

Cyst Detection in Renal Ultrasonographic Images

Computer Vision Final Project - Master in Information Health Engineering

Ignacio Serrano Llabrés and María Postigo Fliquete

Abstract—In this paper we propose a Region-based Convolutional Network model (Faster R-CNN) for cyst detection in kidney ultrasonographic images. For this task we have implemented a custom dataset, which is finetuned from a model pre-trained on COCO dataset[1]. Faster R-CNN introduces a Region Proposal Network (RPN), which enables nearly cost-free region proposals.

I. INTRODUCTION

In the interdisciplinary scientific field of Computer Vision, obtaining information about pictures or videos has been a challenging task. A few years ago, with the appearance of Deep Learning, this task started to include harder demands such as image classification, face recognition and object detection [2].

Object detection is a more challenging task than image classification since detection requires objects' localization. To locate the regions of interest, numerous candidate locations must be processed. Moreover, to reach a precise localization, the obtained regions must be refined [3].

The object detection system that we employ in our project is the Faster R-CNN. This system is composed by two modules: a deep fully convolutional neural network for region proposal together with the Fast R-CNN detector [4].

We base our study in the use this unified network for object detection to develop a rising renal cyst detector.

II. METHODOLOGY

A. Data description

The Signal Theory and Communications department from *Universidad Carlos III* provided us a set of 1986 renal ultrasound images. Within this dataset there are healthy and pathological kidneys, where the pathological ones present different global and/or local pathologies. For each image we also have the corresponding segmentation mask of the kidney and the bounding box coordinates of the local lesions.

Currently, the goal of our approach consists on cyst detection. Therefore, our database is conformed by the set of images which present this pathology. The project was developed with a set of 637 renal ultrasound images with cysts: 537 for training, 50 for validation and 50 for testing the model.

B. Faster R-CNN training and evaluation

The backbone selected for our object detection approach is **Resnet50**. A previous healthy vs pathological classification study on this dataset determined that this network produced the best performance over other architectures.

Most of the object detection algorithms, as well as the one used, focus on detecting different objects. In our case, we wanted just to detect cysts in the given images and locating them afterwards. The most important changes we did for training the model were the following:

- 1) Evaluation method: We based our evaluation code in the one used in the structure of [5]. The main change was the comparison between predictions and real bounding boxes. The original implementation was too complex for our objective, therefore we just took the structure and added a self-made evaluation over the test set.

The pseudocode is listed below:

```
for true_box in true_bboxes:

    count_true += 1

    for pred_box in pred_bboxes:

        get IoU(true_box , pred_box)

        best = pred_box that max(IoU)

    if best_IoU > threshold:

        count_pred += 1

accuracy = count_pred / count_true
```

- 2) Intersection Over Union (IoU):

The IoU refers to the, commonly used in object detection, metric using bounding boxes (bbox in the pseudocode), we base ourselves on [6]. We compare each true bounding box with each predicted one, we select the prediction that has the maximum IoU and check whether it satisfies a previously chosen threshold.

In Figure 1 we can see two bounding boxes, the ground truth (gt) and the predicted one (pred), the metric is computed as follow: We first calculate the intersection (grid part) and

then divide it by the union of both boxes (blue colored part). The bounding boxes are defined by two opposite points:

```
A_gt = (x_topleft_gt, y_topleft_gt)
B_gt = (x_bottomright_gt, y_bottomright_gt)
A_pred = (x_topleft_pred, y_topleft_pred)
B_pred = (x_bottomright_pred, y_bottomright_pred)
```

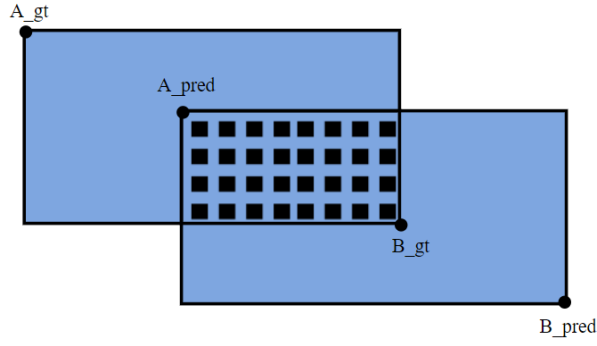


Fig. 1. Caption

III. EXPERIMENTS AND ANALYSIS OF RESULTS

When working with pre-trained models, data augmentation and hyperparameters finetuning play a key role for accuracy improvement. We evaluated the Faster R-CNN performance on our dataset with different transformations, batch sizes and intersection over union thresholds.

