AN EFFICIENT DEEP CONVOLUTIONAL NEURAL NETWORK APPROACH FOR OBJECT DETECTION & RECOGNITION FROM A VIDEO SEQUENCE USING MULTI-SCALE ANCHOR BOX

A THESIS SUBMITTED TO

PARUL UNIVERSITY

FOR AWARD OF DEGREE OF

DOCTOR OF PHILOSOPHY (Ph. D.)

IN

COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING & TECHNOLOGY

SUBMITTED BY

DWEEPNA GARG

UNDER THE GUIDANCE OF

DR KETAN KOTECHA

Head, Symbiosis Centre for Applied Artificial Intelligence (SCAAI)
Dean, Faculty of Engineering,
Director, Symbiosis Institute of Technology
Chief Executive Officer (CEO), Symbiosis Centre for Entrepreneurship and Innovation
Symbiosis International (Deemed University), Pune, Maharashtra, India

PARUL UNIVERSITY

P.O LIMDA, TAL: WAGHODIA, DIST: VADODARA,

GUJARAT STATE, INDIA-391760

SEPTEMBER, 2020

ABSTRACT

Deep learning is a new era of machine learning which trains computers to find the structure from a massive amount of data. Learning is described at multiple levels of representation. This enables us to make sense of the data consisting of text, sound, and images. Many computer vision problems such as object detection, image classification, and semantic segmentation have been solved using convolutional neural networks. Object detection in videos involves confirming the presence of the object in the image or video and then locating it accurately for recognition. Detecting and recognizing the still object from an image has comparatively shown better performance with the use of detection frameworks like R-CNN, Fast R-CNN, Faster R-CNN etc. The challenge is to detect and recognize the moving object from a static camera efficiently and accurately. In the video, modeling techniques suffer from high computation and memory costs which may lead to a decrease in performance measures such as accuracy and efficiency to identify the object accurately. The motive behind this work is to accurately detect and recognize the moving and still object from a video sequence using deep learning in real-time. The existing algorithms of object detection based on the deep convolution neural network worked well for large-size objects as the detection models get better results. However, those models fail to detect the varying size of the objects that have low resolution. This is because the features do not fully represent the essential characteristics of the objects in real-time after going through the repeated convolution operations of existing models. The proposed work improves the accuracy of detection by extracting the features of object at different size and scale by using a multi-scale anchor box. With the help of CNN, the deep knowledge from the dataset is extracted by giving the model a rigorous set of training samples. Our model has achieved 84.49 mAP on the test set of the Pascal VOC-2007 dataset which is higher than the state-of-the-art models. In our work, considering the accuracy as one of the evaluation measures, the objects get detected and recognized at 11 FPS which is comparatively better than other real-time object detection models. Our model is also trained and tested for face detection using the FDDB dataset. Moreover, the model is also able to detect partially covered faces. This also serves as one of the real-time application of our proposed work.