MULTI-CLASS ARTEFACT DETECTION IN VIDEO ENDOSCOPY VIA CONVOLUTION NEURAL NETWORKS

Mohammad Azam Khan, Jaegul Choo

Korea University
Department of Computer Science and Engineering
Seoul, South Korea.
{a_khanss, jchoo}@korea.ac.kr

CEUR-WS.org/Vol-2366/EAD2019_paper_9.pdf

ABSTRACT

This paper describes our approach for EAD2019: Multi-class artefact detection in video endoscopy. We optimized focal loss for dense object detection based RetinaNet network pre-trained with the ImageNet dataset and applied several data augmentation and hyperparmeter tuning strategies, obtaining a weighted final score of 0.2880 for multi-class artefact detection task and mean average precision (mAP) score of 0.2187 with deviation 0.0770 for multi-class artefact generalisation task. In addition, we developed a U-Net based convolutional neural networks (CNNs) for multi-class artefact region segmentation task and achieved a final score of 0.4320 for the online test set in the competition.

Index Terms— Endoscopic artefact, Video endoscopy, artefact generalization, Convolutional neural networks

1. INTRODUCTION

Endoscopic Artefact Detection (EAD) [1, 2] is a core challenge in facilitating diagnosis and treatment of diseases in hollow organs. This Challenge highlights the growing application of artificial intelligence (AI) in general, and specific application of deep learning (DL) techniques for the early detection of numerous cancers, therapeutic procedures and minimally invasive surgery. In this concern, the organizers mainly focused on three sub-tasks for this challenge using the EAD dataset [1, 2]: multi-class artefact detection, region segmentation and detection generalization.

2. OUR APPROACH

For multi-class artefact detection and generalisation tasks, our solution is based on keras-retinanet [3] which is basically an implementation of a popular dense object detection method called RetinaNet [4] using open-source framework Keras [5] with Tensorflow[1] back-end. The RetinaNet is a single-stage convolutional neural network detection architecture, which

[1]https://www.tensorflow.org/

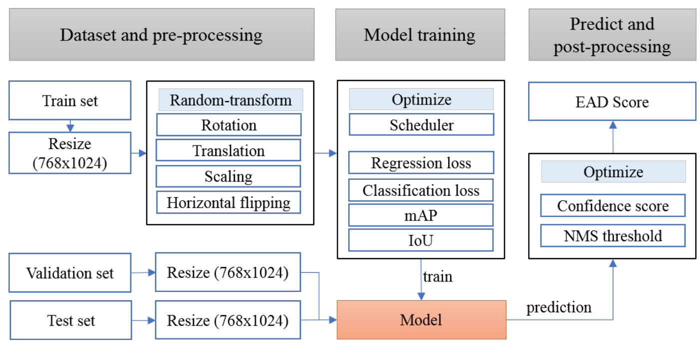

Fig. 1. Overall detection pipeline for multi-class artefact detection and generalisation task.

was really appealing to us for its training simplicity. Overall detection pipeline for two tasks is shown in Figure 1.

2.1. Multi-class artefact detection and generalisation

For multi-class artefact detection task, first of all, we preprocessed the dataset (by resizing the images into 768×1024 pixels), and applied several standard data augmentation techniques including rotation, translation, scaling, and horizontal flipping. We optimized the network with resnet-101 backbone that were pretrained on ImageNet images. Later, we used non-maximum suppression to eliminate some overlapping bounding boxes from predicted bounding boxes as a post-processing step.

In this challenge, the third task was multiclass-artefact generaliation task. Sometimes it is crucial for algorithms to avoid biases induced by specific training dataset. Hence, to be aligned with the organizers' motivation, we tried to optimize the network that we used for artefact detection task above. Our main intuition was to develop more generalized model so that the model can be used across different endoscopic datasets.

Table 1. Segmentation scores.

Dataset	Model	Overlap	F_2	Final
Test (Online)	U-Net	0.4324	0.4310	0.4320

Table 2. Detection results.

