Few Shot Learning for Object Detection: Comparison of Fine-tuning and Meta-learning Approaches

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

1 Problem Statement

A small child visiting a zoo with parents for the first time recognizes a strange animal—a zebra, she is told. She keeps walking and a few minutes later, by the infirmary, she recognizes another animal, only smaller and wobbly this time. "A baby zebra", she exclaims. From just this one example, she will be able to recognize just about every zebra she will ever see. Not only this, she will also be able to make remarkable connections to other animal that are are similar to zebras [1].

This is an ability machines have yet to acquire. Although machines have surpassed humans in visually recognizing objects, they still lack the ability to do so from a few examples. Recently,there have been promising advances towards the goal of making machines generalize from a few examples via a deep learning method called *few-shot learning*. There are many important applications that could benefit from the ability to learn from a few samples. Like any other problem, there are various approaches to this this task. In this project, we propose to evaluate and analyze two of the most promising approaches.

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 [2] and Prototypical Networks [3], 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 [4].

Preprint. Under review.

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 [5] and PASCAL [6] and also extend these approaches towards 3D object detection with the KITTI dataset [7].

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 Work Plan

We plan to work on the meta-learning and fine-tuning based approaches simultaneously. Sayak and Bernard will focus on meta-learning-based approaches, while Om, Chetan and Vikarn will work on fine-tuning approaches. The experimental results will be compiled by Om and Vikarn while Sayak, Bernard and Chetan will work on final report compilation.

References

[1] L. K. Samuelson and L. B. Smith, "They call it like they see it: Spontaneous naming and attention to shape," *Developmental Science*, vol. 8, no. 2, pp. 182–198, 2005.

[2] 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

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

[4] 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

[5] 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

[6] 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.

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