Tang Yi, Zhao Di

IMPROVED REAL-TIME OBJECT DETECTION ALGORITHM

In this paper, the study explores a real-time object detection algorithm that strikes a balance between detection accuracy and speed by employing a lightweight network model, based on the existing classical object detection algorithm YOLOv4, to address the aforementioned challenges.

I. Introduction

As deep learning progresses, object detection has become a vital and fundamental task in computer vision. This technique is widely used in numerous sectors such as face identification, action recognition, and real-time monitoring. Recently, there has been a significant amount of research dedicated to improving the accuracy of object detection through the development of complex network structures and algorithms. Nevertheless, intricate deep learning models frequently encounter constraints in relation to processing resources and software support when utilized in practical situations, such as industrial scenarios. Therefore, they are unsuitable for use in actual production environments that include mobile and edge devices. Hence, the examination of efficient detection algorithms is of great importance in advancing deep learning research in actual real-world situations[1].

II. Method improvement and description

First, backbone is replaced with mobile version of the ViT (Vision Transformer). MobileViT's Transformer structure allows parallel computation with speeds up the inference of the model. It is particularly suitable for real-time object detection tasks. In addition, MobileViT uses a self-attention mechanism to learn the relationship between global and local features in an image, which improves the accuracy and robustness of object detection. Meanwhile, the use of depth-separable convolution in the detection head with fewer parameters is easy to train and tune, making its training process more stable and reliable.

Second, in terms of data processing, two data enhancement techniques, mosaic and mix up, are integrated. Specifically, the probability of mosaic data enhancement is set to 0.5, while the mix up data enhancement technique is further applied to the mosaic-enhanced data, and the probability is also set to 0.5. In this way, more diversified training samples can be generated, which can help the model better adapt to different scenarios and data changes. It can also reduce the error introduced by fuzzy labeling and enhance the interference of complex background on detection. The specific process is shown below in Figure 1.

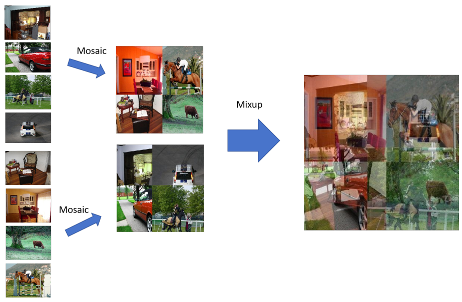

Рис. 1 – Integration of mosaic and mixup augmentation

Thirdly, the loss composition is modified. In particular, for classification loss, the model utilizes Binary Cross-Entropy loss, which can effectively handle multi-label classification tasks. The regression loss is computed using the Complete Intersection over Union, which measures the similarity between the predicted bounding box and the ground truth bounding box, considering their size, aspect ratio, and overlap. Objectivity confidence loss is determined using focal loss, which assigns higher weights to difficult examples and helps the model focus on challenging regions.

III. Conclusion

This paper successfully proposes and implements the improved YOLOv4 algorithm in response to the limitations of the computing power of terminal devices and the urgent need for the response speed of detection algorithms.

1. Kamath V, Renuka A. Deep learning based object detection for resource constrained devices: Systematic review, future trends and challenges ahead[J]. Neurocomputing, 2023, 531: 34-60.

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