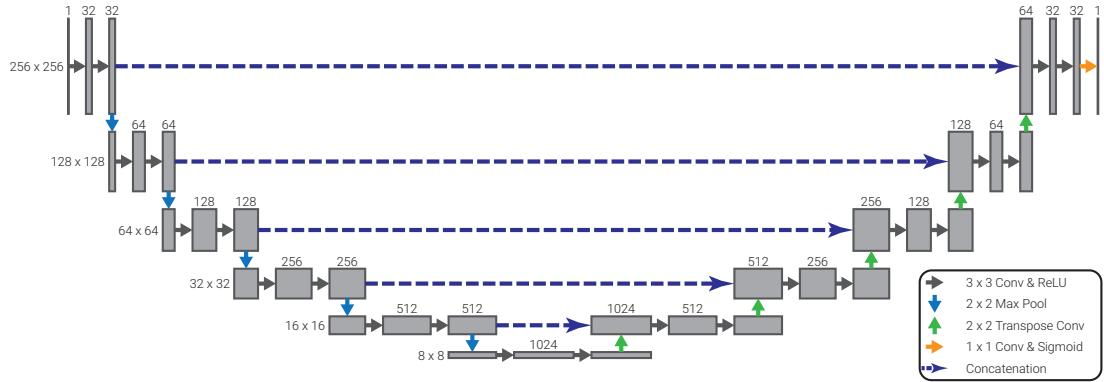


Figure 5.6: The architecture of the network used.

The network was implemented using Keras (v2.2.4) [39] with a TensorFlow backend (v1.13.1) [40] in Python 3.6.9. All training was carried out on an NVIDIA Titan Xp Graphical Processing Unit (GPU) (3840 CUDA cores, 12 GB GDDR5X). The network

### 5.3. Methods

---

uses a Dice score loss function, given by,

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|} = \frac{2TP}{2TP + FP + FN}, \quad (5.1)$$

where  $TP$  is true positive,  $FP$  is false positive and  $FN$  is false negative. A value of 1 implies complete overlap between the automated mask and the manual mask while 0 implies no overlap. This function is ideal for renal segmentation as it does not weight true negatives which represent the majority of voxels input to the network and thus means that while the network is training, it does not become trapped in a local minimum outputting solely background voxels. Training was carried out over 150 epochs using stochastic gradient descent with an initial learning rate of 0.01 and learning rate decay of  $5 \times 10^{-7}$  and momentum of 0.8, these parameters help the optimiser converge quickly while also avoiding overshooting. As seen in Figure 5.7, after 150 epochs the validation Dice score plateaued while the training Dice score was still rising slightly, indicating that any further training would lead to over-fitting. Training took approximately thirty minutes.

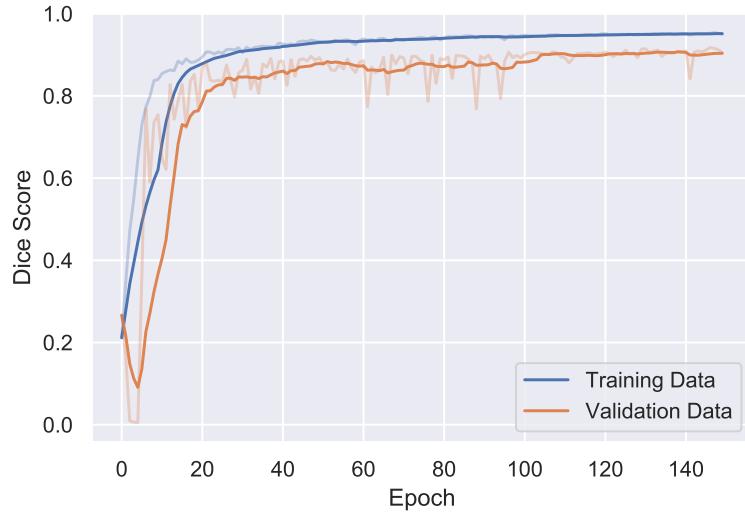


Figure 5.7: Dice score of the network for the training and validation data. Data is shown with a 10 epoch rolling average.

### 5.3.4 Statistical Analysis

Baseline demographics are reported as mean  $\pm$  Standard Deviation (SD). Inter-observer variability in manual segmentation and TKV was calculated by comparing the TKV of the manual masks each observer generated for a given volume, and also assessing the Bland-Altman and regression analysis. Intra-observer variability in manual segmentation was calculated by comparing the TKV of the five masks generated by an observer for a given subject. For each, the mean Coefficient of Variation (CoV); defined as standard deviation/mean and Intraclass Correlation (ICC) were used as measures of repeatability of TKV. Voxel-based metrics (Dice score, Equation (5.1) and Jaccard index, Equation (5.2)) and surface based metrics such as the average distance between the surface of the two masks and Hausdorff Distance 95th percentile, the 95th percentile of the largest distance between the two surfaces, were also calculated between each observer.

$$J(A, B) = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN} \quad (5.2)$$

The performance of the automated segmentation was assessed using the voxel and surface based similarity metrics outlined above and, in addition, sensitivity, specificity, precision and accuracy, Equations (5.3) - (5.6). Performance was further assessed by determining the mean difference in TKV between the automatic and manual methods. Both actual and percentage (%) difference in TKV were evaluated. Bias (mean) obtained from the automatic and manual methods was assessed using a paired sample t-test. The mean CoV and ICC were also used as measures of repeatability of the automated TKV.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5.3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5.4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5.5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.6)$$

## 5.4 Results

### 5.4.1 Characteristics of the Training Cohort

Data was collected using a  $T_2$ -weighted HASTE sequence providing optimal contrast between the kidneys and surrounding tissue, examples shown in Figure 5.8, however there is limited contrast between the left kidney and spleen due to their similar  $T_2$ -weighting. Cysts of variable size are clearly visible in the kidneys of the CKD patient.

