

Article

Deep Learning Assisted Localization of Polycystic Kidney on Contrast-Enhanced CT Images

Djeane Debra Onthoni ^{1,†}, Ting-Wen Sheng ^{2,3,†}, Prasan Kumar Sahoo ^{1,4,*}, Li-Jen Wang ^{2,3,*} and Pushpanjali Gupta ¹

¹ Department of Computer Science and Information Engineering, Chang Gung University, Guishan 33302, Taiwan; d0421008@cgu.edu.tw (D.D.O.); d0521006@cgu.edu.tw (P.G.)

² Department of Medical Imaging and Radiological Sciences, Chang Gung University, Guishan 33302, Taiwan; steven.sheng@gmail.com

³ Department of Medical Imaging and Intervention, New Taipei Municipal TuCheng Hospital, Chang Gung Medical Foundation, New Taipei City 236017, Taiwan

⁴ Division of Colon and Rectal Surgery, Chang Gung Memorial Hospital, Linkou 33305, Taiwan

* Correspondence: pksahoo@mail.cgu.edu.tw (P.K.S.); lijenwang0918@gmail.com (L.-J.W.); Tel.: +886-3-211-8800 (ext. 3804) (P.K.S.)

† These authors contributed equally to this work.

Received: 3 December 2020; Accepted: 16 December 2020; Published: 21 December 2020



Abstract: Total Kidney Volume (TKV) is essential for analyzing the progressive loss of renal function in Autosomal Dominant Polycystic Kidney Disease (ADPKD). Conventionally, to measure TKV from medical images, a radiologist needs to localize and segment the kidneys by defining and delineating the kidney's boundary slice by slice. However, kidney localization is a time-consuming and challenging task considering the unstructured medical images from big data such as Contrast-enhanced Computed Tomography (CCT). This study aimed to design an automatic localization model of ADPKD using Artificial Intelligence. A robust detection model using CCT images, image preprocessing, and Single Shot Detector (SSD) Inception V2 Deep Learning (DL) model is designed here. The model is trained and evaluated with 110 CCT images that comprise 10,078 slices. The experimental results showed that our derived detection model outperformed other DL detectors in terms of Average Precision (AP) and mean Average Precision (mAP). We achieved mAP = 94% for image-wise testing and mAP = 82% for subject-wise testing, when threshold on Intersection over Union (IoU) = 0.5. This study proves that our derived automatic detection model can assist radiologist in locating and classifying the ADPKD kidneys precisely and rapidly in order to improve the segmentation task and TKV calculation.

Keywords: autosomal dominant polycystic kidney disease; contrast-enhanced computed tomography; deep learning

1. Introduction

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is the most common hereditary renal disease with an estimated prevalence of 1:1000 to 1:2500 [1,2]. The ADPKD kidneys are characterized by continuous proliferation and growth of the bilateral renal cysts, which lead to the compression and damage of healthy renal tissue, progressive renal enlargement, and eventually progress to end-stage renal disease in majority of patients [3]. Two important biomarkers need to be examined for predicting the progressive loss of renal function: Glomerular Filtration Rate (GFR) and Total Kidney Volume (TKV) [4]. TKV allows stratification of patients into low and high-risk subgroups to identify individuals who may benefit from the treatment [3,5]. TKV is calculated using common medical imaging tests such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). However, MRI technique is

expensive and takes a long examination time of 30–60 min. Known for its faster and low-cost technique, CT has been used globally and is a popular technique in clinical imaging tests. CT can be performed with or without intravenous contrast enhancement. Contrast-enhanced Computed Tomography (CCT) highlights the blood vessels and enhances organs which provides better contrast resolution on images as compared to Non-contrast-enhanced Computed Tomography (NCCT).

Precise TKV calculation requires manual localization and segmentation by clinical experts or trained personnel. Conventionally, the radiologist delineates the boundary of kidneys on images using semi-automatic tools at the medical image post-processing workstation. The use of these tools on CCT images remains a challenge with ADPKD kidney due to the contrast difference between the kidney and surrounding structure. In CT images, different body components such as air, fat, fluid, soft tissue, hemorrhage, calcification, and bones have different CT Hounsfield Units (HU) values which are displayed from low to high density. The normal kidney has homogeneous density and surrounding by the perirenal fatty tissue which shows lower density compared to soft tissue. However, in ADPKD, the kidney is replaced by multiple cysts of various densities (some of the cysts have high density due to hemorrhage or calcification), which cause the density of the kidney to become heterogeneous due to soft tissue, fluid, hemorrhage, and calcification as shown in Figure 1a. In addition, the kidney is enlarged and closely abutted to the adjacent soft tissue structure resulting in no contrast difference between the kidneys and surrounding structure. As often seen in ADPKD patients, the presence of multiple liver cysts would make segmentation of kidneys more difficult as shown in Figure 1a. It is also observed that the density of ADPKD kidney can be the same as other adjacent organs such as liver and spleen, as shown in Figure 1b. As compared to the renal cyst in non-ADPKD kidney, morphology and density of ADPKD kidney are non-uniform as shown in Figure 1c. Due to these reasons, if we want to calculate the TKV by planimetry, which is the gold standard method, the semi-automatic tool is not useful. Besides, manual segmentation of ADPKD kidney slice by slice is also a very time-consuming job.

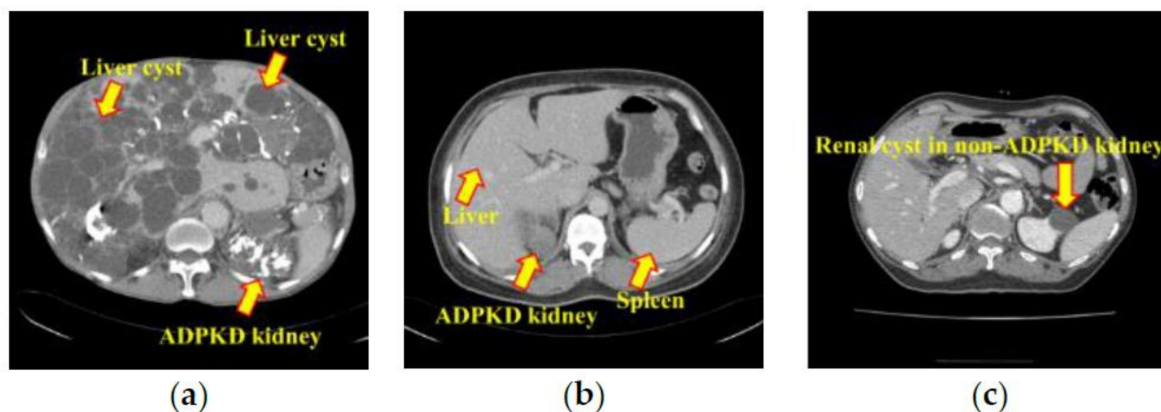


Figure 1. Kidney with different types of cysts: (a) Autosomal Dominant Polycystic Kidney Disease (ADPKD) kidney and liver cyst; (b) ADPKD kidney, liver, and spleen; (c) Renal cyst in non-ADPKD.

