# Kidney and tumor segmentation using combined Deep learning method

Yuliia Kamkova[1,2][0000−0002−9717−2866], Hemin Ali Qadir[1,2][0000−0002−9271−5040], Ole Jakob Elle[1,2][0000−0003−2359−1272], and Rahul Prasanna Kumar,[1][0000−0002−5951−7317]

[1] Oslo University Hospital, Oslo, Norway
[2] Department of Informatics, University of Oslo, Oslo, Norway
juliua.kamkova@gmail.com
https://www.ous-research.no/interventionalcentre/

**Abstract.** This paper presents our method for automatic segmentation for kidney and tumor as part of the Kidney Tumor Segmentation Challenge (KiTS19). The KiTS19 Challenge had released a dataset of 300 unique kidney cancer patients, with manual annotations done by Climb 4 Kidney Cancer (C4KC). Here we have proposed our new combined cascade deep learning (DL) approach for solving the tasks of the challenge. We used deep learning based detection for localising kidney with the tumor, followed by deep learning based segmentation to create the labels for kidney and tumor locally. Our approach resulted in high recall (96.13) and high Jacquard score (95.4) on the randomly selected 30 volumes that were picked as the validation set.

**Keywords:** Kidney segmentation · Tumor segmentation · CT · Deep learning

## 1 Introduction

Kidney tumor is a major concern worldwide with more than 400,000 cases are registered yearly in the field. Among which, Norway is $15^{th}$ worldwide and $13^{th}$ among men [1]. Almost 60% of all people diagnosed with kidney cancer are assigned for the surgical intervention for removing injured tissues or organ. [2]. For planning the surgery, Computer Tomography (CT) imaging is the most common way of screening the patients, due to its cheap cost compared to Magnetic Resonance Imaging (MRI). Furthermore, CT includes almost all information needed for radiologists and surgeons to localise the tumor morphology and specificity. However, this process is extremely time consuming due to the big numbers of slices per patient and low contrast between soft tissues on the CT images and . Thus, availability of the methods that could segment kidneys and tumor automatically could decrease the time per case, and provide volumetric information about tumor morphology in the organ.

Kidney detection is not extensively researched area in medical image processing. There were several approaches discussed in the literature that show good results [3], [4], [5], [6], [7]. The major challenges in kidney segmentation are the size of the initial volume, and organ variability from patient to patient. The variability is even more when tumor is present, where the size, shape and texture changes from case to case. Thus the introduction of KiTS19 challenge is a great contribution to this field.

Currently, all methods in medical image processing for detection and segmentation could be divided into 3 main groups.

1. Manual segmentation performed by an expert, usually by a radiologist or other specialists in the field.
2. Methods performed by semi-automatic and automatic methods based on the statistical information from manually extracted features.
3. Methods based on the DL, requiring a lot of training data.

First method, is the most accurate one, however it is extremely time consuming. Usually this method is used for creating ground-truth for detection and segmentation.

Second method include such methods as atlas, active shape models, active contours, level set models and hybrid level set models. Those methods showed results such as 86% and 88% of average Dice score and correlation accordingly [5,6]. However there are no results from applying these methods to CT images where tumor is present.

The last group is the latest in this field, with Deep Learning is already showing promising results for kidney detection and tumor segmentation [3,4,8]. The advantage of using DL is that it can self-generate the features used to segment kidneys and tumors from the rest of the scan. All DL methods in the literature are based on 2D segmentation, which takes the input slices from the volume and feed them to the network. However, when 2D approach is applied, some volumetric information could be lost. Even radiologists, when annotating the organ should look through several slices before and after for being sure in the decision about the label. Withal, take as an input 3D volume is a challenge, due to the size of those kind of volumes. Also, if resizing is applied, then important information could be lost.

Thus, we proposed a new combined method, that include two neural networks inside. First network is 2D Faster R-CNN network [9] that is used for finding kidney with tumor in the full body CT scan, and then on the extracted volumes we applied 3D V-Net for segmenting kidneys and tumors itself. Faster R-CNN has shown great results in polyp detection in video [10], and V-Net [11] continues to show leading results in segmentation of different organs in medical field including abdominal organs. With the application of the detection method, we could crop from the full CT-scan to only the region of interest, which is normally almost 10 times less in terms of size. Thus, making it possible to fit region of interest to the segmentation network without any down sampling process.

