exeley INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Article | DOI: 10.21307/ijssis-2021-017 Issue 1 | Vol. 14 (2021)

Using explainable deep learning in da Vinci Xi robot for tumor detection

Rohan Ibn Azad, Subhas Mukhopadhyay and Mohsen Asadnia*

School of Engineering, Macquarie University, Sydney, Australia.

*E-mail: mohsen.asadnia@mq.edu.au

This paper was edited by Subhas Chandra Mukhopadhyay.

Received for publication July 29, 2021.

Abstract

Deep learning has proved successful in computer-aided detection in interpreting ultrasound images, COVID infections, identifying tumors from computed tomography (CT) scans for humans and animals. This paper proposes applications of deep learning in detecting cancerous cells inside patients via laparoscopic camera on da Vinci Xi surgical robots. The paper presents method for detecting tumor via object detection and classification/localizing using GRAD-CAM. Localization means heat map is drawn on the image highlighting the classified class. Analyzing images collected from publicly available partial robotic nephrectomy videos, for object detection, the final mAP was 0.974 and for classification the accuracy was 0.84.

Keywords

Convolutional neural network, Tumor detection, YOLOv4, GRAD-CAM, Live surgery, da Vinci Xi.

Introduction

Kidney cancer is not as prevalent as prostate cancer in men, but it is still the ninth most common cancer in men and the 14th most common cancer in women (N. Cancer, n.d.). In terms of occurrence kidney cancer was the 7th most common cancer in Australia in 2016 and it still remained the 7th most common cancer in 2020. The number of people affected by kidney cancer seems to be increasing from 793 in 1982 to 3,627 in 2016. In total, 3,627 news cases of which 2,408 were males and 1,219 were females was diagnosed with kidney cancer in 2016 in Australia. The surgical procedure for treating kidney cancer is called Nephrectomy. In partial nephrectomy, only the portion of the kidney that is diseased is removed. The procedure is done either open or robotically (minimally invasive surgery). In the open nephrectomy method, depending on cases, it might be required to remove a rib bone. The procedure is done using general anesthesia and so, the open nephrectomy method is less desirable for patients. The more preferred method both by surgeons and patients is partial robotic nephrectomy (National Kidney Foundation, n.d.). Artificial intelligence (AI) has significant potential in many engineering applications including manufacturing (Razfar et al., 2010), hydrology (Asadnia et al., 2010, 2014, 2017; Khorasani et al., 2018), sensors (Asadnia et al., 2013; Hagihghi et al., 2020; Kottapalli et al., 2015; Razmjou et al., 2017), and additive manufacturing (Bazaz et al., 2018; Mahmud et al., 2020; Moshizi et al., 2020). Helping surgeons identifying tumors not only in partial robotic nephrectomy, but also in other cancer cases such as bowel, prostate, canine mammary carcinoma.

da Vinci Xi enables robotic surgery using small incisions which can significantly help the surgeons with cancerous tumor removal surgery. da Vinci surgical robot was first commercialized by intuitive in 2000. In 2014, intuitive released da Vinci Xi (David and Samadi, n.d.). The da Vinci Xi robot has three parts, the patient cart, the surgeon console, and the vision cart. The patient cart has four arms to be used during the surgery, the surgeon controls the robotic arms through the console, the vision cart works as the CPU for the system and works as the second screen (American Institute of Minimally Invasive Surgery, 2019). Currently, the surgeons rely on their experience to identify the tumors. Once the tumor's

location has been approximated, da Vinci Xi provides intra operative ultrasound and Indocyanine Green (ICG) with Fluorescence Imaging to further assist the surgeon. Intra operative ultrasound shows the depth of the tumor and makes a 3D reconstruction of the organ on a tablet beside the surgeon's console. Injecting ICG and turning on fluorescent light makes the kidney green and the tumor grey. But if the tumor location cannot be identified then intra operative ultrasound will not work. Not injecting ICG in the correct dose will either make the whole field of view green or will not change color. ICG also comes with side effects which makes it necessary to keep the injection of ICG minimum (Inc and Grove, 2021). There had been remarkable progress made in medical applications of image processing due to the availability of open source large scale annotated datasets. The applications include both pre and post-operative diagnosis. In 2015, support vector machine (SVM) was the most reliable classifier. Papers presented before Chung et al. (2015) only considered one slide from each MRI scan. Chung for the first time considered the spatial information contained in 3D voxels in the MRI scans. After 2015, when deep neural networks gained some insights as to how they work owing to the work of Zeiler and Fergus (2014), convolutional neural networks became popular for image classification. Shin et al. (2016) made use of publicly available CT images for thoraco-abdominal lymph node detection and interstitial lung disease classification. Pantanowitz et al. (2020) fulfilled the need for computer-assisted diagnostics of prostate core needle biopsies (CNBs) by developing an algorithm that takes input as hematoxylin and eosin (H&E) stained slides outputs the result with 0.997 AUC. Deep learning was also used for detecting cancer in animals (Aubreville et al., 2020), agricultural greenhouse detection (Li et al., 2020), analyzing traffic load distribution on a bridge (Ge et al., 2020), airplane detection (Chen et al., 2018), hand gesture recognition (Kharate et al., 2016), automatic vehicle inspection (Nakhaeinia et al., 2016), license plate recognition (Bennet et al., 2017). All these works presented here, only focused on pre and post-operative diagnosis using magnetic resonance imaging, computer tomography scans, ultrasound images. None of the papers consider real-time surgical images to identify tumors. This paper will address this issue and propose using convolutional neural network (YOLOv4 at first, then optimized VGG-16) for giving the surgeons a second opinion during real time tumor removal surgery.

This paper proposed an easy solution to use and reliable deep learning algorithm to help the surgeons to identify the cancerous cells while running the surgery. This will provide a second opinion besides the surgeon's experience in identifying tumors during surgery which is extremely valuable to reduce the errors and to ensure all the potential cancerous tumors are removed.

The proposed process had been carried out in three steps. The first step uses deep learning on a live surgical video to show the locations of the tumors on a global range. The second step is for classifying among cancerous tissue, non-cancerous tissue, fatty tissue and localizing the identified class in close range. If it is preferred to have two class classification, the third step is for classifying between cancerous and non-cancerous tissue with localization with Gradient-based Class Activation mapping (GRAD-CAM) (Selvaraju et al., 2017) in close range. The idea is that once the more aggressive tumor had been identified using object detection in global range, the close range classification will be used to identify if there are any more tumors left inside the patient before closing the wounds.

To the best of our knowledge, this is the first work in real time tumor detection during surgery. We show a comparison of YOLOv3 (Redmon and Farhadi, 2018) and YOLOv4 (Bochkovskiy et al., 2020) object detection for global range detection and then use variations of VGG-16 (Simonyan and Zisserman, 2014), for classification and localization. The base VGG-16 architecture was used changing the output layer to 2/3 and the input image shapes were $224 \times 224 \times 3$. The results were compared with different regularizations, dropout rate and it was found that the chosen VGG-16 generalized the most. Referring to Li et al. (2020) where a comparison between YOLOv3, faster R-CNN, Single Shot Detector (SSD) was done, the mean average precision (mAP) of YOLOv3 was 90.4% and 86%, 84.9% for faster R-CNN and SSD, respectively. Referring to Aly et al. (2021), where a comparison between YOLOv1, v2, v3 was done, it was found that YOLOv3 performed the best with 75.8% mAP compared to 69.52% and 48.1% for YOLOv2 and YOLOv1, respectively. A problem that was carried on to YOLOv3 from v1 and v2 was that small objects were not getting detected. YOLOv4 fixes the issue by incorporating Cross Stage Partial Network (CSPDARKNET53) that extracts the most significant context features without reducing the network operation speed. Li et al. (2020) mention that using YOLOv4 improved their performance from 90.4% mAP for YOLOv3 to 91.8% mAP for YOLOv4. Evaluation metric for object detection was precision, recall, mean Average Precision, frames per second and for classification was Loss VS epochs curve,

confusion matrix, GRAD-CAM (Selvaraju et al., 2017). We do not have a big dataset. So, unsupervised masking, semantic segmentation, was not going to work in our case. Rather, we divided the work in two tasks and the prepared the dataset for two tasks accordingly. For the first task we did global range tumor detection using object detection and then in the second task, we did classification and localization in close range. The training and the testing process were conducted on Intel Core i7 10th gen, equipped with NVIDIA Quadro P4000 GPU and 32 GB of RAM. In this paper, related works are described in the second section, the third section describes the method, results are covered in the fourth section, and the final section covers conclusion and future work.