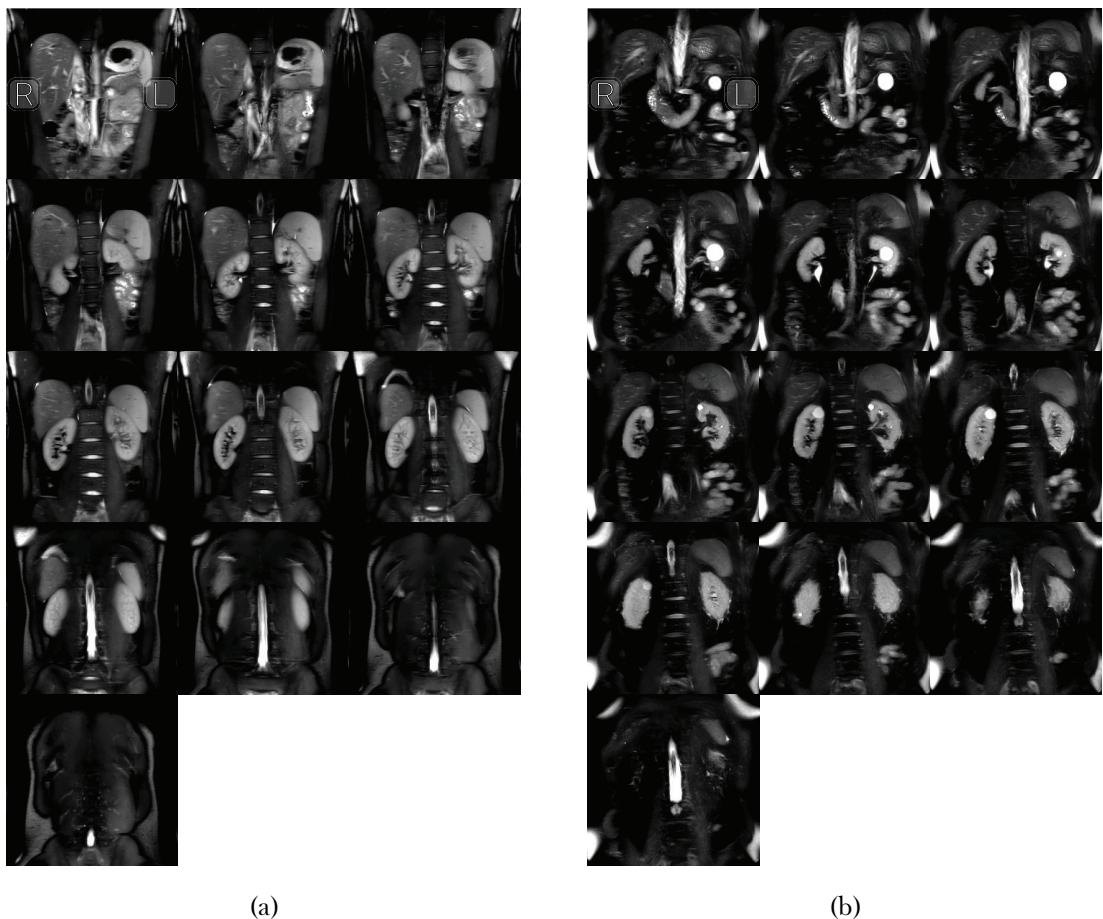


Figure 5.8: All slices of the raw data from representative subjects of the HC cohort, shown in a, and CKD cohort, b.

The training data comprised 25 healthy controls (9 female, 16 male) with a mean age of  $26 \pm 12$  (19–77) years and 25 CKD patients (6 female, 19 male) with a mean age of  $58 \pm 15$  (19–80) years and mean CKD stage  $3.3 \pm 1.1$  (1–5). The manual TKV was  $277 \pm 60$  mℓ, ranging between 145 and 422 mℓ. Including both healthy control subjects and CKD patients meant the kidneys had variable morphology (shape, size and heterogeneous

## 5.4. Results

---

cysts) within the training dataset. Table 5.1 provides the characteristics of datasets used for training and testing of the CNN, whilst Figure 5.9 shows the distribution of TKV within the training and testing data.

	Number of Subjects	Number of Datasets	Number of 2D Slices	Sex (F/M)	Mean Age	TKV (mℓ)
Training HC	25	25	325	9/16	$26 \pm 12$	$296 \pm 38$
Training CKD	25	25	324	6/19	$58 \pm 15$	$258 \pm 72$
Testing HC	5	25	325	1/4	$25 \pm 3$	$330 \pm 35$
Training CKD	5	25	325	0/5	$69 \pm 3$	$274 \pm 56$

Table 5.1: Characteristics of datasets used for training and validation of the 2D U-Net model CNN.

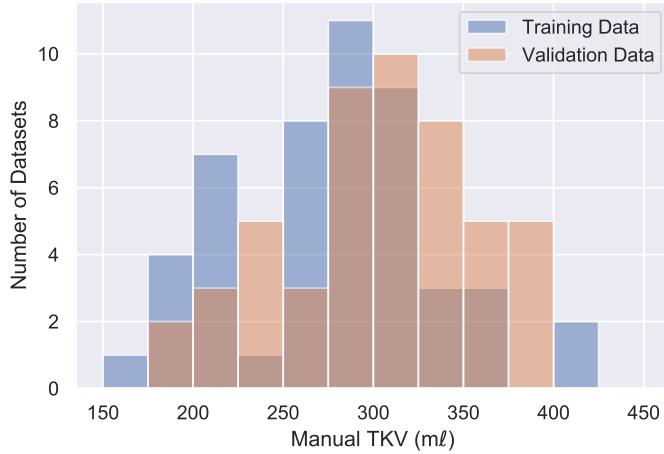


Figure 5.9: Distribution of TKV within the training and testing data.

### 5.4.2 Reducing Acquisition Time

Initial data was collected with a TR of 1800 ms leading to a breath hold of approximately 28 seconds. Some subjects struggled to hold their breath for this long on expiration, therefore the effects of reducing the TR of the sequence were investigated. As can be seen in Figure 5.10, there is no degradation in image quality from the image with TR of 1800 ms to that with at TR of 1300 ms, the differences between these images are mainly due to the small movements between volumes, as can be seen in the difference data where the largest

## 5.4. Results

---

differences are seen around the periphery of the kidneys and in the gut. Moving forward, the TR was reduced to 1300 ms leading to a sequence with a breath hold of approximately 17 seconds.

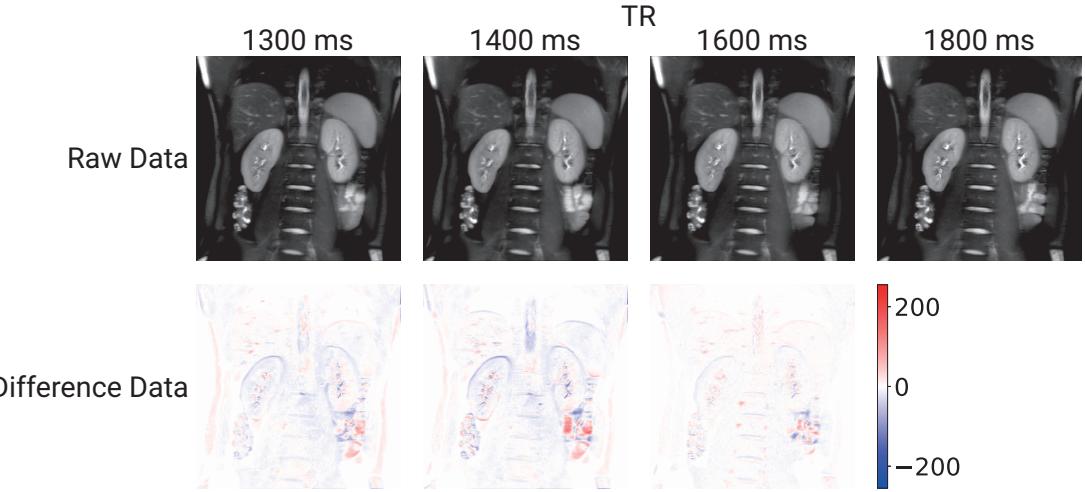


Figure 5.10: The effects of changing the TR of the sequence.

### 5.4.3 Accuracy of Manual Segmentation

Four of the test subjects were each scanned five times, with the left and right kidneys in the 20 datasets each masked by Observers A, B and C. The intra-observer and inter-observer variability for this manual segmentation was computed, as shown in Table 5.2.

Observer	Kidneys	CoV (%)	ICC
Intra A	Total	$2.2 \pm 0.7$	0.939
	Left	$3.2 \pm 0.8$	0.783
	Right	$1.9 \pm 0.5$	0.957
Intra B	Total	$1.9 \pm 0.3$	0.895
	Left	$2.0 \pm 0.5$	0.807
	Right	$2.4 \pm 0.3$	0.892
Intra C	Total	$2.5 \pm 0.9$	0.908
	Left	$2.8 \pm 1.3$	0.769
	Right	$3.1 \pm 1.9$	0.940
Inter	Total	$3.0 \pm 1.0$	0.897
	Left	$4.0 \pm 1.4$	0.713
	Right	$2.9 \pm 1.0$	0.910

Table 5.2: Repeatability of the manual segmentation for left, right and TKV, with coefficient of variation and intraclass coefficient computed.