## 2   DataSet

The dataset in this paper was taken from KITS19 challenge, and include 300 CT volumes. For 210 of them ground-truth is available and 90 left as a testing set with no annotations presented. All data presented in anonymized NIFTI format with fixed width and height(512x512), and with variability in slice thickness from 1mm to 5mm.

For first DL network, dataset was pre-proceed and resampled along the axial axes, in this way we increased amount of samples and went from 3D approach to 2D approach. New data contains 45375 images, from which only slices containing kidneys and tumor were used in the training process.

For the second network, only volumes with detected kidneys were used. Volumes has the same pre-processing as for the first method. In this way, we increased the training set almost two times without even applying augmentation. The random crop and rescaling was also applied as part of augmentation for training the second model.

## 3   Methodology

Our proposed method is based on the cascade approach implemented by combing results of two deep learning networks, Faster R-CNN and V-net. Schematically the workflow process is presented on the Fig. 1. Details about each of the stages are provided below.
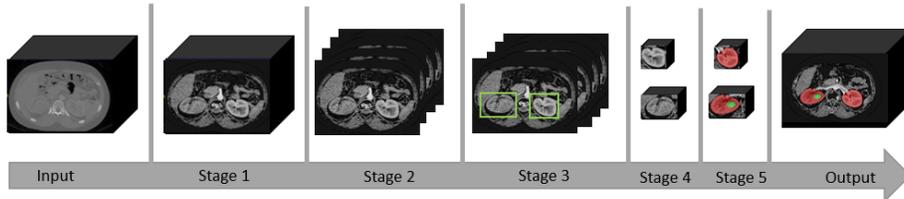


**Fig. 1.** Schematic representation of the proposed method for kidney and cancer segmentation using combined DL approach. (Stage 1 - Image pre-prossessing, Stage 2 - Preparation of the data for the DL detection, Stage 3 - Kidney and tumor detection using Faster R-CNN, Stage 4 - Post-processing of detection result, extraction of the volumes contain kidney and tumor, Stage 5 - V-net segmentation of kidney and tumor, Stage 6 - applying kidney segmentation to the initial volume)

### 3.1    Stage 1&2. Image Preparation

From the challenge dataset, we had a set of 210 volumes in NIFTI format that were available with ground-truth. These volumes were used for training of both our DL models. Initially volumes were thresholded using Hounsfield unit (HU) values in the range of interest: from $-60$ to $300$. Afterwards, the volume were re-sampled along the axial axes, resulting in the input for the first network. For the ground-truth, we modified presented labels into the cubic segmentations, which cover the kidney with the tumor inside as presented on the Fig. 2. In this way we are looking for one box for each kidney with all tumors in it.
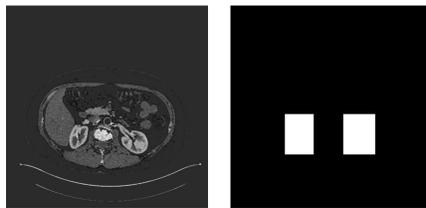


**Fig. 2.** Example of the input for kidney and lesion detection in Faster R-CNN, training stage

### 3.2    Stage 3. Kidney Detection using DL

For our kidneys detection, we adapt a Faster R-CNN [9] architecture shown in Fig. 3. We use ResNet50 [12] as the feature extractor network which was pre-trained on Microsoft's COCO dataset [13]. Faster R-CNN uses two stages: region proposal network (RPN), and a box classifier network. Both stages share a common set of convolutional layers to reduce the marginal cost for detection. The RPN utilizes feature maps in the last layer of the *conv*4 block to generate class-agnostic RoI proposals, each with an objectness confidence value. The proposed RoIs are a grid of anchors titled in different aspect ratios and scales. The classifier network crops these anchors from the feature maps of the *conv*4 block and feeds the cropped features to the remainder of the network in order to predict locations and confidence value of the detection.