Related work

Deep learning had been used in many medical applications including fast identification of COVID infection (Brunese et al., 2020; Junnumtuam et al., 2021), seismic (Hammal et al., 2020), medical segmentation (Wu et al., 2021), Referring to Chung et al. (2015), Pantanowitz et al. (2020), Wang et al. (2018), research work had been done to detect prostate cancer for diagnosis using magnetic resonance imaging (MRI) scans and ultra-sound scans. Shin et al. (2016) studied computer aided detection in thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification. The author used publicly available dataset in Depeursinge et al. (2012) for ILD and publicly available dataset in Roth et al. (2016) and Seff et al. (2014). They found that unlike heart or liver which have a specific orientation, lymph nodes do not have a specific orientation. For this reason, they could not apply segmentation on the images to apply convolutional neural network (CNN) on the segmented region. They had to rely on applying CNN on the entire images. The author used variations of CifarNet, AlexNet, GoogleNet (Szegedy et al., 2015) and shows that fine tuning the network for GoogleNet performed best for both LN and ILD since they had a lot of images in the dataset. The GoogleNet was still overfitting at first (Shin et al., 2016). Analyzing the models with variations including random initialization, transfer learning the best performance was achieved with GoogLeNet random initialization with 0.95 AUC.

Later on, Wang et al. (2020) applied mask on the CT images so that it is easier for the convolutional neural network to focus on the affected regions on the lung CT scan. The authors first used an unsupervised method to first add ground-truth masks on the training set. Then the training set images along with their mask was inputted in a 2D-UNet to train an algorithm to add mask on the test set images. All the training images that had the wrong mask, during unsupervised training for adding mask, was manually removed. They used 499 CT volumes for training and 131 CT volumes for testing. 1 was COVID positive and 0 was COVID negative. Then the trained 2D-UNet was used for adding masks on the testing set 3D CT volumes frame-by-frame. Then, the lung volume masks were concatenated with their CT volume images and the data set was prepared. Then, training was done on a deconvolutional neural network (DeCoVNet) with the labels as 0 or 1. Then the activations from the DeCoVNet with CAM (class activation mapping) along with unsupervised lung segmentations with 3d connected component (3DCC) (Ohira, 2018) were used for lesion localization. They evaluated their model's performance at different threshold, after statistical analysis it was found that at threshold 0.3 the DeCoVNet performed with 0.908 maximum accuracy.

Roy et al. (2020) went a step above from the previous one and besides doing classification on images, they applied classification on video, and they used semantic segmentation to segment the infected regions from ultrasound. The reason for choosing ultrasound as the imaging technique was that it costs less compared to CT scans and clinicians have recently started to use this (Poggiali et al., 2020). Since, interpreting ultrasound is more challenging, the author devised a deep learning (DL) model in the paper to help interpret ultrasound reports for COVID infection. The author suggests a frame level classification, video level grading, and pathological artefact segmentation. The author had in total 58,924 frames to work with. Regularized spatial transformer network (Reg-STN) was used as the network.

In another study, Chung et al. (2015) extracted radiomics features from multi-parametric MRI using a quantitative radiomics feature model. Then, the author uses a support vector machine (SVM) classifier to get initial detection of cancer and then combines the output from SVM with radiomics-driven conditional random field (RD-CRF) framework to get the final detection. Even though this method achieved accuracy more than its predecessors, it is very trivial. Hadjiyski (2020) had used Inception v3 neural network on 3D rendered CT scans to predict the staging of kidney cancer. The images were cropped by using ImageJ making sure the cropped portion included kidney cancer. He achieved an AUC score of 0.90 for the test set (Hadjiyski, 2020).

In canine mammary carcinoma, mitotic count from whole slide images (WSI) of canine breast is analyzed to be used in human breast cancer research (Aubreville et al., 2020). Inaccurate mitotic count can

lead to wrong diagnosis. The WSIs that are available for human breast cancer do not contain annotations for the entire WSIs. Keeping in mind the need of an algorithm to detect mitotic count in WSIs a number of challenges including MITOS 2012 dataset had been released. The best performing model at that time had F1 score of 0.66 and the result from the model was flawed and the algorithm was not considered state-of-the-art anymore as the algorithm had been trained and tested from the same data (Aubreville et al., 2020). The author suggested using a combination of RetinaNet (Lin et al., 2020) and then ResNet (He et al., 2016) to increase the efficiency of identifying mitotic counts in 21 WSIs of Canine Mammary Carcinoma. The author's proposed method achieved a F1 score of 0.791, which is a significant improvement from the supposed to be state-of the-art model for identifying mitotic counts of Canine Mammary Carcinoma.

Charibaldi et al. (2018) proposed using fuzzy learning vector quantization (FLVQ) for Mycobacterium Tuberculosis (MTB) detection. The method provided a faster and cheaper solution as ZN staining method produced unsatisfying results, thorax X-ray irradiation was not suitable for developing countries. The author compared FLVQ and LVQ with three different sensors TGS822, TGS813, and TGS2611. The FLVQ neural network achieved a sensitivity (true positive rate) of 95.83% (Charibaldi et al., 2018).

Wu et al. (2021) proposed a method for joining the output from classification and segmentation for COVID-19 detection form chest CT diagnosis. They suggested to use image mixing technique (Zhang et al., 2018) to ensure the classifier does not focus on the area outside the lesion. For the classification evaluation metric, they used specificity and sensitivity. But since the goal of the project was to identify COVID-19 infected patients from their chest CT diagnosis, in other words, the COVID-19 positive class is of more importance, accuracy would have been a better performance metric.

Similar to Shin et al. (2016), where they focus on lymph node detection which can have random orientation, for our project tumor can have any random orientation. So, segmentation for applying CNN in particular regions on the image cannot be applied. That is why our work cannot rely on unsupervised masking, semantic segmentation. The method section explains how the dataset for object detection and classification was prepared separately.

The method

It will be convenient for the surgeons if the algorithms presented in this paper were able to first detect tumors at a global range inside the patient using object detection and then, to give the surgeons a second opinion in identifying any more tumors left inside the patient using classification and localization.

The proposed method in this paper is aimed to:

- Detect tumors from live surgical videos on a global range inside the patient using object detection.
- Detect tumors inside the patient at close range using images with classification and localization for three class classification.
- Detect tumors inside the patient at close range using images with classification and localization for two class classification.

Classification network

VGG-16 (Simonyan and Zisserman, 2014) convolutional neural network architecture was used for classification.

Figure 1 shows the network architecture used for two class classification. Assigning 'n' to the image size, 'd' to the color channels or depth of the image, 'f' to filter size. The image size is $(n \times n \times d)$ or $(224 \times 224 \times 3)$ and the first filter size is $(224 \times 224 \times 3)$, If convolution operation has valid padding and stride 1, the updated image size becomes $(n-f+1) \times (n-f+1)$ or $(224-224+1) \times (224-224+1) = 1 \times 1$. But with same padding, the updated image size is $(224 \times 224 \times$ number of filters). Since there were 64 filters in the first layer, the output from the first layer is $224 \times 224 \times 64$. For three class classification, it is almost the same except before the SoftMax function dropout at 0.5 is applied and the output contains 3 classes instead of 2. In the figure, the RGB image is one of the test set images getting divided into its color channels.

Before feeding the images into the network, the images had been resized to 224×224. Strides of the convolutional layers were set to 1 with same padding while the strides for the max pooling layers were set to 2, in every convolution operation the non-linearity was set to ReLU (Rectified Linear Unit).

Working principle of some of the layers in Figure 1 is mentioned below:

- MaxPooling2D: the maximum value from each of the 2×2 square of the image is taken and the activation is down sampled to the maximum value.
- Flatten: converts a two dimensional matrix to a one dimensional vector that is used as the input for the densely connected neural network. The vector output is of shape rows × columns (Chollet, 2017).

