Automated renal compartment segmentation of the kidneys

Soham Talukdar

ppxst3@nottingham.ac.uk

Benjamin McLaren

ppybm3@exmail.nottingham.ac.uk

Abstract: *Magnetic resonance imaging (MRI) has become increasingly important in the clinical work-up of renal diseases such as chronic kidney disease. This project shows the use of a UNet to automatically segment the kidney tissue to separate the medulla and cortex based on the T_1-weighted MRI scans in human subjects. The project aims to automate the segmentation process for T_1-weighted (T_1w) scans. First masks are generated of the cortex and medulla from time-consuming (~5 minutes) MPRAGE scans of the kidneys using a K-means clustering technique. These masks are then used to train a UNet to segment the cortex and medullary of T_1w images collected in a single breath-hold. UNet CNN architectures was assessed. The mean IOU score compared to the manual masks for renal cortex and medulla was 0.9974872 and 0.84625 for UNet. The availability of such an automated method will reduce the time required and the difficulty associated with manual segmentation of the cortex and medulla and will have a direct impact on medical trials, such as the ongoing Application of Functional Renal MRI to improve the assessment of chronic kidney disease (AFiRM) study.*

1. Introduction

Medical imaging is used to provide patients with high-quality medical care. Manually reviewing medical images is time-consuming and arduous, requiring medical practitioners to be exceedingly exact and consistent in their analysis and the tools they employ to obtain the images (Östling, 2019). Kidney volume influences kidney activity and function in many illnesses, including polycystic kidney disease, lupus nephritis, renal parenchymal disease, and kidney transplant rejection (Zöllner et al., 2012, pp.1060-1069). Kidney imaging might be used to identify patients and predict outcomes automatically. Magnetic Resonance Imaging (MRI) can provide an image by which to segment the kidneys and offer volume-related information since it does not require ionising radiation, is operator-independent, and has high tissue contrast. Detecting and segmenting the distinct structures of the cortex and medulla from images was previously done manually, which takes a long time. **Figure 1** shows a schematic of the kidney showing the different compartments of the cortex and medulla (Hodneland et al, 2014, p. 23). Deep learning segmentation frameworks surpass standard methods in automatically deciphering the
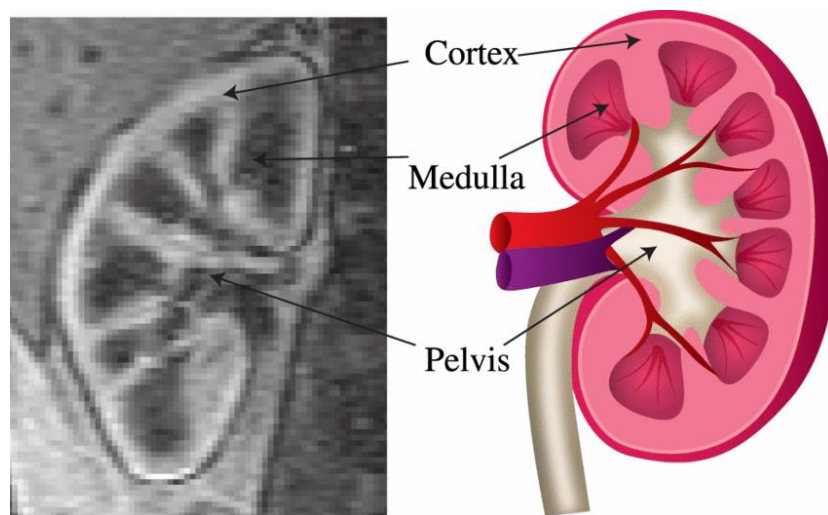

Figure 1. Schematic of the kidney showing the different compartments of the cortex and medulla

kidney in volumetric MRI datasets (Bazgir et al.,2020, pp. 22-25). Various automated techniques have

been created in recent years to do these tasks for different human organs (brain, lungs, liver) and objectives. The most prominent approaches employ neural networks using multiple neural network topologies, both general purpose (ResNet, VGG, AlexNet) and Deep Neural Network (DNNs) created notably for medical imaging (U-Net, V-Net) (Will et al., 2014). Internal renal structures such as the cortex, medulla, and pelvis can be segmented for functional and morphological analyses. It is well known that the renal cortex volume decreases with time in chronically damaged kidneys (Statkevych, Stirenko and Gordienko, 2021, pp. 129-134). As a result, a precise assessment of cortical volume might be a helpful indicator in chronic renal disease or transplanted kidney follow-up tests. Renal length measures utilising ultrasound imaging can provide a basic estimate of kidney volume. This approach, however, has a propensity to underestimate renal volume. Magnetic resonance imaging (MRI) produces spatially high-resolution anatomical pictures, making it a more exact imaging modality for volumetric measurements. Manual segmentation across a large number of slices is a time-consuming technique, as such, exact estimation of the overall volume of internal organs is seldom performed in clinical practice. As a result, for this purpose, automated segmentation methods are preferred. Many approaches, such as thresholding, clustering, region growth, and contour detection, as well as their combinations, (Will et al., 2014) have been used for segmentation. Internal renal structures (cortex, medulla, and pelvis) must be separated using MR methods that provide adequate contrast between these compartments. T_1- and T_2-weighted (w) sequences are commonly used in clinical abdominal imaging.

Using Computed Tomography (CT) images, there are a number of methods for kidney segmentation in the literature. For example, a study done by Kobashi and Shapiro (1995, pp. 475-491) suggested a knowledge-based identification using the anatomical details and CT image properties. Lin and Hsiung (2003, pp. 75-82) obtained the coarsely segmented left and right kidney regions as ROIs, used an adaptive region growing technique with initial seed points and a threshold value, and then modified the regions for precise segmentation. A growing method was proposed by Pohle and Toennies (2001, pp. 1337-1346) that uses the characteristics of the segmentation target area to automatically optimise the homogeneity criterion. Yan and Wang, (2010, pp. 280-284) created a new technique for growing regions by estimating kidney position, using a labelling algorithm, and obtaining seeds with mathematical morphology to extract the kidney regions. A non-rigid registration-based active shape model was proposed by Spiegel et al. (2009, pp. 29-39). Another study by Yang et al. (2014, pp. 5538-5541) suggested an approach based on multiple atlases for registration. Huang (2009, pp. 243-246) developed a shape-based level set method by using the Chen and Vese level set model and connected component analysis to obtain separated components, and then the size feature to identify the kidneys. To fit the model in their work, Tsagaan et al. (2002, pp. 556-563) proposed a B-spline-based deformable model by incorporating the organ's mean and variation into the objective function. An automated kidney segmentation method based on neural networks was put forth by Goceri and Goceri (2015, pp. 1195–1198). This method continues segmentation from a dataset's initial slice that is chosen up until the slice where the kidneys vanish. Since parameter values vary for each image, segmenting each slice of the image was done iteratively rather than using a universal parameter from all data sets. The ability to handle contrast variations and various kidney region shapes accounts for the proposed method's robustness and effectiveness. The iterative processing of each slice and the proposed approach's superior

segmentation performance to conventional thresholding and K-means based approaches were responsible for enabling these capabilities.

Magnetic resonance imaging (MRI) is becoming increasingly relevant in the clinical evaluation of renal disorders. A panel of parameters may now be assessed non-invasively, which can help diagnose and monitor renal disorders. This includes assessing kidney volume and microstructure using diffusion-weighted imaging, hemodynamic characteristics using arterial spin labelling (ASL) or dynamic contrast-enhanced (DCE-) MRI, and finally, oxygenation using blood oxygen level-dependent imaging (BOLD). In individuals with autosomal dominant polycystic kidney disease (ADPKD), total kidney volume (TKV) is an often-accessed statistic (ADPKD). It has been demonstrated that as the illness
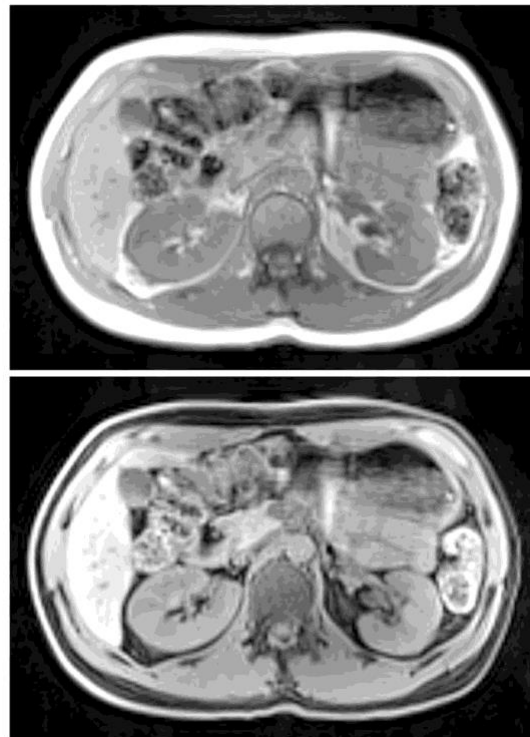
Figure 2. Adjacent organs and skeletal muscle with MR signal similar to kidney

progresses, total kidney volume grows but renal function decreases. The Federal Medication Administration (FDA) recently approved TKV as a biomarker for use in drug development in ADPKD. It is also the first MRI-based biomarker. The similar signal intensities of renal tissue to neighbouring liver, spleen, vertebrae, and sections of the gastrointestinal system can be seen in **Figure 2** (Rusinek et al., 2016, pp.197-206). These similar intensities pose a problem in renal segmentation. This condition is exacerbated in slim people who lack the fat that surrounds the kidney. Thus, signal intensity-based methods, like thresholding, do not give enough information to provide successful segmentation. More advanced techniques, such as kidney shape analysis (Zöllner et al., 2021, pp.71577-71605) are required. This entails regularisation to give the organ the needed resilience to noise, poor contrast, heterogeneity, and significantly changing form. As a result, multi-modal techniques that combine complementary information, such as T_1- and T_2-weighted images, appear promising.

Convolutional networks are commonly used for classification tasks, where the output to an image is a single class label. However, in many visual tasks, particularly in biomedical image processing, the intended output should contain localization, i.e., each pixel should be allocated a class label (Ronneberger, Fischer and Brox, 2015, pp. 234-241). Furthermore, in biomedical activities, thousands of training photos are typically out of reach. Therefore, the network lacks ultimately linked layers and

employs just the valid part of each convolution, i.e., the segmentation map only comprises pixels for which the whole context is accessible in the input picture. The research in this project aims to create a deep learning network that can segment cortex, medulla, and multi-contrast segments using T_2-weighted and T_1-weighted MRI datasets. It has been divided in the following parts: methodology, results and discussion.

2. Methodology

This section outlines the MRI data provided for this project, the generation of masks for training the machine learning networks, and the development of the convolutional neural network architectures employed for automated segmentation of the cortex and medulla of the kidney.

