PATCH-BASED DEEP LEARNING APPROACHES FOR ARTEFACT DETECTION OF ENDOSCOPIC IMAGES

Xiaohong W. Gao[1], *Yu Qian* [2]

[1]Department of Computer Science, Middlesex University, London, NW4 4BT, UK
[2]Cortexcia Vision System Limited, London SE1 9LQ, UK

ABSTRACT

This paper constitutes the work in EAD2019 competition. In this competition, for segmentation (task 2) of five types of artefact, patch-based fully convolutional neural network (FCN) allied to support vector machine (SVM) classifier is implemented, aiming to contend with smaller data sets (i.e., hundreds) and the characteristics of endoscopic images with limited regions capturing artefact (e.g. bubbles, specularity). In comparison with conventional CNN and other state of the art approaches (e.g. DeepLab) while processed on whole images, this patch-based FCN appears to achieve the best.

Index Terms— Endoscopic images, Deep neural networks, Decoder-Encoder neural networks

1. INTRODUCTION

This paper details the work by taking part of the Endoscopic artefact detection challenge (EAD2019) [1, 2] with three tasks, which are detection (task #1), segmentation (task #2) and generalization (task #3). All three tasks are performed using the current state of the art deep learning techniques with a number of enhancements. For example, for segmentation (task #2), patch-based approached are applied. In doing so, each image is divided into 55 non-overlapping patches of equal sizes. Then based on the contents of their counterparts of masks, only patches with non-zero masks are selected for training to limit the inclusion of background information. Each class is trained individually firstly. Then upon the last layer of receptive fields, the features from five classes are trained together using SVM to further differentiate subtle changes between five classes.

For detection of bounding boxes (Tasks #1 and #3), while the above patch-based approach delivers good segmentations, the bounding boxes of those segments do not seem to agree well with the ground truth with Null values of IoU. Hence the state of the art models of faster-RCNN with resNet101 backbone has been applied that gives the ranking position of 12th on the leaderboard, which is build upon tensorflow model. In addition, the models of YOLOv3 [3] by using darknet is also evaluated, which delivers detection ranks between 17 to 21 based the selection of thresholds (0.5 or 0.1).

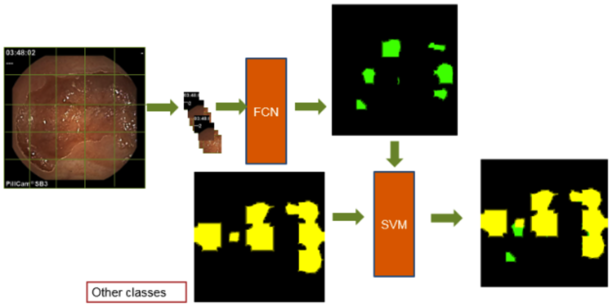

Fig. 1: The steps applied in the proposed patch-based segmentation.

2. METHOD

2.1. Segmentation

Before training, each image undergoes pre-processing stage to be divided into 25 (55) small patches in equal size. As a result, the training samples have width and height sizes varying from 60 to 300 pixels. Those patches with their corresponding masks with zero content are removed from the training to level the influence of background.

For segmentation, the training applies the conventional fully connected neural network [4, 5, 6]built upon Matconvnet[1] that begun with imageNet-vgg-verydeep-16 model. To minimise the influence of overlapping segments, instead of training all the classes collectively, this study trains each segmentation task individually. The final mask for each image is then the integration of five individual segmentation masks after fine tuning using SVM. In other words, the last layer of features from each model are collected first. Then SVM classifier is applied to fine tune each segmentation class to further differentiate each class. Figure 1 illustrates the proposed approach. Firstly, each of five classes is trained on patches independently to take into account of overlapping classes. Then upon connection layer of all five classes, SVM classifier is trained to highlight the distinctions between each class. This classifier will perform the final segmentation for each