In recent studies, computerized techniques have been widely explored for solving the challenges in medical imaging big data analysis using Artificial Intelligence (AI). Powerful AI techniques including Machine Learning (ML) [6] and Deep Learning (DL) have been used to solve problems. DL has been used in several medical applications include feature extraction, classification [7], detection [8], and segmentation [9]. One crucial application in DL is the detection or localization task. This task is essential for localizing the particular Region of Interest (ROI), which could be any nodule, cyst, tumor, or cancer in an image. It is also found that detection can be used for improving the segmentation task of ADPKD kidneys [10–13]. Therefore, the detection task can be considered as an important intermediate technique for segmentation. The detection task can be carried out in three different approaches, which could be (1) Detection by classifying a single ROI, (2) detection by locating and classifying a single ROI and (3) detection by locating and classifying multiple ROIs. One existing

work [14] has used the detection by classification as an intermediate technique to detect the presence of ADPKD kidney on MRI using Convolutional Neural Network (CNN). Similarly, an ADPKD detection model has been designed using CCT images and CNN architecture which was developed based on two training sets: positive and negative kidney training patches [11]. However, the classified results only comprise information about the presence of ROI in an image without indicating the location of the ROI. In the second approach, detection is performed by classifying and locating a single ROI. However, this approach is not efficient when applied to ADPKD as we need to differentiate multiple ROIs that comprise left and right kidneys.

The object detection approach is the combination of classification and localization for detecting multiple ROIs, where it has been used as an intermediate technique for improving the segmentation task. For instance, Region with Convolutional Neural Networks (R-CNN) has been used as an object detection architecture for detecting ADPKD kidneys on MRI images with higher numbers of False Positives (FP) [15]. Several CNN architectures have been used for segmentation and object detection tasks applying Visual Geometry Group 16 layers (VGG-16), and R-CNN, respectively, on MRI [16]. Various applications have been designed using the object detection approach for different purposes. Those applications are universal lesion detection using CT images, VGG-16, and Region Proposal Network (RPN) [17], and breast cancer detection using histopathology images and Fully Convolutional Network (FCN) [18]. Furthermore, we found that less work has been done to solve the detection problem in ADPKD kidneys using an object detection approach.

Current state-of-the-art results for automated object detection can be achieved by applying several CNN DL architectures on various medical imaging techniques such as CT, CCT, NCCT, MRI, histopathology, Fundus image, colonoscopy, etc. However, the applicability of the object detection approach for localizing and classifying ADPKD kidneys using CCT images has not been thoroughly investigated. Therefore, in this work, we propose an automatic detection of ADPKD using DL on CCT images.

2. Materials and Methods

2.1. Data Acquisition

This study was approved by the Chang Gung Memorial Hospital Institutional Review Board, project identification code 201701583B0C501, approved on 18 December 2017. For this study, we collected CCT images of ADPKD patients from the Picture Archiving and Communication System (PACS) of Linkou Chang Gung Memorial Hospital from 2003 to 2019. There was a total of 110 CCT acquisitions retrieved from 97 ADPKD patients, composed of 10,078 CCT raw images. We collected one CT scan from each of 85 patients, two different CT scans from each of 11 patients, and three different CT scans from one patient. Table 1 shows the summarization of collected ADPKD patients' gender, age, and TKV. The section thickness and interval of the collected CCT images were 5 mm and 5 mm, respectively. CCT images in Digital Imaging and Communications in Medicine (DICOM) format were displayed with a window level and width of 35 HU and 350 HU, respectively. Besides, some of the ADPKD patients with liver cysts were verified by expert radiologists.

Table 1. 110 Contrast-enhanced Computed Tomography (CCT) acquisitions from 97 ADPKD patients.

Characteristic	Mean \pm SDs (Range) or Number	
Age at examination (yrs)	54.59 \pm 18.5 (27–88)	
Sex	Male	52
	Female	45
TKV (cm ³)	2734.33 \pm 2312.45 (345.14–13,666.88)	

2.2. Ground Truth Annotation

Two radiologists with more than 10 years of experience annotated the left and right kidney in CCT raw images as ground truth or gold standard. The annotation was done manually by drawing the boundary of kidneys using OSIRIX MD v10.0.5. The annotation was applied to both kidneys in all 110 CCT images. Figure 2 shows the example of CCT raw image (Figure 2a) and the respective ground truth image of right kidney (Figure 2b), left kidney (Figure 2c), and both kidneys (Figure 2d).

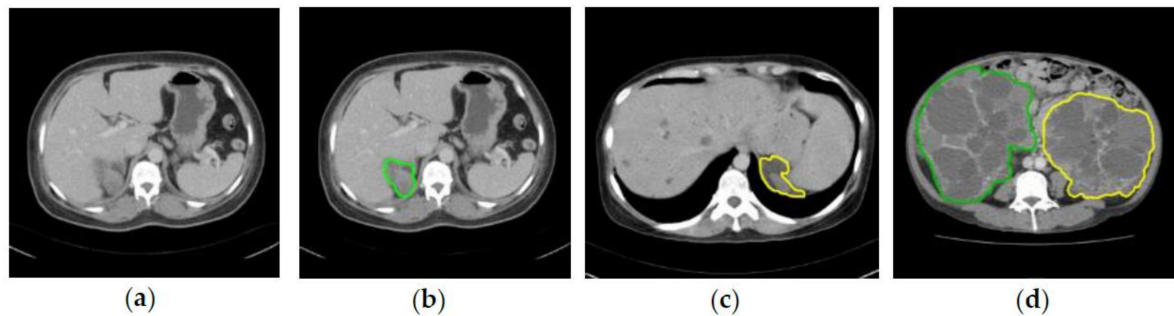


Figure 2. CCT raw image and respective ground truth images: (a) Raw image; (b) Ground truth for right kidney (Green); (c) Ground truth for left kidney (Yellow); (d) Ground truth for both right (Green) and left (Yellow) kidneys.

2.3. Methods

In this section, we describe the proposed method, which is composed of adopted image preprocessing techniques and a Single Shot Detector (SSD) Inception V2 DL model. Figure 3 shows the proposed automatic ADPKD detection model framework.

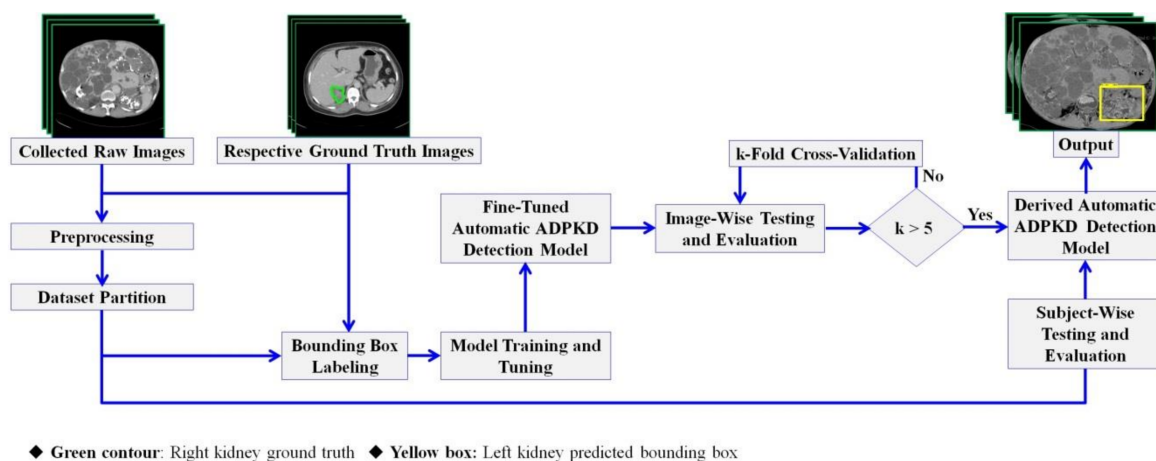


Figure 3. Proposed automatic ADPKD detection model framework.