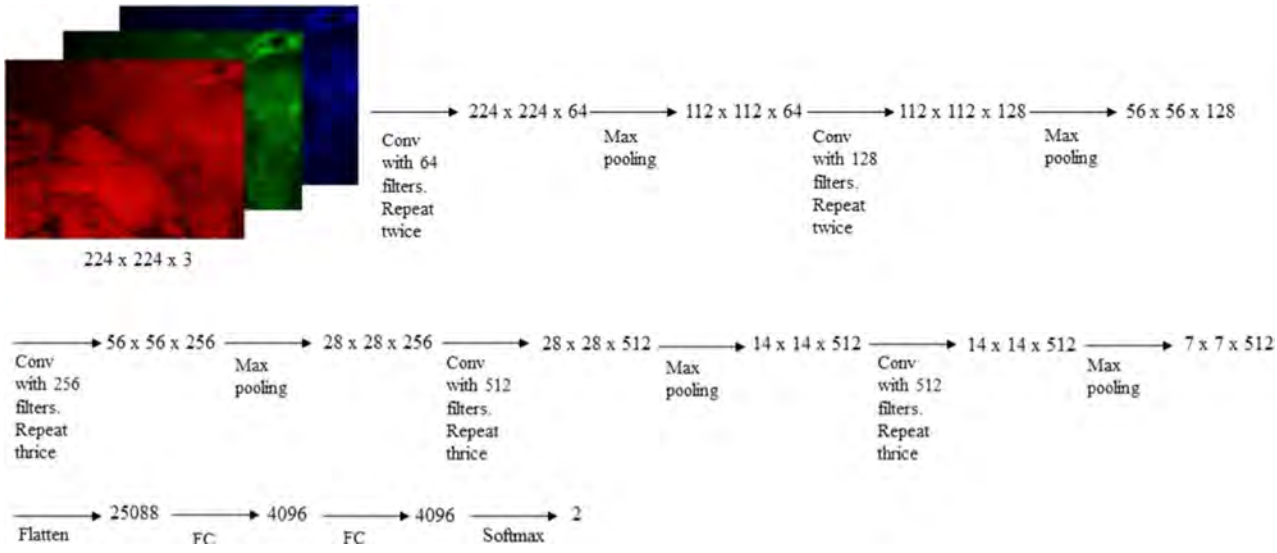

Figure 1: VGG-16 network used for classification.

- Dropout: every neuron in the network gets assigned a probability *p* for getting dropped temporarily. During one step of the training the particular neuron might be active but in the next step might be ignored. Typically, the dropout rate *p* is set as 50%. Using dropout possesses the chance for improving performance because they will have to be sufficient as possible as they cannot co-adapt with their neighboring neurons (Geron, 2019).
- Dense: this works as the output of the network. This reduces a vector of 4,096 elements to two elements (Chollet, 2017).

For debugging the neural network and for providing visual explanation, GRAD-CAM (Selvaraju et al., 2017) had been used. Regarding biological context, it is crucial that the output from deep learning model is reliable considering the deep learning models can be a 'black box'. It is anticipated that using GRAD-CAM algorithm it will be possible to debug the model and visually understand whether the predictions are correct (Brunese et al., 2020). GRAD-CAM makes use of the target class's gradient flowing to the final convolutional layer to produce a heat map to show which portions of the image contributed toward the prediction (Brunese et al., 2020; Chollet, 2017).

Object detection network

Referring to Figure 1, $7 \times 7 \times 512$ image had been flattened to 25,088 dimensional vector for feeding into densely connected neural network. Fully convolutional network (FCN) (Long et al., 2015) suggests instead of flattening the matrix into a vector, to use a 1×1 convolution to preserve spatial information.

YOLO (You Only Look Once) builds up on the idea of FCN. You Only Look Once (YOLO) is a framework for deep learning that has been used for tumor detection in global range from real time surgical images. It had also been used for skin lesion detection (Ünver and Ayan, 2019), breast masses detection (Aly et al., 2021).

This makes YOLO a good option for detecting cancerous tumors real time during surgery. The task in this section consists of determining tumor locations from partial robotic nephrectomy images or videos by drawing bounding boxes around those and also classifying those as cancerous, non-cancerous, and fatty tissue. A comparison was done between YOLOv3 and YOLOv4.

In Figure 2, first is the input image, then comes the backbone as the feature extractor, neck is the subset of the backbone and it enhances the feature discriminability and robustness. Afterwards, comes dense prediction step which does object detection. If it is a two-stage detector, like Faster R-CNN or Mask R-CNN, the next step is sparse prediction (Bochkovskiy et al., 2020).

YOLOv3 uses Darknet-53 as the backbone, Feature Pyramid Netwrok as the neck and YOLO as the detector (Redmon and Farhadi, 2018). YOLOv4 uses CSPDarknet53 as the backbone as the spatial pyramid pooling introduced in this backbone structure can significantly increase the receptive field and extract

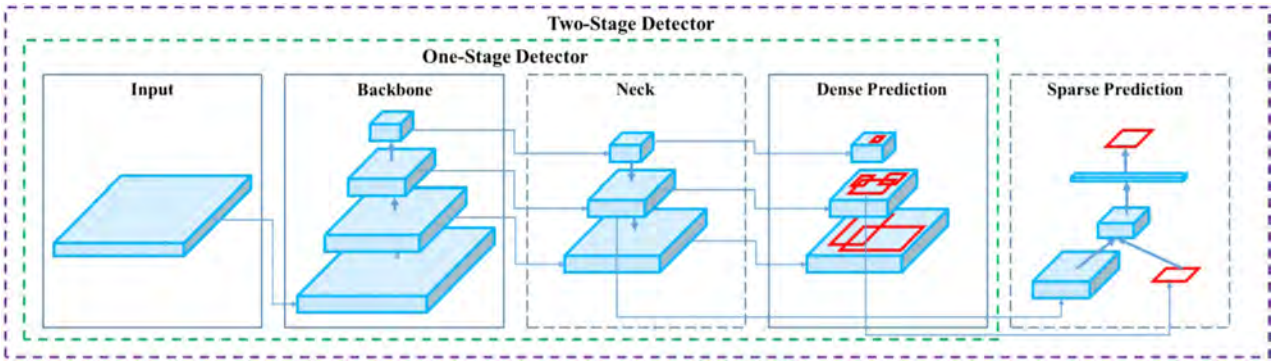

Figure 2: Object detection stages (Bochkovskiy et al., 2020).

the important features, Path Aggregation Network (PANet) as neck and the same detector as YOLOv3 (Redmon and Farhadi, 2018).

The backbone of v3 is deeper than the backbone of v4, which makes v3 slower for training. Small objects are still difficult to identify, a problem that was there from v1 and v2. YOLOv4 improves on the performance of YOLOv3 without requiring any additional hardware which also can be seen in the results section (Redmon and Farhadi, 2018).

All the input images are resized to 608 × 608 and in the detection layer the images are divided into 12 × 12 grid. Previously mentioned, the object detector step for both YOLOv3 and YOLOv4 are the same. Each of the grid cell on the image is linked with one object along with the confidence score and the coordinates for the bounding boxes. There can be more than one box around one object in the detection stage. Non-max suppression is used to get rid of the extra bounding boxes with an intersection over union (IoU) threshold. The stages of yolov4 are shown in Figure 3.

Experimental analysis

This section explains the results of tumor detection from live surgical videos using YOLOv3 and YOLOv4.

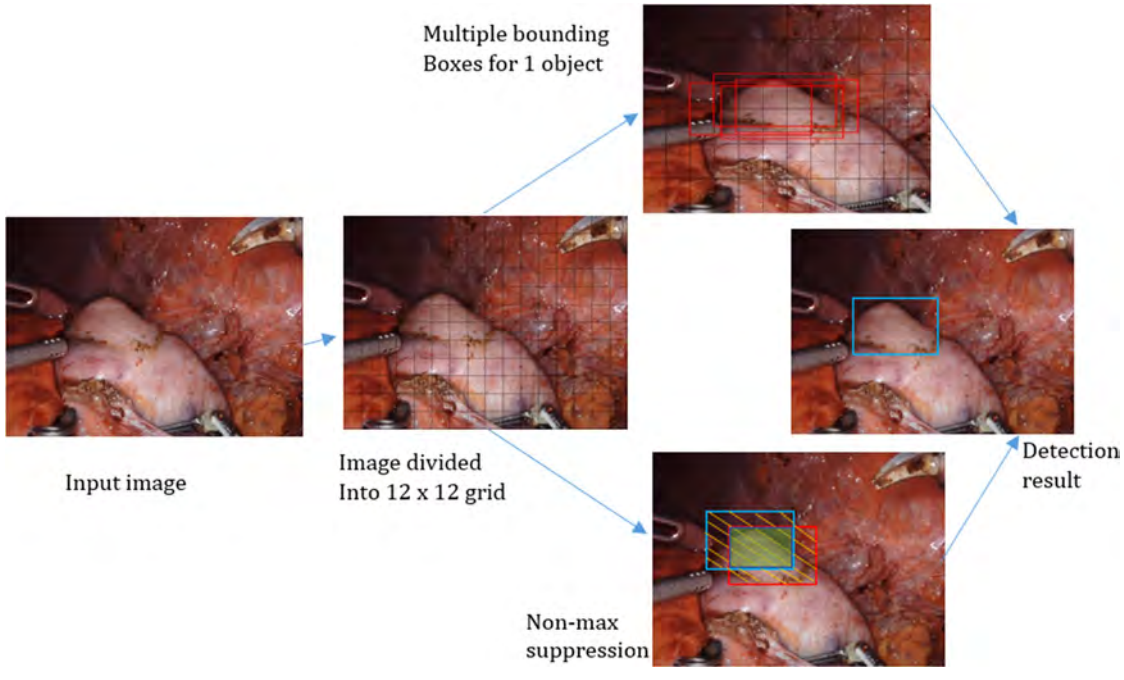

Figure 3: Object detection stages of YOLOv3 and YOLOv4.