[1] https://github.com/vlfeat/matconvnet-fcn

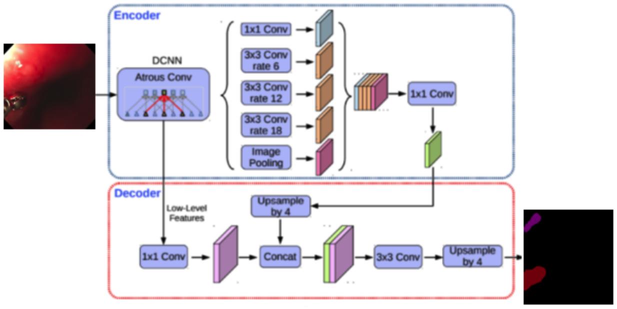

Fig. 2: Segmentation applying deepLabV3 model.

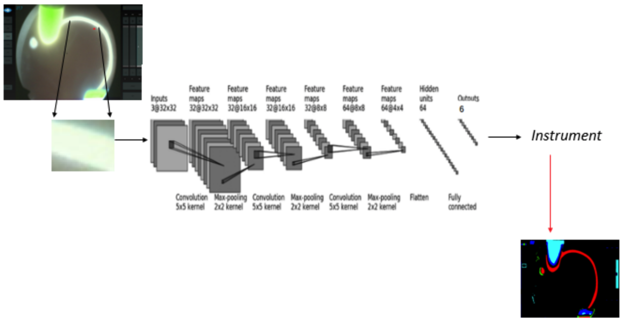

Fig. 3: Caffe classification model while applying 32 × 32 patches.

of five categories, i.e. instrument, specularity, artefact, bubbles, and saturation. In addition, two other popular models are evaluated, which are deepLab [7] and patch-based pixel labeling [8] which is to label every pixel based on centered patch classification result. Table 1 presents the outcome from EAD2019 leaderboard [2] after uploading each result obtained from different deep learning models where our patch-based FCN delivers the best F2 and semantic scores.

Figure 2 demonstrates the steps taken while applying deepLab V3 using tensorflow model [9]. Similarly, Figure 3 represents the procedures while utilizing the patch-based classification model of Caffe. The patch size is selected to be 32x32.

Model	F2-score	Semantic score
Patch-based labeling	0.2300	0.2155
deepLab	0.1638	0.1872
Patch-based FCN	0.2354	0.2434

Table 1: Competition results obtained from EAD2019 after uploading the results.

Model	$\mathbf{IoU_d}$	$\mathbf{mAP_d}$	Overlap	$\mathbf{score_d}$
Fast-RCNN-nas	0.3164	0.2425	0.2107	0.2720
YOLOv3 ($trs = 0.1$)	0.2273	0.1750	0.2331	0.1959
YOLOv3 ($trs = 0.25$)	0.2687	0.1668	0.2331	0.2075
Fast-RCNN-resNet101 ($trs = 0.3$)	**0.3482**	**0.2416**	**0.1638**	**0.2842**

Table 2: Detection results obtained from the leaderboard of EAD2019 for each tested model (trs=threshold).

