Chapter 5
Human Face Mask Detection Using YOLOv7+CBAM in Deep Learning

Xinyi Gao
https://orcid.org/0000-0001-7727-9087
Auckland University of Technology, New Zealand

Minh Nguyen
Auckland University of Technology, New Zealand

Wei Qi Yan
Auckland University of Technology, New Zealand

ABSTRACT

COVID-19 and its variants have affected millions of people around the world. Wearing a mask is an effective way to reduce the spread of the epidemic. While wearing masks is a proven strategy to mitigate the spread, monitoring compliance remains a challenge. In this chapter, the authors propose a mask detection method based on deep learning and convolutional block attention module (CBAM). In this chapter, they extract representative features from input images through supervised learning. In order to improve the recognition accuracy under limited computing resources. They choose YOLOv7 network model and incorporate CBAM into its network structure. Compared with the original version of YOLOv7, the proposed network model improves the mean average precision (mAP) up to 0.3% in face mask detection process. Meanwhile, the method improves the detection speed of each frame 73ms. These advancements have significant implications for real-time, large-scale monitoring systems, thereby contributing to public health and safety.

DOI: 10.4018/978-1-6684-9999-3.ch005

INTRODUCTION

Currently, there are variants of COVID-19 virus (Ciotti et al., 2020) in the throes of the pandemic. In 2022, Monkeypox virus (Rizk et al., 2022) began to appear globally. The current known routes of transmission for both viruses are dropped alot. Wearing a mask is an effective way to stop the spread of the viruses. Wearing masks is required in closed and public spaces. Manually testing whether a mask is worn requires a lot of costs and increases the risk of testing personnel being infected. The mask monitoring through digital cameras can reduce the chance of inspectors being infected (Yan, 2019). In recent years, researchers have proposed several effective mask classification and monitoring algorithms that can detect faces with and without masks (Balaji et al., 2021). In this book chapter, we group mask detection into three cases: Wearing a mask correctly, wearing a mask incorrectly, and not wearing a mask.

Visual object detection (Zou et al., 2023) has always been an enduring research direction in the field of computer vision. The conventional object detection algorithms consist of three stages. Object proposals are firstly generated in the input image. The features in each proposal box are then extracted. Finally, different visual features were extracted by designing a classifier. However, the algorithms were not ideal in terms of accuracy and speed in visual object recognition. In recent years, visual object detection algorithms based on deep learning (Shen et al., 2018) have performed well in terms of accuracy and speed. Among them, You Only Look Once (YOLO) series of algorithms are Superior to others in visual object detection (Redmon et al., 2016).

In this book chapter, we propose a deep learning-based face mask object detection algorithm CBAM-YOLOv7. This model is based on the existing YOLOv7 model (Wang et al., 2023) with the addition of CBAM (Woo et al., 2018). YOLO is a one-stage detector model based on convolutional neural networks. It applies a neural network to the entire image. The network model firstly segments an image into regions and then predicts the bounding box of each region. CBAM consists of two modules: The channel attention module (CAM) (Huang et al., 2020) and the spatial attention module (SAM) (Wang et al., 2019). CAM can make the network pay more attention to meaningful ground truth regions. On the other hand, SAM allows the network to focus on context-rich locations throughout the image (Yin et al., 2023). Through this supervised learning method, the accuracy of mask recognition is effectively improved.

The rest of the book chapter is structured as follows. The second part of this book chapter introduces the related work on face mask detection. In the third part, we introduce the datasets, methods and models used. In Section 4, we compare and analyse our experimental results. Finally, we summarize our work in Section 5.

LITERATURE REVIEW

Deep learning (Yan, 2021) has now been widely harnessed in the field of computer vision. Especially in the recognition of various images, deep learning (Lu et al., 2021) (Liang et al., 2022) is playing an increasingly important role. For mask recognition, a consortium of deep learning (Wang & Yan, 2022) (Lu et al., 2018) models are widely used a number of representative methods are Faster R-CNN (Lin et al., 2020), InceptionV3 (Jignesh Chowdary et al., 2020), MobileNet (Venkateswarlu et al., 2020), YOLO, etc. Among them, the YOLO series are taken account for a large proportion which is the current mainstream.

In recent years, affected by the epidemic, almost all countries require people to wear masks frequently during travel to prevent the spread of the virus. To detect those who are not wearing masks,