Then, the results from classification are discussed with variations of AlexNet and VGG-16.

Dataset

There are no publicly available data for tumor detection during surgery. The dataset for object detection and classification were prepared in different ways. For object detection, there were 56 images in total (49 for training and 56 for test). The training images were loaded onto LabelImg (tzutalin, 2017), an open source image labeling tool and their annotations saved. The dataset for classification were prepared in three different ways mentioned later in this section. The images were collected from YouTube videos on partial robotic nephrectomy. Table 1 shows from which institutions the images were collected from Kibel (2018), Porter (2015), Abaza (2020a, b), P. N. U. Specialist (n.d.), Rogers (2015), Engel et al. (2016), GlobalCastMD (n.d.), Abaza (2020a, b), Hampton (2015).

The images for detection set were not further cropped. It was made sure that when cropping photos from the videos, the robotic arms are kept out of the images as much as possible. Two examples are shown in Figure 4.

Upon further inspection, it can be seen that the images in partial robotic nephrectomy contains images that has portion of cancerous tissue and non-cancerous tissue and some even has fatty tissue. The images were further cropped so that there are images only for cancerous tissue, only for non-cancerous tissue and fatty tissue. The image from Figure 4 had been further cropped to only include the tumor as the cancerous tissue as shown in Figure 5.

Table 1. Institutions that produced the videos.

Source	Country	State
Brigham and Women's Hospital	USA	Massachusetts
Seattle Science Foundation	USA	Washington
Pacific Northwest Urology specialist	USA	Washington
Vattikuti Foundation	USA	Michigan
Urologic Surgeons of Washington	USA	Washington

Cropping the photos similarly in the example showed in Figures 4 and 5, three sets of data were prepared. The first set of data only contained the cropped photos as they were. The first dataset contained three classes including cancerous tissue, non-cancerous tissue, and fatty tissue. In total, 30 cancerous tissue images were used for the training set, nine were used for validation, and five were used for testing. For non-cancerous tissue, there were 40 for training, 13 for validation and 9 for testing. For fatty tissue there were 21 for training, 10 for validation and 6 for testing. This was named as the first dataset.

For the second and third dataset, image augmentation was applied where 1 image was mirrored, rotated 90^0 clockwise, rotated 180^0 clockwise, rotated 45^0 clockwise. So, five images were made from 1 image. For the second dataset, there were still three classes, but for the third dataset there were two classes including cancerous tissue and non-cancerous tissue.

Finally, to keep the number of images same during training through different classes, when two classes were considered, there were 150 images for the cancerous tissue, and 150 images for non-cancerous tissue.

For the dataset with three classes, for training there were 105 cancerous tissue images, 105 non-cancerous tissue images, and 105 fatty tissue images.

The numbers have been summarized in Table 2.

Result evaluation with metrics (object detection/global range detection)

The YOLOv4 (Bochkovskiy et al., 2020) algorithm used for object detection was written by Alexey (n.d.). The open-source code was downloaded from GitHub and using OpenCV as the vision engine the images were loaded into the model. After running the training for 15 hr with image augmentation activated. After training was done, the algorithms were tested with the images from the test set and on the videos from which the test set images had been extracted.

Before training the algorithm on the windows PC, it was implemented in Google Colab virtual machine using YOLOv3 (Redmon and Farhadi, 2018). The evaluation metrics that were used are as follows.

Precision, recall, mean Average Precision (mAP), Frames Per Second (FPS) (For video data).

From Table 3, the mAP of YOLOv4 on windows is better than YOLOv3 on virtual machine. Also, it was not possible to run detection on videos on the virtual machine. Therefore, in terms of evaluation metric, the YOLOv4 is better for tumor detection.

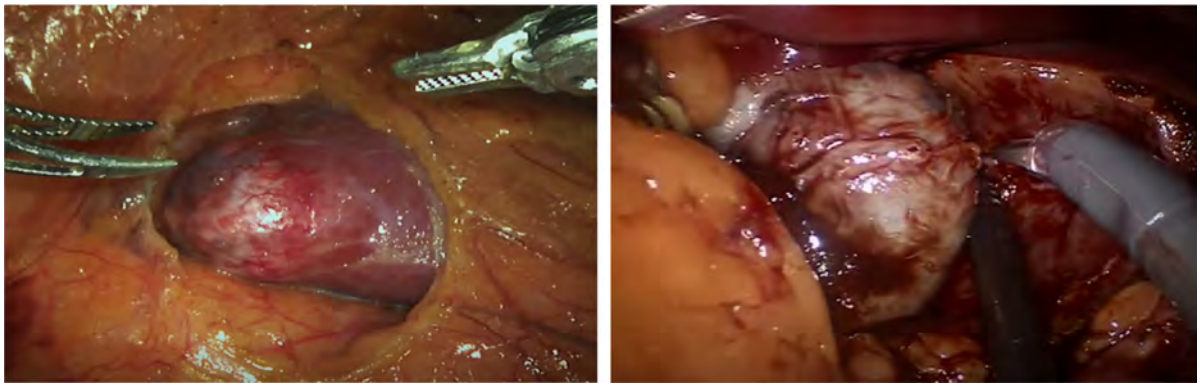

Figure 4: Cropped photo from the surgical video. Contains tumor, portions of kidney, portions of fatty tissue (Abaza, 2020a, b; P. N. U. Specialist, n.d.).

Visual evaluation (object detection/global range detection)

Here are some of the detection images from the video file attached to the document with YOLOv4 in Figure 6. In the algorithm, batch size was set as 16 with subdivision 64. The learning rate was 0.0001.

In Figure 6, the cancerous tissue is pink, non-cancerous tissue is blue, fatty tissue is green bounding box.

In the Supplementary file (https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0), visual evaluation results for YOLOv3, the detections are run at different iteration weights, and the detection at Figure 1s (c) is not entirely correct. From both, evaluation with metrics and visual evaluation YOLOv4 is better for tumor detection at global range inside the patient. Now, we will discuss the results for close range detection.

Evaluation with metrics (classification/close range detection)

The dataset was prepared in three different ways for classification and localization as mentioned in dataset section. The performance was evaluated using the following evaluation matrices:

- Loss VS epochs curve.
- Confusion matrix.

First, AlexNet was implemented with 7×7 filter window for the first layers with four strides. The network was overfitting the data and gave the indication that a deeper network was required. So, VGG-16 was implemented with learning rate 10^{-4}. The network was still overfitting. The learning rate was reduced to 10^{-6} for two class classification and 10^{-5} for three class classification. See

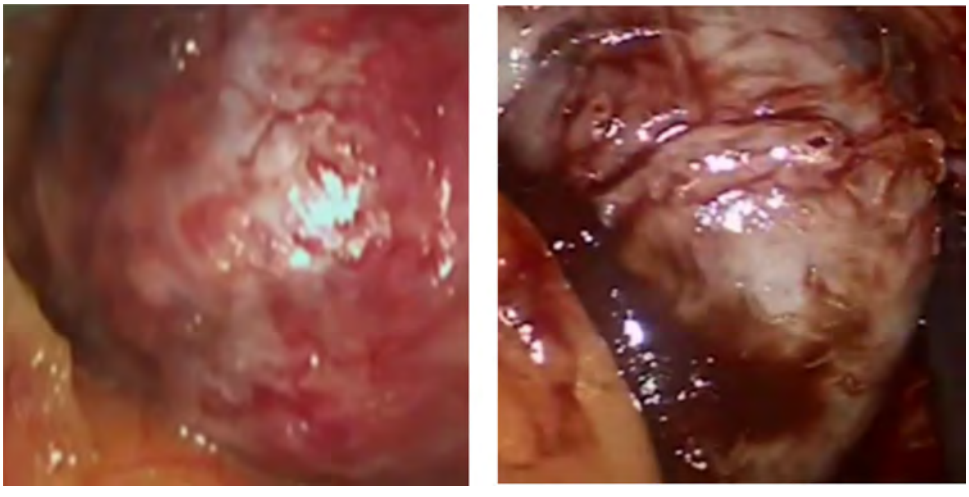

Figure 5: Tumors cropped from Figure 3 (Abaza, 2020a, b; P. N. U. Specialist, n.d.).

Table 2. Train, validation, test division.