Dataset	Backbone	mAP	IoU	Score
Validation	ResNet101	0.4547	0.5167	0.4926
Online	ResNet101	0.2581	0.3330	0.2880

Table 3. Generalisation results.

Dataset	Backbone	mAP	Dev
Test (online)	ResNet101	0.2187	0.0770

2.2. Multi-class artefact region segmentation

The second task of the challenge was multi-class artefact region segmentation. We used an encoder-decoder architecture called U-Net that is designed for biomedical image segmentation [6]. The encoder path identifies the contents of the image while the decoder part localize where the contents are available. More importantly, in a U-Net, the output is an image with the same dimension of the input, but with one channel. Unfortunately, we were not able to make extensive experiments for this task.

3. RESULTS

Model performance of multi-class artefact detection task is shown in Table 2. Table 3 shows the overall performance of multi-class artefact generalization task. As explained in Section 2.2, with the limited experiments, our model performance is shown in Table 1 on the final test set of region segmentation task.

4. DISCUSSION

In the beginning, when we had phase 1 dataset, we tried to develop our model using 3-fold cross validation. Our models relatively worked as well. Later, when dataset 2 had been released, we incorporated these additional data in our models using 5-fold cross validation. However, our model perform a bit worst. After carefully analyzing, we found that the dataset provided in the second phase is more diverse than the first dataset. We were not able to manage this diversity somehow.

Overall, we noticed a significant gap between our local validation score and the leader board score. Then we reviewed the annotation process more carefully. We found that some cases a bit unusual in the training dataset having more than one classes for almost same bounding boxes. It was understandable why some bounding boxes were overlapped for different class artefacts. However, the situation was not the same for all the bounding boxes of the different/same class(s). An example case is shown in figure 2 (overlapping bounding boxes are marked with circle in yellow color).

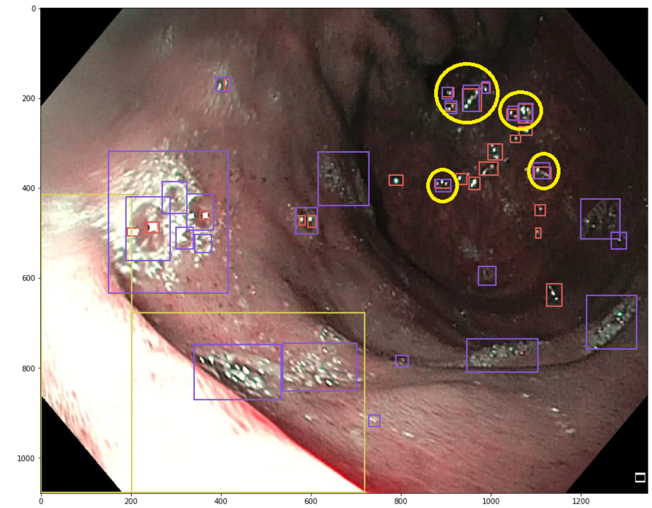

Fig. 2. Sample image having almost same bounding boxes for different classes.

The competition was an exciting and educational experience to solve a problem in real-life settings. We thank the organizers for all their hard work for organizing and annotating the datasets for the competition; large medical image data sets of sufficient size and quality for this purpose are rare.

5. CONCLUSION

Motivated by the no new-net [7], we wanted to demonstrate the effectiveness of well trained state-of-the-art networks in the context of three different tasks of EAD 2019 challenge. While most of the researchers are currently besting each other with minor modification of exiting networks, we instead focused on the training process. The detection of specific artefacts and then precise boundary delineation of detected artefacts, and finally detection generalization of independent of specific data type and source - all would mark critical steps forward for this domain.

6. 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," 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] Hans Gaiser et al., "Fizyr/keras-retinanet: 0.5.0," 2018.

[4] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. oct 2017, IEEE.

[5] François Chollet et al., "Keras," https://github.com/fchollet/keras, 2015, GitHub.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, pp. 234–241. Springer International Publishing, 2015.

[7] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H. Maier-Hein, "No new-net," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 234–244. Springer International Publishing, 2019.