2.3.1. Preprocessing

The preprocessing procedure is composed of four stages, namely; (1) Slice Selection, (2) Joint Photographic Experts Group (JPEG) Conversion, (3) Image Enhancement, and (4) Automatic Cropping. We aimed to prepare the training set with less noise and eliminate the unwanted area. Figure 4 shows an example of preprocessing procedures. Firstly, some image slices were selected from the collected raw images in DICOM format based on the inclusion and exclusion criteria. The selection criterion was decided based on the presence of either left or right kidney in the images. Thus, total number of collected raw images was reduced from 10,078 to 4648 images. Secondly, all selected images were converted to JPEG format using open source software such as RadiAnt DICOM Viewer without changing the dimension, quality of raw images, and disabling patient's information. Thirdly, the image

enhancement method was applied to reduce the noises such as high-density organs in the spine, ilium, and cysts or kidney stones. To cope with variant of intensity, an intensity segmentation based approach was chosen as in [19], where a global thresholding method was adopted. Based on our experiment and observation, we found the preferable value of threshold $T_{\max} = 195$ and $T_{\min} = 45$. The thresholding operation can be expressed as follows:

$$\text{dst}(x, y) = \begin{cases} \text{MaxVal} & \text{if } \text{src}(x, y) > T_{\max} \\ 0 & \text{if } \text{src}(x, y) < T_{\min} \end{cases} \quad (1)$$

where, $\text{MaxVal} = 120$ denotes maximum pixel intensity, $\text{src}(x, y)$ in the source image pixel intensity, and $\text{dst}(x, y)$ refers to the destination image pixel intensity. Fourthly, an automatic cropping mechanism was designed to avoid bias by eliminating unimportant parts in the image without modifying the dimension. As shown in Figure 4, kidneys are located in center of the abdominal cavity surrounded by the black pixel intensity, where $\text{src}(x, y) = 0$. Dilation morphological [20] method was applied to emerge the shape of abdominal cavity. Considering the enhanced image as a binary input image B , and kernel size $S = 4 \times 4$, the dilation of $B \oplus S$ was performed. Moreover, we automatically cropped the whole abdomen cavity area by finding the maximum contour area [21] using hierarchy contour. Then rectangular shape was drawn by taking the maximum area of the contour. Lastly, we cropped the rectangular area by using the coordinates $y_{\min}(y) : y_{\max}(y + h)$, $x_{\min}(x) : x_{\max}(x + w)$, where x, y are coordinates, and h, w represents the height and width, respectively.

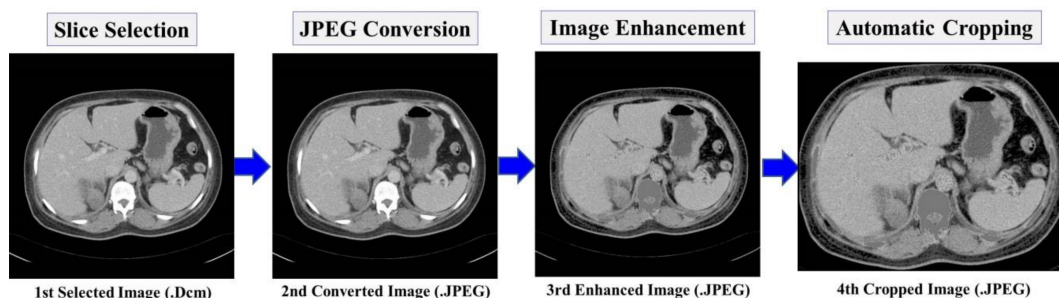


Figure 4. Preprocessing procedures.

2.3.2. Dataset Partition

Initially, the dataset of 110 subjects (e.g., 4648 slices) is partitioned randomly into training and testing set with a ratio of 80:20. Accordingly, 88 subjects (i.e., 3718 slices) and 22 subjects (i.e., 930 slices) were selected for the training and testing set, respectively. Then, we carried out the model training, tuning, and testing. It is to be noted that the set of images in the training set were not included in the testing set. The 88 subjects (i.e., 3718 slices) of the training set were used to train and tune the model. We performed the image-wise testing and evaluation using 5-fold cross validation [22], where each k_i , an element of k -subsets, will have an approximately equally sized image, resampled randomly from all considered subjects. Then, we assigned randomly each k_i for the testing set and the rest ($k-1$) were assigned for the training set. For all k -rounds of training and testing, the final image-wise testing results were obtained from the average results of k_i testing sets. This technique is used to assess the overfitting problem and evaluate the robustness of our fine-tuned trained model, as given in Section 2.3.5. Lastly, to test our derived detection model with the unseen data, we performed subject-wise testing on 22 subjects (i.e., 930 slices) testing set.

2.3.3. Bounding Box Labeling

Bounding box labeling was performed after completing the preprocessing procedure as given in Section 2.3.1 and partitioning the dataset as described in Section 2.3.2. this is required to learn the coordinates and classes of the kidneys during the training process. The respective ground truth

images were used as reference for the bounding box labeling. The labeling was carried out by drawing boxes around the kidneys and assigning classes within an image using open source Labelling software v1.8.4 [23]. By doing so, the coordinates and classes of the kidneys were saved in Extensible Markup Language (XML) files in PASCAL VOC format. To verify our bounding box labeled with the annotated ground truth, we redrew the bounding box on the annotated ground truth. Then, we checked the coordinates, where our derived bounding box labeled coordinates should be greater than the annotated ground truth coordinates.

2.3.4. Automatic ADPKD Detection Model

We proposed an automatic ADPKD detection model using a DL object detection approach, specifically a regression-based approach. We selected SSD [24] Inception V2 [25] as the adopted network based on pre-trained model performance speed, using the Microsoft Common Objects in Context (COCO) dataset [26], a large-scale dataset for DL applications such as object detection, segmentation, and captioning. Based on the single feed-network approach, SSD comprises two main layers: extraction layer and detection layer. SSD is categorized as a one stage detector, where the detection is performed directly after extracting the feature maps through the CNN backbone. Thus, SSD can work faster as compared to a three stages detector such as Viola and Jones, Histogram of Oriented Gradient (HOG), etc., and a two stages detector such as R-CNN, Faster R-CNN, etc. The architecture of the proposed detection model is illustrated in Figure 5.

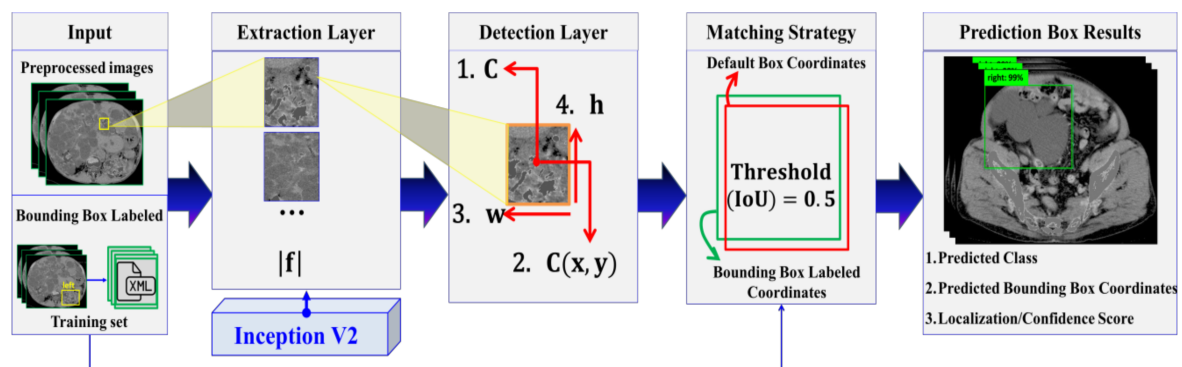


Figure 5. The architecture of automatic ADPKD detection model, where $|f|$, $C(x, y)$, w , h , and $V2$ refer to as total number of feature maps, center bounding box, width, height, and version 2, respectively.