To train our Faster R-CNN model, we use a multi-task loss on each anchor. For each anchor $a$, we find the best matching ground-truth box $b$. If there is a match, anchor $a$ acts as a positive anchor, and we assign a class label $y_a = 1$, and a vector $(\phi(b_a; a))$ encoding box $b$ with respect to anchor $a$. If there is no match, anchor $a$ acts as a negative sample, and the class label is set to $y_a = 0$. The loss for each anchor $a$, then consists of two losses: location-based loss $\ell_{loc}$ for the predicted box $f_{loc}(I; a, \theta)$ and classification loss $\ell_{cls}$ for the predicted class
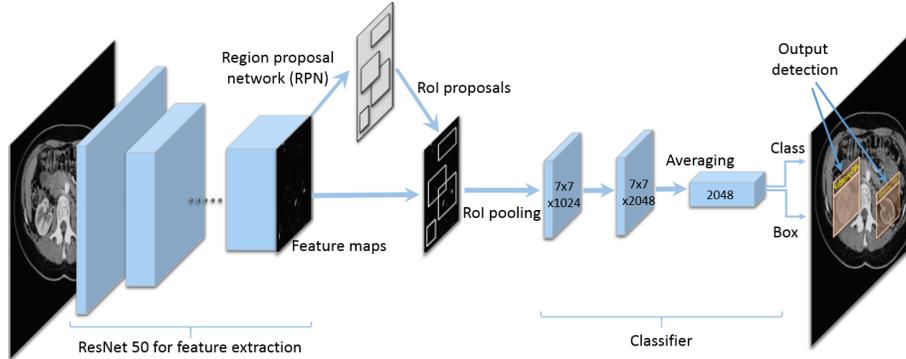
**Fig. 3.** Our Faster R-CNN framework used for kidney detection as the first stage of our proposed method. In the first stage, we use Resnet50 to extract features from the input image. Region proposal network (RPN) utilizes feature maps at one of the intermediate layers (usually the last convolutional layer) of the CNN feature extractor networks to generate box proposals (300 boxes in our study). The proposed boxes are a grid of anchors tiled in different aspect ratios and scales. The second stage predicts the confidence value, the offsets for the proposed box for each anchor.

$f_{cls}(I; a, \theta)$, where $I$ is the image and $\theta$ is the model parameter,

$$
\begin{aligned}
\mathcal{L}(a, I; \theta) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \sum_{j=1}^{N} &\, 1[a \, is \, positive] \, . \, \ell_{loc}\Big(\phi(b_a; a) \\
&- f_{loc}(I; a, \theta)\Big) + \ell_{cls}\Big(y_a, f_{cls}(I; a, \theta)\Big)
\end{aligned}
\tag{1}
$$

where $m$ is the size of mini-batch and $N$ is the number of anchors for each frame. We use the following loss functions: Smooth L1 for the localization loss and softmax for the classification loss. We set the ratio between negatives and positives to 1:1 to avoid imbalance training and the input image size to 512x512.

For checking the performance of the network, we took randomly 30 samples as a validation set and calculate Recall and Preccision as follows,

$$
Recall = \frac{TP}{TP + FN},
\tag{2}
$$

$$
Precision = \frac{TP}{TP + FP},
\tag{3}
$$

Where $TP, FP, TN$ and $FN$ are a number of true positives, false positives, true negatives and false negatives respectively. Precision will show us the percentage of the results which are relevant, and recall refers to the percentage of total relevant results correctly classified by our algorithm.

### 3.3    Stage 4. Kidney extraction and volume cropping

After our detection step, the post-proccesing was applied. Initially from labelled slices, the labelled volume were recreated. Afterwards two largest component were specified, then the size of each of them were checked. We assume that kidney, even with the tumor on it could not be variate more that three times in terms of volume. In this way, we take care of cases where only one kidney is presented, or where kidneys are merged together.

Using segmentation obtained in the previous step, the largest width, height and length of all cropping box for each individual kidney were defined. New volume limited by those parameters then was extracted from the original CT images.

### 3.4    Stage 5. Kidney and tumor segmentation using DL

After the volumes were extracted, these were fed into the V-net segmentation network [11,14,15], where the architecture of V-net from the original paper [11] was used in the proposed method that is presented on the Fig.5. For evaluation of the segmentation performance, we used Jaccard and Accuracy value, and for the loss - Dice loss was used.

$$Jaccard = \frac{|AB|}{|A \cup B|}, \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

$$Diceloss = 1 - \frac{2|AB|}{A + B}, \tag{6}$$

Where $TP, FP, TN$ and $FN$ are a number of true positives, false positives, true negatives and false negatives respectively. And A and B are the volumes from ground-truth and predicted from the model one.

From the previous step, we obtain volumes with different sizes, which were randomly cropped and resized into volumes of (64,64,32) as input to this network.