Dataset	Label	Training	Validation	Test
1st dataset	Cancerous tissue	30	9	5
	Non-cancerous tissue	40	13	9
	Fatty tissue	21	10	6
2nd dataset	Cancerous tissue	105	9	5
	Non-cancerous tissue	105	13	9
	Fatty tissue	105	10	6
3rd dataset	Cancerous tissue	150	9	5
	Non-cancerous tissue	150	23	15

Table 3. Result comparison.

Detection algorithm	Precision	Recall	Mean average precision	Frames per second
YOLOv3 on virtual machine	0.88	0.62	0.758	Not applicable
YOLOv4 on windows	0.98	0.99	0.974	21.4

Supplementary file for the plots (Figs. 2s and 3s, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0). We have also applied callback (to stop iteration when there are no more improvements), dropout to combat overfitting, also dropout with callback. Loss VS Epoch and Accuracy VS epoch curves. The Loss VS epoch curves for the models mentioned before have very similar curves. That is why confusion matrix was employed for further evaluation. The confusion matrix for the best performing models for second dataset and third dataset will be shown here. For the rest of the confusion matrix refer to Supplementary file (Figs. 4s and 5s, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0).

Figure 7A, B shows that in the *y*-axis is the true label, in the *x*-axis is the predicted label. Out of nine cancerous tissue, seven were correctly classified for both cases. In the Supplementary file (Figs. 4s and 5s, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0), other models with poor performing Loss VS epochs curve was not shown because of their poor performing confusion matrix.

Looking into Supplementary file (Figs. 4s and 5s, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0), comparison cannot be done between VGG16 second dataset lower lr callback and VGG16 second dataset lower lr Dropout 0.5 callback for three class classification, between VGG-16 third dataset lower lr using callback and VGG-16 third dataset lower lr dropout 0.5 callback. Then, we opt for visual evaluation to eliminate the poor performing models.

Visual evaluation (localization/close range detection)

Visual evaluation was done by:

- Gradient based Class Activation Mapping (GRAD-CAM).

There are five cancerous tissue images in the test set. Here, it will be tested which algorithm can give correct prediction for the cancerous tissue and also highlight the cancerous tissue portion on the image. The networks output 2 or 3 probabilities for each image depending on whether 2 (third dataset) or 3

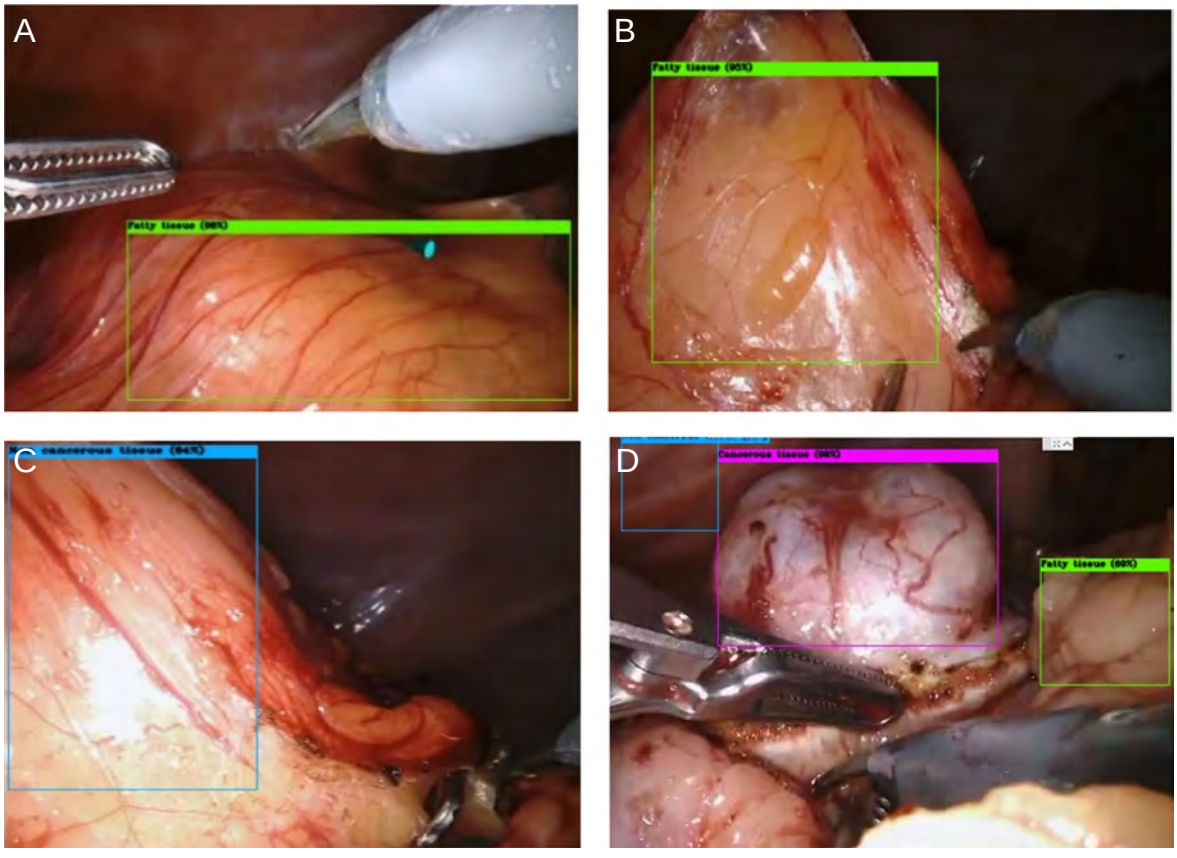

Figure 6: Tumor detection on windows machine on videos (A) fatty tissue, (B) fatty tissue, (C) non-cancerous tissue, (D) cancerous tissue, non-cancerous tissue, fatty tissue.

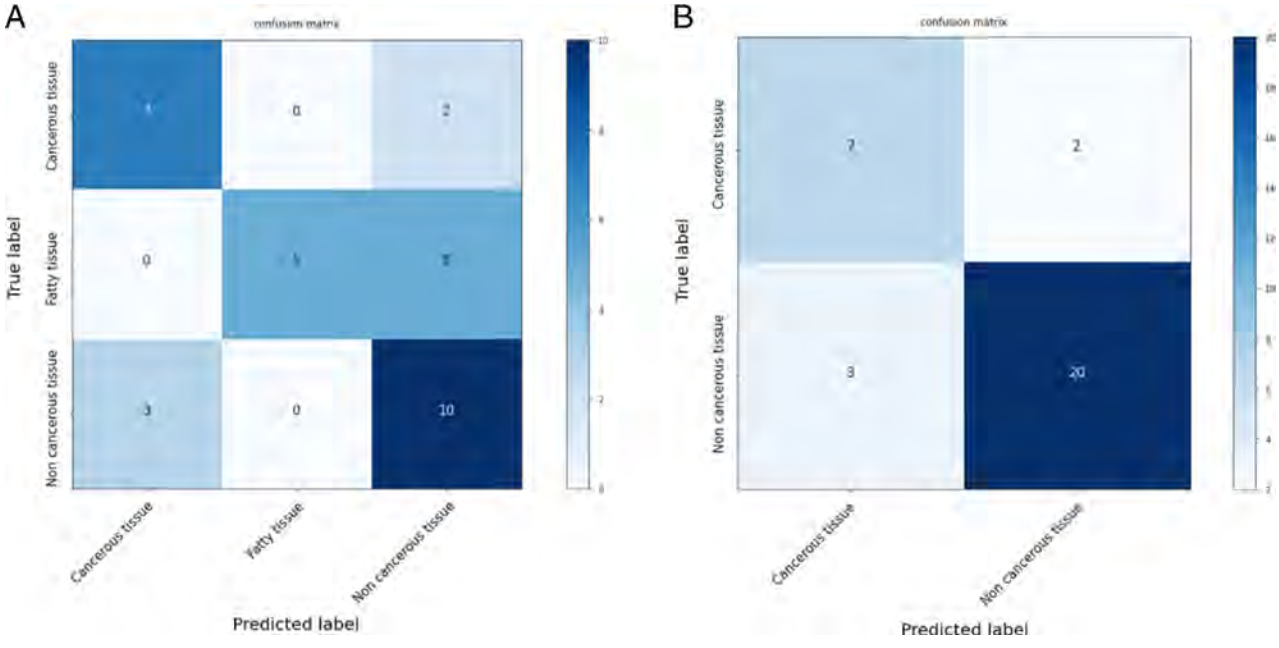

Figure 7: (A) VGG-16 2nd dataset, lower lr, Dropout 0.5, callback (B) VGG-16 3rd dataset, lower lr callback.