Additionally, similarity metrics were used to assess the overlap between each observer's

## 5.4. Results

manual masks, Table 5.3. Due to the large difference between in-plane and out-of-plane resolution ( $1.5 \text{ mm}^3$  vs  $5.5 \text{ mm}^3$ ) the Hausdorff distance is very susceptible to inaccuracies in the anterior – posterior direction; this metric is highly sensitive to noise and as such the 95th percentile is used to generate a more representative value. Bland-Altman plots and regression analysis of inter-observer variance in measured TKV are provided in Figure 5.11.

Observer	Kidney	Dice Score	Jaccard Index	Average Distance (mm)	Hausdorff Distance (mm) (95th Percentile)	Volume Difference (ml)
A – B	Both	$0.93 \pm 0.03$	$0.87 \pm 0.05$	$0.81 \pm 0.58$	$5.59 \pm 2.77$	$20.84 \pm 9.33$
	Left	$0.92 \pm 0.07$	$0.85 \pm 0.10$	$0.94 \pm 1.12$	$5.53 \pm 3.65$	$13.36 \pm 5.76$
	Right	$0.94 \pm 0.01$	$0.88 \pm 0.02$	$0.65 \pm 0.14$	$4.75 \pm 1.15$	$7.48 \pm 5.63$
A – C	Both	$0.93 \pm 0.01$	$0.87 \pm 0.02$	$0.79 \pm 0.18$	$5.83 \pm 1.86$	$16.01 \pm 8.56$
	Left	$0.93 \pm 0.01$	$0.87 \pm 0.02$	$0.84 \pm 0.27$	$6.83 \pm 3.12$	$6.93 \pm 5.78$
	Right	$0.93 \pm 0.01$	$0.87 \pm 0.02$	$0.72 \pm 0.17$	$4.82 \pm 1.25$	$9.08 \pm 5.41$
B – C	Both	$0.94 \pm 0.04$	$0.89 \pm 0.06$	$0.63 \pm 0.62$	$3.59 \pm 2.74$	$-4.83 \pm 9.92$
	Left	$0.93 \pm 0.08$	$0.88 \pm 0.11$	$0.78 \pm 1.22$	$4.31 \pm 3.58$	$-6.44 \pm 6.17$
	Right	$0.95 \pm 0.01$	$0.90 \pm 0.02$	$0.48 \pm 0.14$	$3.39 \pm 1.15$	$1.61 \pm 6.56$

Table 5.3: Metrics comparing each combination of observers manual masks (A – B, A – C and B – C). All values are quoted as mean  $\pm$  standard deviation.

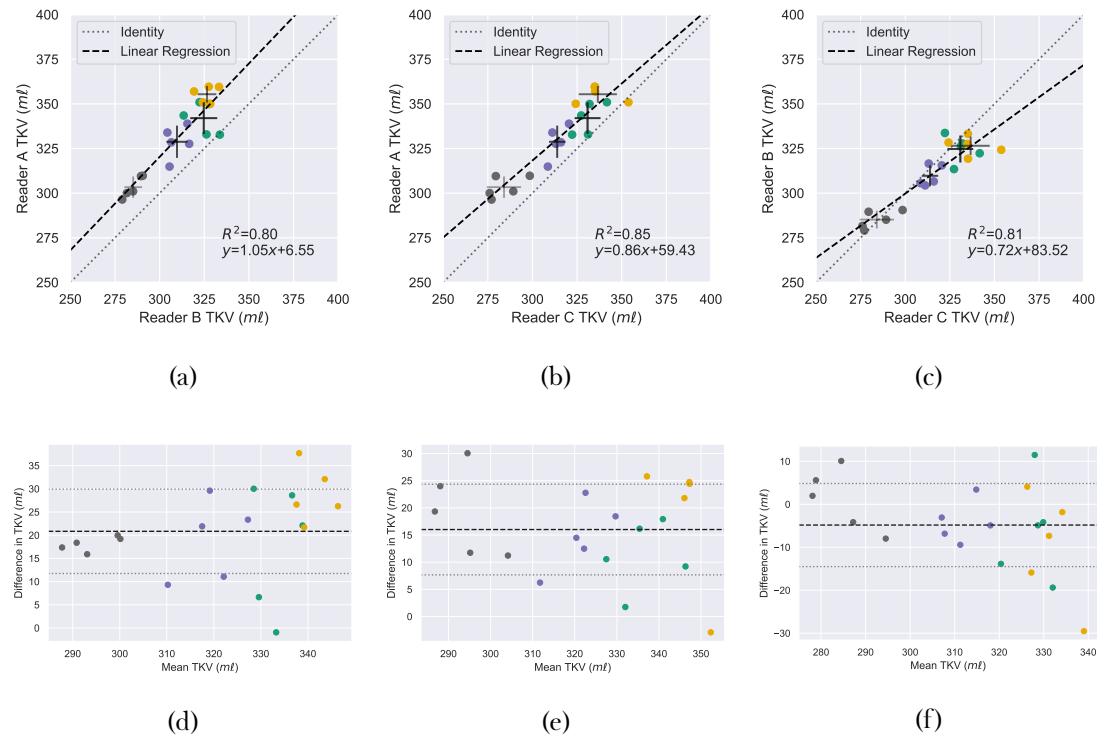


Figure 5.11

#### 5.4.4 Network Testing

To verify that the trained network is behaving as expected saliency maps were produced, Figure 5.12, this is especially important given the black box nature of machine learning methods. This map shows the areas the network is using most in its classification [41]. It verifies that the networks is using the outside areas of the kidney to make its prediction with areas of a similar intensity receiving some attention to distinguish them from the kidney. While this is precisely what is expected of the algorithm, it is important to check this as it is possible for such a method to have learnt a slightly different mechanism for the segmentation, one that is more prone to errors if new data is presented to it.

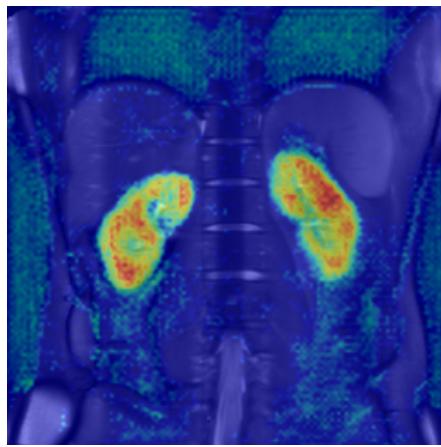


Figure 5.12: An example saliency map of the areas the network uses most when segmenting the kidney.

The trained network was used to predict segmentations of the 2D kidney slices and compute TKV for each of the unseen test volumes. The mean Dice score over the 50 test volumes was  $0.93 \pm 0.01$  ( $0.94 \pm 0.02$  for HC and  $0.92 \pm 0.01$  for CKD patients). The TKV predicted by the network was, on average,  $1.2 \pm 16.2$  mℓ less than the manually segmented TKV and thus not significantly different ( $p = 0.615$ ) (Figure 5.13) This accuracy was comparable for the HC and CKD cohorts, with automated CNN TKV measurements of  $4.7 \pm 17.7$  mℓ greater than manual and  $7.0 \pm 12.4$  mℓ less than manual respectively. A summary of the CNN accuracy when evaluated using similarity metrics and volume difference from manual measures is shown in Table 5.4. Note a slightly larger discrepancy for the left compared to the right kidney. Figure 5.13 shows plots of the difference in volume between manual segmentation and automated segmentation of the test dataset.

