Few Shot Object Detection Using Fine-tuning

Jean-Bernard Uwineza
University of California, Riverside
Riverside, CA 92501
buwineza@ee.ucr.edu

Chetan Reddy Mudireddy
University of California, Riverside
Riverside, CA 92501
cmudi001@ucr.edu

Om Shankar Ohdar
University of California, Riverside
Riverside, CA 92501
oohda001@ucr.edu

Sayak Nag
University of California,Riverside
Riverside, CA 92501
snag005@ucr.edu

Vikarn Bhakri
University of California, Riverside
Riverside, CA 92501
vbhak001@ucr.edu

Abstract

Abstract goes here

0.1 Some rules for collaboration

- DO NOT delete text you did not write, unless the person who wrote it asks you. If you have issues with some of the text, discuss with the author first.
- Commit and push to GitHub often. This helps others staying updated.
- Apply other common sense rules of collaboration.

1 Introduction

This text is from the proposal. It will be changed.
Few-shot learning has received significant interest in the past few years, but mainly for the tasks of classification and rarely for object detection. In computer vision, the task of object detection is more challenging since the detector not only has to perform recognition of the different kinds of objects present, it also has to localize them. This is already a challenging task that relies heavily on the availability of massive amounts of labeled training data. Now when a new data-point is obtained belonging to a novel category, adapting the model becomes a very difficult task especially when the new category contains a few samples. Recently, meta learning techniques have been proposed for adapting deep models to novel categories. However, they are not easily extendable to the task of object detection. Take for example the Matching [1] and Prototypical Networks [2], building prototypes of objects is much more difficult than building prototype of the categories. Another approach that is being explored by researchers is to provide ways to fine-tune the detection layers of deep models to adapt to the new categories [3].

In this project we aim to do a comparative study of meta-learning and fine-tuning approaches towards object detection. We aim to experiment on benchmark datasets such as COCO [4] and PASCAL [5] and also extend these approaches towards 3D object detection with the KITTI dataset [6].

Preprint. Under review.

In addition, we plan on examining how many shots are necessary to reach comparable accuracy relative to conventional detection approaches. To this end, we will attempt to develop a metric that assesses the model's knowledge, and requests additional labeled examples if it has not reached a certain accuracy. This will allow us to further compare the performance of the two approaches.

2 Other Sections

3 Experimental Results

4 Conclusion

References

[1] O. Vinyals, C. Blundell, T. P. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *CoRR*, vol. abs/1606.04080, 2016. [Online]. Available: http://arxiv.org/abs/1606.04080

[2] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in neural information processing systems*, 2017, pp. 4077–4087.

[3] X. Wang, T. E. Huang, T. Darrell, J. E. Gonzalez, and F. Yu, "Frustratingly Simple Few-Shot Object Detection," *arXiv preprint arXiv:2003.06957*, 2020. [Online]. Available: http://arxiv.org/abs/2003.06957

[4] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: http://arxiv.org/abs/1405.0312

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[6] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *International Journal of Robotics Research (IJRR)*, 2013.