MidTerm Report: Bully Images Recognition

Haotian Deng and Qingbo Lai

[1] School of Computing, Clemson University
[2] hdeng@clemson.edu and qingbol@clemson.edu

1 Introduction

Detecting In the mid-term project, we classified the bully pictures and nonbully pictures which are totally 10 categories through the modified VGG16 model[4] , and subdivided into 9 categories in the bully picture. The project achieved a satisfactory accuracy. In the final project we will inherit the success of the mid-term project and implement object detection on the basis of the mid-term project. In the final project, we implement object detection based on Single Shot MultiBox Detector (SSD)[2].

In the report, section 2 talks about our project setting and models, section 3 section 4 illustrate our experiments, section 5 describe the future work and plans of final project, section 6 is a conclusion.

2 Project Description

The project's goal is develop deep learning models to detect images with bullying actionsModel able to identify perpetrators which do bully action and victims. In our whole project design, We first analyzed the well-known object detection model and found that the Single Shot MultiBox Detector(SSD)[2] model is the most suitable for us. we use the VGG16 model for image classification in our mid-term project. We know more about the VGG16 model and SSD is based on a modified VGG- 16 network. In the project, We don't have a ready data set for this projectthe data set provided by the CPSC 8810 class which do not have the information and labels about objects like bully and victim, so preprocessing data and make own dataset is also a important procedure. SSD model combines the regression idea in YOLO [5] and the Anchor mechanism in Faster-RCNN, and uses the multi-scale regional features of each position of the whole graph to perform regression, which not only maintains the fast YOLO characteristics, but also guarantees the window prediction and the Faster-RCNN.

The core of the SSD is to use a convolution kernel on the feature map to predict the category and coordinate offset of a series of Default Bounding Boxes. In order to improve the detection accuracy, SSD predicts on different scales of feature maps. More details about model like the specific accuracy and performance, parameters in Section 3.

2.1 Model Architecture

The idea of SSD network main body design is feature layer extraction, and the frame regression and classification are performed in turn. Because different levels of feature maps can represent different levels of semantic information, low-level feature maps can represent low-level semantic information (with more details), can improve the quality of semantic segmentation, and are suitable for small-scale target learning. High-level feature maps can represent high-level semantic information, smooth segmentation results, and are suitable for in-depth learning of large-scale goals. Therefore, the network proposed by the author of SSD can be suitable for target detection at different scales.

The object detection algorithm is divided into one-stage and two-stage. SSD is a one-stage algorithm, which has many advantages. SSD model combines the regression idea in YOLO [5] and the Anchor mechanism in Faster-RCNN, and uses the multi-scale regional features of each position of the whole graph to perform regression, which not only maintains the fast YOLO characteristics, but also guarantees the window prediction and the Faster-RCNN. SSD model use VGG16 network to get features, On the basis of the VGG model, remove the last fully connected layer, add 6 convolution layers (0, 1, 2, 3, 4, 5) and take two convolution layers Conv4-3 and Conv7 in the VGG network. (Fc7), take 4 (2, 3, 4, 5) in the next 6 convolutional layers. If the resolution of the input image is 300X300, then the size of the 6 selected features is (38X38, 19X19, 10X10, 5X5, 3X3, 1X1), compared to the original image, the reduced steps are (8, 16, 32, 64, 128, 256). We draw fixed-scale frames on different feature maps. Small frames on large resolutions are helpful for detecting small targets. Boxes on small resolutions are useful for detecting large targets. Therefore, it is possible to obtain predicted values of multiple scales.

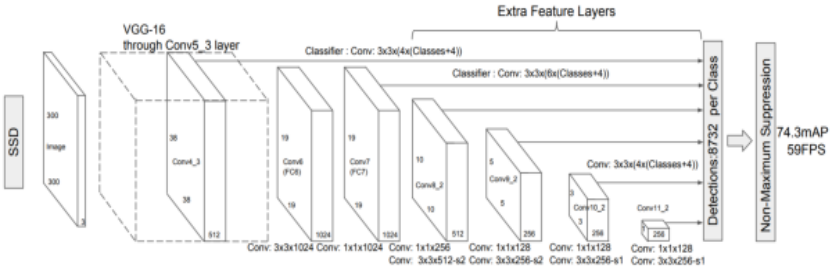

Fig. 1. Single Shot MultiBox Detector(SSD) architecture[2]

