**Hand Gesture Recognition Using CNN and Real-Time Detection**

Popescu Matei Calin

Rus Alexandru

**Abstract**

Hand gesture recognition is a key technology enabling human-computer interaction through non-verbal communication.[1] Our project presents a system that utilizes a convolutional neural network (CNN) to classify six distinct hand gestures in real-time, supported by image preprocessing and gesture segmentation using OpenCV. The model achieved effective results through efficient training on labeled datasets and demonstrated its applicability in real-world scenarios through live webcam feed integration.

The demand for intuitive interfaces has led to an increased focus on gesture recognition technologies. The Gesture Recognition Market size is estimated at USD 31.35 billion in 2025, and is expected to reach USD 99.51 billion by 2030, at a CAGR of 25.99% during the forecast period (2025-2030)[2] This paper details a solution that integrates deep learning with real-time video processing to recognize hand gestures. The system identifies 3 gestures with their action:

* Palm/Fist -> Play/Pause
* Thumbsup -> Next Song
* Thumbsdown -> Previous Song

**Description**

The goal is to design a robust gesture recognition model capable of real-time classification and user feedback. The key challenges include creating an effective dataset, designing a lightweight yet accurate CNN, and integrating live video input processing.

**Setup**

1. **Dataset**

The dataset comprises 1500 images per gesture class, resized to 100×100100 \times 100100×100 grayscale images. Each image is preprocessed to reduce noise and improve model performance.

1. **Model Architecture**

The proposed CNN consist of:

* Convolutional Layers: Feature extraction using 64 filters with ReLU activation.
* Batch Normalization: Stabilization of training and faster convergence.
* Pooling Layers: Dimensionality reduction.
* Dropout Layers: Prevention of overfitting.
* Fully Connected Layers: Mapping extracted features to class probabilities.

1. **Integration**

The trained model is deployed for real-time classification using OpenCV. A region of interest is segmented and processed for prediction, and feedback is displayed on a live video feed.

**Experimental setup**

1. Data Preprocessing:

* Images are resized to 100 x 100
* Converted to grayscale
* Normalized between 0 and 1 for compatibility with the CNN

1. Gesture Recognition:

* Experiment using MediaPipe and Pygame -> providing an inuitive hands free interaction, but often giving wrong results as the hand landamarks calculations was not very accurate
* Training on images the model achieved a validation accuracy of ~90% after 10 epochs and loss convergence demonstrated stable training with no overfitting, supported by dropout layers.
* The real-time system achieved a near-instantaneous response with accurate gesture classification under standard lighting conditions. However, accuracy dropped in poor lighting or occlusion scenarios.
* Well-lit room, Dim lighting, Background clutter, Varying hand distances and angles.

**Conclusion**

The real-time system achieved a quick response with accurate gesture classification under standard lighting conditions. However, accuracy dropped in poor lighting or occlusion scenarios. Hand gesture recognition, as a subset of computer vision, holds significant potential across various domains, including assistive technology, smart homes, gaming, and automotive systems.

**References**

[1] "Gesture Recognition." Wikipedia, Website. Accessed 21 Jan. 2025.

[2] "Gesture Recognition Market - Growth, Trends, and Forecasts (2023 - 2028)." Mordor Intelligence. Website. Accessed 21 Jan. 2025.