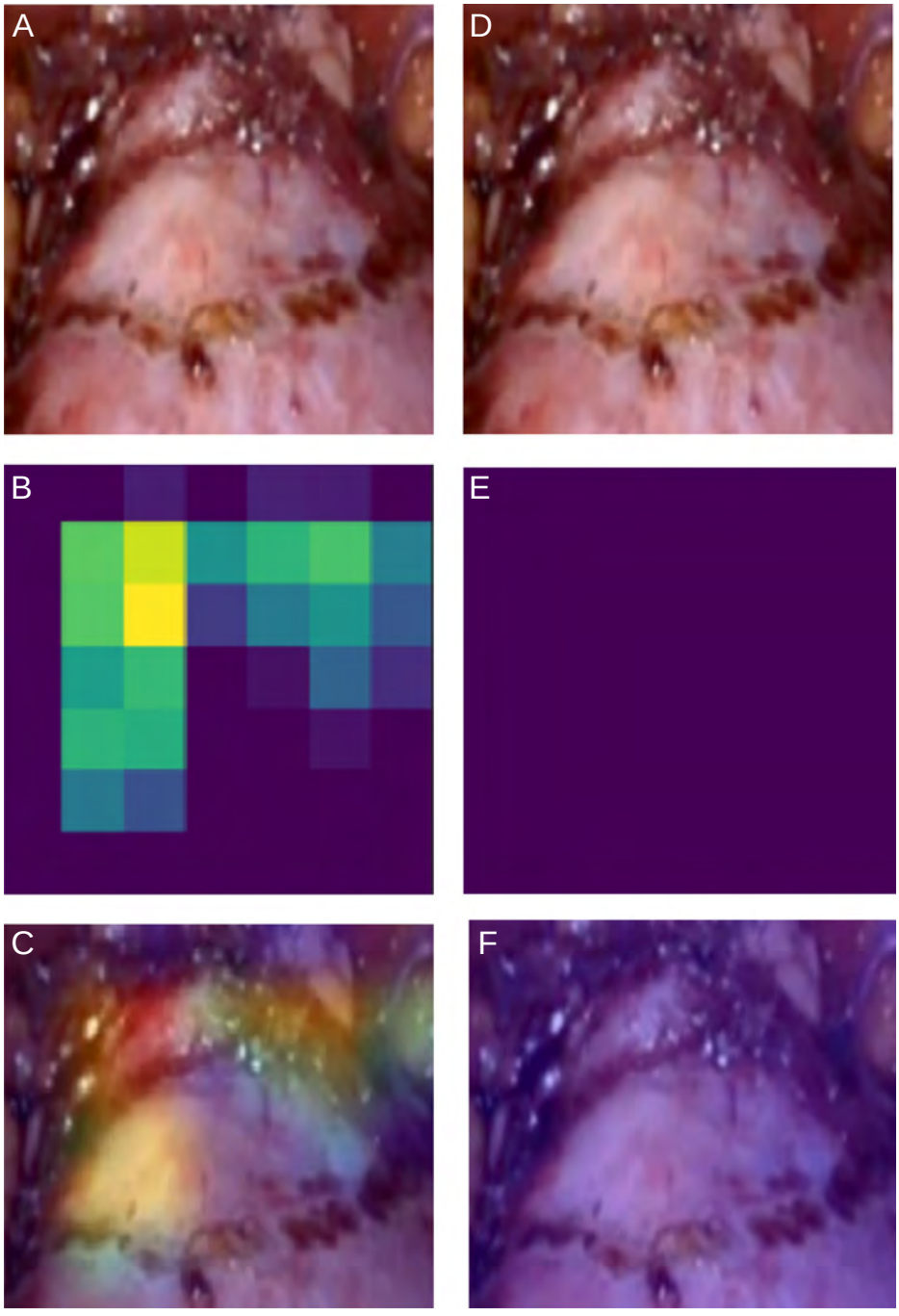

Figure 8: Comparison of VGG-16 with lower lr, dropout, callback (left column) and VGG-16 with lower lr, callback (right column) for 5th dataset (3 class classification). (A) One of the images from the test set, (B) Coarse heatmap for the image from lower lr, dropout, callback, (C) Heatmap for the image from lower lr, dropout, callback. On the right column it was not detected. (D) The same image from the test set, (E) Network was not able to detect the image, that is why it is purple (F) No heatmap got detected on the image.

(second dataset) class classification is being done. The highest probability region gets highlighted as red in the image. For example, the figure shown in Figure 8 gets three probability output as [0.971,0.00035,0.028] when three class classification is done. The first probability is for cancerous tissue, the second number is for fatty

tissue, the third is for non-cancerous tissue. Since the cancerous tissue probability is high, that gets highlighted with red in the image.

From Figure 8, VGG16 second dataset lower lr Dropout 0.5 callback was the model that was selected for three class classification.

Comparison was done the same way for two class classification and VGG-16 third dataset 2 lower lr using callback was selected as the model for two class classification. Look into Supplementary file Visualizing heatmap and prediction outputs section, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0.

The comparison of this study with other related studies are summarized in Table 4.

Figure 9 shows a flowchart of how the two-stage detectors are being used for cancerous tumor detection.

Conclusion and future work

Considering there is no live tumor detection technology currently in the da Vinci Xi robot, this paper proposes a CNN approach to help surgeons detect tumors live during surgery. Global range tumor detection inside the patient was done via YOLOv4. The close range detection approach is built on VGG-16 base model. Two main models were considered for the paper. Variation of the models was also considered. For global range detection, there was comparison between YOLOv3 and YOLOv4. For classification, comparison was between two classes (cancerous tissue, non-cancerous tissue) and three classes (cancerous tissue, fatty tissue, non-cancerous tissue) and for the two variations five different models of VGG-16 were considered. The other model classified between two classes which included cancerous and non-cancerous tissue. Also, the areas where tumor was detected was highlighted depending on the output of the CNN model (more details of this in the Supplementary file, https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0). For two class classification, with 150 cancerous tissue images and 150 non-cancerous tissue images in the training set, the final accuracy was 0.84. For three class classification, with 105 images for each of cancerous, non-cancerous, and fatty tissue in the training set, the final accuracy was 0.69. The proposed method is for identifying tumors in global range at first, and then, when the tumor had been cut off, close range (in 'Method' section it was mentioned that the images were cropped to include the cancerous and non-cancerous portion) detection will come into play to give the surgeons a second opinion in terms of identifying if there were any more tumors that the surgeon had missed. Looking at the results in this paper, it is hoped that surgeons will

Table 4. Comparison with other studies.

Method	Image type	AI technique used	Total images (TI)	Evaluation metric	Validation performance (VP)	$\frac{VP}{TI}$
Hadjiyski (2020)	CT scans	Inception v3	4,200	AUC	86%	0.02
Aubreville et al. (2020)	Whole Slide Images	RetinaNet with ResNet-50	13,907	F1 score	79.1%	0.01
Wang et al. (2018)	Multi parametric MRI	V-net	79 cases in total. About 790 images	Accuracy	89.4%	0.11
Chung et al. (2015)	Multi parametric MRI	SVM with RD-CRF	20 cases in total. About 200 images	Accuracy	59%	0.29
Brunese et al. (2020)	Chest X-ray	VGG-16	9,326	Accuracy	98%	0.01
Wu et al. (2021)	Chest CT scan	VGG-16 with segmentation	3,855	Sensitivity	95%	0.03
This study	Live partial robotic nephrectomy	Object detection with VGG-16	143	Accuracy	84%	0.59