In input layer, all preprocessed images of size = 224×224 and the corresponding labeled bounding box coordinates of XML files were fed into the extraction layer, where the total number of feature maps $|f|$ was extracted using a pre-trained Inception V2 CNN network. The feature maps were extracted by implementing minimum depth of convolutional layers = 16 and the Rectified Linear Unit (ReLU) activation function. In the next step, each extracted feature map was passed to the detection layer, where four categories need to be detected: the total number of classes C , center bounding box $C(x, y)$ coordinates, width w , and height h . To detect these four categories, default detector boxes were used in each feature map, where different scales, scale max = 0.2, scale min = 0.95, and a set of aspect ratio = $\{1, 2, 3, 1/2, 1/3\}$ were considered in our implementation. The detection was implemented with total number of default boxes = 4 and kernel size 3×3 . Furthermore, a matching strategy was applied based on the best fit in terms of aspect ratio, scale, and location. The comparison was based on the Jaccard index or Intersection over Union (IoU), where threshold = 0.5 was considered in our experiment. For a particular class, a matching method was carried out by comparing the labeled bounding box with all generated default boxes' coordinates as shown in Equation (2).

$$\text{Match}(\text{Bounding Box Labeled}[Y_{\min}(y) : Y_{\max}(y + h), X_{\min}(x) : X_{\max}(x + w)], \quad (2)$$

$$\text{Default Boxes}[Y_{\min}(y) : Y_{\max}(y + h), X_{\min}(x) : X_{\max}(x + w)]),$$

Based on the comparison, prediction boxes were produced, where each predicted box comprises the predicted class, predicted bounding box coordinates [$Y_{\min}(y):Y_{\max}(y + h)$, $X_{\min}(x):X_{\max}(x + w)$] and localization or confidence score. The prediction results were categorized into three metrics: True Positive (TP) if localization score \geq threshold with correct classification, False Positive (FP) if localization score $<$ threshold with wrong classification, and False Negative (FN) if labeled bounding box is not detected by the model. Moreover, Non-Maximum Suppression (NMS) was applied to find the best prediction box results for each class. The smooth L1 was used for localization loss L_{L_0} and Softmax was applied for classification loss L_{co} .

2.3.5. Training and Tuning Model

We performed the training and tuning model without cross-validation strategy on 88 subjects (i.e., 3718 slices) of training data. As reported in [27], cross-validation is known as internal validation. Using both techniques together can lead to high variance and non-optimized hyper-parameters. Moreover, the random search technique, one of the hyper-parameter tuning strategies which outperforms the grid search technique, was applied. To improve the robustness of our designed model, selection of the best hyper-parameters is performed through model training. Based on our experiment, a fine-tuned trained model was achieved by setting the optimal value of the hyper-parameters: initial learning rate = 0.004, decay = 0.9, epsilon = 0.001, momentum optimizer = 0.9, batch size = 24, and step = 6000. Furthermore, two data augmentation techniques, random horizontal flip, where an input image is flipped from left to right with probability of 0.5, and SSD random crop, where subsets of patches are cropped from an input image randomly, were applied.

2.3.6. Image-Wise and Subject-Wise Testing and Evaluation

To verify how robust our fine-tuned trained model is, image-wise testing and evaluation using 5-fold cross-validation were performed, with k -rounds of training and testing, where one fold k_i was assigned to the testing set in each round, and the remaining $(k - 1)$ subsets were assigned to the training sets. After testing and evaluating our fine-tuned trained model, performance of detection was evaluated by taking the average evaluation metrics of whole k -rounds. Lastly, we assessed and evaluated subject-wise testing to finally test our fine-tuned trained model using 22 subjects (i.e., 930 slices).

2.4. Experimental Setup

The automatic ADPKD detection model was built using the OpenCV v3.2.0 image preprocessing tool for image enhancement and automatic cropping, and TensorFlow-GPU v1.12 image analysis tool for detection model. Python v3.6.9 programming language was used as an interface for both tools. Moreover, high-performance hardware and software are required to execute the experiments. The hardware specification was GPU TITAN RTX 24 GB \times 4 and 256 GB memory. For software, we used Ubuntu v18.04.3 operating system including python libraries Pandas v1.0.5, Numpy v1.19.1, Matplotlib v3.3.0, Pillow v7.2.0, osmnx v0.15.1, lxml v4.2.1, imageio v2.5.0, Urllib3 v1.22, Sys v3.6.9, and Zipfile v0.5.1.

2.5. Evaluation Metrics

IoU was used to evaluate the performance of our ADPKD detection model. This metric is commonly used for object detection and segmentation evaluation, where IoU is defined by the intersection between predicted box coordinates and labeled bounding box coordinates divided by their union. Based on IoU calculation, TP, FP, and FN can be defined. Moreover, metrics including

Accuracy, Localization loss (L_{lo}), Classification loss (L_{co}), Precision, Recall/Sensitivity, and F1-Score were considered and calculated as follows:

$$\text{Accuracy} = \frac{TP}{TP + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1 - Score/Dice - Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (6)$$

After defining other metrics, Average Precision (AP) and mean AP (mAP) were calculated. Both metrics were calculated based on precision-recall calculation metrics. The precision-recall curve can be plotted graphically into a 2D graph, where X-axis denotes the recall and Y-axis represents the precision range [0, 1]. AP for predicted class can be calculated by finding precision-recall Area Under the Curve (AUC). Based on the AP value, mAP value can be calculated by taking average AP of both kidney classes. In each round, precision and recall were calculated as given in Equations (4) and (5) with IoU threshold set to be 0.5. Then AP and mAP values were computed. These procedures were carried out until k-rounds were completed.

2.6. Evaluation Procedures

Firstly, we performed image-wise testing and evaluation on our fine-tuned ADPKD detection model using k-fold cross-validation technique. Secondly, we conducted subject-wise testing. Both testing evaluations were carried out by comparing our method with other well-established DL pre-trained object detection models including CNN backbones, which have been adopted in similar current approaches for pneumonia detection in chest X-ray [28] and for malignant pulmonary nodule detection in CT scans [29]. These models and backbones were SSD Inception V2, Faster-RCNN, ResNet, and MobiNet. Alongside, we included DL object detection models, which have been trained using large-scale datasets including COCO, Kitti, Open Image, etc., developed by TensorFlow [30]. Correspondingly, we selected models such as original SSD Inception V2, SSD MobileNet V1, Faster R-CNN NAS, Faster R-CNN Inception ResNet V2, and R-FCN ResNet 101. We trained all selected pre-trained models by partitioning the data set as given in Section 2.3.2 and evaluated using k-fold cross validation techniques as given in Section 2.3.6.

3. Results

3.1. Evaluation Results of Image-Wise Testing

We conducted the experiments on image-wise testing, and compared the results with other detection architectures as shown in Table 2. It was observed that our model achieved the highest performance in various evaluation metrics with an accuracy of 0.90 for right and 0.91 for left kidney. It was also found that our model outperformed over other architectures with an optimal confidence F1-Score of 0.86 for right kidney and 0.88 for left kidney.

Table 2. Evaluation metrics results on image-wise testing.

