**Machine Learning Report**

**Lingtao Jiang**

Machine learning has seen great development since 2004 because of several factors. One of them is that the application areas grow so rapidly. Different kinds of technologies, such as spam filters, search engines and intrusion detection systems are all developed during the recent years. These technologies have gradually become a regular part of human life.

As we all known, machine learning consists of a lot of branches. My interest of machine learning is image recognition.

The image recognition market is estimated to grow from USD 15.95 Billion in 2016 to USD 38.92 Billion by 2021. Companies in different fields such as gaming, e-commence and health care are used to adopt this technology to their regular plan. From my point of view, image recognition refers to an identification of a specific object such as human, animals, vehicles and other variable things from images or videos.

I’m trying my best to understand the working principle of image recognition. Images are actually a kind of data in the form of two-dimensional matrices. In this way, image recognition is a process of classifying data. The key step of image recognition is to gather the data and then create a module to predict and recognize the images.

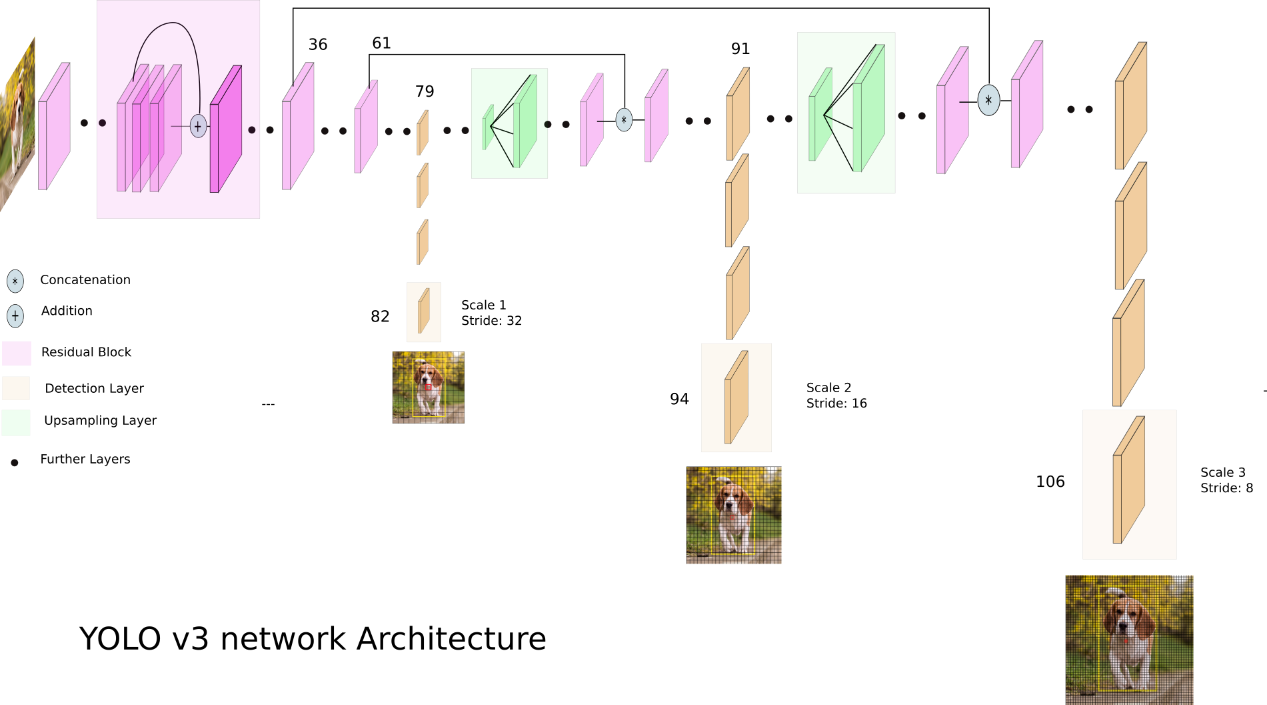
Image recognition tries to mimic the process that human eyes receive the image. As we all know, our eyes perceive an image as a signal and then the signal would be processed by the visual cortex in our brain. This process can make a vivid scene associated with objects and concepts recorded in our memory. Computer perceives an image as either a raster or a vector image. Raster images are a sequence of pixels with discrete numerical values for colors while vector images are a set of color-annotated polygons.[1]

In order to analyze the images, the encoding is transformed into constructs to depict the physical features of the objects. After then computers will analyze these constructions automatically by classification and feature extraction. Image classification consists of several steps and the first step is to simplify the image by extracting the key features of the objects and leave out the rest. During this process, images will eventually transform into feature vectors for the next processing. In the second step, these feature vectors will be used as input and then output a class label such as human or animals. Before testing the classification algorithm, thousands of images of the label and not the label are shown to the system to train it.

A predictive module is needed in image recognition. It’s necessary to use neural network to build a complete predictive module. Neural networks are a set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns. They interpret sensory data through a kind of machine perception, labeling or clustering raw input. The patterns they recognize are numerical, contained in vectors, into which all real-world data, be it images, sound, text or time series, must be translated.[2]

After simplifying the images and building the predictive module, it’s not that difficult to implement the image recognition any more. Training data will be fed into the module to recognize the image. I have learnt the general process of recognize the images, however the specific algorithm is still very hard to me. And that is also what I want to learn in mini project 2. Try to know more about this technology, I also find some information of completed image detector systems and I will talk about them in the following content.

The image detector algorithm is YOLO, which stands for you only look once. The latest version of YOLO, YOLO v3, has been considered as one of the fastest object detection algorithms. Prior image detection algorithms apply a module to the specific image at multiple scales and locations. Then regions with high scores will be considered detections. However, YOLO use a different method. It applies a single neural network to the image and the neural network divides the image into a grid of 13 by 13 cells. Each cell will take responsibility for five bonding boxes. A bonding box is used to describe the object in it. The network can predict these bonding boxes and probabilities of each region in this way. The following diagram perfectly shows the network architecture of YOLO v3.



*Source:*<https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b>

It can be seen from the diagram that YOLO uses only convolutional layers and combine these layers into a fully convolutional network. (FCN) Yolo has dozens of convolutional layers with up sampling layers and skip connections. Typically, the features learned by the layers are passed onto a classifier to make the detection prediction. In YOLO, the one by one convolutions are used to make the predictions. What we need to know is that the output of YOLO is a feature map. The prediction map will be the same size as the feature map because of the 1 x 1 convolutions. Interpreting the prediction map in YOLO v3 is to predict a fixed number of bonding boxes by each cell.

Mathematically, there are (B x (5+C)) entries in the feature map, where B represent the number of bonding boxes predicted by each cell. These bonding boxes are used to detect specific objects and each of them has (5+C) attributes. which describe the center coordinates, the dimensions, the object score and *C* class confidences for each bounding box. Each of the bonding boxes will be responsible for detect the object in it.

The next processing is a little bit complicated and I’m not going to describe it in this paper, but I must admit that I was attracted by a series of operations to recognize the image and it’s really fantastic. Generally, as a rookie in machine learning, I know a lot of the process and mathematical method of the object detection. However, I’m still weak in creating the code to implement these operations and that is also the goal of my next phase of the study.