### 3.5    Postprocessing

After segmentations were obtained, the biggest connected component was extracted from it. Following which, the volume was padded back to the initial size, and the second kidney segmentation (if any) was added. Here, the spacing was taken from the original volumes.

## 4    Results and Discussion

### 4.1    Detection results

It is important to know the performance of the detection model before we implement the segmentation model. Because we did not have access to the ground-truth of the test dataset, we used validation dataset to pre-evaluate our detection model. Table 1 shows that the Faster R-CNN was able to detect 96.13% of
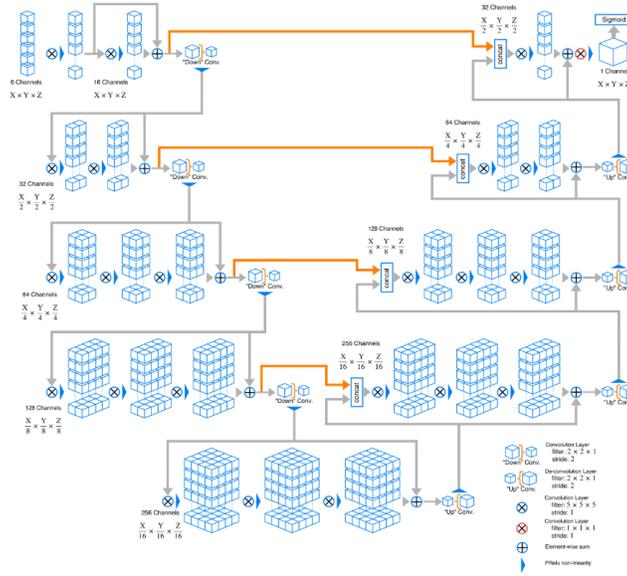
**Fig. 4.** Original V-Net architecture for medical image segmentation [11]

the kidneys. It also generated a lot of FP detection, resulting in low precision (50.21%). In the detection stage (stage 3), we are more interested to detect kidneys in the input images. Later, some of these FPs will be eliminated by the postprocessing step.

**Table 1.** Kidneys detection from our Faster R-CNN on the validation set.

| Models | Recall | Precision |
|---|---|---|
| Faster R-CNN with ResNet50 | 96.13 | 50.21 |

Below we can see that the network shows promising results even in cases where big tumors are present, or where just one kidney is on the screen.

### 4.2  Results of V-Net on the validation set

The curves shown below (Fig. 6, Fig. 7 and Fig. 8) is the dice loss and Jaccard metric on the training and validation sets. We can clearly see that our approach enables the network to achieve high results quite fast. Also, on the validation set, the segmentation is performed with high Jaccard value and Accuracy.

To obtain weights used for producing segmentation from the testing set, the network was trained on the whole set of images to achieve maximum possible result.
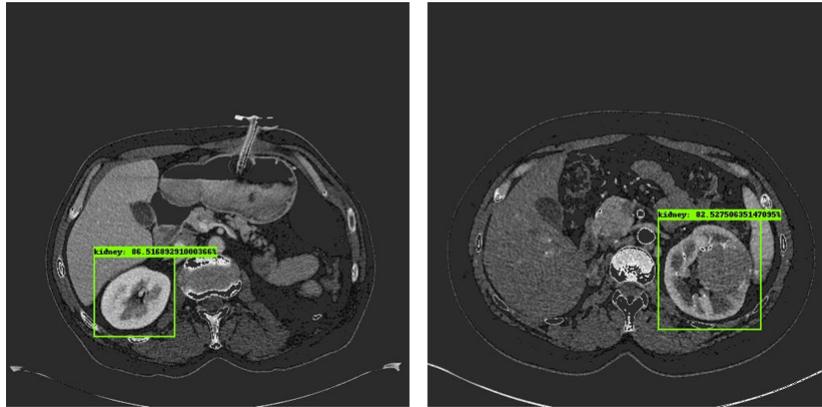
**Fig. 5.** Example of kidney detection on the testing set. Number on the top show the probability of correct detected kidney according to the network
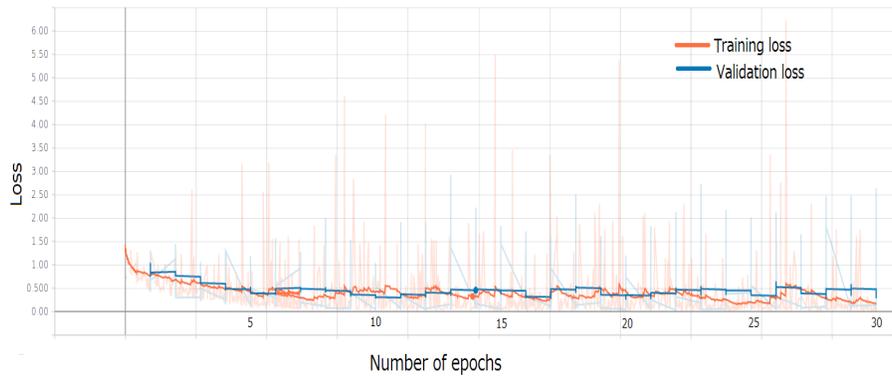