MR Image data

The dataset provided consisted of ten healthy individual scans of T_1-weighted (T_1w), T_2-weighted (T_2w), and MPRAGE (Magnetization Prepared Rapid Gradient Echo) images comprising 17 coronal-oblique slices of the abdomen to cover the kidneys. The full kidney segmentation mask derived using a U-Net on the T_2-weighted scans, as described by Daniel et al. (2021, pp.1125-1136), was provided by our supervisors. The T_1w acquisition is a multi-slice gradient echo with a TE of 2.3 ms, TR of 155 ms, flip angle of 42 degrees, bandwidth of 787 Hz, SENSE factor of 2, partial Fourier factor of 0.625, field of view of 384×384 mm, voxel size of 1.5 x 1.5 x 5.5 mm, and 17 slices. It was obtained in a single 16-second breath hold. The abdomen is imaged using three different T2 fast spin echo (FSE) imaging sequences: segmented FSE, which can be respiratory-triggered or respiratory-gated; breathhold FSE with a long echo-train length; and single-shot FSE (SSFSE), which has a partial Fourier transform and is typically performed during a breath-hold. All of the k-space is acquired using breath-hold FSE imaging with a long echo-train length in 15 to 25 seconds over several TRs. Unless the cardiac cycle is a whole multiple of the repetition time (TR), in which case pseudogating will happen, more time of flight (TOF) losses are distributed across all slices (Lisanti and Douglas, 2009, pp. 1348-1358). Hence, it removes respiratory motion artifacts but significantly slows down data acquisition. This restriction lowers the spatial resolution and image signal-to-noise ratio.

The MPRAGE acquisition is a 3D rapid gradient echo acquisition with a TE of 2.6 ms, TI of 960 ms, shot interval of 6000 ms, turbo factor of 40, flip angle of 8 degrees, bandwidth of 287 Hz, compressed SENSE factor of 4, and 17 slices. The MPRAGE data is collected using a respiratory triggered scan which is activated by breathing through a respiratory bellows placed around the subject. The MPRAGE scan provides improved definition of the cortex and medulla but is time consuming taking ~5 minutes to collect a 17 slice image data set. Example images are shown in **Figure 3(a, b)**.
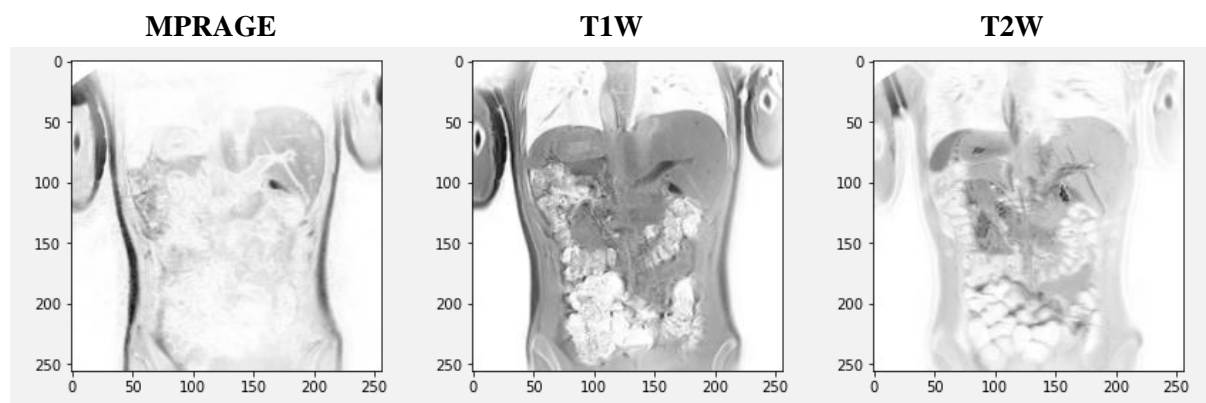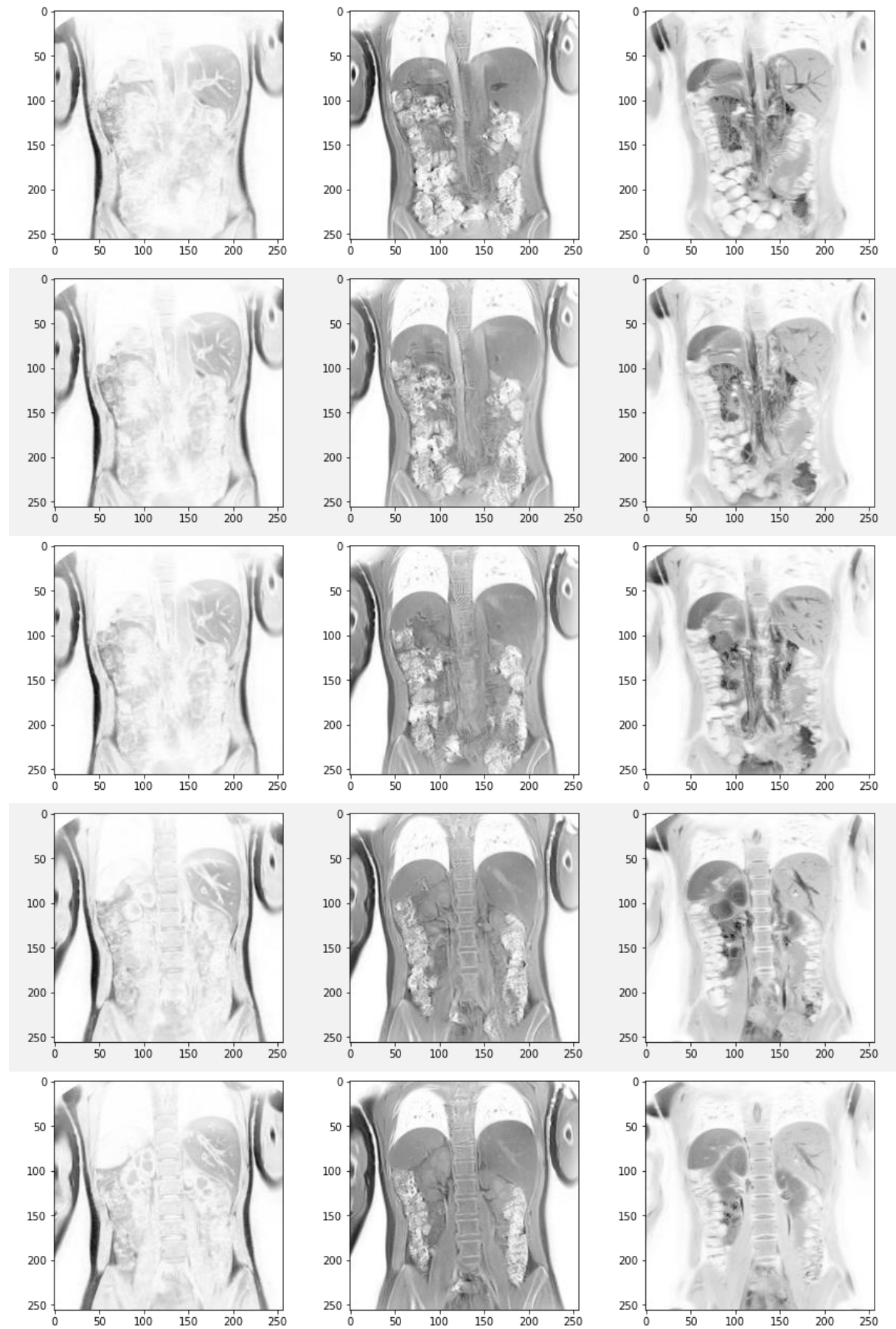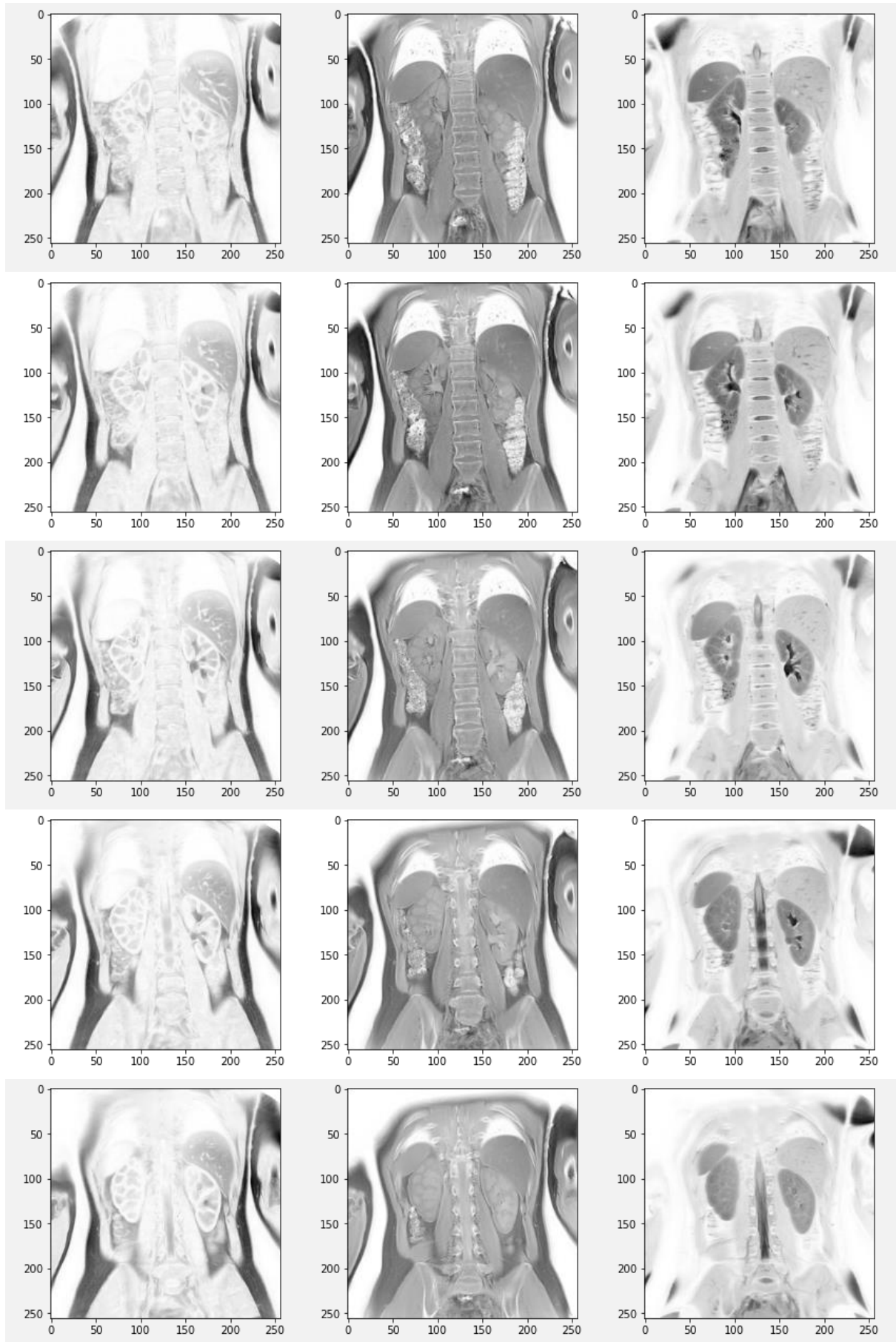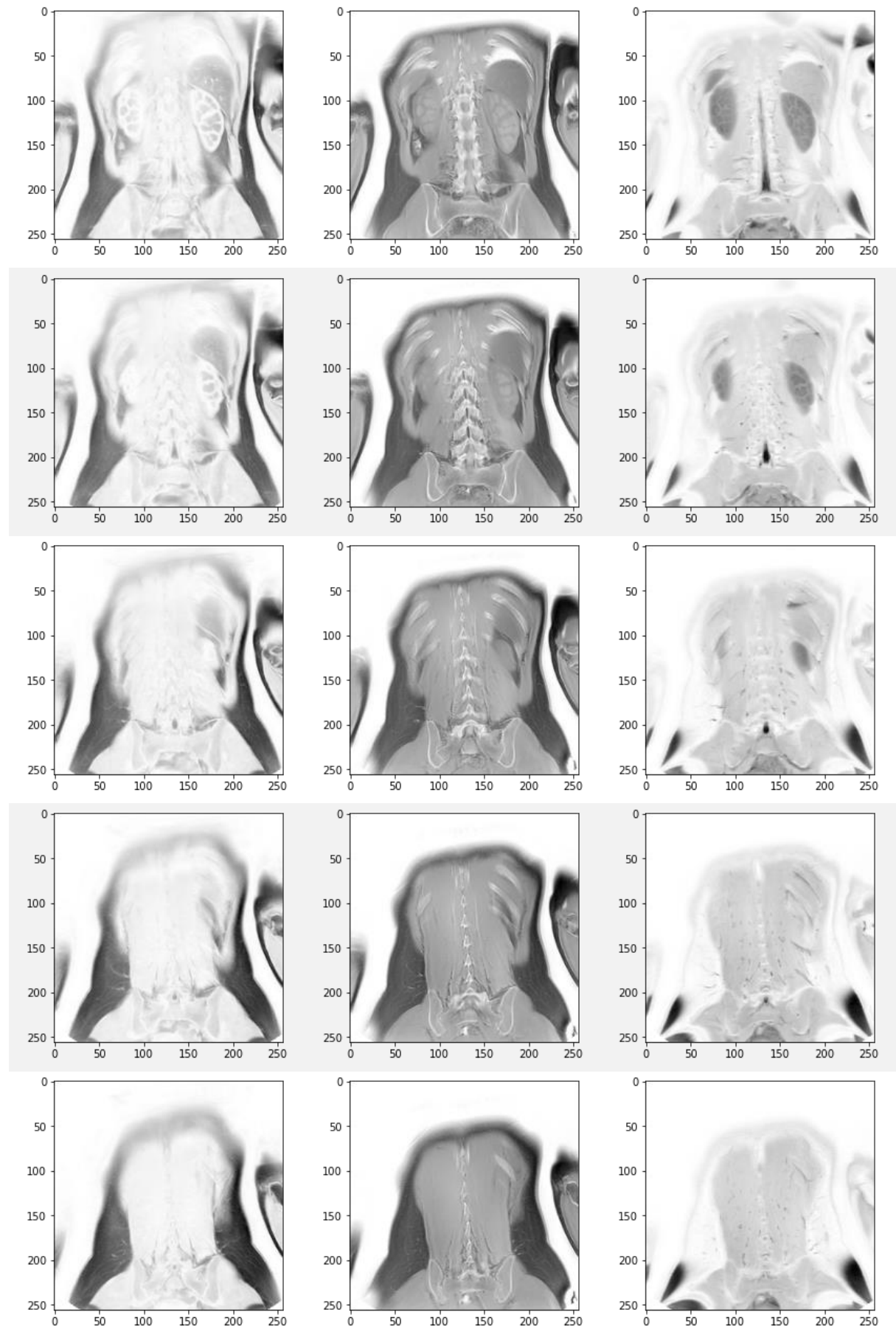
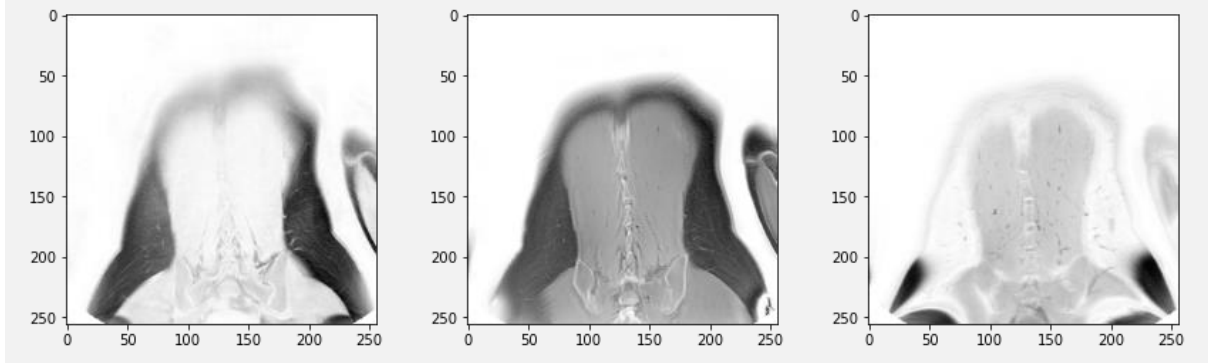