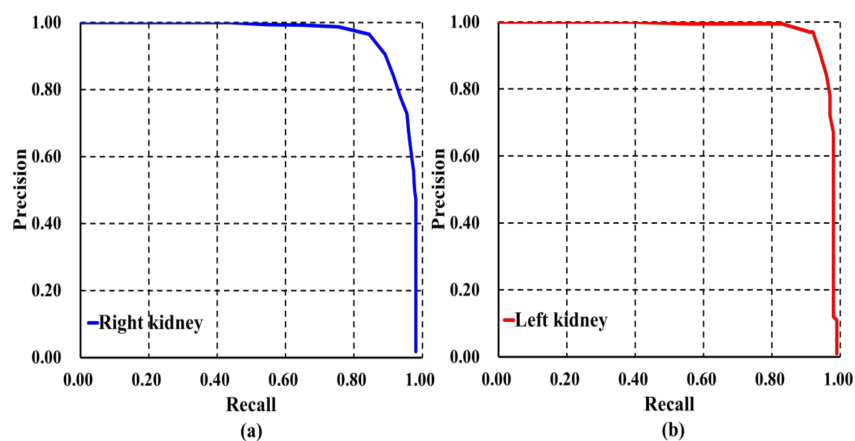
Detection Architectures	Class	Evaluation Metrics			
		Accuracy	Precision	Recall	F1-Score
Our Model	right	0.90 (± 0.06)	0.92 (± 0.07)	0.82 (± 0.02)	0.86 (± 0.04)
	left	0.91 (± 0.06)	0.92 (± 0.06)	0.84 (± 0.06)	0.88 (± 0.05)
Single Shot Detector (SSD) Inception V2	right	0.86 (± 0.04)	0.90 (± 0.03)	0.80 (± 0.04)	0.84 (± 0.02)
	left	0.86 (± 0.04)	0.91 (± 0.03)	0.82 (± 0.03)	0.86 (± 0.03)
SSD MobileNet V1	right	0.73 (± 0.1)	0.75 (± 0.01)	0.72 (± 0.1)	0.72 (± 0.09)
	left	0.71 (± 0.1)	0.81 (± 0.06)	0.66 (± 0.1)	0.71 (± 0.1)
Faster Region with Convolutional Neural Networks (R-CNN) NAS	right	0.57 (± 0.1)	0.51 (± 0.04)	0.69 (± 0.07)	0.58 (± 0.01)
	left	0.46 (± 0.06)	0.69 (± 0.02)	0.43 (± 0.06)	0.52 (± 0.04)
Faster R-CNN Inception ResNet V2	right	0.27 (± 0.05)	0.43 (± 0.09)	0.26 (± 0.09)	0.31 (± 0.07)
	left	0.52 (± 0.04)	0.50 (± 0.02)	0.67 (± 0.1)	0.57 (± 0.06)
Region-FCN (R-FCN) ResNet 101	right	0.37 (± 0.09)	0.57 (± 0.1)	0.34 (± 0.08)	0.42 (± 0.07)
	left	0.53 (± 0.08)	0.58 (± 0.05)	0.67 (± 0.06)	0.62 (± 0.05)

NAS: Neural Architecture Search.

In addition to different evaluation metrics, AP and mAP were calculated to analyze the robustness of our localization model, the results of which are presented in Table 3 and Figure 6. It is observed that our model was able to locate and classify both kidneys with AP right = 0.934 and AP left = 0.944 as shown in Figure 6a,b, respectively. We also found that our model was able to localize and classify both kidneys with the highest value of mAP = 0.94 in comparison to other well-established detection architectures.

Table 3. Comparison of Average Precision (AP) and mean Average Precision (mAP) with other architectures on image-wise testing.

Detection Architectures	Evaluation Metrics (Average (\pm SD))		
	AP		mAP
	Right	Left	Both Classes
Our Model	0.934 (± 0.01)	0.944 (± 0.01)	0.94 (± 0.01)
SSD Inception V2	0.844 (± 0.03)	0.866 (± 0.04)	0.855 (± 0.03)
SSD MobileNet V1	0.679 (± 0.10)	0.705 (± 0.10)	0.692 (± 0.09)
Faster R-CNN NAS	0.594 (± 0.03)	0.625 (± 0.03)	0.609 (± 0.03)
Faster R-CNN Inception ResNet V2	0.491 (± 0.08)	0.568 (± 0.06)	0.53 (± 0.07)
R-FCN ResNet 101	0.506 (± 0.11)	0.649 (± 0.11)	0.578 (± 0.11)

**Figure 6.** Precision-recall curve of our model on image-wise testing set: (a) Right kidney; (b) Left kidney.

It is observed that our model has higher average value of Classification loss $\overline{L}_{co} = 3.137$ and Localization loss $\overline{L}_{lo} = 0.464$ as compared to the original SSD Inception V2 Classification loss $\overline{L}_{co} = 2.511$ and Localization loss $\overline{L}_{lo} = 0.353$. However, our model has lower average value of Classification and Localization loss as compared to SSD MobileNet V1, where SSD MobileNet V1 Classification loss $\overline{L}_{co} = 3.423$ and Localization loss $\overline{L}_{lo} = 0.52$. Figure 7 shows the comparison of loss calculation graphically.

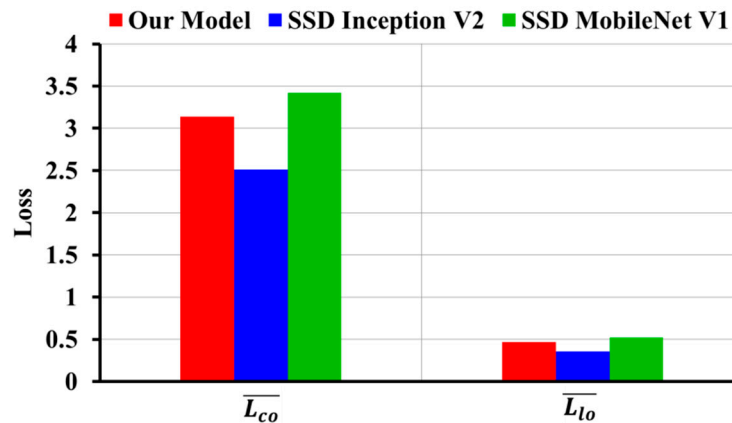


Figure 7. Comparison of classification and localization loss on image-wise testing set.

Figure 8a shows the example of results generated by our derived ADPKD detection model on CCT. It can be seen that our model works robustly though ADPKD kidneys are heterogeneous in density, in particular when multiple liver cysts are present. Similarly, adjacent to the liver and spleen, our model was able to localize and classify the kidneys as shown in Figure 8b.

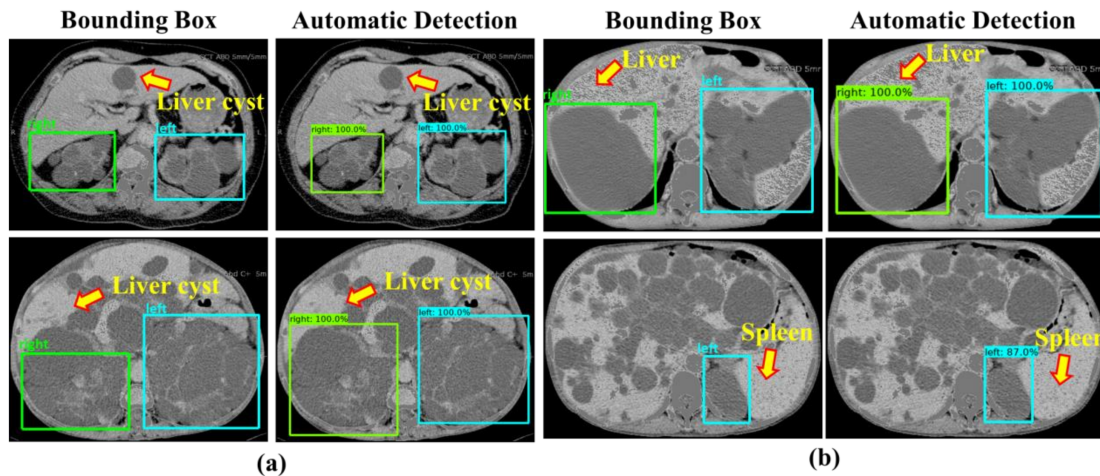


Figure 8. Detection results: (a) ADPKD kidneys associated with liver cysts; (b) ADPKD kidneys with adjacent organs.