## 5.4. Results

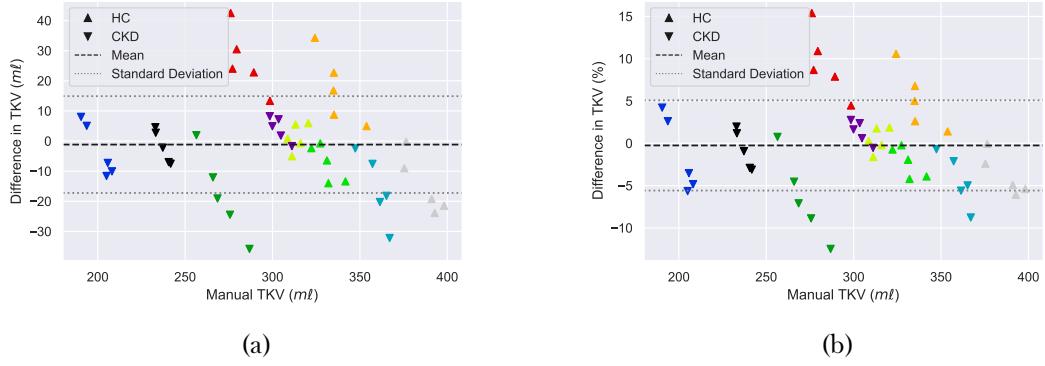


Figure 5.13: The difference between the TKV predicted by the CNN and the manually segmented true TKV. Mean and standard deviation TKV difference are shown as dashed and dotted lines respectively. Each subject is shown in a different colour. (a) shows the absolute volume difference (b) shows the percentage volume difference.

Cohort	Kidney	Dice Score	Jaccard Index	Sensitivity	Specificity	Precision	Accuracy	Mean Surface Distance (mm)	Hausdorff Distance (mm) (95th Percentile)	Volume Difference (p)
All	Total	0.93 ± 0.01	0.87 ± 0.03	0.93 ± 0.03	0.997 ± 0.001	0.93 ± 0.02	0.995 ± 0.001	0.65 ± 0.21	4.33 ± 1.64	-1.16 ± 16.23 (0.615)
	Left	0.92 ± 0.02	0.86 ± 0.04	0.91 ± 0.05	0.997 ± 0.001	0.94 ± 0.03	0.994 ± 0.002	0.76 ± 0.31	4.42 ± 1.52	-3.95 ± 12.38 (0.029)
	Right	0.94 ± 0.02	0.89 ± 0.03	0.95 ± 0.03	0.997 ± 0.001	0.93 ± 0.03	0.996 ± 0.001	0.54 ± 0.21	3.66 ± 1.76	2.79 ± 6.84 (0.006)
HC	Total	0.94 ± 0.02	0.88 ± 0.03	0.95 ± 0.05	0.997 ± 0.001	0.93 ± 0.03	0.995 ± 0.001	0.68 ± 0.27	4.50 ± 1.97	4.66 ± 17.72 (0.201)
	Left	0.93 ± 0.02	0.87 ± 0.04	0.94 ± 0.05	0.997 ± 0.001	0.93 ± 0.03	0.994 ± 0.002	0.79 ± 0.37	4.47 ± 1.81	1.91 ± 12.93 (0.467)
	Right	0.95 ± 0.02	0.90 ± 0.03	0.96 ± 0.03	0.997 ± 0.001	0.94 ± 0.02	0.996 ± 0.001	0.56 ± 0.26	3.81 ± 2.11	2.75 ± 7.70 (0.087)
CKD	Total	0.92 ± 0.01	0.86 ± 0.02	0.91 ± 0.02	0.998 ± 0.001	0.94 ± 0.02	0.995 ± 0.001	0.63 ± 0.14	4.16 ± 1.24	-6.98 ± 12.38 (0.009)
	Left	0.92 ± 0.02	0.85 ± 0.03	0.89 ± 0.04	0.998 ± 0.001	0.95 ± 0.02	0.994 ± 0.002	0.73 ± 0.24	4.37 ± 1.21	-9.81 ± 8.62 (0.00001)
	Right	0.93 ± 0.01	0.88 ± 0.02	0.94 ± 0.02	0.997 ± 0.001	0.92 ± 0.03	0.996 ± 0.001	0.51 ± 0.13	3.51 ± 1.34	2.83 ± 6.02 (0.027)

Table 5.4: The accuracy of the CNN compared to manual segmentations using a variety of metrics stratifying the testing data by cohort and left vs right kidney. All values are given as mean ± standard deviation.

## 5.4. Results

---

In Figure 5.14, the TKV predicted by the CNN is plot against the manual TKV, in ninety percent of subjects, the standard deviation of TKV measurements between each volume for a subject was smaller when the TKV was measured using the CNN as oppose to manually. The mean CoV and ICC were  $2.7 \pm 0.9\%$  and 0.979 respectively across the 5 repeats of the manually segmented test data (using masks from observers A, B and C), compared to a value of  $1.5 \pm 0.5\%$  and 0.993 respectively for the automatic segmentations of the 5 repeats of test data. The CNN produced a significantly lower CoV than the manual segmentations ( $p = 0.008$ ).

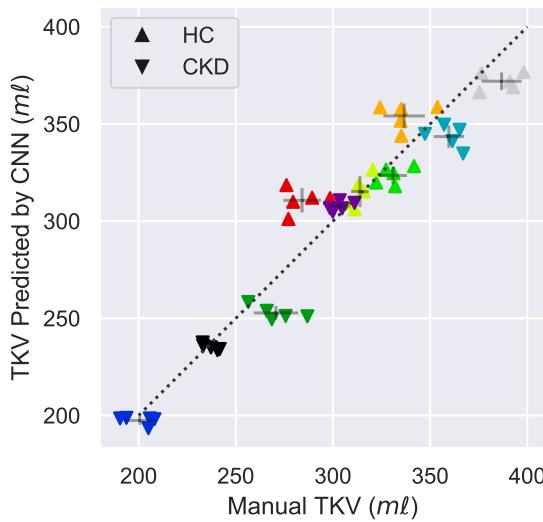


Figure 5.14: The TKV predicted by the CNN plot against the manually segmented true TKV with each subject plot in a different colour. The standard deviation measured using both methods is shown as error bars originating from the mean of each subject. The dotted line represents perfect correlation between the CNN and manual segmentation.

Representative examples of the output from the network for both HC and CKD data are shown in Figure 5.15. The automated CNN accurately segments the kidneys, and for CKD patients, often omits cysts from the masks.