**Fig. 6.** Loss progress on training and validation datasets during training V-Net.
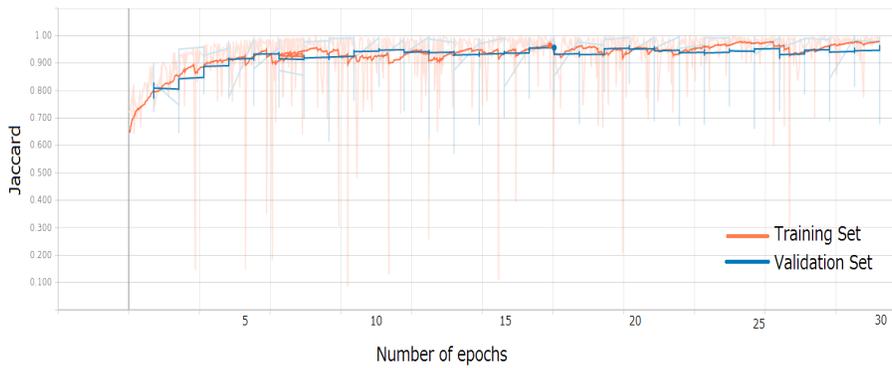


**Fig. 7.** Jaccard index on training and validation datasets during training V-Net.
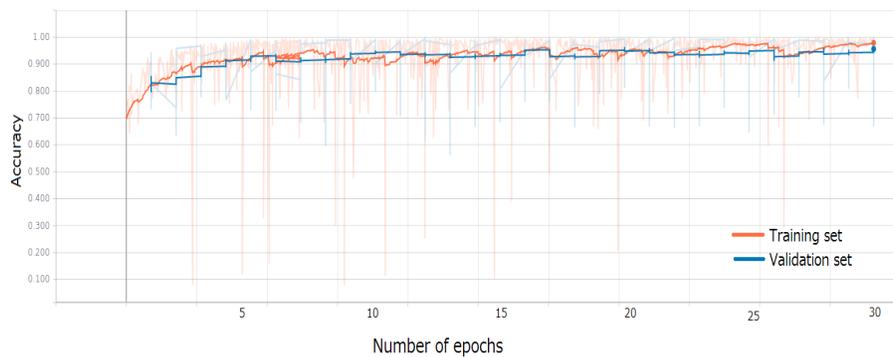
**Fig. 8.** Accuracy on training and validation datasets during training V-Net.



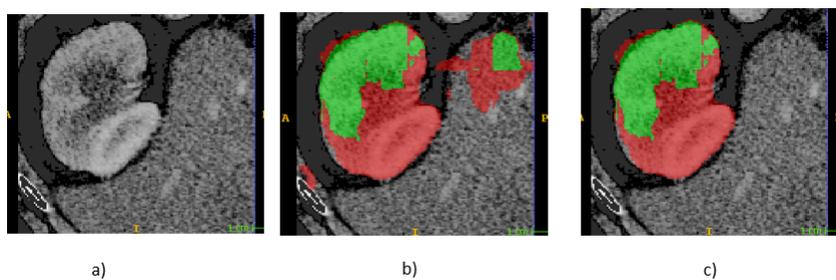a)                                    b)                                    c)