| MPRAGE | T1W | T2W |

Figure 3a.Three image types (T1w, T2w and MPRAGE) for a given subject showing the 17 slices for each technique
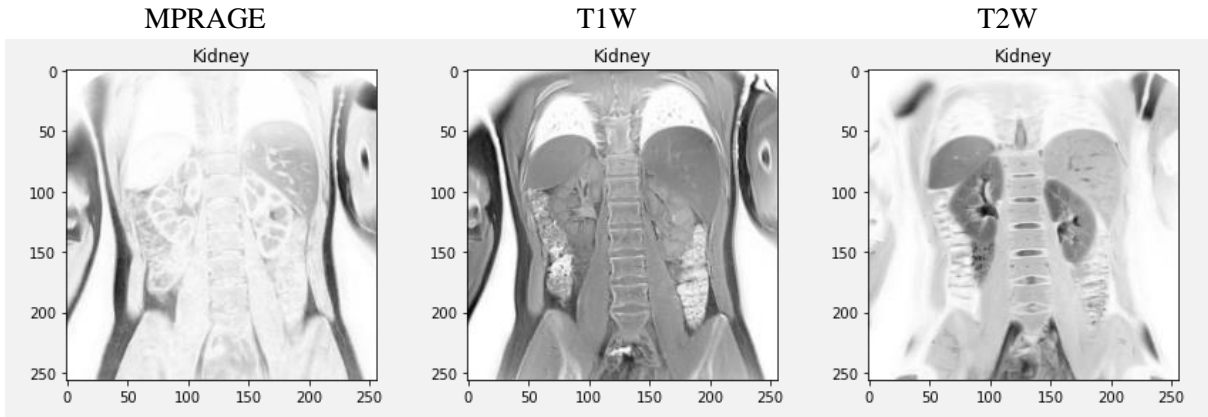

Figure 3b. Central slice for three image types for a given subject

Generating ground truth masks

The initial effort was spent trying fundamental thresholding approaches to segment the cortex and medulla from the MPRAGE images. Methods attempted included those such as Otsu thresholding, k-means clustering, and Gaussian mixture models. However, we struggled to analyse the quality metrics of our basic segmentations without some ground-truth masks to compare against, and we found it challenging to measure boundaries ourselves due to our lack of knowledge.

After this, we experimented with third-party segmentation tools. FAST (FMRIB's Automated Segmentation Tool) (Zhang, Brady and Smith, 2001, pp. 45-57) was the major tool we employed. It typically deals with T_1-weighted and T_2-weighted MR images of the brain. FAST uses a hidden Markov random field model alongside an Expectation-Maximisation algorithm to differentiate tissues based on their intensities while also correcting radio frequency (RF) inhomogeneities that lead to variations in signal intensity across the field of view. The T_2w masks were first used to separate the background noise and to get a clear image of the whole kidney. The T_1w and MPRAGE images were first registered to T_2w to minimize the affect of motion between aquisition. We then spent some time manually adjusting settings and from this generated ground-truth masks ready for our DL implementation.

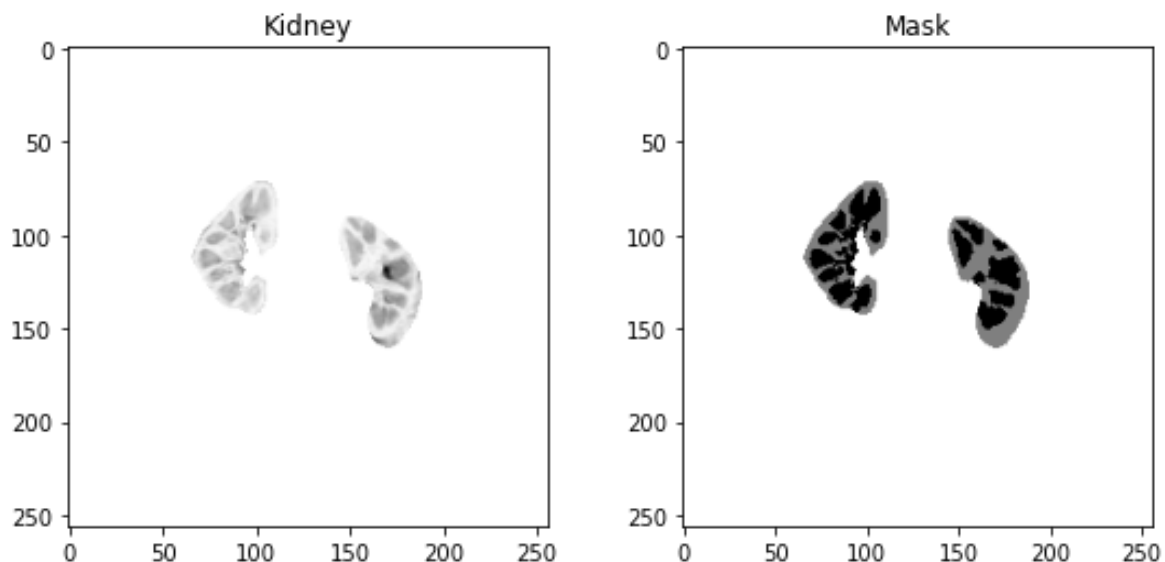**Figure 4** shows an example MPRAGE image and the cortex and medulla masks generated using FAST.

Figure 4. (Left) shows an example MPRAGE image, (Right) the cortex and medulla masks generated using FAST

Segmentation Model

Figure 5 (Ronneberger, Fischer and Brox, 2015, pp. 234-241) depicts the U-Net model used in the automation of the cortex and medulla segmentation. The essential specifications and architectural components of the U-Net architecture's initial implementation from the original paper (Ronneberger, Fischer and Brox, 2015, pp. 234-241) remains the same. However, there will be a few minor alterations from the original content that will enhance the model's speed, simplicity, and modern efficiency. Using the value of Convolution as "same" is one of the adjustments that will be made to the structure because a number of studies conducted (Hashemi, 2019, pp. 1-13) have demonstrated that this particular change had no adverse effects on the structure's construction. Additionally, the original architecture didn't use batch normalisation (Ioffe and Szegedy, 2015, pp. 448-456) since the idea was introduced in 2015. But because batch normalisation typically produces the better results, it will be a part of how our model is implemented. The primary function of the convolution operation block is to process a double layer of convolution operations using the input parameters. The convolution layer's input and the function's number of filters—64 by default—are its two arguments. In place of unpadded or valid convolutions, we will use the same padding value as previously discussed to preserve the same shapes. Batch Normalisation is the layer that comes after these convolutional layers. These modifications to the original model have been made to achieve better results. As described in the research paper, a ReLU activation layer is added to the mix. Building the encoder and decoder blocks will come next. Consecutive inputs will be used in the encoder architecture, starting from the top layer and working your way down. The convolutional block, or two convolutional layers, will be followed by their respective batch normalisation and ReLU layers in the encoder function that we have defined. As stated in the research paper, after passing these elements through the convolution blocks, we will downsample them. Finally, we'll employ a max-pooling layer and adhere to the strides equal to two as described in the paper. Then, since we require the initial output for implementing the skip connections, we will also return the max-pooled output. The receiving inputs, the input of the skip connection, and the number of filters in the specific building block are the three arguments that the decoder block will consider. Next, we will use the Conv2DTranspose layers in our model to upsample the input that has been entered. The receiving input and the newly upsampled layers will then be concatenated to obtain the final value of

