Tang Yi, Zhao Di

WEAKLY SUPERVISED OBJECT DETECTION METHOD

This research mainly introduces the existing challenges on the topic of weakly supervised object detection (WSOD) and proposes a new network to enhance feature representation for object detection.

I. Introduction

Compared with fully supervised learning, weakly supervised learning uses limited, noisy or inaccurately labeled data to train model parameters.supervisions can be divided into three categories: incomplete supervision, inexact supervision, and inaccurate supervision[1]. Below in Pic.1 shows three categories of supervision.

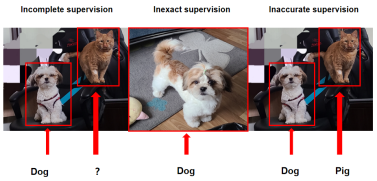

Fig. 1 Categoriesof threesupervision

II. Advantages and Disadvantages of existing method and model

The most popular weakly supervised object detection currently consists of three main steps: candidate region extraction, candidate region feature extraction, and candidate region classification.

Multi-instance learning is trained by viewing the images as packages and the candidate regions cropped from images as instances. However, this approach is only designed for the third step of weakly supervised object detection. For other steps, reliance on existing methods makes the speed of detection highly limited.

III. Method improvement and description

The CAM (Class Activation Mapping) is calculated as shown in the following equation, where equation (1) represents the scores of class c and equation (2) represents the class activation map for class c, W_k^c represent the weight of the unit k corresponded to class c, $f_k(x, y)$ represent the activation of unit k in the last convolutional layer at spatial location (x, y).

$$S_c = \sum_k \mathbf{W}_k^c \sum_{x,y} f_k(x,y) = \sum_k \sum_{x,y} \mathbf{W}_k^c f_k(x,y). \tag{1}$$

$$M_c(x,y) = \sum_{x,y} \mathbf{W}_k^c f_k(x,y) \tag{2}$$

After calculating the score of each class and its corresponding class activation map, the corresponding candidate regions are generated based on the hotspots.

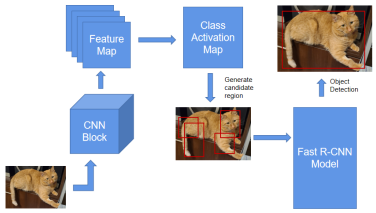

Fig. 2 Generalmodel structure

Above in Fig.2 shows the general structure of the whole network model. The original image is input to a convolution module to obtain feature map. By generating a series of candidate frames in the class activation map, the obtained candidate regions are input to the Fast R-CNN model to filter the candidate frames and continuously adjust the position coordinates of the candidate regions.

IV. Conclusion

This paper proposes a new network for weakly supervised object detection. By using class activation map to generate the candidate regions, we can use image-level annotations to implementing instance-level object detection.

1. Zhang D, Han J, Cheng G, et al. Weakly supervised object localization and detection: A survey[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(9): 5866-5885.

Tang Yi, master student in the Faculty of Information Technology and Control of BSUIR, tangyijcb@163.com.

Zhao Di, phD student in the Faculty of Information Technology and Control of BSUIR, 189124246@qq.com.

Supervisor, Gourinovitch Alevtina, Associate Professor, PhD in Physics and Mathematics, the Belarusian State University of Informatics and Radioelectronics gurinovich@bsuir.by.