**Fig. 9.** Example of obtained segmentation on one of the sample from the testing set. In red colour - kidney , in green - tumor (a. cropped volume using DL detection; b. segmentation obtained with DL segmentation; c. final segmentation).
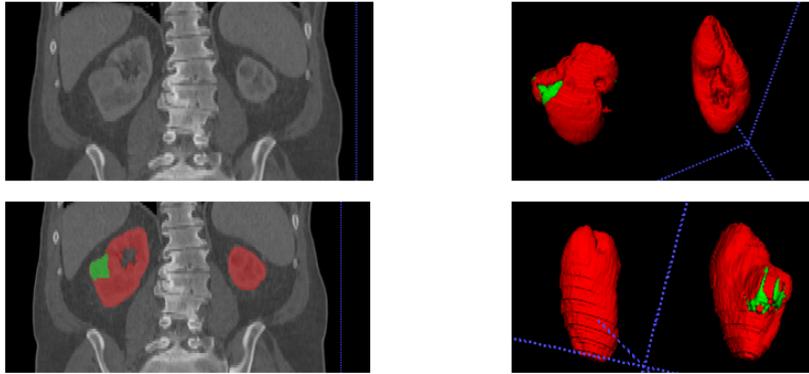
**Fig. 10.** Example of obtained segmentation on one of the sample from the testing set. Volumetrical representation of three different kidney samples. In red colour - kidney , in green - tumor

As seen from Fig. 9 and Fig. 10, results obtained with deep learning resembles close to reality. Observations through different samples make us believe that this method could give high accuracy metrics using ground-truth data on the testing dataset.

## 5   Conclusions

In this work, we proposed a novel combined approach for kidney and tumor segmentation. Our approach combine 2D and 3D methods, and include two deep learning methods for kidney detection and segmentation. With the use of Faster R-CNN with ResNet50, we were able to obtain fast and high recall results for kidney detection. Also, this network created 3D cropping boxes to extract only region of interests. Followed by V-Net with these 3D volumes, resulted in our accurate segmentation of kidney and tumor.

## References

1. Kidney      cancer      statistics.      world      cancer      research      fund. www.wcrf.org/dietandcancer/cancer-trends/kidney-cancer-statistics, 2018.
2. Cancer   diagnosis   and   treatment   statistics.   stages   —   mesothelioma   — cancer   research   uk.      www.cancerresearchuk.org/health-professional/cancer-statistics/diagnosis-and-treatment, year = 2017.
3. Qian Yu, Yinghuan Shi, Jinquan Sun, Yang Gao, Jianbing Zhu, and Yakang Dai. Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in ct images. *IEEE Transactions on Image Processing*, 2019.

4. Timothy L Kline, Panagiotis Korfiatis, Marie E Edwards, Jaime D Blais, Frank S Czerwiec, Peter C Harris, Bernard F King, Vicente E Torres, and Bradley J Erickson. Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. *Journal of digital imaging*, 30(4):442–448, 2017.
5. Daw-Tung Lin, Chung-Chih Lei, and Siu-Wan Hung. Computer-aided kidney segmentation on abdominal ct images. *IEEE transactions on information technology in biomedicine*, 10(1):59–65, 2006.
6. Andrzej Skalski, Katarzyna Heryan, Jacek Jakubowski, and Tomasz Drewniak. Kidney segmentation in ct data using hybrid level-set method with ellipsoidal shape constraints. *Metrology and Measurement Systems*, 24(1):101–112, 2017.
7. Paola Campadelli, Elena Casiraghi, and Stella Pratissoli. A segmentation framework for abdominal organs from ct scans. *Artificial Intelligence in Medicine*, 50(1):3–11, 2010.
8. Tomas Sakinis, Fausto Milletari, Holger Roth, Panagiotis Korfiatis, Petro Kostandy, Kenneth Philbrick, Zeynettin Akkus, Ziyue Xu, Daguang Xu, and Bradley J Erickson. Interactive segmentation of medical images through fully convolutional neural networks. *arXiv preprint arXiv:1903.08205*, 2019.
9. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015.
10. Younghak Shin, Hemin Ali Qadir, Lars Aabakken, Jacob Bergsland, and Ilangko Balasingham. Automatic colon polyp detection using region based deep cnn and post learning approaches. *IEEE Access*, 6:40950–40962, 2018.
11. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.
12. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
13. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
14. Miguel Monteiro. Vnet-tensorflow: Tensorflow implementation of the v-net architecture for medical imaging segmentation. https://github.com/MiguelMonteiro/VNet-Tensorflow, 2018.
15. Jacky KL Ko. Implementation of vnet in tensorflow for medical image segmentation. https://github.com/jackyko1991/vnet-tensorflow, 2018.