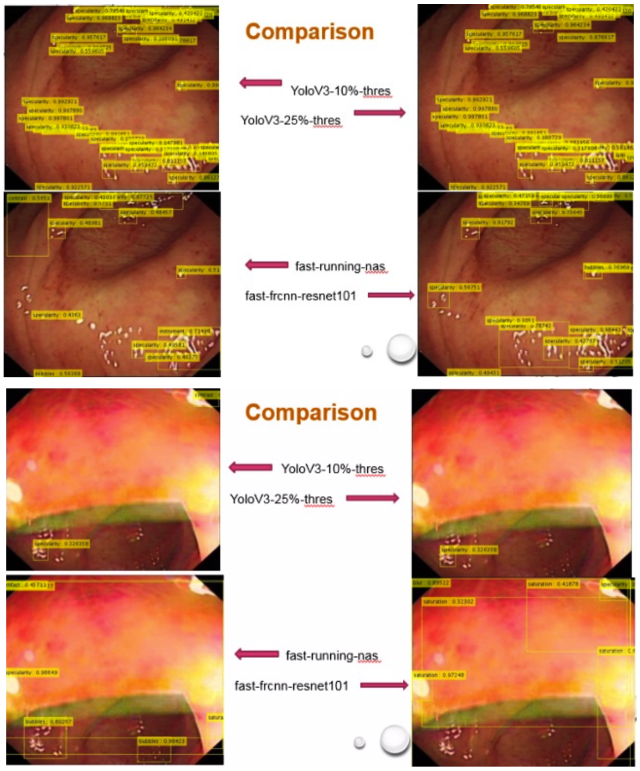

Fig. 4: The comparison results for the four models of fast-rcnn-nas, YOLOV3 (threshold=0.1), YOLOV3(threshold=0.25), and fast-rcnn-resnet101 (threshold=0.3).

2.2. Detection of artefact

While the above patch-based segmentation model appears to perform well for segmentation, when it comes to detection of bounding boxes of intended segments, for some unknown reasons, the detected value of IoU$_d$ appears to be NULL. Hence a number of existing state of the art models are evaluated due to time constraint, comprising fast-rcnn-nas[3] and fast-rcnn-resnet101 [5] using tensorflowand YOLOV3 [3] using darknet [3]. Table 2 presents the evaluation results of the above models. The fast-rcnn-resnet101 model with the threshold of

[2] https://ead2019.grand-challenge.org/

[3] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md

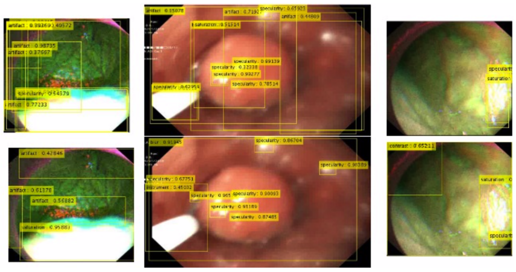

Fig. 5: The comparison results of generalization task (task 3) using two models: fast-rcnn-nas (top) and fast-rcnn-resnet101 (bottom).

0.3 appear to perform the best, which is the one given on the leaderboard of EAD2019 with a rank of 12.

3. RESULTS

Table 2 presents the evaluation results of the above models. The fast-rcnn-resnet101 model with the threshold of 0.3 appear to perform the best, which is the one given on the leaderboard of EAD2019 with a rank of 12. Figuratively, Figure 4 demonstrates the comparison results between the above four models for 2 images. Figure 5 compares the generation results (Task #3) between models Fast-rcnn-nas (top) and fast-rcnn-resnet101 (bottom).

4. CONCLUSION AND DISCUSSION

It has been a very enjoyable experience while taking part in this EAD2019 competition. Due to the late participation (two weeks before the initial deadline), implementation of several ideas could not be fully completed. However, the final position of 12 is better than expected, which is quite uplifting. After initial evaluation of existing models (both in-house and in the public domains), it is found that, no model performs significantly better than the other. Semi-supervised approach will be recommended coupled with clinical knowledge.

Contribution includes patch-based training. While several existing models incorporate regions of interest for training, some regions appear to be overwhelmingly larger than the intended targets ($> 95\%$), hence introducing too much background information, leading to the sampling distribution substantially unbalanced. Because of the varying size of training datasets, from 300 to 1400 pixels along both width and height directions, fixed patch size may instigate under or over sampling. Hence in this study for segmentation (task #2), each image is divided into 25 equal sized patches non-overlapping, which appears to give good segmentation results. However, it is foreseen that sampling with overlapping regions collectively might deliver even better results, which will be investigated in the future. Figure 6 depicts the learning information of whole-image-based (top) and patch-based segmentation as well as whole-image-based detection (bottom).