the skip connections. Then, using this combined function, we'll carry out our convolutional block operation, move on to the following layer, and then return this result.
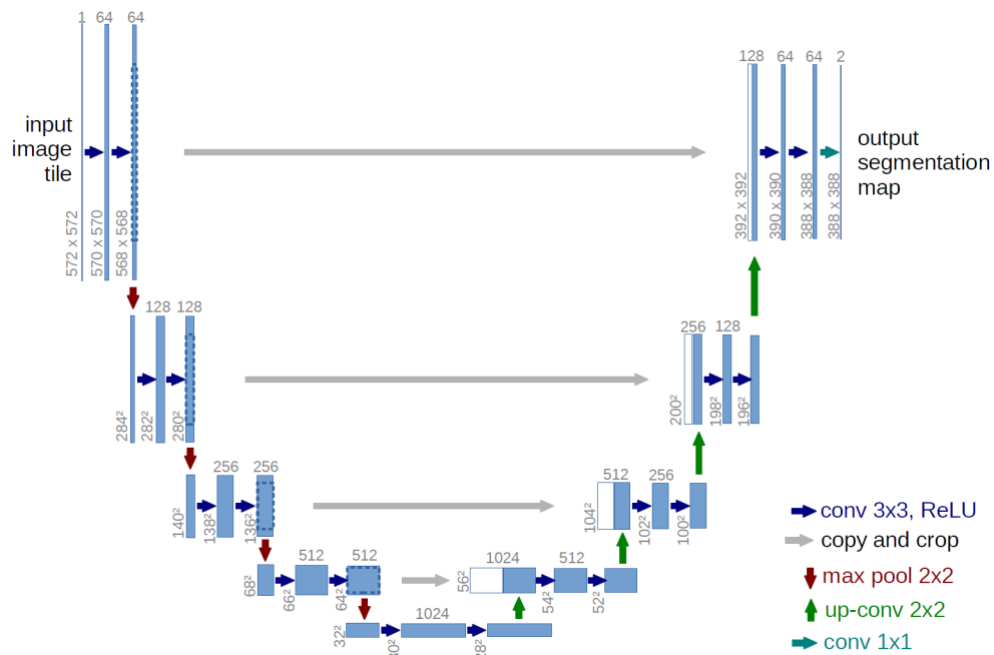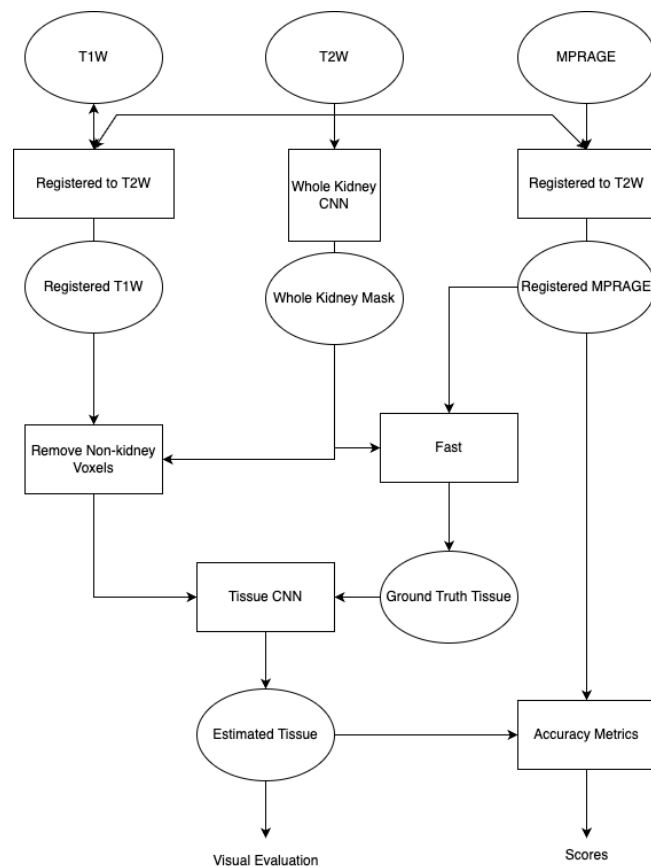

Figure 5. UNet Architecture

Figure 6. Flowchart of the entire segmentation pipeline

Training the model

The network was built in Python 3.7.13 with Keras (v2.8.0) and a TensorFlow backend (v2.8.2). All training was done on an NVIDIA Tesla P100 GPU (NVIDIA Pascal architecture, CUDA Cores 3584, Memory of 16GB CoWoS HBM2 at 732 GB/s or 12GB CoWoS HBM2 at 549 GB/s). As for the loss function, the network employs a sparse categorical cross-entropy as show in **Equation 1**.

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i\,log(\hat{y}_i) + (1-y_i)\,log(1-\hat{y}_i)]$$

Equation 1. Sparse Categorical Cross-Entropy

The training was done over 110 epochs with a batch size of 2 using a learning rate of 1e-4. Adaptive Moment Estimation (Adam) was used as the optimiser because it is a method for efficient stochastic optimization that requires first-order gradients and uses little memory. Individual adaptive learning rates for distinct parameters are calculated using estimations of the gradient's first and second moments. Nadam (Nesterov-accelerated Adaptive Moment Estimation) was used in another experiment. Adam comprises two primary hyperparameters, a momentum component and an adjustable learning rate component. Regular momentum may be theoretically and experimentally inferior to a similar technique known as Nesterov's accelerated gradient (NAG) (Dozat, 2016).
A trial with generalized dice loss was carried out without much success. In order to effectively balance out the foreground and background, dice loss **Equation 2** optimises networks based on the dice overlap coefficient between the predicted segmentation result and the ground truth annotation (Zhao et al., 2020, pp. 851-860). However, it only considers one aspect of this imbalance and ignores the disparity in data difficulty. In particular, if the ratios of easy and hard cases are substantially out of balance during the training process, the sum of the contributions from easy examples will overwhelm the segmentation model and have an effect on the preference of anticipated results in the test stage. The weighted sum of the products over the weighted sum of the sums between the ground-truth and predicted probabilities was computed for the generalised Dice loss in the generalised Dice instead of computing the Dice loss of each label, with the weights being inversely proportional to the label frequencies. In actuality, the Dice coefficient is unfavourable to small structures since even a small mistake of a few pixels can result in a significant drop in the coefficient, regardless of the relative sizes of the various structures. In light of dice losses, balancing by label frequencies is not the best option. The train-test split was done in 90%-10% of the T_1w data. For this limited data set this was the ideal split, though training with 80% didn't make much difference, however with less than 80% of data it didn't provide satisfactory results.

$$DL(y,\hat{p}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1}$$

Equation 2. Dice Loss

3. Result

Figure 7. From left to right. (Left most) training and validation loss. (middle) training and validation loss of dice coefficient, (right) training and validation accuracy. Using Adam optimizer.
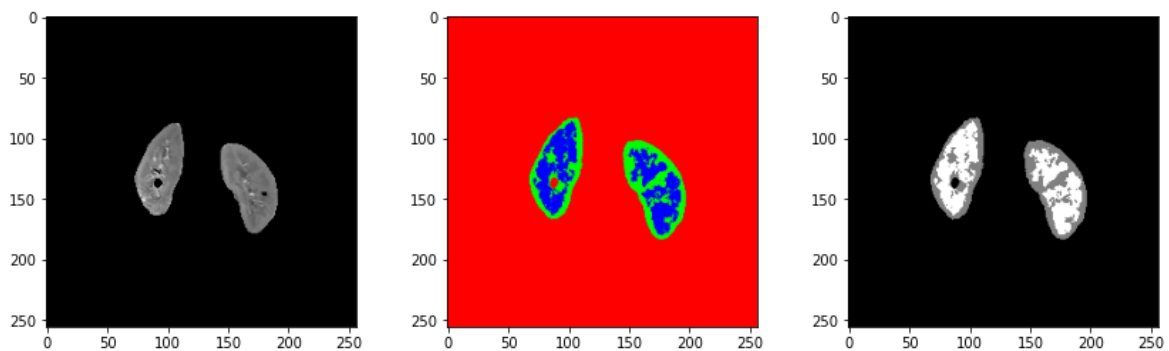

Figure 8. From left to right. For an example subject, (Left most) the T1w image input to the tissue segmentation algorithm. (middle) the predicted image generated from applying FAST to an MPRAGE image, (right) the predicted mask. Using Adam optimizer applied to t_1w image

Optimizer	Loss	Accuracy	Jaccard Coefficient	Dice Coefficient	Mean IOU	Mean IoU for all slices
Adam	0.058	0.9924	0.3879	0.0376	0.72864395	slice 0 = 0.74830914 slice 1 = 1.0 slice 2 = 1.0 slice 3 = 0.733556 slice 4 = 0.5449839 slice 5 = 0.64035535 slice 6 = 1.0 slice 7 = 1.0 slice 8 = 0.78182095 slice 9 = 0.63636565 slice 10=0.40274432 slice 11= 0.85755205 slice 12 = 0.8116662 slice 13= 0.73635817 slice 14 = 1.0 slice 15= 0.70997864

						slice 16= 0.65551966
Nadam	0.50	0.9928	0.0376	0.3834	0.7373338	slice 0 = 0.73926425
						slice 1 = 1.0
						slice 2 = 1.0
						slice 3 = 0.7349798
						slice 4 = 0.5751345
						slice 5 = 0.654766
						slice 6 = 1.0
						slice 7 = 1.0
						slice 8 = 0.7912619
						slice 9 = 0.6488395
						slice 10 = 0.5778026
						slice 11 = 0.8648946
						slice 12 = 0.8128759
						slice 13= 0.75743675
						slice 14 = 1.0
						slice 15 = 0.7083936
						slice 16= 0.69033664

Table 1 Performance metrics of the algorithm

Figure 7 shows how the loss, accuracy and dice changed over the epoch. A zoomed in version of the images are illustrated in the appendix as **Figure 10**. The corresponding graphs for Nadam is also highlighted in **Figure 9**. **Figure 8** takes an example subject and highlights the input and corresponding predicted image and mask. **Table 1** shows the performance metrics, highlighting loss, accuracy, jaccard coefficient, dice coefficient, mean IOU across individuals and mean IoU for all slices in an individual.

4. Discussion of the Results

The study was done to segment the medulla and cortex using an UNet, a convolutional neural network and seems to provide the most promising result. Similar research in semantic segmentation has taken place with other advanced variations of UNet. However, in this case, a trial with different other algorithms (Attention UNet, UNet++, Resunet, Bi-directional ConvLSTM U-Net with Densely connected convolutions (BCDU-Net)) was taken place. Among all these nothing stood out and compared well enough with the UNet. A hypothesis that can be generated is that the more advanced and complex algorithms were designed to solve a very niche problem and their architecture is too complex to use directly on other problems. A much deeper understanding of the architecture and experimentation are required to conclude whether it is good enough. Resource constraint is something that needs to be considered in this situation. When working with limited GPU resources, training with a larger batch size makes it much more difficult and hence it reduces the scope of experimentation. Working with models with a large number of parameters is again not viable. Different loss functions are taken into consideration, however, apart from sparse categorical cross entropy none of the others (Dice, Jaccard) seemed to work out in this situation. Optimising the data to work for the corresponding loss function is important. Several studies took place where Dice and Jaccard were successfully implemented. However, their work was mostly on binary classification. When working with multiclass a lot more stress needs to be put on augmenting the data and rescaling it. Another instance which might be in this case is again the lack of resources. When working with larger batch sizes with higher learning rates, showed an improvement in the loss. However, it went out of memory after a few epochs. SGD optimizer fits the data better than Adam; however, it requires a lot of

tweaking to get the perfect fit. Here it's much more of a time constraint. Working with Adam gives an opportunity to quickly fit the data. Due to a lack of time and resources, it was decided to work around Adam. Nadam was hoped to be promising. However, the visual result didn't make much of a difference.

