Tanta University
Faculty of Engineering
Computer and Control Engineering Department

High Performance Hand Gesture Recognition Framework in Real-Time

A Thesis submitted in partial fulfilment of the requirements for the Doctor of Philosophy Degree in Engineering (Computer and Control Engineering)

By

Eng. Nehal Fathi Attia Abdelkawi khalifa

Under Supervision of :

Prof. Dr. Mohammed Talaat Faheem Saidahmed
Emeritus Professor at Computer Engineering and Automatic Control Department and Former Dean of Faculty of Engineering, Tanta University

Assoc. Prof. Dr. Mahmoud A. M. Alshewimy
Associate Professor at Computer and Control Engineering Department, Faculty of Engineering, Tanta University

2024

Acknowledgments

All praise and thanks be to **Allah**, the Lord of the world, the Ever-Magnificent; the Ever-Thankful for his support and bless to accomplish this thesis.

First, I would like to express my deep gratitude to my principal supervisor of this thesis **Prof. Dr. Mohamed Talaat Faheem,** Emeritus Professor and Former Dean, Computers and Control Engineering Department of the Faculty of Engineering at Tanta University, for his invaluable guidance, patience, encouraging, and immense knowledge.

I also wish to show my sincere gratitude to my co-advisor Dr. **Mahmoud Alshewimy**, Assoc. Professor at Computers and Control Engineering Department of the Faculty of Engineering at Tanta University, for his valuable suggestions and continuous help on the research topic. I appreciate his honest opinions, his calmness and clarity of advice, and his understanding, support, and patience.

Finally, I express my special appreciation to my family and friends. They have provided me the encouragement and continuous inspiration while working on this thesis.

Abstract

Recent technological advancements, particularly in the fields of computer vision and machine learning, have created new opportunities for sign language recognition. These technologies aim to enable real-time sign language interpretation, breaking down communication barriers and fostering greater inclusivity for the deaf and hard-of-hearing. The development of robust sign language recognition systems has enormous potential in a variety of domains, including assistive technologies, educational tools, and human-computer interaction. These technologies aim to empower people with hearing impairments by providing them with tools for effective communication and broader societal participation.

The main objective of this thesis is to develop and evaluate a real-time sign language recognition system that leverages the YOLOv5 architecture alongside attention mechanisms and various activation functions as well. This system aims to improve the accuracy of sign language detection and recognition, making it viable for use in diverse settings. By focusing on the integration of these technologies, the thesis will contribute to the field of assistive communication technologies, ultimately aiding in the seamless integration of sign language users into the broader community.

In the context of this thesis, four YOLOv5-based models were developed, three were enhanced with distinct attention algorithms, particularly CBAM and SE techniques, while the fourth model was established by improving the activation functions inside YOLOv5 architecture. Various well-known datasets were used in evaluating the proposed models such as public American Sign Language (ASL), Arabic Sign Language (ArSL), Bangla Sign Language (OkkhorNama: BdSL) and American Sign Language Letters (ASLL) datasets.

Based on the experimental results, the proposed models in this study achieved high recognition rates exceeding 99% and a high F1 score reaching 98% as well. So, it can be concluded that these models are efficiently capable of recognizing the different Sign Language gestures in various datasets. The results demonstrate that the presented models outperform their competitors in the literature. Therefore, these models could be highly competitive candidates for real-time gesture detection. Furthermore, these models mainly rely on the YOLOv5 architecture, which renders them lightweight and swift enough for deployment on any edge-based platform.