3.2. Evaluation Results of Subject-Wise Testing

Upon performing the image-wise testing, we conducted subject-wise testing. As reported in Table 4, our proposed model outperformed all object detection architectures, where the average of all metrics is 0.8 for right kidney and 0.816 for left kidney.

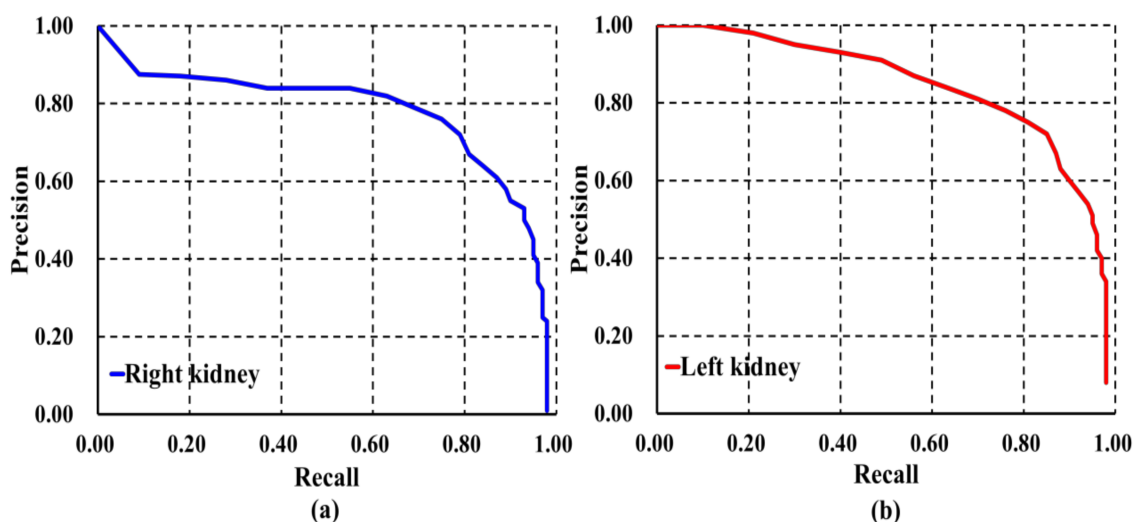
These results were supported by other values of the evaluation metrics as shown in Table 5. It is evidently observed that our model can detect and classify ADPKD kidneys with an AP of 0.80 for right kidney and 0.852 for left kidney, shown in Figure 9a,b, respectively. As compared to other detection architectures, our model achieved the highest mAP, 0.824.

Table 4. Evaluation metrics results of subject-wise testing.

Detection Architectures	Class	Evaluation Metrics			
		Accuracy	Precision	Recall/Sensitivity	F1-Score
Our Model	right	0.8	0.8	0.8	0.8
	left	0.817	0.8	0.84	0.81
SSD Inception V2	right	0.434	0.552	0.512	0.531
	left	0.511	0.582	0.632	0.606
SSD MobileNet V1	right	0.563	0.574	0.652	0.611
	left	0.549	0.654	0.558	0.602
Faster R-CNN NAS	right	0.483	0.394	0.728	0.511
	left	0.248	0.549	0.232	0.326
Faster R-CNN Inception ResNet V2	right	0.268	0.291	0.266	0.278
	left	0.513	0.464	0.694	0.556
R-FCN ResNet 101	right	0.3	0.464	0.245	0.332
	left	0.529	0.572	0.676	0.620

Table 5. AP and mAP comparison with other architectures on subject-wise testing.

Detection Architectures	Evaluation Metrics		
	AP		mAP
	Right	Left	Both Classes
Our Model	0.80	0.852	0.824
SSD Inception V2	0.42	0.497	0.458
SSD MobileNet V1	0.567	0.536	0.551
Faster R-CNN NAS	0.48	0.549	0.515
Faster R-CNN Inception ResNet V2	0.483	0.612	0.548
R-FCN ResNet 101	0.474	0.681	0.577

**Figure 9.** Precision-recall curve of our model on subject-wise testing: (a) Right kidney; (b) Left kidney.

In addition to the loss calculation, we analyzed and plotted the Classification loss and Localization loss as shown in Figure 10. It was observed that our model has lower average value of Classification loss $\overline{L_{CO}} = 3.89$ as compared to SSD Inception V2 $\overline{L_{CO}} = 4.859$ and SSD MobileNet V1 $\overline{L_{CO}} = 4.424$, though Localization loss of our model was $\overline{L_{lo}} = 0.463$.

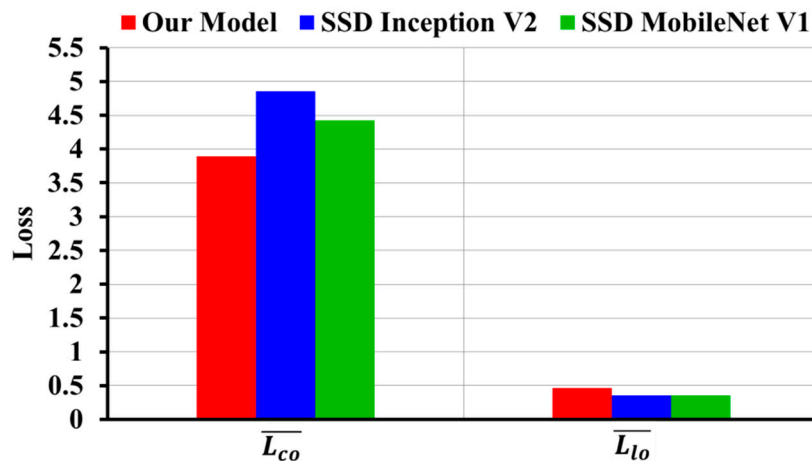


Figure 10. Classification and localization loss comparison on subject-wise testing set.

Furthermore, the localized outputs obtained from the subject-wise testing are depicted in Figure 11a,b. It was observed that our model could detect and classify either a small or big size of the ADPKD kidneys.