Limitation

Google Colab Pro's GPU is not powerful enough to work with such data. Working with data of such scale will require much more powerful GPU which will make it easier to experiment and try out other techniques. Probably with more data the training of the model would be better and hence will make better predicton.

Future Work

A lot could have been done. Two models UNet++ and BCDU-Net might probably provide a far better result. working with SGD and implementing different decay functions might provide a better fit of the model and provide a better result on unseen data.

5. Conclusion

A Unet has been successfully applied to segment the kidneys medulla from the cortex from T1-weighted renal MRI data. In the future, this project could directly impact clinical trials including those studying COVID-19, and a trial we are currently running in chronic kidney disease where 450 patients are planned to be scanned longitudinally to assess disease progression.

References

Östling, A., 2019. Automated Kidney Segmentation in Magnetic Resonance Imaging using U-Net.

Zöllner, F.G., Svarstad, E., Munthe-Kaas, A.Z., Schad, L.R., Lundervold, A. and Rørvik, J., 2012. Assessment of kidney volumes from MRI: acquisition and segmentation techniques. American Journal of Roentgenology, 199(5), pp.1060-1069.

Hodneland, E., Hanson, E.A., Lundervold, A., Modersitzki, J., Eikefjord, E. and Munthe-Kaas, A.Z., 2014. Segmentation-driven image registration-application to 4D DCE-MRI recordings of the moving kidneys. IEEE transactions on image processing, 23(5), pp.23

Bazgir, O., Barck, K., Carano, R.A., Weimer, R.M. and Xie, L., 2020, March. Kidney segmentation using 3D U-Net localized with Expectation Maximization. In 2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI) (pp. 22-25). IEEE

Statkevych, R., Stirenko, S. and Gordienko, Y., 2021, July. Human kidney tissue image segmentation by u-net models. In IEEE EUROCON 2021-19th International Conference on Smart Technologies (pp. 129-134). IEEE.

Will, S., Martirosian, P., Würslin, C. and Schick, F., 2014. 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

Rusinek, H., Lim, J.C., Wake, N., Seah, J.M., Botterill, E., Farquharson, S., Mikheev, A. and Lim, R.P., 2016. A semi-automated "blanket" method for renal segmentation from non-contrast T1-weighted MR images. Magnetic Resonance Materials in Physics, Biology and Medicine, 29(2), pp.197-206.

Zöllner, F.G., Kociński, M., Hansen, L., Golla, A.K., Trbalić, A.Š., Lundervold, A., Materka, A. and Rogelj, P., 2021. Kidney segmentation in renal magnetic resonance imaging-current status and prospects. IEEE Access, 9, pp.71577-71605.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Dozat, T., 2016. Incorporating nesterov momentum into adam.
Goceri, N. and Goceri, E., 2015, December. A neural network based kidney segmentation from MR images. In 2015 IEEE 14th international conference on machine learning and applications (ICMLA) (pp. 1195-1198). IEEE.

Höhne, K.H. and Hanson, W.A., 1992. Interactive 3D segmentation of MRI and CT volumes using morphological operations. Journal of computer assisted tomography, 16(2), pp.285-294.

Chang, Y.L. and Li, X., 1994. Adaptive image region-growing. IEEE transactions on image processing, 3(6), pp.868-872.

Adams, R. and Bischof, L., 1994. Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence, 16(6), pp.641-647.

McInerney, T. and Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. Medical image analysis, 1(2), pp.91-108.

Jones, T.N. and Metaxas, D.N., 1997, June. Automated 3D segmentation using deformable models and fuzzy affinity. In Biennial International Conference on Information Processing in Medical Imaging (pp. 113-126). Springer, Berlin, Heidelberg.

Cagnoni, S., Dobrzeniecki, A.B., Poli, R. and Yanch, J.C., 1999. Genetic algorithm-based interactive segmentation of 3D medical images. Image and Vision Computing, 17(12), pp.881-895.
Chalana, V., Sannella, M. and Haynor, D.R., 2000, June. General-purpose software tool for serial segmentation of stacked images. In Medical Imaging 2000: Image Processing (Vol. 3979, pp. 192-203). SPIE.

Ma, F. and Xia, S., 1998. A multiscale approach to automatic medical image segmentation using self-organizing map. Journal of Computer Science and Technology, 13(5), pp.402-409.

Ahmed, M.N. and Farag, A.A., 1997. Volume segmentation of ct/mri images using multiscale features, self-organizing principal components analysis (sopca), and self-organizing feature map (sofm). In Proc. of the ICNN97 (Vol. 3, pp. 1373-1378).

Pham, D.L. and Prince, J.L., 1999. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. Pattern recognition letters, 20(1), pp.57-68.

Wegner, S., Harms, T., Oswald, H. and Fleck, E., 1996, August. The watershed transformation on graphs for the segmentation of CT images. In Proceedings of 13th International Conference on Pattern Recognition (Vol. 3, pp. 498-502). IEEE.

Sijbers, J., Scheunders, P., Verhoye, M., Van der Linden, A., Van Dyck, D. and Raman, E., 1997. Watershed-based segmentation of 3D MR data for volume quantization. Magnetic Resonance Imaging, 15(6), pp.679-688.

Bueno, G., Musse, O., Heitz, F. and Armspach, J.P., 2000, June. 3D Watershed-based segmentation of internal structures within MR brain images. In Medical Imaging 2000: Image Processing (Vol. 3979, pp. 284-293). SPIE.

Kobashi, M. and Shapiro, L.G., 1995. Knowledge-based organ identification from CT images. Pattern Recognition, 28(4), pp.475-491.

Lin, D.T., Lei, C.C. and Hsiung, S.Y., 2003. An efficient method for kidney segmentation on abdominal CT images. In 8th Australian and New Zealand Intelligent Information Systems Conference, Sydney, Australia (pp. 75-82).

Pohle, R. and Toennies, K.D., 2001, July. Segmentation of medical images using adaptive region growing. In Medical Imaging 2001: Image Processing (Vol. 4322, pp. 1337-1346). SPIE.

Yan, G. and Wang, B., 2010, October. An automatic kidney segmentation from abdominal CT images. In 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems (Vol. 1, pp. 280-284). IEEE.

Spiegel, M., Hahn, D.A., Daum, V., Wasza, J. and Hornegger, J., 2009. Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. Computerized Medical Imaging and Graphics, 33(1), pp.29-39.

Yang, G., Gu, J., Chen, Y., Liu, W., Tang, L., Shu, H. and Toumoulin, C., 2014, August. Automatic kidney segmentation in CT images based on multi-atlas image registration. In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 5538-5541). IEEE.

Huang, C.L., 2009, August. Shape-based level set method for image segmentation. In 2009 Ninth International Conference on Hybrid Intelligent Systems (Vol. 1, pp. 243-246). IEEE.

Tsagaan, B., Shimizu, A., Kobatake, H. and Miyakawa, K., 2002, September. An automated segmentation method of kidney using statistical information. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 556-563). Springer, Berlin, Heidelberg.

Daniel, A.J., Buchanan, C.E., Allcock, T., Scerri, D., Cox, E.F., Prestwich, B.L. and Francis, S.T., 2021. Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network. *Magnetic Resonance in Medicine*, *86*(2), pp.1125-1136.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham

Zhang, Y., Brady, M. and Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, *20*(1), pp.45-57.

Hashemi, M., 2019. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data*, *6*(1), pp.1-13.

Ioffe, S. and Szegedy, C., 2015, June. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). PMLR.

Dozat, T., 2016. Incorporating nesterov momentum into adam

Zhao, R., Qian, B., Zhang, X., Li, Y., Wei, R., Liu, Y. and Pan, Y., 2020, November. Rethinking dice loss for medical image segmentation. In *2020 IEEE International Conference on Data Mining (ICDM)* (pp. 851-860). IEEE.

Lisanti, C.J. and Douglas, D.B., 2009. Effects of breath-hold and cardiac cycle on the MRI appearance of the aorta and inferior vena cava in t2 abdominal imaging. *American Journal of Roentgenology*, *192*(5), pp.1348-1358.