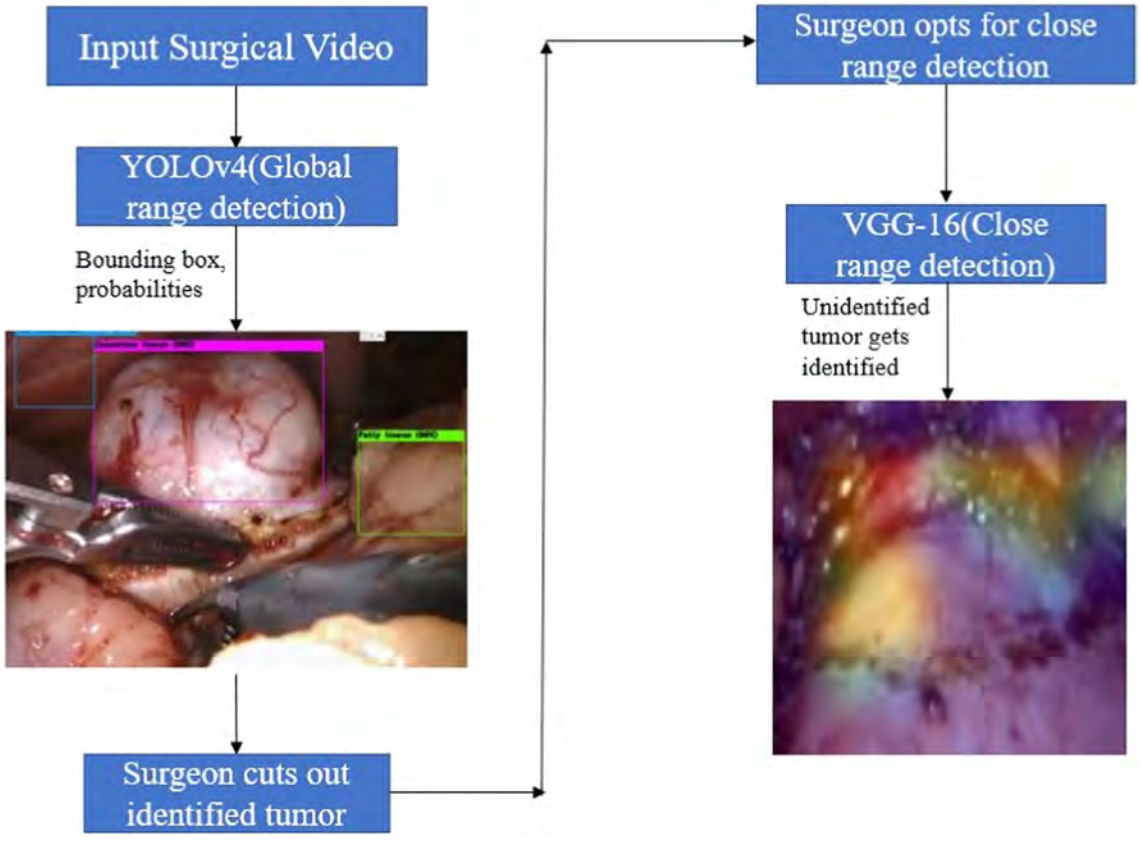

Figure 9: Research methodology flowchart.

be more interested in making dataset on live surgery images available online. Provided about 5,000 images (1,500 cancerous tissue for training, 1,500 non-cancerous tissue for training, 500 cancerous tissue for validation, 500 non-cancerous tissue for validation, 500 cancerous tissue for testing, 500 non-cancerous tissue for testing) can be made available, will enable the results to be more promising and will allow detection on a more customized scale. If more data are made publicly available, then the semantic segmentation can be applicable and the close range detection will be more accurate and reliable.

Literature Cited

Abaza, R. 2020a. *Avoiding Positive Margins During Robotic Partial Nephrectomy presented by Ronney Abaza*, Seattle Science Foundation, Seattle, Washington, Available at: https://www.youtube.com/watch?v=C3VTbb_1GAM&ab_channel=SeattleScienceFoundation (Accessed March 20, 2021).

Abaza, R. 2020b. *Robotic Partial Nephrectomy for Complex Tumors presented by Ronney Abaza*, Seattle Science Foundation, Seattle, Washington, Available at: https://www.youtube.com/watch?v=vvf16vBrgxQ&t=662s&ab_channel=SeattleScienceFoundation (Accessed March 20, 2021).

American Institute of Minimally Invasive Surgery 2019. *DA VINCI XI*, American Medical Center, Available at: https://www.aimisrobotics.com/da-vinci-xi/ (Accessed June 26, 2021).

Alexey, A. B. n.d. *darknet*, Available at: https://github.com/AlexeyAB/darknet.

Aly, G. H., Marey, M., El-Sayed, S. A. and Tolba, M. F. 2021. YOLO based breast masses detection and classification in full-field digital mammograms. *Computer Methods and Programs in Biomedicine* 200: 105823. Available at: https://doi.org/10.1016/j.cmpb.2020.105823.

Asadnia, M., Kottapalli, A. G. P., Miao, J., Benson, R. A., Sabbagh, A., Kropelnicki, P. and Tsai, J. 2013. High temperature characterization of PZT (0.52/0.48) thin-film pressure sensors. *Journal of Micromechanics and Microengineering* 24(1): 015017.

Asadnia, M., Chua, L. H., Qin, X. and Talei, A. 2014. Improved particle swarm optimization–based artificial neural network for rainfall-runoff modeling. *Journal of Hydrologic Engineering* 19(7): 1320–1329.

Asadnia, M., Yazdi, M. S. and Khorasani, A. 2010. An improved particle swarm optimization based on neural network for surface roughness optimization in face milling of 6061-T6 Aluminum. *International Journal of Applied Engineering Research* 5(19): 3191–3201.

Asadnia, M., Khorasani, A. M. and Warkiani, M. E. 2017. An accurate PSO-GA based neural network to model growth of carbon nanotubes. *Journal of Nanomaterials* 2017.

Aubreville, M., Bertram, C. A., Donovan, T. A., Marzahl, C., Maier, A. and Klopfleisch, R. 2020. A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. *Scientific Data* 7(1): 417, doi: 10.1038/s41597-020-00756-z.

Bazaz, S. R., Mehrizi, A. A., Ghorbani, S., Vasilescu, S., Asadnia, M. and Warkiani, M. E. 2018. A hybrid micromixer with planar mixing units. *RSC Advances* 8(58): 33103–33120.

Bennet, M., Thamilvalluvan, B., Alphonse, P. P., Thendralarasi, D. R., Sujithra, K. J. I. J. O. S. S. and Systems, I. 2017. Performance and analysis of automatic license plate localization and recognition from video sequences. *International Journal on Smart Sensing and Intelligent Systems* 10: 330–343.

Bochkovskiy, A., Wang, C. and Liao, H. 2020. YOLOv4: optimal speed and accuracy of object detection. *Computer Vision and Pattern Recognition* 1.

Brunese, L., Mercaldo, F., Reginelli, A., Santone, A. J. C. M. and Biomedicine, P. I. 2020. Explainable deep learning for pulmonary disease and Coronavirus COVID-19 detection from x-rays vol. 196: 105608–105608.

Charibaldi, N., Harjoko, A., Azhari, Hisyam, B. J. I. J. O. S. S. and Systems, I. 2018. A new HGA-FLVQ model for Mycobacterium Tuberculosis detection,". *International Journal on Smart Sensing and Intelligent Systems* 11: 1–13.

Chen, Z., Zhang, T. and Ouyang, C. 2018. End-to-end airplane detection using transfer learning in remote sensing images. *Remote Sensing* 10(1): 139, doi: 10.3390/rs10010139.

Chollet, F. 2017. *Deep Learning with Python*. Manning Publications, Shelter Island, NY.

Chung, A. G., Khalvati, F., Shafiee, M. J., Haider, M. A. and Wong, A. 2015. Prostate cancer detection via a quantitative radiomics-driven conditional random field framework. *IEEE Access* 3: 2531–2541, doi: 10.1109/ACCESS.2015.2502220.

David, B. and Samadi, M. D. n.d. *History and The Future of Robotic Surgery*, Robotic Oncology, Available at: https://www.roboticoncology.com/history-of-robotic-surgery/.

Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P. -A. and Müller, H. 2012. Building a reference multimedia database for interstitial lung diseases. *Computerized Medical Imaging and Graphics* 36(3): 227–238, Available at: https://doi.org/10.1016/j.compmedimag.2011.07.003.

Engel, D., Jason, D. and Engel, M. D. 2016. *Robotic Partial Nephrectomy*, Oroogic Surgeons of Washington, Available at: https://www.youtube.com/watch?v=UXWjNqTwb_4&ab_channel=JasonD.Engel%2CM.D (Accessed March 20, 2021).

Ge, L., Dan, D. and Hui, L. 2020. An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision. *Structural Control and Heath Monitoring* 27.

Geron, A. 2019. *Hands-on Machine Learning with Scikit-learn, Keras & TensorFlow*. o'Reiley Media, Inc, Sebatopol, CA.

GlobalCastMD. n.d. 02 Robotic partial nephrectomy-course tips for retroperitoneral partial nephrectomy James Porter HD, Available at: https://www.youtube.com/watch?v=S80t7cnFLus&ab_channel=GlobalCastMD.

Hadjiyski, N. 2020. Kidney cancer staging: deep learning neural network based approach. 2020 International Conference on e-Health and Bioengineering (EHB), October 29–30, pp. 1–4, doi: 10.1109/EHB50910.2020.9280188.

Hagihghi, R., Razmjou, A., Orooji, Y., Warkiani, M. E. and Asadnia, M. 2020. A miniaturized piezoresistive flow sensor for real-time monitoring of intravenous infusion. *Journal of Biomedical Materials Research Part B: Applied Biomaterials* 108(2): 568–576.

Hammal, S., Bourahla, N. and Laouami, N. 2020. Neural-network based prediction of inelastic response spectra. *Civil Engineering Journal* 6(6): 1124–1135.

