Towards a Generic Object Detection Algorithm

Ashvin Pidaparti

Advisor: Prof. Junaed Sattar

Summary

In the field of Autonomous Robotics, object detection is a key element for a robot to interact with the world. A robot must be aware of its surroundings in order to take the appropriate action for those surroundings. Through extensive research, algorithms to detect objects and determine their location in images have been developed, however, these algorithms require a large amount of high quality images of the objects. These images must be annotated, requiring a human to explicitly state where an object is in an image. For several objects, images are not available, and there are simply too many objects in our world to effectively detect all of them.

I put forward a method of detecting objects without a significant number of images, dubbed Zero Shot Detection [2]. This method functions by creating binary feature extractors and using the output of several feature extractors to create a vector representation of an object. This vector representation is then matched against a list of several objects that were not in the original dataset and their expected output from those feature extractors to find the closest match. This match is the classification of that object.

At a high level, feature extractors can be as simple as whether or not an object is made of biological material, or whether or not an object can tangle the propeller of an underwater robot. The Faster R-CNN neural network [1] was attempted to train to locate objects in an image and classify them as biological or non-biological, able to tangle or not able to tangle, and metal or nonmetal. These models would be applied to images and their output was vectorized. The vector

representations of objects in the image would be matched against hard-coded representations of known but unseen objects via the K-Nearest Neighbor algorithm. The closest neighbor in feature space will be treated as the classification.

Discussion

Despite the poor results, I believe this work was successful. A long term goal in the field of Robotics and Vision is to require less image data for training neural networks, and my work has created documentation for this lab to continue this work. I found tremendous difficulty in training the algorithm. The high performance computing resources at the Minnesota Supercomputing Institute were used because of the availability of Graphics Processing Units, which allow for simple calculations to be carried out very quickly. However, the GPUs at the Minnesota Supercomputing Institute are not supported by PyTorch [3], the deep learning library I was using. At the point in the project that I began training the algorithm, there was not enough time to refocus my work to use another library. As a result, the training process took several hours, even days, for a single epoch to complete for a single feature extractor, as well as the evaluation of the model not being able to complete. This imposed significant limitations on the accuracy of the feature extractors, in turn, resulting in low accuracy in object detection.

If I were to start this project over with this knowledge, I would have applied an entirely different approach. I would have used the You Only Look Once (YOLO) algorithm [4] for feature extraction due to its improved accuracy, higher frame rate support in video object detection, and faster training time, as well as the support for older GPUs.

Despite the technical portion of this project not functioning, I left Professor Sattar's lab documentation of this project and I will be continuing my work in the fall. Moving forward, they will be able to continue this project and avoid the road blocks I faced. In this sense, my work has laid groundwork for this lab to detect objects without significant data on them.

Evaluation

My experience in undergraduate research was positive. There were significant resources to assist me in my work. I certainly have a better understanding of the research process as well as a greater confidence in my ability to ask the right questions for how a project should proceed. When this project started, I had some theoretical knowledge of Deep Learning and minimal knowledge of the practical applications, but as a result of this work, I am better equipped for future projects.

This project has prepared me for a career in Computer Vision and Robotics. Specifically, this upcoming summer, I will be working heavily with Deep Learning to analyze video and detect objects. Working through the obstacles of this project has equipped me to approach my work this summer with more clarity and confidence. Moreover, I will be returning in the Fall to continue this work in the Interactive Robotics and Vision Lab.

References

1. Ren, Shaoqing, et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017, pp. 1137–1149., https://doi.org/10.1109/tpami.2016.2577031.
2. C. H. Lampert, H. Nickisch, and S. Harmeling. "Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer". In CVPR, 2009 (pdf)
3. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In

Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc.

4. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.