Appendix

Figure 9. Training and validation graphs for loss, accuracy and dice, using Nadam optimizer is shown below.
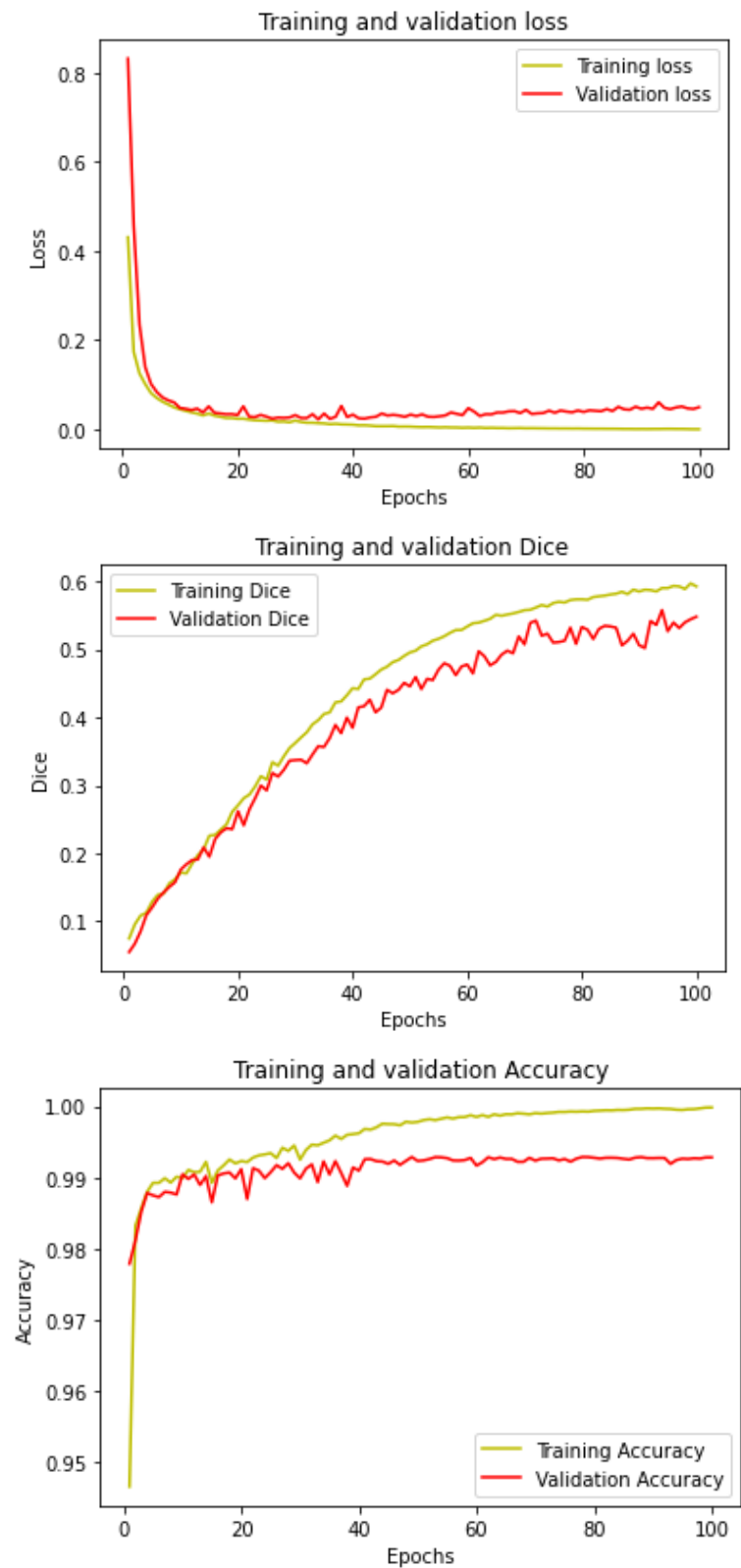
Figure 10. Training and validation graphs for loss, accuracy and dice, using Adam optimizer is shown below (Zoomed out)
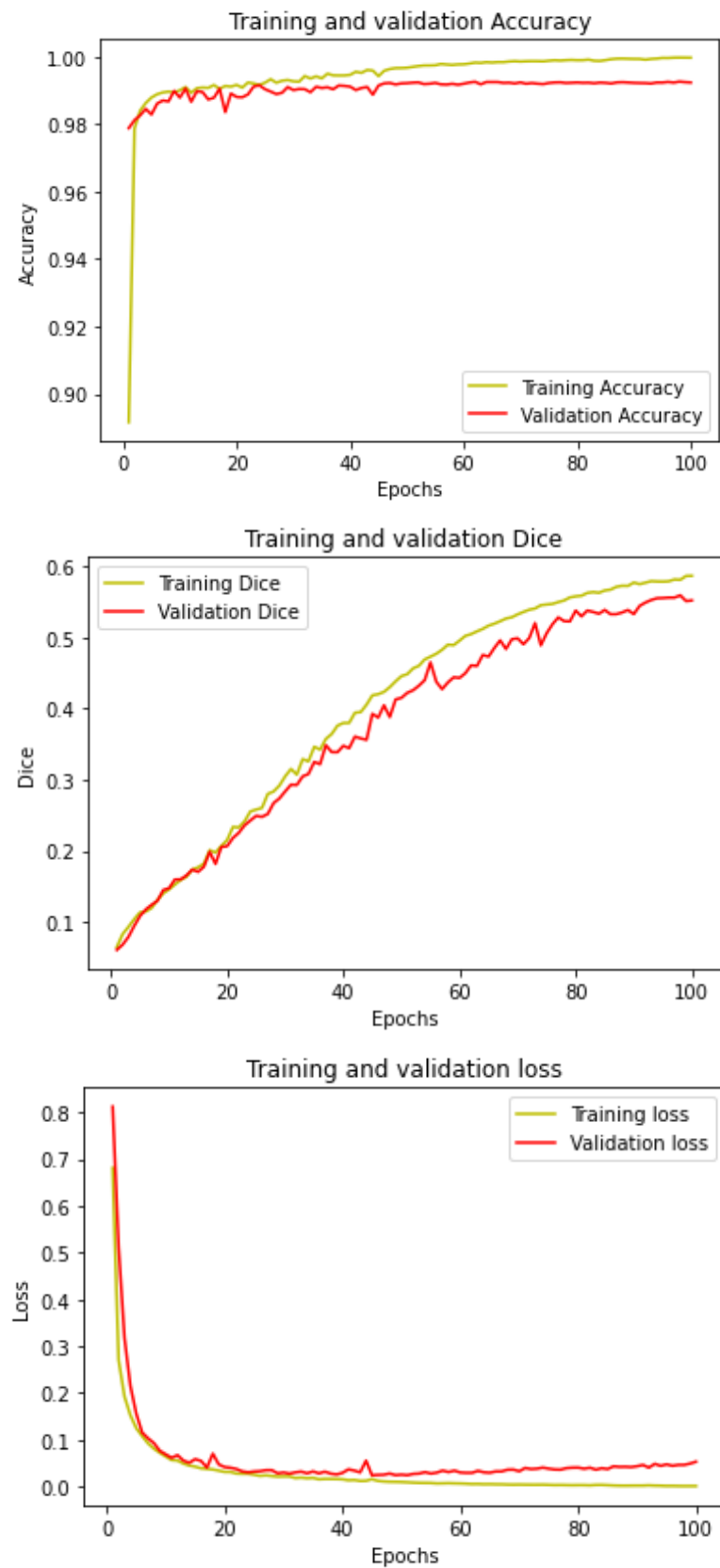
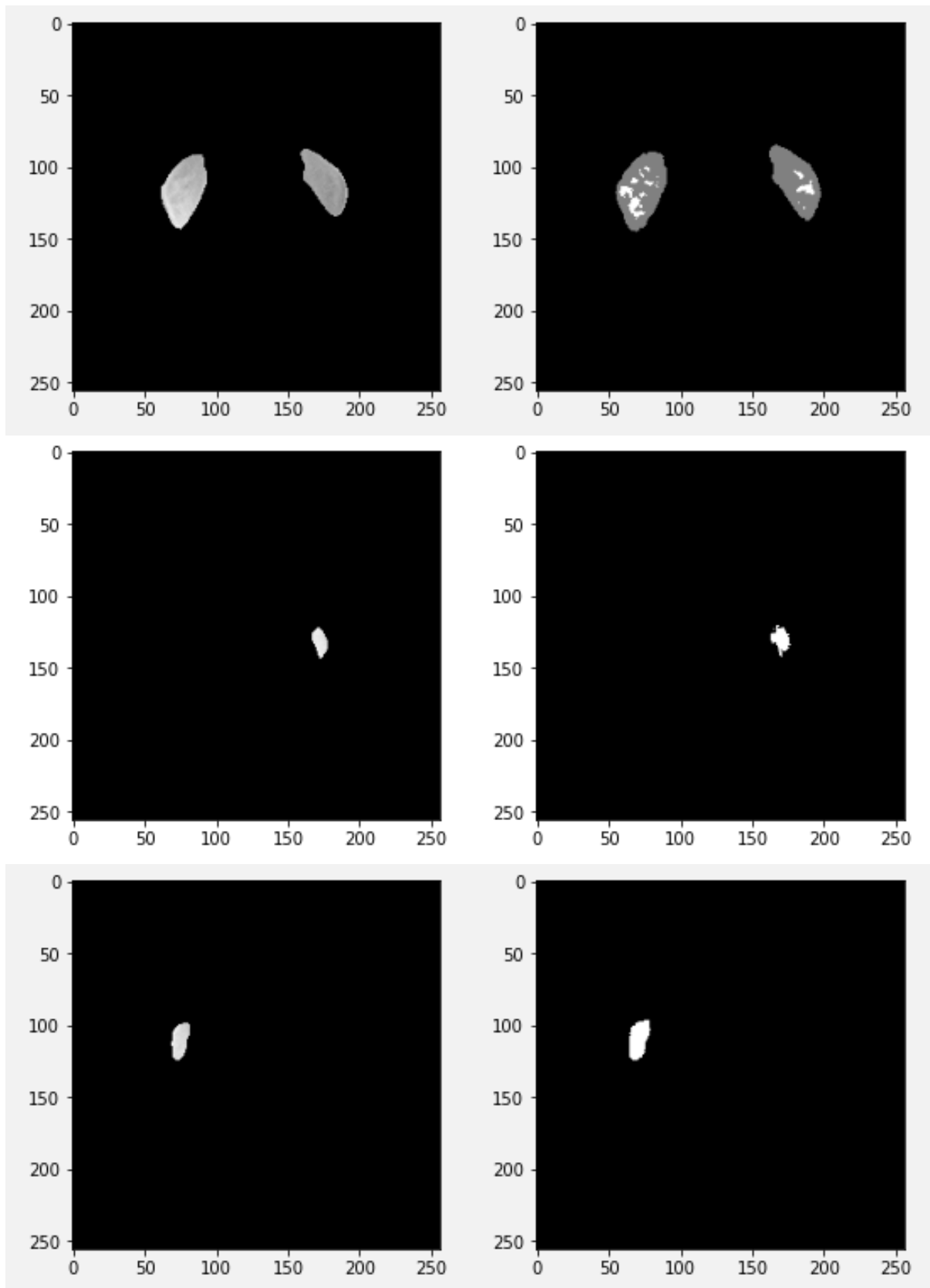**Figure 11**. Slices for all individuals and their corresponding masks