A. Data Transforms

When applying custom transforms to our dataset, the same transformations were applied over the segmentation masks and the bounding boxes of the lesions. The employed transforms are: the horizontal flip, a crop around the segmentation mask, rotation and brightness increase, which are visible in Figure 2. These image variations were combined differently in order to reach the best training and detection precision.

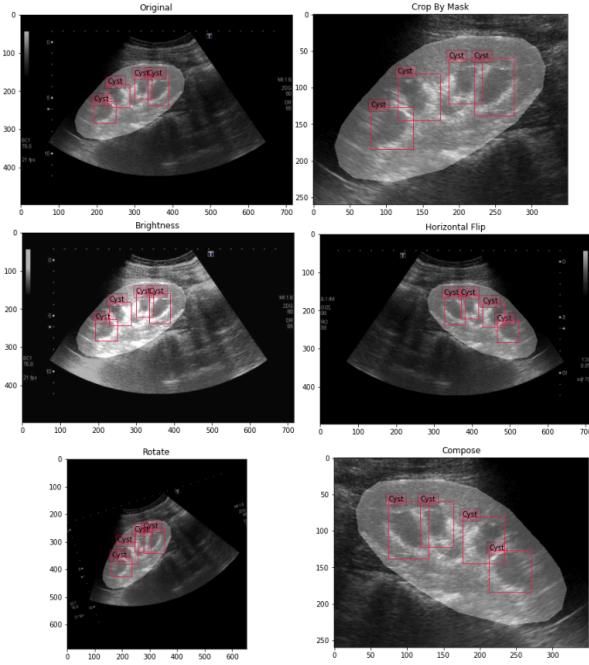


Fig. 2. Image transforms

The model was finetuned for 10 epochs with different transformations. During the training, the model was validated over the 50 images of the validation set, with and without

saving the network weights that provided the best validation accuracy. Then, we tested the resulting model over the test set (50 images). The accuracy per epoch evolution is visible in Figure 3 and the prediction accuracies over the test set are in Table I. The acronyms appearing in both table and figure stands for: Horizontal Flip (HF), Crop By Mask (Mask), Rotation (rot), Increase Brightness (bright) and saving the best weights (BW) during train and validation.

Transforms	Accuracy not BW	Accuracy BW
HF	0.8023	0.8372
HF+Mask	0.7325	0.8604
HF+Mask+rot	0.5814	0.6279
HF+Mask+Brighth	0.1395	0.6976

TABLE I
PREDICTION ACCURACIES ON THE TEST SET WITH DIFFERENT TRANSFORMS

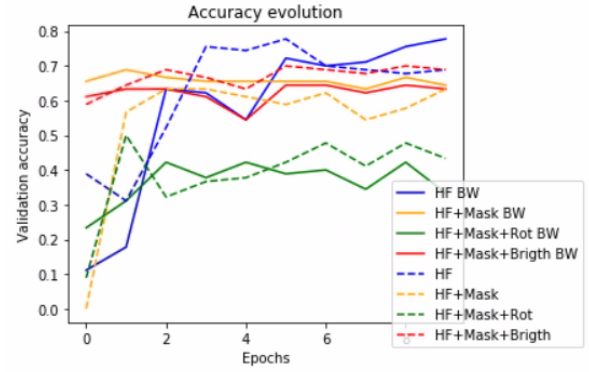


Fig. 3. Validation accuracy evolution with different transforms

The obtained results reveal us that saving the **best model weights** per epoch, according to the best validation accuracy, provides a better performance when detecting cysts in the test set.

It seems that there was a problem with the rotation and brightness transforms, since they significantly drop the accuracy. Nevertheless, the Crop by Mask and the Horizontal Flip work pretty well, providing the bests accuracies in validation and test. Although the Mask transform reaches an accuracy of **0.8604**, we consider that only applying the Horizontal Flip (**0.8372**) will lead to a better detection due to the accuracy evolution during training (See the blue lines in Figure 3). This Hypothesis would be confirmed/rejected if we train the model with different train, validation and test partitions.

B. Finetuning Hyperparameters

The main approach in this section is to finetune the hyperparameters of the network. This is done by comparing different results for the experiments made. In the following table we expose the most interesting results.