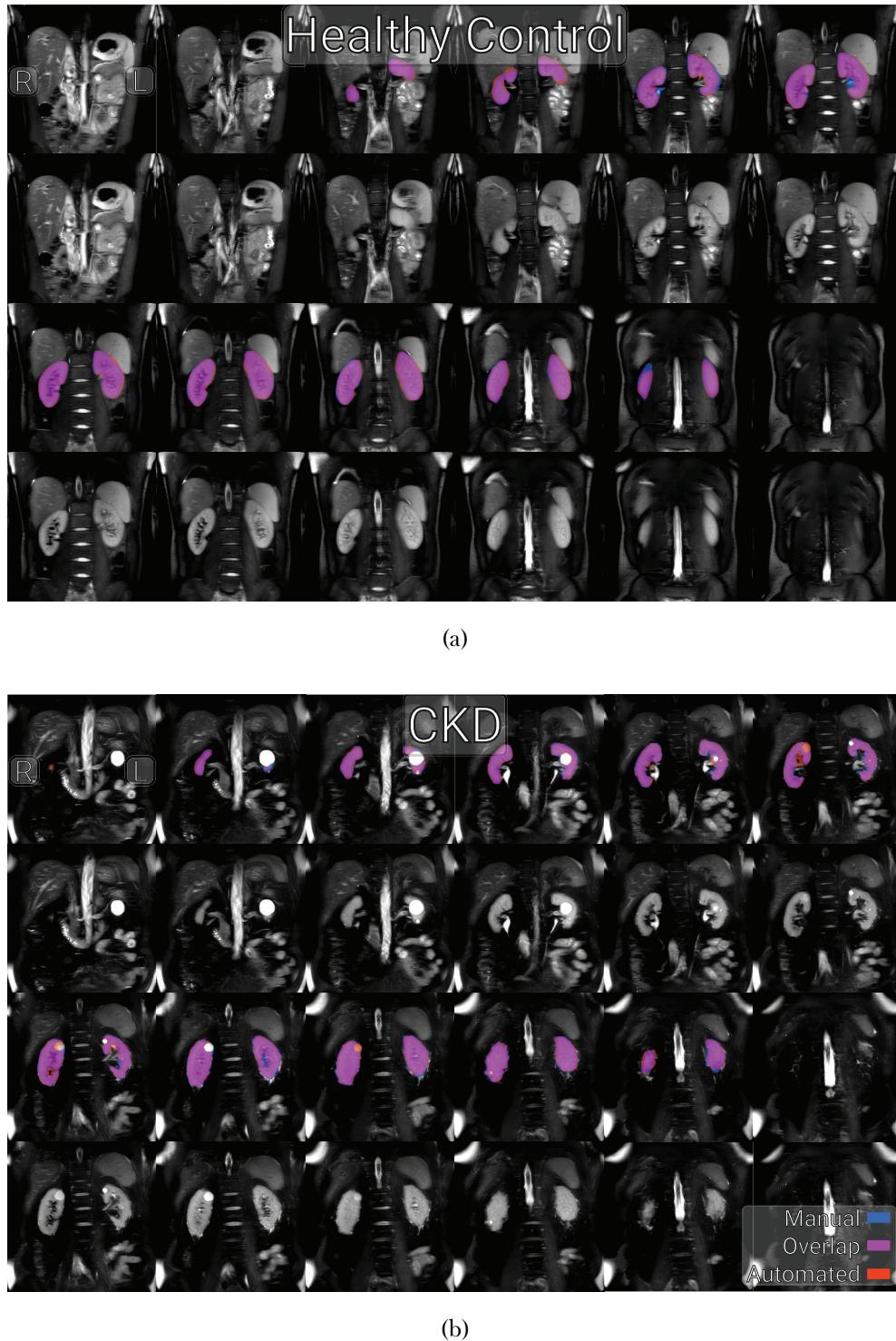


Figure 5.15: Representative raw test data and corresponding masks of a HC, (a), and CKD subject, (b). Manually generated masks are shown in blue, automatically generated masks are shown in red and the overlap of the two is shown in magenta.

Since this is a 2D CNN, it is important to assess the accuracy across the anterior – posterior 2D slices of the kidney. This was achieved by comparing the Dice score of the

## 5.4. Results

---

CNN to the inter-reader Dice scores, Figure 5.16. A decrease in accuracy in the outer slices can be seen in both the CNN and manual masks.

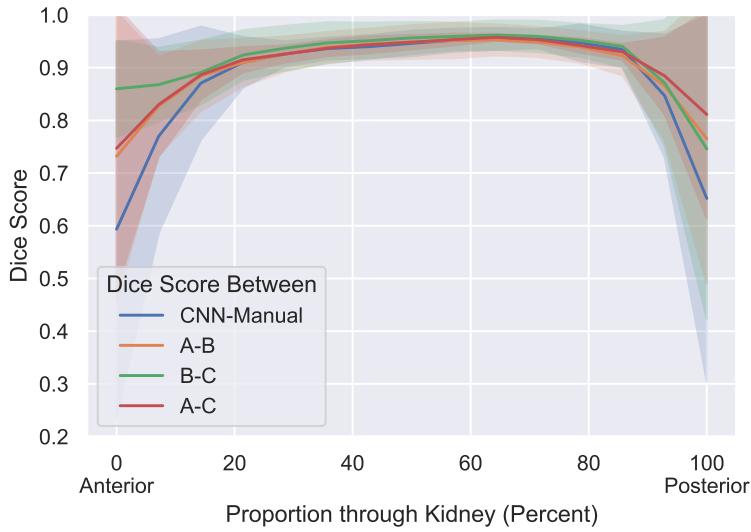


Figure 5.16: Mean Dice score for 2D slices from anterior to posterior. The shaded area represents one standard deviation from the mean Dice score.

This decrease in accuracy manifests itself on the outer slices of the volume, where the proportion of kidney per slice is smaller and as such the 2D network, with a lack of spatial context in the anterior – posterior direction, finds these outer slices more challenging. This decrease in accuracy can partly be explained by the fact that larger structures (in terms of number of voxels) will in general produce higher scores for comparable errors because the vast majority of errors are on the perimeter of the kidney in each slice, slices with fewer voxels of kidney have a smaller area to perimeter ratio, Figure 5.17.

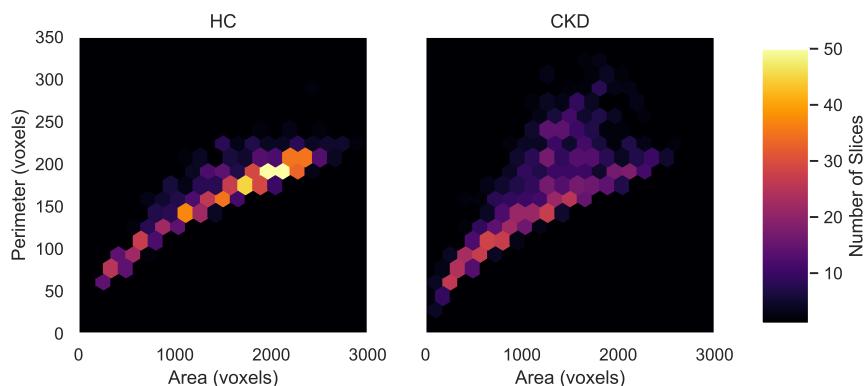


Figure 5.17: A 2D histogram of the perimeter and area of each slice for both the HC and CKD cohorts.

## 5.5 Discussion

In this chapter, a 2D CNN has been trained to generate automatic segmentations of healthy control and CKD patients. Segmentations of the left and right kidneys are computed from which total kidney volume is estimated. The CNN was trained on both healthy control and CKD kidneys with a range of TKV (144.76 – 422.49 m $\ell$ ) which included the presence of cysts. The automated segmentation by the CNN yielded a mean Dice score of  $0.93 \pm 0.01$  and took an average time of 9 s to measure TKV compared to 15 – 30 minutes [13] for manual segmentation. The automated CNN can be run as a self-contained Graphical User Interface (GUI) with the data and programme freely available [[daniel\\_alexander654/renal\\_segmentor\\_2020](#)] and thus avoid the need for complicated software setup. Note the software released at present can only be used to process coronal HASTE images and will not be accurate with other geometries/contrasts, for this, training of the network with a different dataset would be required and thus necessitate the use of a GPU.

