

ECE590-11-Project-Proposal

Hand Gestures Classification

1. Team Members

- Yiming Feng
-

2. Problem Statement

This project aims to develop a machine learning model to classify static hand gestures using both RGB images and 2D hand landmark data. The system will learn to identify gestures such as "thumbs-up," "stop," "OK," and "fist" from hand images.

This problem is interesting because it can be applied to smart home control — allowing users to interact with lights, appliances, or multimedia systems using simple hand gestures. Gesture-based control is intuitive, touchless, and accessible, making it useful in everyday settings (e.g., when hands are wet, full, or gloved) and for individuals with limited mobility. The project explores how combining visual (RGB) and structural (landmark) features can improve classification accuracy and reliability under different lighting and backgrounds.

3. ML Task(s)

- Type: Multi-class classification
 - Objective: Given an RGB image and its corresponding hand landmarks, classify the hand gesture into one of several predefined categories (e.g., thumbs-up, open hand, stop, etc.).
-

4. Data Involved

Datasets:

- HaGRID (HAnd Gesture Recognition Image Dataset) – A large-scale dataset with over 552,000 RGB images across 18 gesture classes, collected from 34,000 individuals under varying lighting conditions, backgrounds, and skin tones.

Source:

- <https://github.com/hukenovs/hagrid>

Features:

- RGB pixel data (512 pixels per image).
- 21 points 2D handmark

Size:

- HaGRID: ~552,000 labeled samples × (18 classes).
-

5. Models to Consider

Potential Models

- Vision Transformer (ViT-Tiny) → for RGB image
- Multi-Layer Perceptron (MLP) → for hand landmarks
- Fusion Model → RGB and landmark features

Baseline for comparison:

- MobileNetV3-Small
-

6. Expected Results

Success Metrics:

- Top-1 Accuracy
- Macro-F1 Score
- Inference Speed (FPS)

Goals:

- Achieve $\geq 90\%$ Top-1 Accuracy
 - Achieve ≥ 0.85 Macro-F1 Score
 - Fusion model accuracy and Macro-F1 by $\geq 5\%$ over single-modality models
 - Maintain real-time inference speed (≥ 25 FPS)
-

7. Relevant Paper

Reference

- Anton Nuzhdin, , Alexander Nagaev, Alexander Sautin, Alexander Kapitanov, Karina Kvanchiani. "HaGRIDv2: 1M Images for Static and Dynamic Hand Gesture Recognition." (2024).

Description

- This paper introduces the large-scale dataset used in this project and presents baseline CNN results for gesture recognition, laying a key foundation for model training and evaluation.