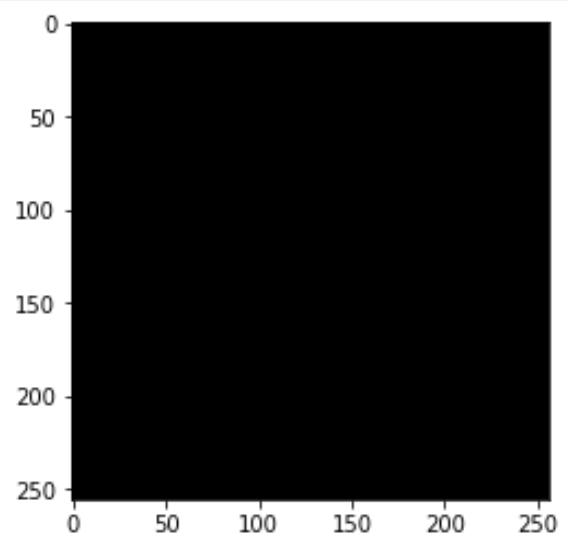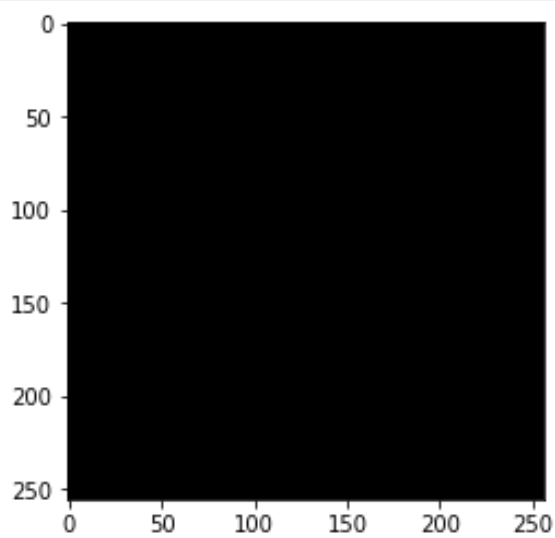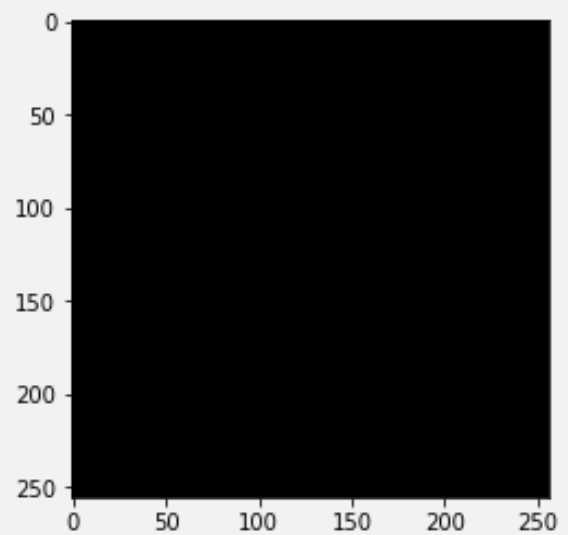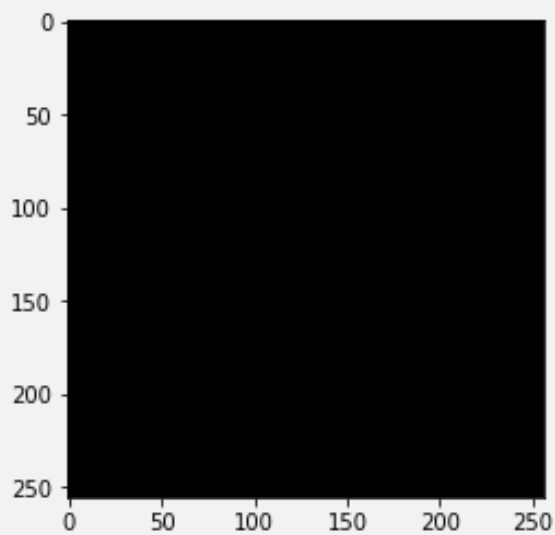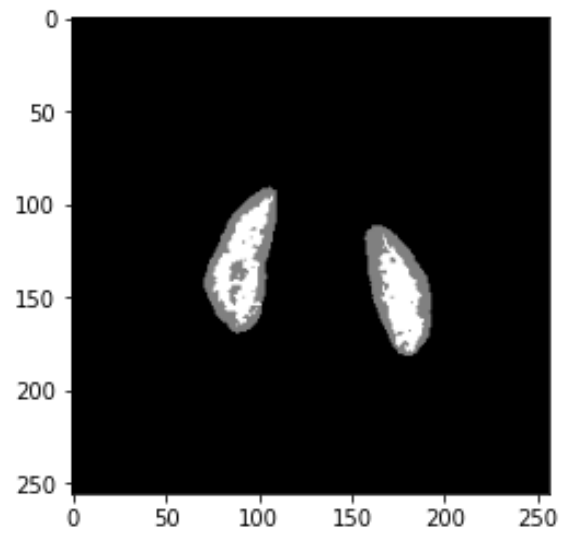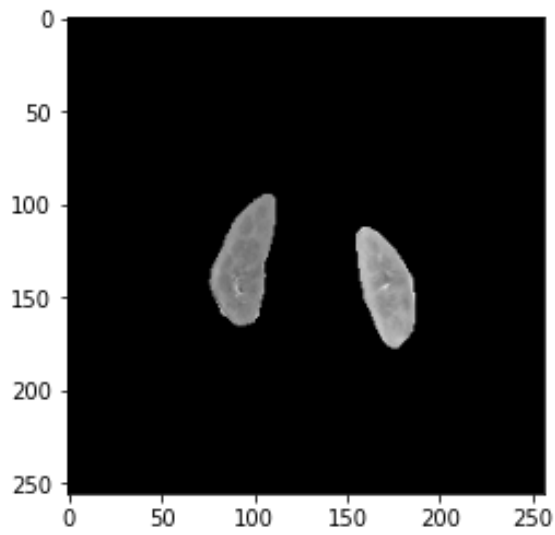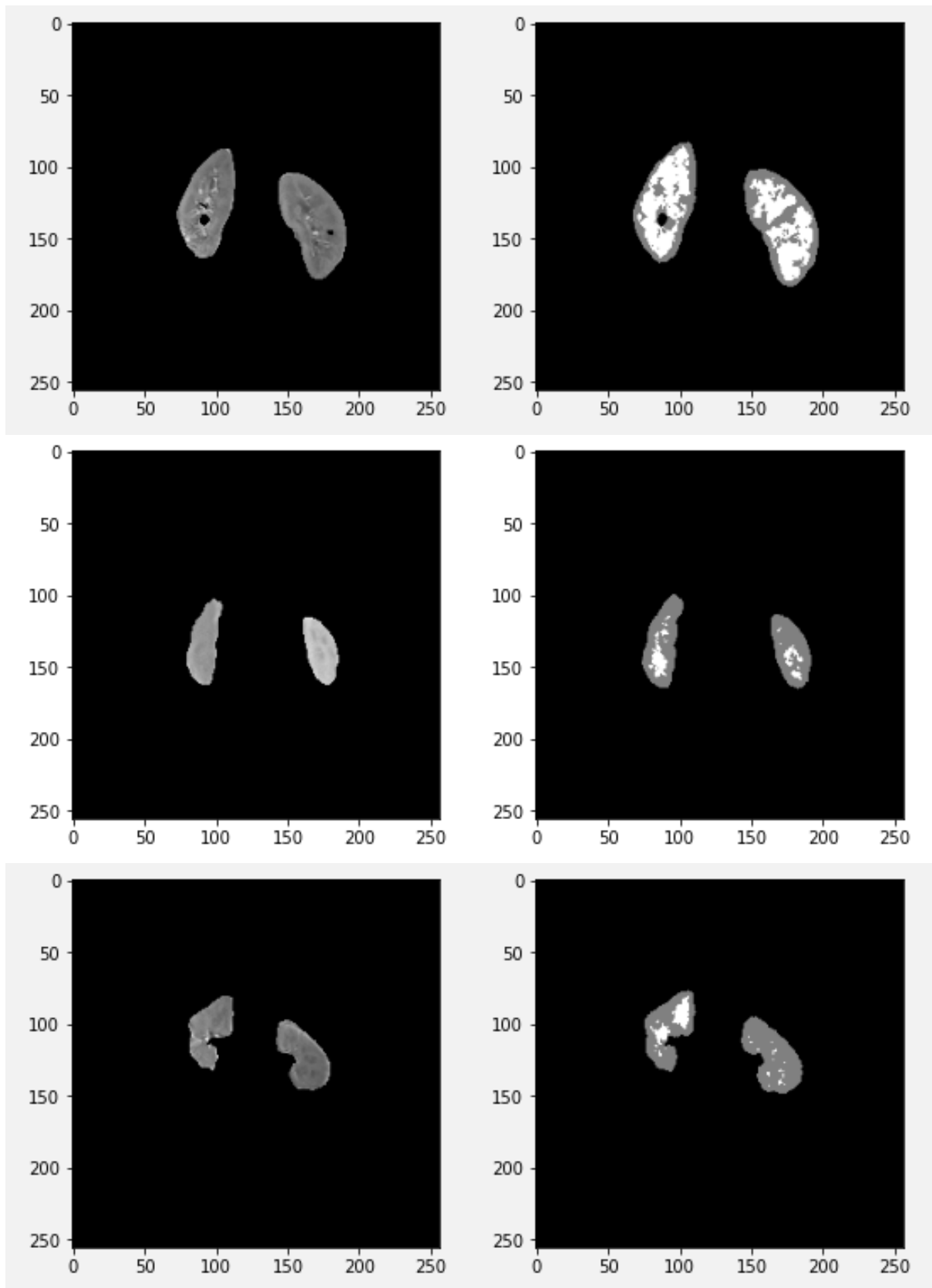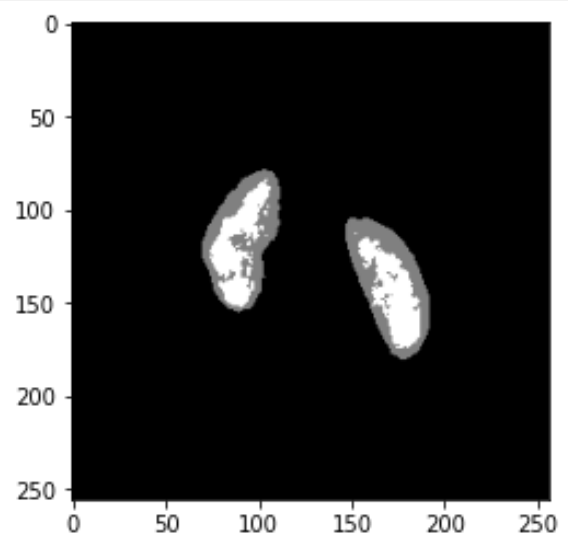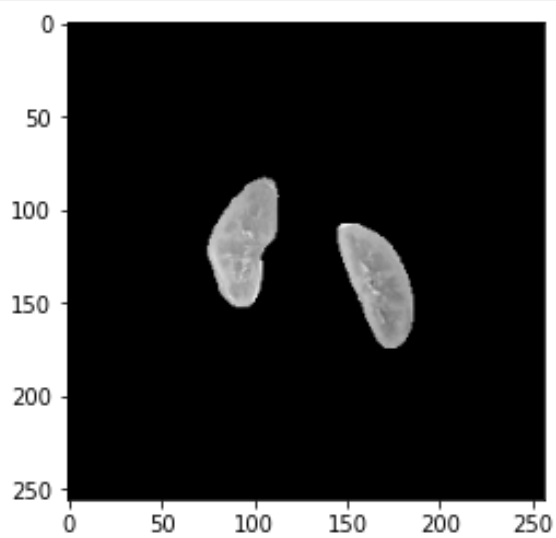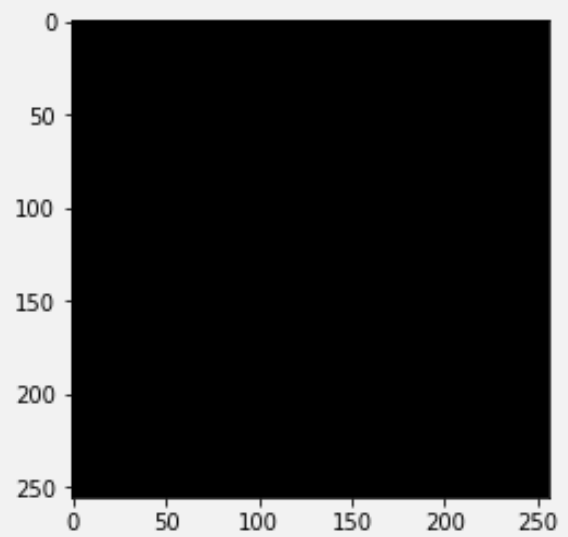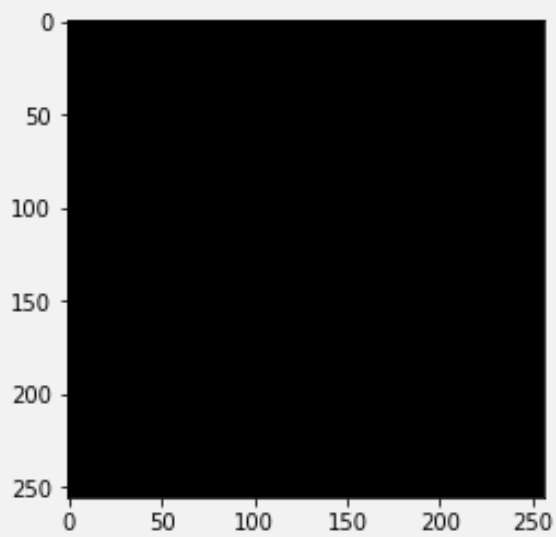