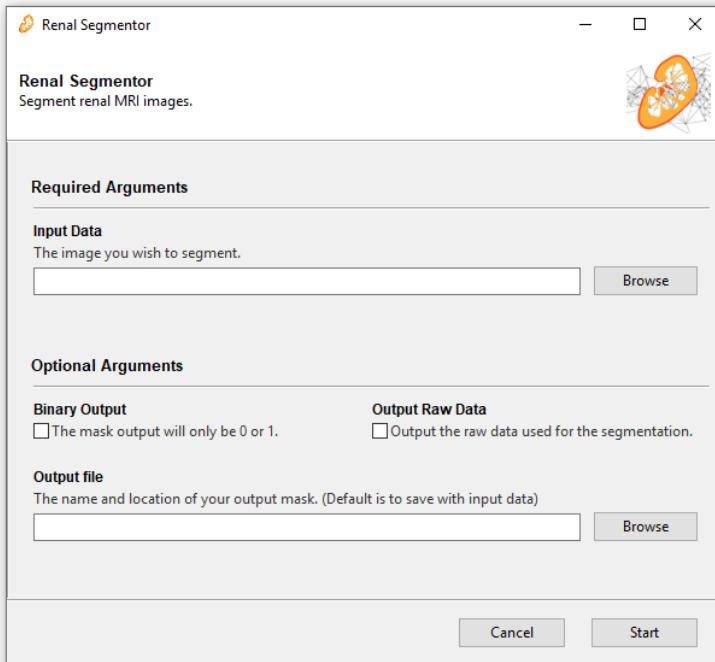


Figure 5.18: The Graphical User Interface used to segment kidneys.

### 5.5.1 Evaluation of Methodology

The network performed with high precision on the test data with a  $1.2 \pm 16.2$  m $\ell$ , statistically insignificant, discrepancy between manual and automated TKV measurements. Table 5.4 shows the agreement between the CNN and manual masks is higher for the right than left kidney, this is in part due to the proximity and lack of contrast between the left kidney and the spleen making distinguishing this boundary difficult for the CNN. This difficulty also leads to inconsistencies in manual masks, borne out by the increased CoV and decreased ICC and similarity metrics of the left kidney when compared to the right kidney in Table 5.2 and Table 5.4 assessing the variability in manual masks between observers. From Table 5.4 it can also be seen that the agreement between the CNN and manual masks is greater for the healthy control cohort than the CKD cohort, this is expected due to the increased variation in kidney morphology and the presence of cysts in the CKD cohort. Figure 5.13 shows that the difference between the manual TKV and CNN predicted TKV is not dependent on the true TKV, therefore the training data is balanced and well augmented as the network is able to accurately perform over the full range of kidney size in the test data.

Here, five volumes of test data were collected for each subject by repositioning the subject in the scanner within an hour scan session, and therefore any variance in measured TKV is purely due to inaccuracies in the kidney ROI definition. On assessing the correlation between manual and CNN measured TKV in Figure 5.14, it can be seen that, in 90 % of subjects the intra-observer variance in manual TKV between the segmentation of the five volumes collected in each subject is larger than using the CNN to estimate TKV, as reflected by the lower CoV and increased ICC of the TKV measured using the CNN (CoV  $1.5 \pm 0.5$  %, ICC 0.993) compared to the manual measures (CoV  $2.7 \pm 0.9$  %, ICC 0.979). As the network is trained on the kidney segmentations from three observers (A, B and C), it has been optimised by inheriting the most accurate tendencies of each observer e.g. one observer may have been very accurate when excluding cysts but not as accurate at defining the kidney-spleen boundary. The network will have learnt to exclude cysts from this observer but to delineate between kidney and spleen from another observer. Thus the network can become more precise than each individual observers manual segmentations. This increased precision can be seen in Figure 5.13 when compared to Figure 5.14 where the variance in difference in TKV is driven by the larger variance in manual TKV. The smallest TKV per subject is consistently overestimated when compared to its manual mask

and vice versa the largest manual TKV per subject is often an underestimation compared to the manual TKV.

Figure 5.15 illustrates the masks produced by the manual segmentation and the CNN for both a HC and CKD patient. For the HC, the CNN includes more voxels around the edge of its mask than manual segmentation, and the network is more anatomically accurate e.g. where the interface between the kidney and spleen is very narrow, the CNN predicts the kidney is adjacent to the spleen whilst the observers manual segmentation leaves a gap. The CKD data shown in Figure 5.15b includes a cyst in each of the kidneys. The network was trained on a combination of healthy and CKD data, with 19 of the 25 CKD training datasets containing at least one cyst. The CNN can be seen to segment out the cysts, despite their highly variable morphology and prevalence in the overall training data.

The amount of augmentation applied to the training data was empirically derived (random shifts up to 25 % of the image in both the horizontal and vertical direction, zooms between 0.75 and 1.25 times magnification, rotations within a 20 degree range, and sheers within a 5 degree range) and led to the potential for large transforms being applied to the data and masks if the extremes of each transform were randomly selected. This large degree of augmentation was advantageous as it mirrors the large variation in acquisition planning in abdominal imaging.

A 2D CNN was used to process each 2D slice of a full volume, rather than a 3D volume. This was advantageous for the relatively small training dataset the network was optimised on, as it avoids overfitting and allows the network to easily be used on volumes of variable slice number. However, this can come at the expense of accuracy as 2D CNNs do not leverage the information from adjacent slices in the segmentation as is done in 3D CNN, but 3D CNN come with a computational cost as a result of the increased number of parameters used. 3D networks have successfully been implemented on neural data using patching methods where the image volume is divided up into smaller cubes [27] to reduce memory requirements and allow for differing input shapes. While this works well in the brain, there are a number of reasons why this method may not be as successful for body applications. The out-of-plane resolution is significantly less than the in-plane resolution; this results in far fewer slices in one direction than the other two. To avoid overfitting for a certain number of slices e.g. training on a 11 slice image with a 113 patch, and subsequently the network not performing well when the patch is applied to a 16 slice image, the patch would

need to be much smaller than the number of slices, thus diminishing the benefits of the 3D methodology. Additionally, the extra memory requirements for a 3D network limit the ease of use of the software for inference on many standard office computers.

### 5.5.2 Future Directions

Future work will explore alternative network architectures. As the main source of inaccuracy with the current network is its lack of slice-to-slice context, there are multiple architectures that can address this. A relatively simple method would be exploiting the fact that the Keras framework is designed to work with colour images. By combining three slices into a single image where each slice represents either the red, green or blue channel, Figure 5.19, the network would be able to use contextual information about the adjacent slices to help inform its predictions. This method result in two slices from each volume being becoming impossible to process and as such the FOV would have to be increased. Alternatively multiple network architectures such as Recursive Neural Network (RNN) [42], and Long Short-Term Memory (LSTM) [43] are designed to have a memory therefore enabling them to retain contextual information from slice to slice. Alternatively, if more training data were available, a 3D CNN could be explored to ascertain if the any improved accuracy is worth the increase in hardware requirements and reduced generalisability.

