National Institute of Technology Rourkela

Defence Seminar

Seminar Title	: Robust Hand Gesture Recognition System for Real-time Multimedia Applications
Speaker	: Abir Sen (Rollno : 518cs1007)
Supervisor	: Dr. Tapas Kumar Mishra
Venue	: Convention Hall, CS Department
Date and Time	: 11 May 2024 (18:00)

Abstract

: Hand gesture recognition (HGR) is now a viable alternative for interaction between humans and machines. It has been applied to various fields, such as sign language interpretation, medical fields, virtual reality (VR) environments, and robotics. It plays a vital role in developing effective human-machine interfaces (HMIs) that enable direct communication between humans and machines. The research challenges such as poor lighting, occlusion, cluttered backgrounds, etc., make gesture identification difficult in real-time scenarios. The dissertation focuses on the development of a vision-based hand gesture recognition system using deep learning approaches, followed by the design of a low-cost human-machine interface (HMI) to control multimedia applications using gesture commands in real-time scenarios. The first contribution includes a low-cost human-machine interface via a hand gesture recognition system based on the ensemble of Convolutional Neural Network (CNN) models. Though this technique exhibits good results in terms of detection accuracy, but, in the case of real-time inferencing tasks, the speed is reduced in terms of frames per second (fps) due to the accuracy-speed trade-off. This problem is addressed in the second contribution, where five pre-trained CNN models and a vision transformer (ViT) are used for gesture classification tasks. The best model among those used models is then utilized to operate multimedia applications like the VLC player and Spotify music player using gesture commands in real-time. However, this system fails to produce promising results under cluttered backgrounds, various lighting conditions, etc. To address those issues in the third
contribution, we have divided our work into two parts by proposing two lightweight deep learning-based models. Here, we have chosen a small version of the &lsquoyou only look once version 5&rsquo (YOLOv5) model for its small size and fast inference speed. In the first part, we have frozen some convolutional layers in the baseline model. As a result, the number of trainable parameters, floating-point operations per second (FLOPS), and model size (in MB) have been reduced. Consequently, the inference speed is increased during real-time inference tasks without sacrificing significant detection accuracy. Next, the model was used to develop a robust hand gesture recognition system that enables physically challenged individuals to interact with systems. In the second section, we have proposed a lightweight YOLOv5s model by employing a channel pruning algorithm to reduce the model size, number of parameters, and FLOPs. Then, the channel-pruned YOLOv5s model is further utilized to build a gesture-controlled HMI to control two multimedia applications (VLC
player, Spotify music player) in the presence of the background environment, low light, and various light conditions.