So the SSD network is divided into 6 stages, each stage can learn a feature map, and then carry out border regression and classification. The SSD network uses the first five-layer convolutional network of VGG16 as the first stage, and

then converts the two fully connected layers of fc6 and fc7 in VGG16 into two convolutional layers, Conv6 and Conv7, as the second and third stages of the network. . Then on this basis, the SSD network continues to add Conv8, Conv9, Conv10 and Conv11 four-layer networks to extract higher-level semantic information. The network structure of the SSD is shown in Figure 3.1 below. In each stage operation, the network contains multiple convolutional layer operations, each of which is essentially a small convolution.

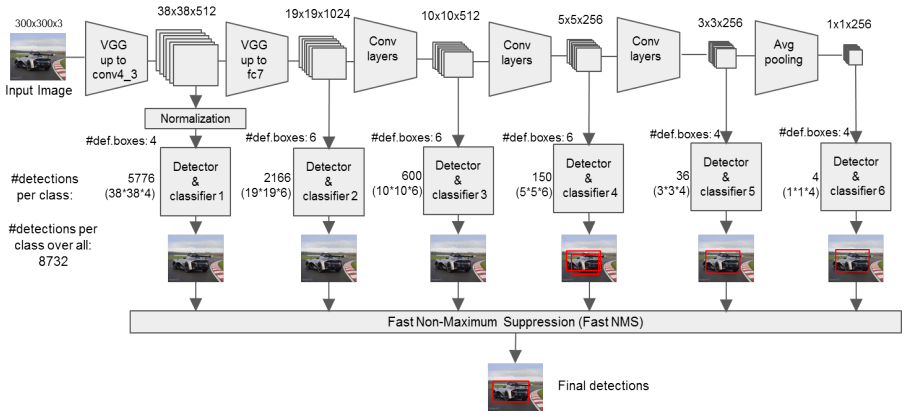

Fig. 2. SSD model architecture[2]

3 Dataset

We use the dataset provided in the classroom as the basis, but for the model training, the dataset in the classroom can not meet the requirements, and further processing, such as labeling the object, we use the "labelImg"[6] tool to mark the file. we can get "xml" files about images which contains information of labels and regions. And then annotations' formula same as the PASCAL VOC data set. Our data mainly contain two objects, one is bully, the other is victim.

```
<annotation>
    <folder>JPEGImages</folder>
    <filename>000050.jpg</filename>
    <path>/home/naruto/Documents/19Spring/Deep Learning/Project/JPEGImages/000050.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>360</width>
        <height>240</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>bully</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>32</xmin>
            <ymin>26</ymin>
            <xmax>180</xmax>
            <ymax>174</ymax>
        </bndbox>
    </object>
    <object>
        <name>victim</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>186</xmin>
            <ymin>69</ymin>
            <xmax>312</xmax>
            <ymax>201</ymax>
        </bndbox>
    </object>
</annotation>
```

Fig. 3. Annotations about Images[2]

4 Experimantal setup

Our enviroments are Python3.6 and Tensorflow basic framework 1.12all the codes implemented on Palmetto Cluster[3] "https://palmetto.clemson.edu/palmetto/" which provided by Clemson University. Our bully picture dataset provided by class and nonbully dataset[1], we download from website. The totally dataset image amount are 2387.

4.1 Parameters setting

We test our model so many times and fix the parameters, the following tables are our model parameter setting include Hyperparameters and parameter of networks.

Table 1. Hyperparameters

Parameters	Values
Learning rate	0.001
iteration times	10000 s
batch size	16
Image size	300

4.2 Train Process

The image is about train process.