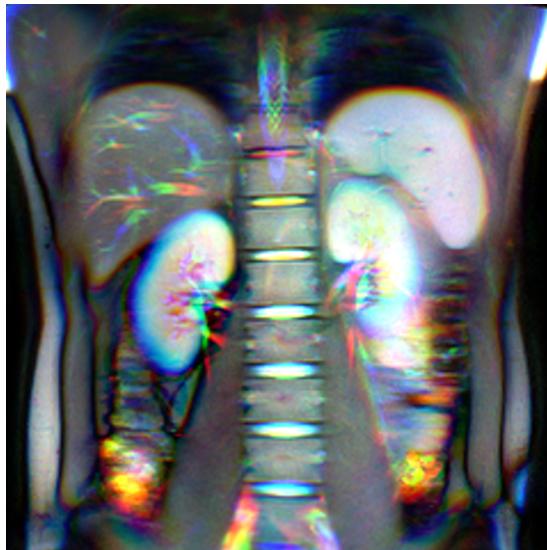


Figure 5.19: An example of using colour information to represent adjacent slices for processing by a CNN. Vessels in the liver can clearly be seen travelling through the three slices as they change from red to green to blue.

## 5.5. Discussion

---

This image contrast was chosen as a result of recent publications comparing  $T_1$ - and  $T_2$ -weighted images for TKV assessment reporting that  $T_2$ -weighted images provide better quality to enable TKV measurements, leading to improved reproducibility with lower intra- and inter-reader variability [44]. Other contrasts e.g. a  $T_1$ -weighted scan, could also be collected, registered to the  $T_2$ -weighted images and used as another channel to inform segmentation.

This network was validated on healthy subjects and CKD patients, but has not been trained and validated on subjects with ADPKD. These subjects have many more cysts in their kidneys, while the CNN was able to segment cysts encountered in the CKD cohort, it would be beneficial for future work on ADPKD to retrain the network with HC, CKD and ADPKD data, where TKV is a recognised biomarker of disease progression.

Another common segmentation task in renal imaging is generating an ROI for the renal cortex and medulla. There are some automated methods of achieving this once a total kidney mask has been produced [2, 45], however there has been no work on the application of deep learning to this task. In addition to the acquisition of the  $T_2$ -weighted dataset used here, a  $T_1$ -weighted dataset designed to optimise the contrast between cortex and medulla was also collected on each subject [87], an example of which is shown in Figure 5.20. Using this data, it may be possible to develop this method further such that an automated mask for each tissue type is produced.



Figure 5.20: An example of the data collected to enable segmentation of the renal cortex and medulla.

## 5.6 Conclusions

A CNN has been shown to successfully segment the kidneys of both HC and CKD subjects from  $T_2$ -weighted data delivering a mean Dice score of  $0.93 \pm 0.01$  leading to a mean volume of  $1.2 \pm 16.2 \text{ ml}$  less than the manually segmented TKV and mean surface distance of  $0.65 \pm 0.21 \text{ mm}$ . The CNN produces higher than human precision, with a CoV and ICC of  $1.5 \pm 0.5 \%$  and  $0.993$  respectively. The accuracy of the network could be further increased via the acquisition of more training data, something the renal group at Sir Peter Mansfield Imaging Centre (SPMIC) are actively pursuing.

The methods developed here can easily be deployed via the self contained, easy to use GUI, thus moving renal segmentation from a 15 to 30 minute skilled task, to a 10 second task for anyone, on any hardware. Additional, this executable can be called from a terminal, making it suitable for use in scripting applications and pipelines.

Future development will focus around exploring different network architectures, use of additional contrasts and expanding the tool to produce masks of both the cortex and medulla.

## 5.7 Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

## 5.8 References

1. Daniel, A. *et al.* Automated Renal Segmentation in Healthy and Chronic Kidney Disease Subjects Using A Convolutional Neural Network in *Proc. Intl. Soc. Mag. Reson. Med.* 28 International Society of Magnetic Resonance in Medicine Annual Meeting (Online, Aug. 2020).
2. Cox, E. F. *et al.* Multiparametric Renal Magnetic Resonance Imaging: Validation, Interventions, and Alterations in Chronic Kidney Disease. *Frontiers in Physiology* **8**. ISSN: 1664-042X (2017).
3. Cohen, E. I., Kelly, S. A., Edye, M., Mitty, H. A. & Bromberg, J. S. MRI Estimation of Total Renal Volume Demonstrates Significant Association with Healthy Donor Weight. *European Journal of Radiology. Rheumatoid Disease and Imaging* **71**, 283–287. ISSN: 0720-048X (1st Aug. 2009).
4. Van den Dool, S. W., Wasser, M. N., de Fijter, J. W., Hoekstra, J. & van der Geest, R. J. Functional Renal Volume: Quantitative Analysis at Gadolinium-Enhanced MR Angiography—Feasibility Study in Healthy Potential Kidney Donors. *Radiology* **236**, 189–195. ISSN: 0033-8419 (1st July 2005).
5. Chapman, A. B. *et al.* Kidney Volume and Functional Outcomes in Autosomal Dominant Polycystic Kidney Disease. *Clinical Journal of the American Society of Nephrology* **7**, 479–486. ISSN: 1555-9041, 1555-905X (1st Mar. 2012).
6. Tangri, N. *et al.* Total Kidney Volume as a Biomarker of Disease Progression in Autosomal Dominant Polycystic Kidney Disease. *Canadian Journal of Kidney Health and Disease* **4**, 2054358117693855. ISSN: 2054-3581 (1st Jan. 2017).
7. Grantham, J. J. *et al.* Volume Progression in Polycystic Kidney Disease. *New England Journal of Medicine* **354**, 2122–2130. ISSN: 0028-4793 (18th May 2006).
8. Buchanan, C. E. *et al.* Quantitative Assessment of Renal Structural and Functional Changes in Chronic Kidney Disease Using Multi-Parametric Magnetic Resonance Imaging. *Nephrology Dialysis Transplantation* (29th June 2019).
9. Stevens, L. A., Coresh, J., Greene, T. & Levey, A. S. Assessing Kidney Function — Measured and Estimated Glomerular Filtration Rate. *New England Journal of Medicine* **354**, 2473–2483. ISSN: 0028-4793 (8th June 2006).
10. Gong, I. H. *et al.* Relationship Among Total Kidney Volume, Renal Function and Age. *The Journal of Urology* **187**, 344–349. ISSN: 0022-5347 (Jan. 2012).
11. Di Leo, G., Di Terlizzi, F., Flor, N., Morganti, A. & Sardanelli, F. Measurement of Renal Volume Using Respiratory-Gated MRI in Subjects without Known Kidney Dis-