Fixed number of epochs			
Model	Batch Size	IoU Th	Accuracy
Faster RestNet50	4	0.5	0.862
Faster RestNet50	4	0.7	0.827
Faster RestNet50	8	0.5	0.75
Faster RestNet50	8	0.7	0.25

We found an obstacle with the threshold. The validation accuracy was extremely low, which is normal if the conditions are too strict for the network. Anyhow, the predictions were good. This is probably due to the way IoU is calculated. The predictions are sometimes good but the bounding box is small compared to the true one, therefore the IoU is low and doesn't reach the threshold as we can see in Figure 4.

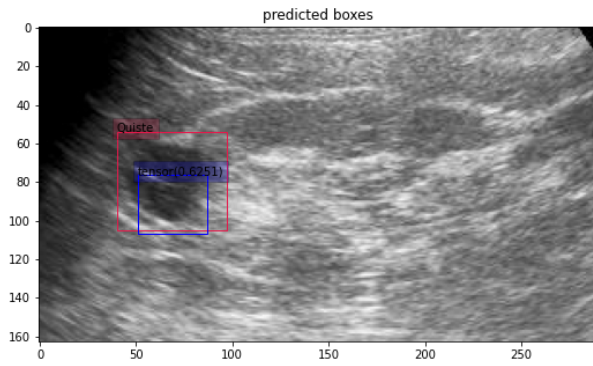


Fig. 4. Problematic output

On the other hand, we can observe some of the predictions made by the model on test data in Figure 5.

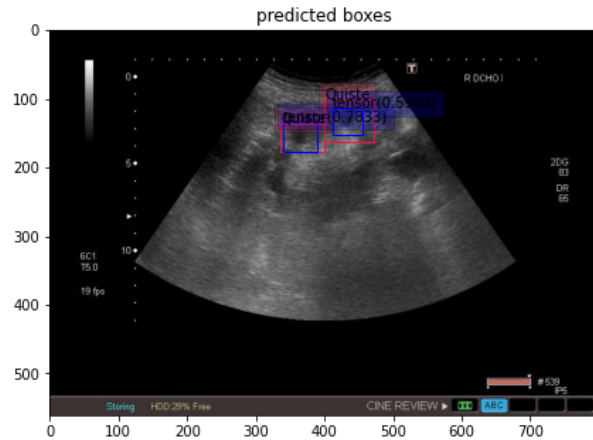


Fig. 5. Test output

IV. CONCLUSIONS AND FURTHER WORK

We can conclude that our approach can properly detect cysts in the kidney US images of the given test set. The best transform to apply is the Horizontal Flip by its own, the best

threshold is 0.5 for the IoU metric and the best batch size is 4 samples.

Nevertheless, as future work the same training should be performed over different data partitions in order to confirm the robustness of the method. Moreover, since this detection problem is hard due to the artifacts in the image and the low amount of data, applying transformations is necessary. The Crop by Mask transform could be applied with a greater dilation, since some cysts are outside the segmentation mask.

Finally, this study only detect cysts whilst the whole dataset contains different pathologies. Therefore, a multiple object detection approach will be a future study.

ACKNOWLEDGMENTS

We would like to extend our sincere thanks to Universidad Carlos III, its Signal Theory and Communications department and to Iván González Díaz to allow this project to happen.

Moreover, we are very grateful to Miguel Molina Moreno helping us during the study.

Finally we would like to express our gratitude to the Master in Information Health Engineering for giving us all the knowledge that we applied in this project.

REFERENCES

- [1] COCO dataset <http://cocodataset.org/home>
- [2] Lecture Notes on Computer Vision, Rob Fergus. <http://cs.nyu.edu/fergus/teaching/vision/index.html>
- [3] Girshick, Ross. "Fast R-CNN." 2015 IEEE International Conference on Computer Vision (ICCV), 2015, doi:10.1109/iccv.2015.169.
- [4] Ren, Shaoqing, et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, 2017, pp. 1137–1149., doi:10.1109/tpami.2016.2577031.
- [5] Pytorch Vision. <https://github.com/pytorch/vision/blob/master/references/detection/engine.py>
- [6] Evaluating performance of an object detection model <https://towardsdatascience.com/evaluating-performance-of-an-object-detection-model-137a349c517b>