**Machine Learning:** Cheating Tool of FPS Game Based on Object Detection

*Wenjun Zhang*

*Department of Electrical and Computer Engineering*

*Boston University*

*Boston, MA*

[*wjz@bu.edu*](mailto:wjz@bu.edu)

***Abstract­ – along with the time changing, technology of machine learning also obtain a great advancement. Such as like Computer Vision and VQA (Visual Question Answering), the innovative technologies are making AI (Artificial Intelligence) even smarter. Increasingly high accuracy models’ algorithms come out to apply into object detection field. Majority of the popular object detection algorithms are mature enough to help people in real life. The goal of this project is to review those popular training models, and to find the best model to make a cheating tool of FPS (First Person Shooter) game.***

***Index Terms --- machine learning, visual question answering, computer vision, artificial intelligence, YOLO, first person shooter.***

1. Introduction

What is Object Detection? Object detection is a new computer technology to detect objects in real world based on some high accuracies trained models. If there are dog, bicycle, car, trees, roads, or something else in a single picture, a high accuracy well trained object detection model should discriminate those objects. Sometimes, a light (small) object detection model will not be trained to detect all objects. For example, if one trained object detection model is responsible to detect car, dog, bicycle, then in last example, if it has successfully detected these three objects, it is a well-trained object detection model, Figure 1. Nowadays, human need smart AIs and high accuracy trained models work together to improve their daily life. Along with the deep learning developments, those cutting-edge fields will be increasingly important, such as like autopilot, security camera, military strike, photo repairment, delivery robot, etc. To realize this technology is not easy, because machine only learn 0 and 1. To make machine “see” (capable to detect object), first step is to figure out how human eyes “see”. This will be related to the topic, computer vision.

Human can see things and figure out what is that things mean to them, even though, not everyone process a seen object within same features of that object. If people want to let machine also capable to detect and process things, the best way is not making them act smarter (deploy more machine cores), is to make them act like a real human being. When human look at an object, they are not simply put whole image into brain, they are also label features of the object into their mind, the image of input will be split into data, and those data will be process in human brain. There is also an important part called memory. Because human already known somethings, so that when they see somethings, they can figure out which one corresponds to the memory in their mind.

狗站在自行车后面

描述已自动生成 图示, 示意图

描述已自动生成

*Figure2. Source:* [*1311.2524v5.pdf (arxiv.org)*](https://arxiv.org/pdf/1311.2524v5.pdf)

*Figure1. Source:*[*Blog by Matthijs Hollemans*](https://machinethink.net/blog/object-detection-with-yolo/)

1. R-CNN, Fast-R-CNN, Faster-R-CNN
2. ***R-CNN***

R-CNN (Convolutional Neural Network with Regional Proposal), this well-known object detection model was once published in 2014 by Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik [1]. It is a typically two-stage object detection which is using Convolutional Neural Network as its fundamental algorithm. To mention two-stages, it does not mean that it only has two stages. The difference of one stage object detection and two stage (multiple stages) detection is two stages object detection will have a RoI (Region of Interest) pooling layer. In the figure 2, this is how R-CNN works basically. It will first select multiple proposal regions, such as like 1000 -2000 regions, and it will label their bounding and features. Later, this proposal region will all transform to the size that the pre-trained conventional neural network needed and get extracted by forward propagation [1].

1. **Fast R-CNN**

The biggest problem of R-CNN is that every picture in the training stage need to take a large amount of time to train. Just imagine this, every picture will have up to thousands of proposal regions, and each region will take thousands of forward propagations to access object detection. This is an impossible mission. The insufficiency of R-CNN is that every forward propagation is individual. In the next year of R-CNN, Fast R-CNN came out to solve this insufficiency. It has higher detection quality (mAP) than R-CNN.

It is trained in single stage (the model is in two stage detection), using multi-task loss, its training can update all network layers, and no disk storage is needed for feature caching [2]. Fast-R-CNN will first extract and later do the selective search. There is a RoI (Region of Interest) pooling layer between CNN and full connect layer, and that’s also where selective search taking place.

1. **Faster-R-CNN**

Even though Fast-R-CNN solves many problems of R-CNN, it will ultimately generate a tremendous amount of proposal regions to keep its high accuracy target detection result. Faster R-CNN proposes that it will use selective search to replace region proposal network to reduce the amount of proposal region generations but keeping its accuracy. To realize it, Faster R-CNN will have its own convolutional layer to process data after CNN. To generate the proposal region, it will have a slider, it slides a small network over the convolutional feature map, this small network will be input as a N x N spatial network [3]. After NMS (Non-Maximum Suppression), the similar predictions are eliminated, and then it will go to the pooling. The features after pooling will be eventually input into two full connect layers, one is box-regression layer and box-classification layer.

