MATH 6380P Project 2: Can Object Detectors Generalize?

PANG, Hong Wing [1] and WONG, Yik Ben [2] {hwpang, ybwong}@ust.hk
[1]: Department of Computer Science and Engineering, HKUST [2]: Department of Electronic and Computer Engineering, HKUST

1. Introduction

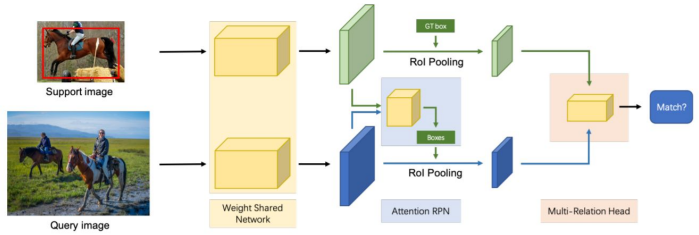
Conventional object detection approaches require large amount of data to reach satisfactory performance but lack of adaptability. Since some of the objects is difficult to obtain sufficient data for the conventional approaches, Few-Shot Object Detection(FSOD) task is proposed to detect the novel class objects with only few labelled data on the novel classes and abundant based class data.

2. Methodology

Three sets of experiments are conducted to evaluate the effectiveness of the FSOD method and apply it on real-life application. We will first recreate the result using the same datasets, then investigate the performance vs number of sample for novel classes. At the end, we will test it on the dataset that completely irrelevant on base class.

Attention-RPN and Multi-Relation Detector

Fan *et al*. [1] proposed a Attention-Based Region Proposal Network which provide the supporting information of the object through the attention mechanism to provide relevant region proposal. Multi -Relation Detector is a combination of global-relation head, local-correlation head and patch-relation head to evaluate the similarity of the input image and support object.



3. Training setting

In the first experiment, we compare two models, one being the weights provided by the authors, trained on 80 classes of **MS COCO**[2]. The other model are trained by us on the **FSOD**[1] dataset also compiled by Fan et. al, which has 800 base classes and another 200 distinct novel classes for evaluation. For the latter set of weights, we trained for 60000 iterations over 21 hours on two GeForce 2080 Ti GPUs.

All experiments are evaluated on the FSOD dataset novel classes. Note that the 200 novel classes cannot be used at once for evaluation due to the large size of the support set images. Thus, we randomly select two sets of 20 classes (i.e. “Set 1” and “Set 2”).

4. Experiments

A. Diversity of base classes

We first compare the detection performance of the object detector when trained on the COCO and FSOD datasets. In all experiments, we follow common procedure in evaluating object detectors by using the average precision (AP) metric at different intersection-over-union (IoU) thresholds.

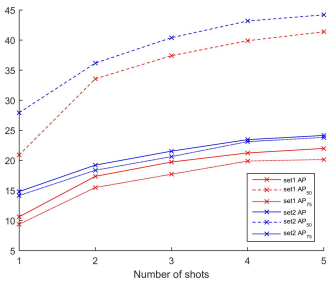
Metric	COCO / Set 1	FSOD / Set 1	COCO / Set 2	FSOD / Set 2
AP*	11.92	21.99	4.61	24.15
AP50	24.74	41.38	9.57	44.20
AP75	10.28	20.15	3.95	23.82

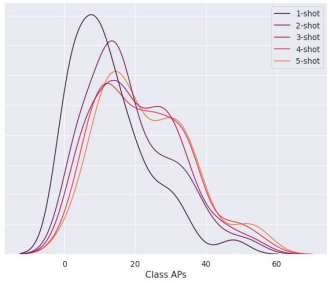
*averaged over AP values from 50% to 95% IoU

Comparing the results we see that the model trained on FSOD has significant performance improvement over COCO. Thus, it is crucial for the detector to be trained on a wide range of classes in order to successfully generalize to unseen classes.

B. No. of support images per class

We evaluate the model at different no. of shots to investigate its impact on the detection accuracy. The improvement in performance starts to slow down at around 4 images. The kernel density plot of all 40 class-wise APs also supports this observation; though it also shows that the variance in performance among classes is large.





C. Face mask detection

To further evaluate the generalization ability of the FSOD network, we detect object instances of face masks using a collection of 30 internet images collected from Kaggle. We perform 5-shot detection with weights trained on the FSOD dataset, using the first 5 images in the collection. Qualitatively, the model achieved the following performance - AP: 20.17, AP50: 37.02, AP75: 18.53. Below are some of the support images chosen.

4. Experiments (cont.)

Some visualizations of the mask detection are shown below. In general, the FSOD network is able to detect masks with a variety of colors, with accurate localization ability. Below are some examples of good detection results. Note that the black / grey masks in the third image can be detected even when one instance of black mask is provided in the support set.

There are quite a number of failed detections with some examples shown below. Typically, we find that if the mask in the image is too small, the detector will fail. Even though the support image include object instance with smaller sizes, the backbone detection network (i.e. Faster-RCNN) might be unable to capture features from smaller objects or the base class dataset did not include sufficient small size objects. Additionally, the detector is also unable to detect masks with similar color to the human skin.

5. Conclusion

The generalizability of the FSOD is still need to be researched to reach a robust level of object detection, but the result is sufficient for dataset construction without extensive human resource to annotate all the objects. Then, the dataset can be used for conventional object detection algorithm to improve their performance.

6. References and contribution

[1] Fan Qi, Zhuo Wei, Tang Chi-Keung and Tai Yu-Wing. Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector. in CVPR, 2020
[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In ECCV, 2014

PANG, Hong Wing: Network training, Result evaluation, Poster
WONG, Yik Ben: Data annotation, Poster, Video