## 5.8. References

---

- ease: Intraobserver, Interobserver, and Interstudy Reproducibility. *European Journal of Radiology* **80**, e212–e216. issn: 0720-048X (1st Dec. 2011).
12. Bae, K. T., Commean, P. K. & Lee, J. Volumetric Measurement of Renal Cysts and Parenchyma Using MRI: Phantoms and Patients with Polycystic Kidney Disease. *Journal of Computer Assisted Tomography* **24**, 614–619. issn: 0363-8715 (July–Aug. 2000).
  13. Zöllner, F. G. *et al.* Assessment of Kidney Volumes From MRI: Acquisition and Segmentation Techniques. *American Journal of Roentgenology* **199**, 1060–1069. issn: 0361-808X (1st Nov. 2012).
  14. Sharma, K. *et al.* Kidney Volume Measurement Methods for Clinical Studies on Autosomal Dominant Polycystic Kidney Disease. *PLOS ONE* **12**, e0178488. issn: 1932-6203 (30th May 2017).
  15. Simms, R. J. *et al.* A Rapid High-Performance Semi-Automated Tool to Measure Total Kidney Volume from MRI in Autosomal Dominant Polycystic Kidney Disease. *European Radiology* **29**, 4188–4197. issn: 1432-1084 (1st Aug. 2019).
  16. Cheong, B., Muthupillai, R., Rubin, M. F. & Flamm, S. D. Normal Values for Renal Length and Volume as Measured by Magnetic Resonance Imaging. *Clinical Journal of the American Society of Nephrology* **2**, 38–45. issn: 1555-9041, 1555-905X (1st Jan. 2007).
  17. Spithoven, E. M. *et al.* Estimation of Total Kidney Volume in Autosomal Dominant Polycystic Kidney Disease. *American Journal of Kidney Diseases* **66**, 792–801. issn: 0272-6386 (1st Nov. 2015).
  18. Seuss, H. *et al.* Development and Evaluation of a Semi-Automated Segmentation Tool and a Modified Ellipsoid Formula for Volumetric Analysis of the Kidney in Non-Contrast T2-Weighted MR Images. *Journal of Digital Imaging* **30**, 244–254. issn: 1618-727X (1st Apr. 2017).
  19. Magistroni, R., Corsi, C., Martí, T. & Torra, R. A Review of the Imaging Techniques for Measuring Kidney and Cyst Volume in Establishing Autosomal Dominant Polycystic Kidney Disease Progression. *American Journal of Nephrology* **48**, 67–78. issn: 0250-8095, 1421-9670 (2018).
  20. Coulam, C. H., Bouley, D. M. & Sommer, F. G. Measurement of Renal Volumes with Contrast-Enhanced MRI. *Journal of Magnetic Resonance Imaging* **15**, 174–179. issn: 1522-2586 (2002).

## 5.8. References

---

21. Karstoft, K. *et al.* Different Strategies for MRI Measurements of Renal Cortical Volume. *Journal of Magnetic Resonance Imaging* **26**, 1564–1571. issn: 1522-2586 (2007).
22. Gloger, O. *et al.* Prior Shape Level Set Segmentation on Multistep Generated Probability Maps of MR Datasets for Fully Automatic Kidney Parenchyma Volumetry. *IEEE Transactions on Medical Imaging* **31**, 312–325. issn: 1558-254X (Feb. 2012).
23. Kim, Y. *et al.* Automated Segmentation of Kidneys from MR Images in Patients with Autosomal Dominant Polycystic Kidney Disease. *Clinical Journal of the American Society of Nephrology* **11**, 576–584. issn: 1555-9041, 1555-905X (7th Apr. 2016).
24. Ronneberger, O., Fischer, P. & Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) (Springer International Publishing, Cham, 2015), 234–241. isbn: 978-3-319-24574-4.
25. Lu, F., Wu, F., Hu, P., Peng, Z. & Kong, D. Automatic 3D Liver Location and Segmentation via Convolutional Neural Network and Graph Cut. *International Journal of Computer Assisted Radiology and Surgery* **12**, 171–182. issn: 1861-6429 (1st Feb. 2017).
26. Sharma, K. *et al.* Automatic Segmentation of Kidneys Using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease. *Scientific Reports* **7**, 2049. issn: 2045-2322 (1 17th May 2017).
27. Wachinger, C., Reuter, M. & Klein, T. DeepNAT: Deep Convolutional Neural Network for Segmenting Neuroanatomy. *NeuroImage. Segmenting the Brain* **170**, 434–445. issn: 1053-8119 (15th Apr. 2018).
28. Fu, Y. *et al.* A Novel MRI Segmentation Method Using CNN-Based Correction Network for MRI-Guided Adaptive Radiotherapy. *Medical Physics* **45**, 5129–5137. issn: 2473-4209 (2018).
29. Hassanzadeh, T., Hamey, L. G. C. & Ho-Shon, K. Convolutional Neural Networks for Prostate Magnetic Resonance Image Segmentation. *IEEE Access* **7**, 36748–36760. issn: 2169-3536 (2019).
30. Li, X. *et al.* H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging* **37**, 2663–2674. issn: 1558-254X (Dec. 2018).
31. Kline, T. L. *et al.* Performance of an Artificial Multi-Observer Deep Neural Network for Fully Automated Segmentation of Polycystic Kidneys. *Journal of Digital Imaging* **30**, 442–448. issn: 1618-727X (1st Aug. 2017).

## 5.8. References

---

32. Van Gastel, M. D. A. *et al.* Automatic Measurement of Kidney and Liver Volumes from MR Images of Patients Affected by Autosomal Dominant Polycystic Kidney Disease. *Journal of the American Society of Nephrology* **30**, 1514–1522. issn: 1046-6673, 1533-3450 (1st Aug. 2019).
33. Shin, T. Y. *et al.* Expert-Level Segmentation Using Deep Learning for Volumetry of Polycystic Kidney and Liver. *Investigative and Clinical Urology* **61**, 555–564. issn: 2466-0493 (Nov. 2020).
34. Fukushima, K. Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition. *Neural Networks* **1**, 119–130. issn: 0893-6080 (1st Jan. 1988).
35. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* **86**, 2278–2324. issn: 1558-2256 (Nov. 1998).
36. Petzold, K. *et al.* Building a Network of ADPKD Reference Centres across Europe: The EuroCYST Initiative. *Nephrology Dialysis Transplantation* **29**, iv26–iv32. issn: 0931-0509 (suppl\_4 1st Sept. 2014).
37. Will, S., Martirosian, P., Würslin, C. & Schick, F. Automated Segmentation and Volumetric Analysis of Renal Cortex, Medulla, and Pelvis Based on Non-Contrast-Enhanced T1- and T2-Weighted MR Images. *Magnetic Resonance Materials in Physics, Biology and Medicine* **27**, 445–454. issn: 1352-8661 (1st Oct. 2014).
38. Rorden, C. *Neurolabusc/MRIcron* 18th Jan. 2021.
39. Chollet, F. *Keras* version 2.2.4. Keras, 2015.
40. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, 19 (9th Nov. 2015).
41. Mahapatra, D. & Buhmann, J. M. Visual Saliency-Based Active Learning for Prostate Magnetic Resonance Imaging Segmentation. *Journal of Medical Imaging* **3**, 014003. issn: 2329-4302, 2329-4310 (Feb. 2016).
42. Chen, J., Yang, L., Zhang, Y., Alber, M. & Chen, D. Z. in *Advances in Neural Information Processing Systems* 29 (eds Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R.) 3036–3044 (Curran Associates, Inc., 2016).
43. Stollenga, M. F., Byeon, W., Liwicki, M. & Schmidhuber, J. in *Advances in Neural Information Processing Systems* 28 (eds Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 2998–3006 (Curran Associates, Inc., 2015).

## 5.8. References

---

44. Van Gastel, M. D. A. *et al.* T1 vs. T2 Weighted Magnetic Resonance Imaging to Assess Total Kidney Volume in Patients with Autosomal Dominant Polycystic Kidney Disease. *Abdominal Radiology* **43**, 1215–1222. ISSN: 2366-0058 (1st May 2018).
45. Morris, D. *et al.* Segmentation of the Cortex and Medulla in Multiparametric Magnetic Resonance Images of the Kidney Using K-Means Clustering in Proc. Intl. Soc. Mag. Reson. Med. 27 ISMRM. **27** (Montreal, 13th May 2019), 1915.