1. YOLO: A One-Stage Object Detection Model

As we know, FPS (First Person Shooter) game is only about few seconds of frames. The fight begins when players see their enemies, the fight finishes until either side down. So, the cheating tool must be acting extremely fast. Obviously, faster R-CNN is not the best model to 在笼子里的狗

描述已自动生成

*Figure3.Source: http://xxx.itp.ac.cn/pdf/1506.02640v5*

apply into this project. We need a real-time, fast, quick response object detection model to help us build this cheating tool. In 2016, we have YOLO (You Only Look Once), a one-stage object detection algorithm that is not using R-CNN (Region-based) way to fitter information from the image. After two years updating, this could finally be realized until 2018, YOLOV3. Version 3 of YOLO (Even though, it’s on V5 now, V3 is the FPS cheating tool topic first came out) obtains both accuracy and the speed, it wins the role of light development of small data set tool.

The stagey of YOLO is to cut grids from image. It divides it into s^2 same high and width grids. Different from sliding strategy, it needs each grid responsible to detect the center point of the object. Each grid will predict B amounts of bounding boxes, each bounding box will have five variables, coordinate (x, y), high(h), width(w), and confidence (Pr (Object)). If we have S^2 grids, every grid has B amounts of bounding boxes, the classifier can recognize C different objects, then the whole ground truth will be S \* S \* (B \* 5 + C). In the Figure 3, there are thick lines, and thin lines, thicker means more confident that there is something fit to objects which user looking for.

The formula of Confidence is C= Pr(obj) \* IOUtruthpred. IOU is to predict bounding box real object position rate, IOU = B1 AND B2 / B1 UNION B2. Pr(obj) is a probability that it has object in current grid, when it has object, the ground truth will be 1, if not then the ground truth will be 0 [4]. The reason that this formula using not only Pr(object) is because even Pr(object) is enough to represent, but multiplied by IOU can make those unnecessary grid disappear in our result, because bounding box using center point coordinates, height, and width to present, each prediction will be required its center point inside the grid, if those points are not in the grid, its IOU definitely will be small, thus, the confidence will be small too. And later, YOLO will use NMS to drop unnecessary box and assemble same genre box.

1. Cheating Tool Based On YOLO

For now, we finally figure out the best object detection model for our cheating tool. To deploy this cheating tool, we first need to make a similar human body recognition model out. This cheating tool will be divided into two parts. The main part is the human recognition. We do not need to detect all objects in a single frame. We just need to make the machine learn what is human body, and when the objects are moving the box need to stick with them. Except that, we need to teach the machine to know what the differences human body are. The hit box of FPS gaming basically will be divided into limbs, stomach, chest, and head.

The second part will be even more challenge, we need a tool that link the host monitor to the well-trained machine, and a script that can apply auto aiming signal from trained model.

The game we are going to choose will be CSGO (Counter Strike: Global Offensive), based on its popularity and high-level anti-cheating defense, CSGO is the best target to deploy this project.

1. Conclusion
2. **Project review**

The most challenge of FPS gaming is about relocating enemy, because in about 60% of the time, you will not see the enemy, and once you see the enemy, you need to move aim quickly on their body, better on their head (In FPS gaming, this is called Location).

The greatest skill in this game is one shot enemy if you have a good aiming. This technique needs players to spend a great amount of time. Except that, the upper bound of a player is decided when they are born, sometimes, train hard cannot narrow the gap between players. Except that most players will have a time called “Zone”, it’s like when you are playing very well, your reactive nerve and muscle impulse are just in the right time, you can headshot any enemy you want. However, nobody can hold their zone or skill forever, except robot. And that’s purpose of this project, is to make an aim bot based on the new object detection technology and see what we can do to this cheating tool. Furthermore, hope we can find a solution to detect what’s the differences between traditional cheating tool and object detection cheating tool, and try to do the anti-cheat to the new cheating tool.

1. **Outlook**

In these few years, AI is becoming increasingly mature, AI in few years ago are to be consider as a joke. What about now? Jason Allen’s A.I.-generated work, “Théâtre D’opéra Spatial,” took first place in the digital category at the Colorado State Fair. With deep learning technology AI become smarter.

Those accumulated data from Big Data have their path to go to help AI become stronger. Object detection gives AI eyes to see things, mechanic and materials field are going to make AI a new body, I believe that in the future we will have an extraordinarily strong robot, an omni-function robot that can help people finish most things.

1. Reference

[1] *Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation.* Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580–587).

[2] *Girshick, R. (2015). Fast r-cnn.*Proceedings of the IEEE international conference on computer vision*(pp. 1440–1448)*

[3] *Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks.*Advances in neural information processing systems*(pp. 91–99).*

[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.*