2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC 2018)

Object Detection Based on YOLO Network

Chengji Liu[1] ,Yufan Tao[1] ,Jiawei Liang[1] ,Kai Li[1] ,Yihang Chen[1]
1. School of Mechanical Electronic and Information Engineering,
China University of Mining and Technology(Beijing), Beijing, China
{l_chengji, T_yufan787, zgkdljw, yihangc, LiBronKai}@163.com

***Abstract*—Object detection based on the deep learning has achieved very good performances. However, there are many problems with images in real-world shooting such as noise, blurring and rotating jitter, etc. These problems have an important impact on object detection. Using traffic signs as an example, we established image degradation models which are based on YOLO network and combined traditional image processing methods to simulate the problems existing in real-world shooting. After establishing the different degradation models, we compared the effects of different degradation models on object detection. We used the YOLO network to train a robust model to improve the average precision (AP) of traffic signs detection in real scenes.**

Keywords—YOLO network, image processing, object detection

I. Introduction

Object detection is one of the most important research directions for computer vision. Object detection algorithms are mainly divided into the traditional methods [1,2,3]and the deep learning methods [4,5,6]. In addition, the latter can be roughly divided into two categories. Some are based on region proposal object detection algorithms, including RCNN [7], SPP-net [8], Fast-RCNN [9] and Faster-RCNN [10], etc., which are algorithms that generate region proposal network and classify these region proposals afterwards. The others are based on the regression object detection algorithm, including SSD [11] and YOLO [12,13,14], etc. These algorithms generate region proposal network and classify these region proposals at the same time. All algorithms have good performances in object detection.

However, these algorithms are not tested with degraded images. In other words, they are trained with academic data sets, including ImageNet [15], COCO [16] and VOC [17], etc., but not well tested with randomly captured data sets. The main issues of images captured in the real scene include:

1) Due to the instability of the camera, the captured images can be blurred.

2) The images are not clear enough because the object can be obstructed.

3) The images may have poor quality as a result of bad weather, overexposure or low resolution. (see Fig. 1)

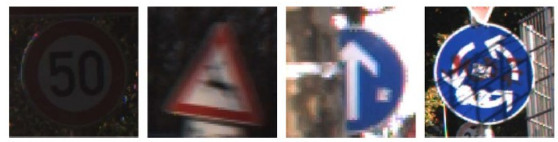

Fig. 1. Image problems: underexposure, rotation, blur, noise

The same object with small deviations in pixels caused by actual shooting may be divided into different classifications by neural networks, even though they are identical to human eyes [18].

In this paper, we simulated different degenerative processes of images for analysis and research. Firstly, we established the models of degraded images. We mainly used mathematical models to generate degraded images which are based on standard data sets. Then, we used these models to train the network to adapt to the complex real-world environment. Finally, we improved the ability of the model to generalize complex images. We took the traffic signs as the research object and used the YOLO [14] neural network to analyze. The experiment was based on the Darknet-53 [14] network structure. As a result, we made the following contributions:

1) We established a new image degradation model and used different degraded images as test sets. Then, we compared the effect of different degraded images on the standard model.

2) We modified the source network and performed different degradation processes for the training set. We also compared the accuracy of test sets on different models. Then, we performed more complex degradation processes on training sets and obtained a more generalized detection network. Afterwards, we compare it with 1) test performances.

3) Based on the above, we optimized the object detection method. To conclude, the generalization ability of the model has been enhanced, and the accuracy of object detection has been improved.

II. Image Degradation

We made some rules for ease of description: First, we assumed the coordinates origin (0,0) to be the lower left corner of the image. Second, the images were all in the first quadrant. What's more, the x-axis was the width of the image and the y-axis was the height of the image.