Regarding to the detection tasks utilising existing models, the challenge here is to find the right threshold for the last fully connected layer of probability. Higher thresholds might miss some intended regions. However, lower thresholds tend to not only over segment but also repeat some regions a number of times. For example, to delineate one single contrast region using YOLOv3 [3] model from one test image, lower threshold (0.4) delivers to three bounding boxes, with each bigger one surrounding smaller one as illustrated in Figure 7. In summary, for medical images, medical knowledge needs to be incorporated in order to generate more accurate results.

5. REFERENCES

[1] Sharib Ali, Felix Zhou, Christian Daul, Barbara Braden, Adam Bailey, Stefano Realdon, James East, Georges Wagnires, Victor Loschenov, Enrico Grisan, Walter Blondel, and Jens Rittscher, "Endoscopy artifact detection (EAD 2019) challenge dataset," *CoRR*, vol. abs/1905.03209, 2019.

[2] Sharib Ali, Felix Zhou, Adam Bailey, Barbara Braden, James East, Xin Lu, and Jens Rittscher, "A deep learning framework for quality assessment and restoration in video endoscopy," *CoRR*, vol. abs/1904.07073, 2019.

[3] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *CoRR*, vol. abs/1804.02767, 2018.

[4] Evan Shelhamer, Jonathan Long, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[5] 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*, 2015, pp. 91–99.

[6] 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*, 2016, pp. 770–778.

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.

[8] Andrew Janowczyk, Scott Doyle, Hannah Gilmore, and Anant Madabhushi, "A resolution adaptive deep hierarchical (radhical) learning scheme applied to nuclear seg-

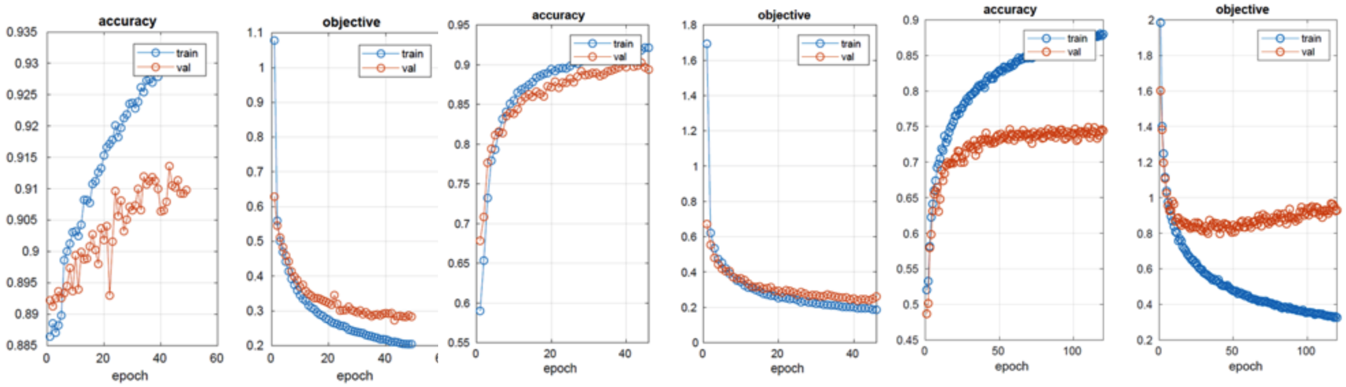

Fig. 6: Learning information for segmentation based on whole image (left), patch (middle) and detection based on whole image (right).

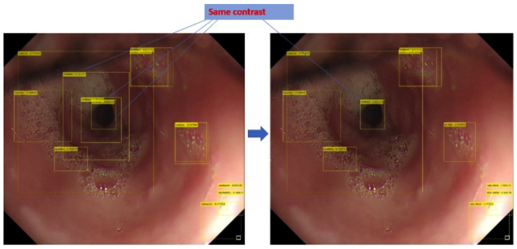

Fig. 7: The impact of thresholding of model of fast-rcnn-nas. Left: threshold= 0.1; right threshold= 0.3.

mentation of digital pathology images," *CMBBE: Imaging & Visualization*, vol. 6, no. 3, pp. 270–276, 2018.

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII*, 2018, pp. 833–851.