ENDOSCOPIC ARTEFACT DETECTION IN MMDETECTION

Hongyu Hu[1], *Yuanfan Guo*[2]

[1] Hongyu Hu , Shanghai Jiaotong University, mathewcrespo@sjtu.edu.cn
[2]Yuanfan Guo , Shanghai Jiaotong University, gyfastas@sjtu.edu.cn

1. METHODS

1.1. Architecture

We use Cascade-RCNN [1], which is a multi-stage object detection architecture as our base model and adopt ResNeXt [2] as backbone with Feature Pyramid Networks (FPN) [3] for feature extraction.

1.2. Implement details

- **Mmdetection toolbox** Mmdetection [4] is toolbox for object detection with many state-of-the-art and pre-trained models, which is very practical in this task.
- **Data augmentation** Each image has 50 percent chance to be flipped horizontally.
- **Soft-nms** We use soft-nms [5] rather than nms to avoid objects being directly ignored by mistake. We carry out a series of experiments on soft-nms threshold and maximum number of bounding boxes to better avoid over-detected objects.
- **Multi-scale detection** Test images and training images are of different scales. When training, images are resized randomly from (512, 512) to (1024, 1024). We are able to have a closer look on small objects.

2. RESULTS

We use 4/5 of the data set for training and the rest for evaluation.

2.1. Object detection of different sizes

As baseline result is shown in Table 1, AP^{small} is much smaller than AP^{medium} and AP^{large}. Accurate detection for small object is the bottleneck of this task. After introducing multi-scale detection, performance on small objects improves by **0.008**, as is shown in Table 2. Notably, the boost of AP mainly comes from performance on medium and large objects. We infer that medium and large objects are also zoomed out and the model has better global cognition over the image.

Table 1. Baseline performance on validation data set

AP	$AP^{IoU=.50}$	$AP^{IoU=.75}$	AP^{small}	AP^{medium}	AP^{large}
0.260	0.514	0.228	0.060	0.127	0.323

Table 2. Performance on validation data set with multi-scale detection

AP	$AP^{IoU=.50}$	$AP^{IoU=.75}$	AP^{small}	AP^{medium}	AP^{large}
0.277	0.539	0.250	0.068	0.152	0.335

Table 3. Results on 100% test data set with different parameters

threshold	0.030	0.030	0.050	0.050	0.100	0.100	0.200	0.200
max	100	20	100	20	100	20	100	20
dscore	0.184	0.194	0.189	0.195	0.116	0.215	0.2115	0.2202

Table 4. Final result on 100% test data set

Score_d	dscore	dstd	gmAP	gdev
0.2202±0.0562	0.2202	0.0562	0.1671	0.0879

2.2. Trade-off on bounding box's number

In given training data set and test data set, each image mainly has about few to tens of bounding boxes [6][7][8]. When inference, threshold in soft-nms and maximum number of bounding boxes in each image decide the number of bounding boxes. In Table 3, we list experiment results on this pair of parameters and decide threshold and maximum number set as **0.2** and **20**.

2.3. Final result

We mainly use multi-scale detection and proper parameter settings in soft-nms to solve the problems mentioned above. Final result on 100 % test set is shown in Table 4. This result ranks 8th in final leader board.

3. REFERENCES

[1] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: High quality object detection and instance segmentation. *arXiv preprint arXiv:1906.09756*, 2019.

[2] Saining Xie, Ross Girshick, Piotr Dollr, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *arXiv preprint arXiv:1611.05431*, 2016.

[3] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie. Feature pyramid networks for object detection. *CoRR*, abs/1612.03144, 2016.

[4] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[5] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S. Davis. Soft-nms – improving object detection with one line of code. 2017.

[6] Sharib Ali, Felix Zhou, Barbara Braden, Adam Bailey, Suhui Yang, Guanju Cheng, Pengyi Zhang, Xiaoqiong Li, Maxime Kayser, Roger D. Soberanis-Mukul, Shadi Albarqouni, Xiaokang Wang, Chunqing Wang, Seiryo Watanabe, Ilkay Oksuz, Qingtian Ning, Shufan Yang, Mohammad Azam Khan, Xiaohong W. Gao, Stefano Realdon, Maxim Loshchenov, Julia A. Schnabel, James E. East, Geroges Wagnieres, Victor B. Loschenov, Enrico Grisan, Christian Daul, Walter Blondel, and Jens Rittscher. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. *Scientific Reports*, 10, 2020.

[7] Sharib Ali, Felix Zhou, Christian Daul, Barbara Braden, Adam Bailey, Stefano Realdon, James East, Georges Wagnieres, Victor Loschenov, Enrico Grisan, et al. Endoscopy artifact detection (EAD 2019) challenge dataset. *arXiv preprint arXiv:1905.03209*, 2019.

[8] 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. *arXiv preprint arXiv:1904.07073*, 2019.