DEEP LAYER AGGREGATION APPROACHES FOR REGION SEGMENTATION OF ENDOSCOPIC IMAGES

Qingtian Ning, Xu Zhao, Jingyi Wang

Department of Automation, Shanghai Jiao Tong University

ABSTRACT

This paper contains our approaches in EAD2019 competition. For multi-class region segmentation (task 2), we utilize deep layer aggregation algorithm to achieve the best results compared to U-net. For the completeness of the competition, we employ the Cascade R-CNN framework to finish multi-class artefact detection (task 1) and multi-class artefact generalization tasks (task 3).

1. INTRODUCTION

In this paper, we will introduce our methods and results of the Endoscopic artefact detection challenge (EAD2019) [1, 2] in detail. The competition consists of three tasks, which are artefact detection (task 1), region segmentation (task 2) and generalization (task 3). For task 1, it aims to get localization of bounding boxes and class labels for 7 artefact classes for given frames. For task 2, Algorithm should obtain the precise boundary delineation of detected artefacts. And for task 3, it aims to verify the detection performance independent of specific data type and source.

2. DETAILS ON OUR METHOD

2.1. Detection and generalisation tasks

2.1.1. Cascade R-CNN

In object detection, we need an intersection over union (IoU) threshold to define positives and negatives. An object detector usually generates noisy detections if it is trained with low IoU threshold, e.g. 0.5. But detection performance degrade with increasing the IoU thresholds [3]. So the Cascade R-CNN is proposed to address two problems: 1) over-fitting during training, due to exponentially vanishing positive samples, and 2) inference-time mismatch between the IoUs for which the detector is optimal and those of the input hypotheses [3]. It consists of a sequence of detectors trained with increasing IoU thresholds, to be sequentially more selective against close false positives. The detectors are trained stage by stage, leveraging the observation that the output of a detector is a good distribution for training the next higher quality detector [3]. So we simply apply the Cascade R-CNN framework and use L1 loss to optimize network.

2.2. Region segmentation

2.2.1. Deep Layer Aggregation

Visual recognition requires rich representations that span levels from low to high, scales from small to large, and resolutions from fine to coarse [4]. Even with the depth of features in a convolutional network, a layer in isolation is not enough: compounding and aggregating these representations improves inference of what and where [4]. Deep layer aggregation (DLA) structures iteratively and hierarchically merge the feature hierarchy to make networks with better accuracy and fewer parameters [4]. For region segmentation, we used DLA-60 model provided. In addition, we use post processing, such as conditional random field [5], to optimize segmentation results. In particular, this is the case that one pixel corresponds to multiple categories in ground truth label. In order to avoid this, we manipulate a simple process to make each pixel correspond to only one classes. To overcome the class imbalance problem, we propose to use a weighted multi-class dice loss as the segmentation loss.

$$L_{Dice} = 1 - 2\sum_{c=1}^{C} \frac{w^c \hat{Y}_n^c Y_n^c}{w^c(\hat{Y}_n^c + Y_n^c)}, \quad (1)$$

where $\hat{Y}_n^c$ denotes the predicted probability belonging to class c (i.e. background, instrument, specularity, artifact, bubbles, saturation), Y_n^c denotes the ground truth probability, and w^c denotes a class dependent weighting factor. Empirically, we set the weights to be 1 for background, 1.5 for instrument, 2.5 for specularity, 2 for artifact, 2.5 for bubbles and 2 for saturation.

3. EXPERIMENTS

3.1. Detection and generalisation tasks

For Detection and generalisation tasks, experiments are built with Caffe framework on a single NVIDIA TITAN X GPU. We use the Adam optimizer with the learning rate $6.25 * 10^{-4}$ and a weight decay of 0.0001 for 250000 iterations with batch

Table 1. Results on EAD2019 detection and generalisation.

Method	Score$_d$	IoU$_d$	mAP$_d$
Cascade R-CNN	0.2330	0.1222	0.3068

Table 2. Score gap for generalisation tasks.

Method	dev$_g$	mAP$_g$
Cascade R-CNN	0.0515	0.3154

Table 3. Results on EAD2019 region segmentation.

Methods	Score$_s$	Overlap	F2-score
DLA-60(crf)	0.5320	0.5206	0.5661
DLA-60	0.4460	0.4352	0.4784

Table 4. Results on our validation set.

Methods	Dice	Jaccard
DLA-60	0.517	0.480
U-net	0.339	0.296

size 1. Table 1 shows the evaluation results on EAD2019 detection and Table 2 shows the score gap for generalisation tasks. What surprises us is that our detection algorithm has good generalization performance.

3.2. Region segmentation

For region segmentation, experiments are built with Pytorch framework on two NVIDIA 1080ti GPUs. We use the SGD optimizer with a weight decay of 0.0001, and adopt the poly learning rate $(1 - \frac{epoch-1}{totalepoch})$ with momentum 0.9 and train the model for 200 epochs with batch size 64. The starting learning rate is 0.01 and the crop size is chosen to be 256. Table 3 shows the evaluation results on EAD2019 region segmentation. Table 4 shows the comparison results of U-net[6] and DLA on our validation set.

4. CONCLUSION

Overall, EAD2019 is a very meaningful competition. We have gained a lot in the process of completing the competition. Finally, we ranked 20th, 11th and 3th for detection, segmentation and generalization, respectively. The final result exceeded our expectations, which is considerably delightful. Of course, we still have a lot of shortcomings. For example, for segmentation tasks, we make each pixel correspond to only one classes, which will lead to some holes in the results. In addition, we can also employ data augmentation, etc. All in all, we still have a lot to improve.

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] Zhaowei Cai and Nuno Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6154–6162.

[4] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell, "Deep layer aggregation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2403–2412.

[5] Philipp Krähenbühl and Vladlen Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.