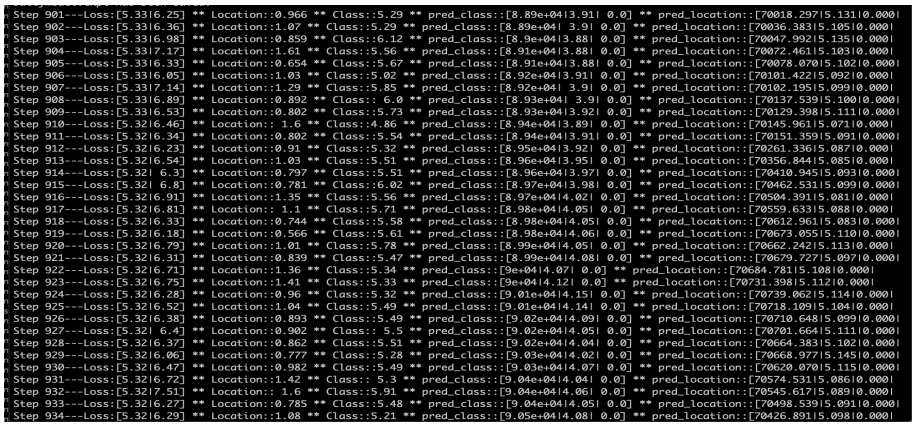

Fig. 4. Process

4.3 Results

We design two Api (script) to develop two functions of our models, we test on the dataset which provided by class.

Our model can predict single image or a group images, we can see the ground truth and we are predicted class indicated the object (Arrange by possibility), predicted class valuepossibility and predicted location (region of object).

We can find our models successfully predict the image and it contains objects.

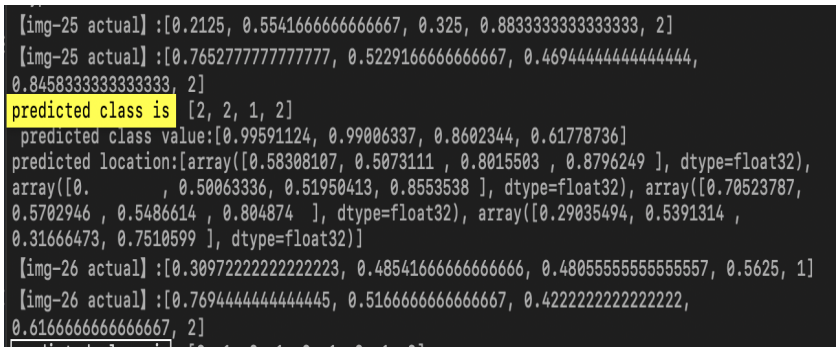

Fig. 5. Predict image

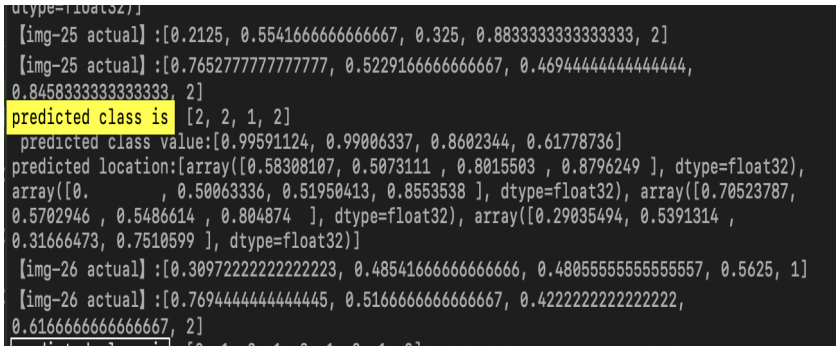

Fig. 6. Predict image

5 Conclusion

We have put a lot of effort into this project. From the very beginning, we have blurred the concept of deep learning. Now we can write our own model, adjust the hyperparameters, and complete the project completely. In this, we have learned a lot and improved the ability to solve problems.

Bibliography

[1] Actions, S. . (2015). Stanford 40 actions. In *Stanford 40 Actions*.

[2] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). SSD: Single shot multibox detector. In *ECCV*.

[3] palmetto (2015). palmetto. In *https://www.palmetto.clemson.edu/palmetto/*.

[4] Qassim, H., Verma, A., and Feinzimer, D. (2018). Compressed residual-vgg16 cnn model for big data places image recognition. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 169–175.

[5] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*.

[6] Tzutalin (2015). Labelimg. In *https://github.com/tzutalin/labelImg*.