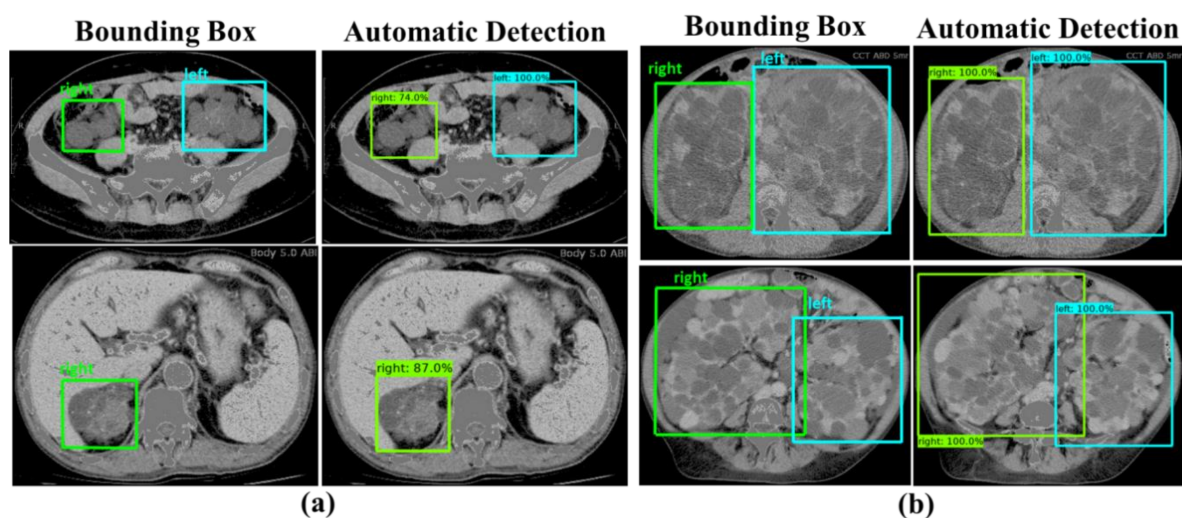


Figure 11. Detection results: (a) Small size of ADPKD kidneys; (b) Big size of ADPKD kidneys.

4. Discussion

In this paper, we demonstrated that the derived automatic ADPKD detection model of CCT images is robust. We performed image-wise testing evaluation on our fine-tuned trained detection model using a k-fold cross-validation technique and compared it with other pre-trained models. It is found that our fine-tuned model outperformed other pre-trained object detection models, though pre-trained models have been designed using large-scale images [31]. Our model can localize and classify ADPKD kidneys with mAP = 0.94. To verify the robustness of our derived detection model, we conducted subject-wise testing evaluation, where our model can obtain mAP = 0.824. It was observed that our model works precisely, using CCT images instead of using MRI images for detecting ADPKD kidneys [15]. Although only 370 subjects for training and 78 for testing have been considered in this work [11], our model has several advantages in solving ADPKD kidney detection without segmentation. These advantages can be described from clinical and technical points of view. From the clinical point of view, our detection model can automatically locate and classify kidneys associated with multiple liver cysts though it has been found that automatic ADPKD kidney segmentation using CCT has over-estimated in the presence of liver cysts [10]. In addition, the automatic detection model can

localize the kidney's area, which overlaps with adjacent organs such as liver and spleen. Therefore, it was also proved from the results that our detection model can work robustly in the case of non-uniform morphology and density of ADPKD kidneys.

By minimizing time-consuming steps and reducing the labor of radiologists, our model achieved a fewer number of False Positives with AP = 0.939 on image-wise testing and AP = 0.826 on subject-wise testing for both classes as compared to [15], where the AP value was 0.78 for both classes. Thus, our model can perform well in assisting radiologists in locating ADPKD kidneys, calculating the TKV and reducing the time-consuming steps.

From the technical point of view, our derived detection model can successfully be implemented on CCT images. Moreover, our model can replace any manual cropping procedure or software that is used to crop the ROI of kidneys and become an intermediate technique before segmentation. Based on our experiment, we also discovered that the accuracy of our detection model could be optimized by training numerous CCT images, of varied morphology and density, of the ADPKD kidneys.

However, our study has limitations. In the testing phase, a high average of losses, misclassification and misdetection can be found. Although it can classify correctly as a "right" kidney, the bounding box is not precisely located in the right kidney. Based on these characteristics, some misclassification and misdetection is inevitable as shown in Figure 12a,b. In addition to data acquisition, repeated scans have been performed to collect the CCT slices from the same subject. Therefore, we plan to conduct more experiments with greater numbers of NCCT and CCT images from ADPKD patients through independent scans. Although our model can locate and classify the ROI of ADPKD kidneys, exact kidney area cannot be extracted perfectly. It is to be noted that, though our model works precisely in detecting the bounding box that can cover all ADPKD kidney areas adjacent to the liver cysts, a few portions of the liver cysts could still be included in the predicted bounding box. Therefore, we plan to extend this work to exactly localize and distinguish the ADPKD kidneys from the liver cysts by segmenting the ADPKD kidney areas. We plan to use our automatic ADPKD detection model as an intermediate technique before segmentation to improve the accuracy of ADPKD segmentation and achieve precise calculation of TKV. For further confirmation of this proposed model, we plan to conduct the experiment with prospective data in future.

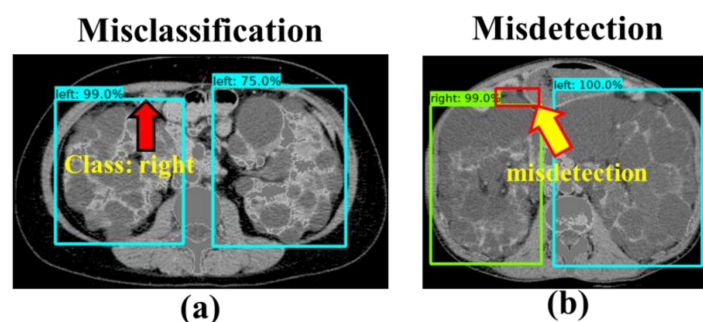


Figure 12. Misclassification and misdetection example: (a) Misclassification; (b) Misdetection.

5. Conclusions

In this study, an automatic detection model for ADPKD kidney on CCT images is designed. The designed model is built using image preprocessing and Deep Learning techniques. We used various image preprocessing techniques such as global thresholding, morphological dilation operation, and contouring. We also adopted and tuned the DL SSD Inception V2 architecture. Based on the performance evaluation, our model outperformed other well-established detection architectures with mAP of 94% for image-wise testing, and mAP of 82% for subject-wise testing. Furthermore, our contribution is to design a robust automatic ADPKD kidneys detection model on CCT images, which can accelerate manual localization and classification tasks with higher accuracy and can establish an intermediate technique for segmentation task. Hence, we believe that the automatic detection

model for ADPKD could be a promising intermediate technique in assisting the radiologists to locate, segment, and analyze ADPKD kidneys for TKV measurement on CCT image big data.

Author Contributions: Conceptualization, D.D.O., T.-W.S., and P.K.S.; methodology, D.D.O., T.-W.S., P.K.S., and L.-J.W.; software, D.D.O., P.G., and P.K.S.; validation, D.D.O., T.-W.S., and P.K.S.; formal analysis, D.D.O., T.-W.S., P.K.S., and L.-J.W.; investigation, L.-J.W., and P.K.S.; resources, T.-W.S., P.K.S., and L.-J.W.; data curation, L.-J.W., and T.-W.S.; writing—original draft preparation, D.D.O., T.-W.S., and P.G.; writing—review and editing, D.D.O., T.-W.S., P.K.S., and L.-J.W.; visualization, D.D.O., and P.G.; supervision, P.K.S., L.-J.W., and T.-W.S.; project administration, P.K.S., L.-J.W., and T.-W.S.; funding acquisition, T.-W.S., and L.-J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CHANG GUNG MEDICAL FOUNDATION, TAIWAN grant number CMRPG3H0021, CMRPG3H0022 and CMRPD2J0142 and MINISTRY OF SCIENCE AND TECHNOLOGY (MOST), TAIWAN grant number 109-2221-E-182-014.