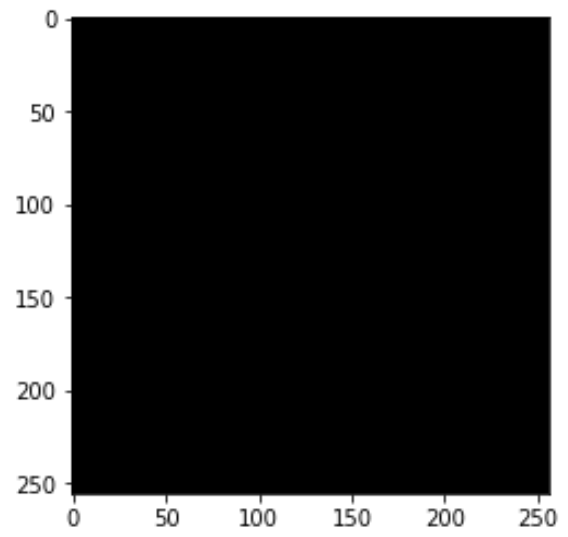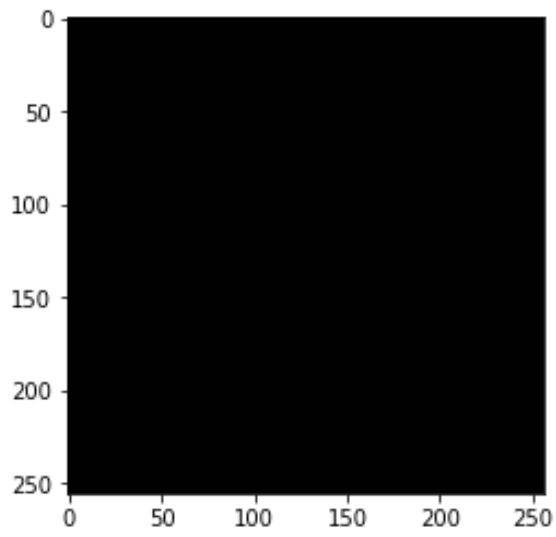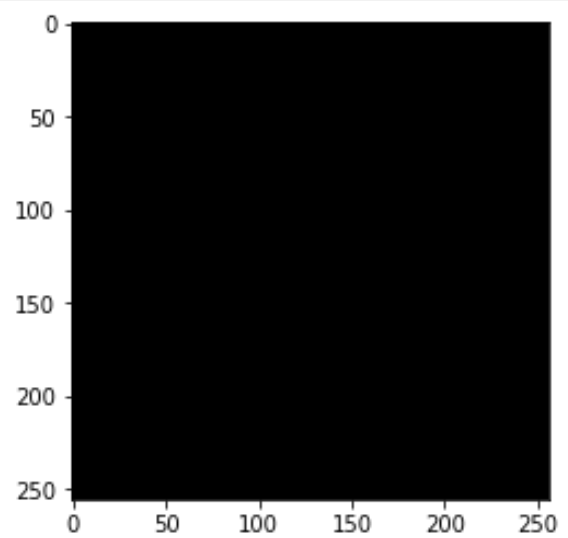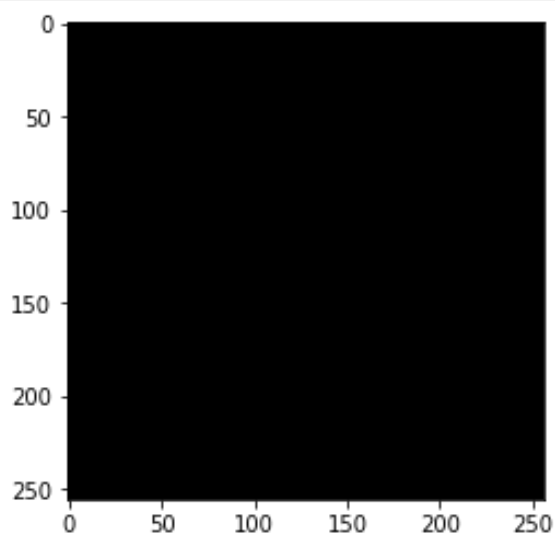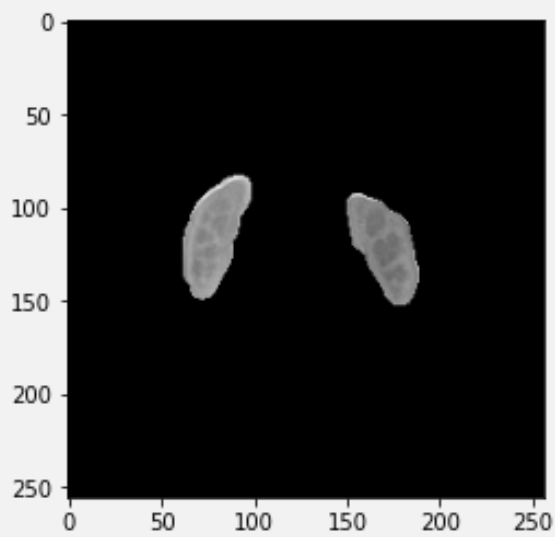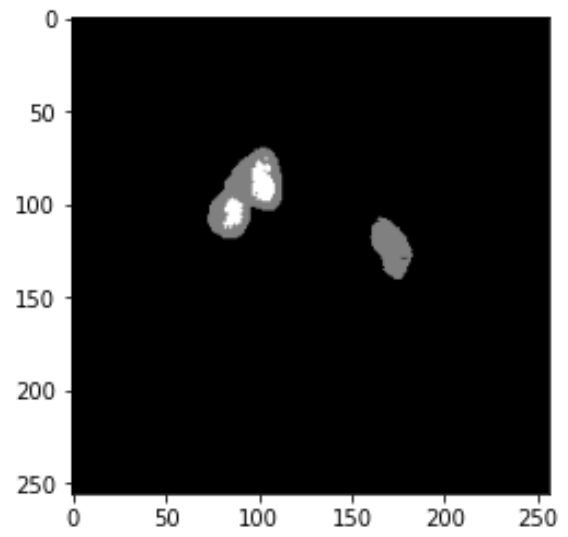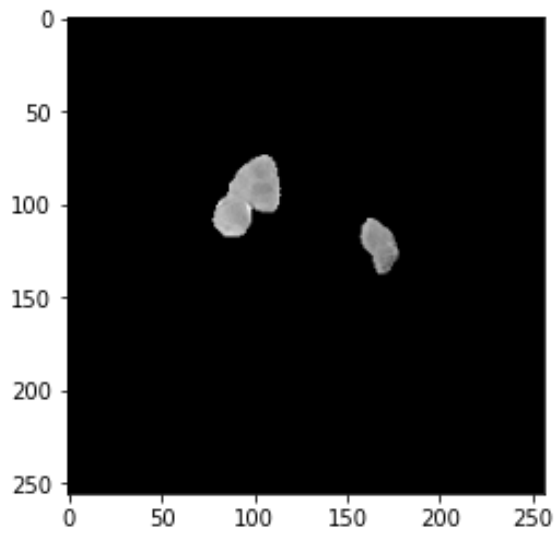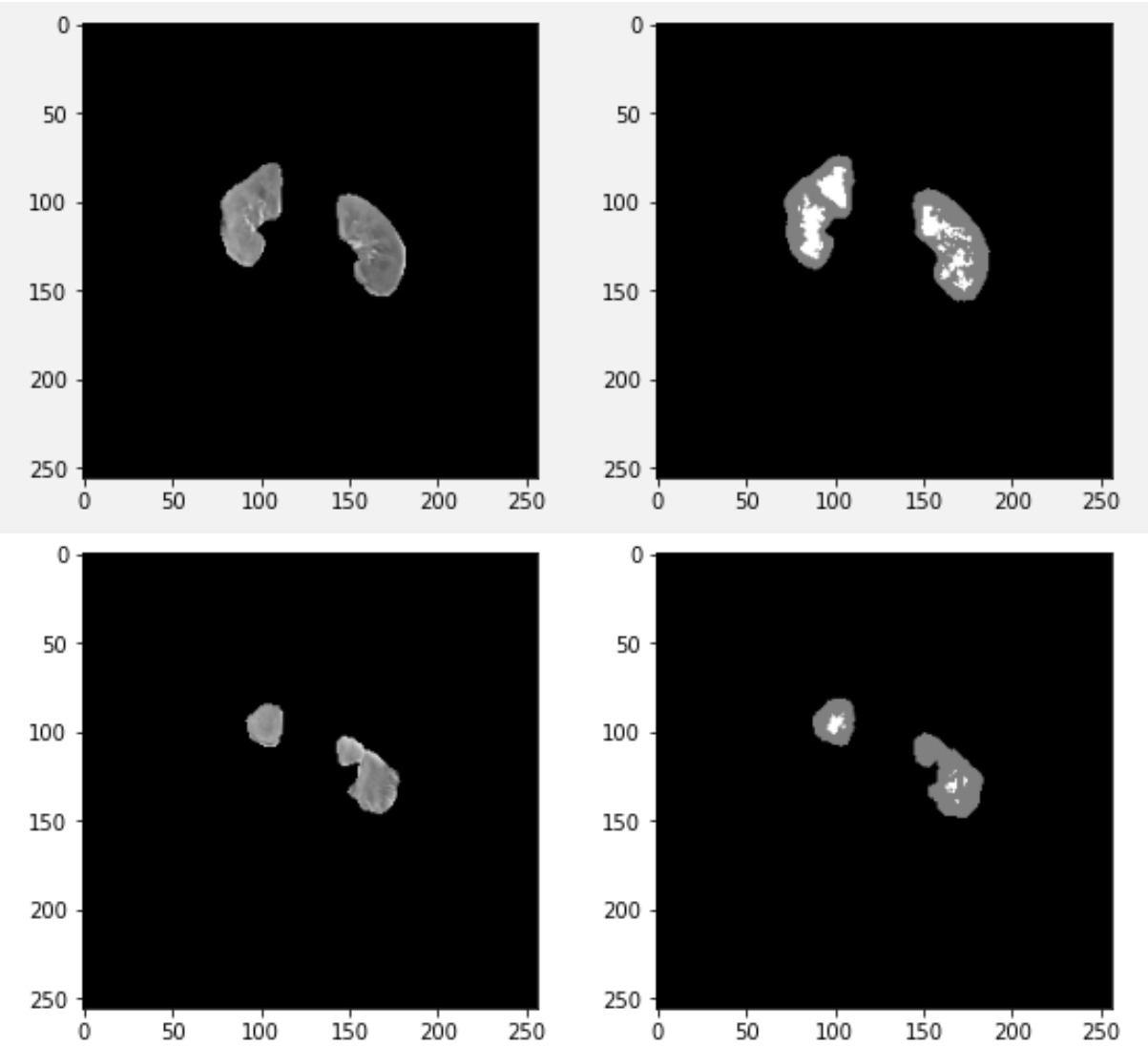
Table 2.

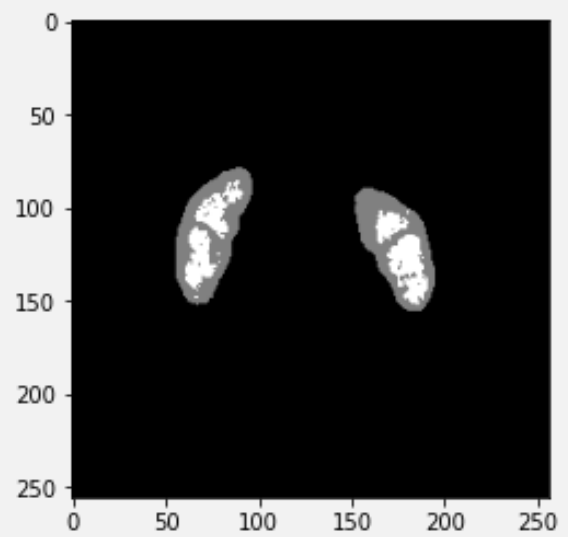| IOU for Cortex	IOU for Medulla
0.9974872	0.84625