Summary for Multi-stage Reinforcement Learning for Object Detection

Siddharth Nayak

1 Introduction

The paper focuses on using reinforcement learning to detect objects within an image. The agent zooms and changes the aspect ratio of the image to tightly fit the bounding box onto the object. The authors use three different reward metrics to increase the number of object-detection in the image. This is motivated by the way humans extract visual information sequentially. Generally in computer vision, brute force approaches like sliding window and region proposal methods are used. This paper leverages the recent successes of reinforcement learning in sequential decision-making.

2 Model

The authors introduce two models: 1-stage model and a 2-stage model which is an extension of the 1-stage model and can fit much better bounding boxes than the 1-stage model. The state for the agent is a concatenation of an encoded feature representation of the image from a VGG network and the history vector of the last four actions of the agent. The agent can choose to either terminate the process or choose one of the five zoom actions: top-left, top-right, bottom-left, bottom-right and center where each zoom action shrinks the bounding box to 75% of its width and height. So by iteratively zooming the agent can zoom until it finally fits the bounding box. In the 2-stage model, the agent can choose to refine the bounding box by shifting it up, down, right or left by 10% of its width and height for five successive steps. The authors come up with 3 different schemes for the rewards. They use the common IOU, Ground Truth Coverage(GTC) and Centre Deviation(CD) to give the reward. GTC is the percentage of the ground truth that is covered by the current bounding box. CD is the distance between the centre of the bounding box and the ground truth. The quality function is defined as:

$$r(b, g, d) = \alpha_1 IOU(b, g) + \alpha_2 GTC(b, g) + \alpha_3(1 - CD(b, g, d))$$

where b is the bounding box, g is the ground truth and d is the image diagonal. Therefore the reward for zooming is $R_z(b^{'}, b, g, d) = sign(r(b^{'}, g, d) - r(b, g, d))$ where $b^{'}$ is the bounding box after zooming and b is before zooming. The terminal reward is given as:

$$R_t(b, g) = \begin{cases} +\eta & \text{if iou(b,g)} \geq \tau \\ -\eta & otherwise \end{cases}$$

They also try a sigmoid version for each of the terms in the zoom-reward. The refinement reward is defined as:

$$R_m(b, b^{'}, g, d) = \begin{cases} sign(cd(b^{'}, g, d) - cd(b, g, d)) & \text{if } cd(b^{'}, g, d) \neq cd(b, g, d) \\ -1 & \text{if } cd(b^{'}, g, d) = cd(b, g, d) \text{ while CD-decreasing refinements possible} \\ +1 & \text{if } cd(b^{'}, g, d) = cd(b, g, d) \text{ while no CD-decreasing refinements possible} \end{cases}$$

In the 1-Stage setting the reward is getting just after the action is taken whereas in the 2-Stage setting the reward is given after all the refinements are done. A Deep Q-Network is used for training the agent.

3 Experiments

The authors perform the experiments on images of 224 x 224 pixels size. The authors have given all the parameters/hyperparameters they have used to obtain the results. They evaluate their performance on both the 1-Stage and the 2-Stage model. The 2-Stage model performs much better than the 1-Stage model. Although they have mentioned the performance of their model, they have not compared their object detection performance with other existing state-of-the-art models.

4 Possible Improvements

Although the F1 score does increase with the use of this model, the reward function is quite complicated and has been manually engineered for the specific purpose of zooming and bounding box refinements. The refinement reward depends on the future actions too. Also, the refinement reward can saturate at a non-optimal value as the reward is +1 if there are no further CD decreasing refinements possible. This CD decreasing action is not necessarily a known parameter and thus we may not necessarily be rewarding the most optimal action. Another possible problem could be that for multiple objects in the image we need to pass the image multiple times through the network which may slow down the training time as well as evaluation time. Also, the network may not be able to classify between different objects like an airplane and a rabbit as it was just trained to detect objects with no information about the class of the object. Thus, it would require a very large network and a large dataset to generalize well.

5 Conclusion

Overall the paper seems to be a good work to improve the object detection performance of a given neural network. The reward system seems a bit tedious to come up with. The paper is well-written and easy to understand.

6 References

Konig, Malberg, Martens, Niehaus, Grimberghe, Ramaswamy. Multi-stage Reinforcement Learning for Object Detection. Computer Vision Conference, 2019