Hampton, L. 2015. *da Vinci Xi Right Robotic Partial Nephrectomy-Unedited*, VCUrobotics, Richmond, VI, Available at: https://www.youtube.com/watch?v=6eyZzoScc54&ab_channel=VCUrobotics (Accessed March 20, 2021).

He, K., Zhang, X., Ren, S. and Sun, J. 2016. Deep residual learning for image recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27–30, pp. 770–778, doi: 10.1109/CVPR.2016.90.

Inc, A. and Grove, B. 2021. *Indocyanine Green Side Effects*, Available at: https://www.drugs.com/sfx/indocyanine-green-side-effects.html.

Junnumtuam, S., Niwitpong, S. -A. and Niwitpong, S. 2021. The Bayesian confidence interval for coefficient of variation of zero-inflated poisson distribution with application to daily COVID-19 deaths in Thailand. *Emerging Science Journal* 5: 62–76.

Kharate, G., Ghotkar, A. J. I. J. O. S. S. and Systems, I. 2016. Vision based multi-feature hand gesture recognition for Indian sign language manual signs. *International Journal on Smart Sensing and Intelligent Systems* 9: 124–147.

Khorasani, A. M., Gibson, I., Asadnia, M. and O'Neill, W. 2018. Mass transfer and flow in additive manufacturing of a spherical component. *International Journal of Advanced Manufacturing Technology* 96: 3711–3718.

Kibel, A. 2018. *Robotic Assisted Laparoscopic Partial Nephrectomy*, Brigham and Women's

Hospital, Boston, MA, Available at: https://www.youtube.com/watch?v=GQm90mWVMJM&ab_channel=BrighamAndWomen%27sHospital (Accessed March 20, 2021).

Kottapalli, A. G. P., Asadnia, M., Miao, J. and Triantafyllou, M. 2015. Soft polymer membrane microsensor arrays inspired by the mechanosensory lateral line on the blind cavefish. *Journal of Intelligent Material Systems and Structures* 26(1): 38–46.

Li, M., Zhang, Z., Lei, L., Wang, X. and Guo, X. 2020. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: comparison of faster R-CNN, YOLO v3 and SSD. *Sensors* 20(17), 10.3390/s20174938.

Lin, T., Goyal, P., Girshick, R., He, K. and Dollár, P. 2020. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(2): 318–327, doi: 10.1109/TPAMI.2018.2858826.

Long, J., Shelhamer, E. and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7–12, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

Mahmud, M. A. P., Azadi, F. E. S., Myers, M., Pejcic, B., Abbassi, R., Razmjou, A. and Asadnia, A. 2020. Recent progress in sensing nitrate, nitrite, phosphate, and ammonium in aquatic environment. *Chemosphere* 259: 127492.

Moshizi, S. A., Azadi, S., Belford, A., Ramjou, A., Qu, S., Han, Z. J. and Asadnia, M. 2020. Development of an ultra-sensitive and flexible piezoresistive flow sensor using vertical graphene nanosheets. *Nano-micro Letters* 12.

Nakhaeinia, D., Payeur, P., Aragon, A. C., Cretu, A-. M., Laganiere, R. and Macknojia, R. 2016. Surface following with an rgb-d vision-guided robotic system for automated and rapid vehicle inspection. *International Journal on Smart Sensing and Intelligent Systems* 9: 419–447.

National Kidney Foundation. n.d. Nephrectomy, Available at: https://www.kidney.org/atoz/content/nephrectomy.

N. Cancer. n.d. Kidney cancer: stages Available at: https://www.cancer.net/cancer-types/kidney-cancer/stages.

Ohira, N. 2018. Memory-efficient 3D connected component labeling with parallel computing. *Signal, Image and Video Processing* 12(3): 429–436, doi: 10.1007/s11760-017-1175-7.

Pantanowitz, L., Garza, G., Bien, L., Heled, R., Laifenfeld, D., Linhart, C., Sandbank, J., Shach, A. and Shalev, V. 2020. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *The Lancet Digital Health* 2(8): e407–e416, doi: 10.1016/S2589-7500(20)30159-X.

Poggiali, E., Dacrema, A. and Bastoni, D. 2020. Can Lung US Help Critical Care Clinicians in the early diagnosis of Novel Coronavirus (COVID-19) pneumonia? *Radiology* 295.

Porter, J. 2015. *LIVE SURGERY: Retroperitoneal Robotic Partial Nephrectomy*, Seattle Science Foundation, Seattle, Washington, Available at: https://www.youtube.com/watch?v=nwrbKNbLCv8&t=5045s&ab_channel=SeattleScienceFoundation (Accessed March 2021).

P. N. U. Specialist. Robotic partial nephrectomy comparisons, Available at: https://www.youtube.com/watch?-v=epvKkH3ekRo&ab_channel=PacificNorthwestUrologySpecialists%2CPLLC.

Razfar, M., Asadnia, M., Haghshenas, M. and Farahnakian, M. 2010. Optimum surface roughness prediction in face milling X20Cr13 using particle swarm optimization algorithm. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 224(11): 1645–1653.

Razmjou, A., Asadnia, M., Ghaebi, O., Yang, H-. C., Warkiani, M. E., Hou, J. and Chen, V. 2017. Preparation of iridescent 2D photonic crystals by using a mussel-inspired spatial patterning of ZIF-8 with potential applications in optical switch and chemical sensor. *ACS Applied Materials & Interfaces* 9(43): 38076–38080.

Redmon, J. and Farhadi, A. 2018. YOLOv3: an incremental improvement. *Computer Vision and Pattern Recognition* 1.

Rogers, C. 2015. *Dr. Craig Rogers: da Vinci Partial Nephrectomy*, Vattikuti Foundation, Bangalore, Available at: https://www.youtube.com/watch?v=gdg7EhsKki8&ab_channel=VattikutiFoundation (Accessed March 20, 2021).

Roth, H. R., Le, L., Liu, J., Yao, J., Seff, A., Cherry, K., Kim, L. and Summers, R. M. 2016. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Transactions on Medical Imaging* 35(5): 1170–1181, doi: 10.1109/TMI.2015.2482920.

Roy, S., Menapace, W., Oei, S., Luijten, B., Fini, E., Saltori, C., Huijben, I. A. M., Chennakeshava, N., Mento, F., Sentelli, A., Peschiera, E., Trevisan, R., Maschietto, G., Torri, E., Inchingolo, R., Smargiassi, A., Soldatti, G., Rota, P., Passerini, A., Sloun, R. J. G. V., Ricci, E. and Demi, L. 2020. Deep learning for classification and localization of COVID-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging* 39(8): 2676–2687, doi: 10.1109/TMI.2020.2994459.

Seff, A., Cherry, K. M., Roth, H., Liu, J., Wang, S., Hoffman, J., Turkbey, E. B. and Summers, R. M. 2014. 2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers. *Medical Image Computing and Computer-Assisted Interval* 17(Pt 1): 544–552, doi: 10.1007/978-3-319-10404-1_68.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), October 22–29, pp. 618–626, doi: 10.1109/ICCV.2017.74.

Shin, H-. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. and Summers, R. M. 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging* 35(5): 1285–1298, doi: 10.1109/TMI.2016.2528162.

Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhouchke, V. and Rabinovich, A. 2015. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7–12, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

tzutalin. 2017. *LabelImg*, Available at: https://github.com/tzutalin/labelImg.

Ünver, H. M. and Ayan, E. 2019. Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut Algorithm. *Diagnostics (Basel, Switzerland)* 9(3), doi: 10.3390/diagnostics9030072.

Wang, Y., Zheng, B., Gao, D. and Wang, J. 2020. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Transactions on Medical Imaging* 39(8): 2615–2625, doi: 10.1109/TMI.2020.2995965.

Wang, Y., Zheng, B., Gao, D. and Wang, J. 2018. Fully convolutional neural networks for prostate cancer detection using multi-parametric magnetic resonance images: an initial investigation. 2018 *24th International Conference on Pattern Recognition (ICPR)*, August 20–24, pp. 3814–3819, doi: 10.1109/ICPR.2018.8545754.

Wu, Y. -H., Gao, S. -H., Mei, J., Xu, J., Fan, D. -P., Zhang, R. -G. and Cheng, M. -M. 2021. JCS: an explainable COVID-19 diagnosis system by joint classification and segmentation 30: 3113–3126.

Zeiler, M. D., Fergus, R. 2014. "Visualizing and understanding convolutional networks", In Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T. (Eds), *Computer Vision – ECCV 2014* Cham: Springer International Publishing, pp. 818–833.

Zhang, H., Cisse, M., Dauphin, Y. N. and Lopezpaz, D. 2018. Mixup: Beyond empirical risk minimization. presented at the in Proc. Int. Conf. Learn. Represent.