Acknowledgments: We are thankful to Healthy Aging Research Center, Chang Gung University for supporting the GPU platform for our analysis from the Featured Areas Research Center Program within the Framework of the Higher Education Sprout Project (EMRPD1I0491) by the Ministry of Education (MOE), Taiwan.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lanktree, M.B.; Haghighi, A.; Guiard, E.; Iliuta, I.A.; Song, X.; Harris, P.C.; Paterson, A.D.; Pei, Y. Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing. *J. Am. Soc. Nephrol.* **2018**, *29*, 2593–2600. [CrossRef] [PubMed]
- Wiley, C.J.; Blais, J.D.; Hall, A.K.; Krasa, H.B.; Makin, A.J.; Czerwiec, F.S. Prevalence of autosomal dominant polycystic kidney disease in the European Union. *Nephrol. Dial. Transplant.* **2017**, *32*, 1356–1363. [CrossRef] [PubMed]
- Cornec-Le Gall, E.; Alam, A.; Perrone, R.D. Autosomal dominant polycystic kidney disease. *Lancet* **2019**, *393*, 919–935. [CrossRef]
- The US Food and Drug Administration. Total Kidney Volume Qualified as a Biomarker. Available online: <https://www.raps.org/news-articles/news-articles/2016/9/total-kidney-volume-qualified-as-a-biomarker-by-fda-for-adpkd-trials?feed=Regulatory-Focus> (accessed on 8 August 2020).
- Irazabal, M.V.; Rangel, L.J.; Bergstralh, E.J.; Osborn, S.L.; Harmon, A.J.; Sundsbak, J.L.; Bae, K.T.; Chapman, A.B.; Grantham, J.J.; Mrug, M.; et al. Imaging classification of autosomal dominant polycystic kidney disease: A simple model for selecting patients for clinical trials. *J. Am. Soc. Nephrol.* **2015**, *26*, 160–172. [CrossRef] [PubMed]
- Gupta, P.; Chiang, S.F.; Sahoo, P.K.; Mohapatra, S.K.; You, J.F.; Onthoni, D.D.; Hung, H.Y.; Chiang, J.M.; Huang, Y.L.; Tsai, W.S. Prediction of Colon Cancer Stages and Survival Period with Machine Learning Approach. *Cancers* **2019**, *11*, 2007. [CrossRef] [PubMed]
- Riaz, H.; Park, J.; Choi, H.; Kim, H.; Kim, J. Deep and Densely Connected Networks for Classification of Diabetic Retinopathy. *Diagnostics* **2020**, *10*, 24. [CrossRef] [PubMed]
- Pehrson, L.M.; Nielsen, M.B.; Lauridsen, C.A. Automatic Pulmonary Nodule Detection Applying Deep Learning or Machine Learning Algorithms to the LIDC-IDRI Database: A Systematic Review. *Diagnostics* **2019**, *9*, 29. [CrossRef] [PubMed]
- Ünver, H.M.; Ayan, E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics* **2019**, *9*, 72. [CrossRef] [PubMed]
- Sharma, K.; Rupprecht, C.; Caroli, A.; Aparicio, M.C.; Remuzzi, A.; Baust, M.; Navab, N. Automatic Segmentation of Kidneys using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease. *Sci. Rep.* **2017**, *7*, 2049. [CrossRef] [PubMed]
- Zheng, Y.; Liu, D.; Georgescu, B.; Xu, D.; Comaniciu, D. Deep learning based automatic segmentation of pathological kidney in CT: Local versus global image context. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 241–255.
- Kline, T.L.; Korfiatis, P.; Edwards, M.E.; Blais, J.D.; Czerwiec, F.S.; Harris, P.C.; King, B.F.; Torres, V.E.; Erickson, B.J. Performance of an Artificial Multi-observer Deep Neural Network for Fully Automated Segmentation of Polycystic Kidneys. *J. Digit. Imaging* **2017**, *30*, 442–448. [CrossRef] [PubMed]
- Keshwani, D.; Kitamura, Y.; Li, Y. Computation of Total Kidney Volume from CT Images in Autosomal Dominant Polycystic Kidney Disease Using Multi-task 3D Convolutional Neural Networks. In Proceedings

- of the International Workshop on Machine Learning in Medical Imaging, Granada, Spain, 16 September 2018; Springer: Cham, Switzerland, 2018; pp. 380–388.
14. Brunetti, A.; Cascarano, G.D.; De Feudis, I.; Moschetta, M.; Gesualdo, L.; Bevilacqua, V. Detection and Segmentation of Kidneys from Magnetic Resonance Images in Patients with Autosomal Dominant Polycystic Kidney Disease. In Proceedings of the International Conference on Intelligent Computing, Nanchang, China, 3–6 August 2019; Springer: Cham, Switzerland, 2019; pp. 639–650.
 15. Bevilacqua, V.; Brunetti, A.; Cascarano, G.D.; Palmieri, F.; Guerriero, A.; Moschetta, M. A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on magnetic resonance images. In Proceedings of the International Conference on Intelligent Computing, Wuhan, China, 15–18 August 2018; Springer: Cham, Switzerland, 2018; pp. 643–649.
 16. Bevilacqua, V.; Brunetti, A.; Cascarano, G.D.; Guerriero, A.; Pesce, F.; Moschetta, M.; Gesualdo, L. A comparison between two semantic deep learning frameworks for the autosomal dominant polycystic kidney disease segmentation based on magnetic resonance images. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 1–12. [[CrossRef](#)] [[PubMed](#)]
 17. Yan, K.; Wang, X.; Lu, L.; Summers, R.M. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging (Bellingham)* **2018**, *5*, 036501. [[CrossRef](#)] [[PubMed](#)]
 18. Li, C.; Wang, X.G.; Liu, W.Y.; Latecki, L.J.; Wang, B.; Huang, J.Z. Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Med. Image Anal.* **2019**, *53*, 165–178. [[CrossRef](#)] [[PubMed](#)]
 19. Sahoo, P.K.; Soltani, S.; Wong, A.K. A survey of thresholding techniques. *Comput. Vis. Graph. Image Process.* **1988**, *41*, 233–260. [[CrossRef](#)]
 20. Maragos, P. Tutorial on advances in morphological image processing and analysis. *Opt. Eng.* **1987**, *26*, 267623. [[CrossRef](#)]
 21. Suzuki, S. Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* **1985**, *30*, 32–46. [[CrossRef](#)]
 22. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009.
 23. Tzutalin, L. Git Code. Available online: <https://github.com/tzutalin/labelImg> (accessed on 8 January 2020).
 24. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of European Conference on Computer Vision; Springer: Cham, Switzerland, 2016; pp. 21–37.
 25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
 26. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755.
 27. He, Q.P.; Wang, J. Application of Systems Engineering Principles and Techniques in Biological Big Data Analytics: A Review. *Processes* **2020**, *8*, 951. [[CrossRef](#)]
 28. EL-Bana, S.; Al-Kabbany, A.; Sharkas, M. A Two-Stage Framework for Automated Malignant Pulmonary Nodule Detection in CT Scans. *Diagnostics* **2020**, *10*, 131. [[CrossRef](#)] [[PubMed](#)]
 29. Hashmi, M.F.; Katiyar, S.; Keskar, A.G.; Bokde, N.D.; Geem, Z.W. Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. *Diagnostics* **2020**, *10*, 417. [[CrossRef](#)] [[PubMed](#)]
 30. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S. Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7310–7311.
 31. Li, H.; Chaudhari, P.; Yang, H.; Lam, M.; Ravichandran, A.; Bhotika, R.; Soatto, S. Rethinking the Hyperparameters for Fine-tuning. *arXiv* **2020**, arXiv:2002.11770.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).