Image Classification using CNN on the Hagrid 150k dataset

Introduction

Image classification is a fundamental task in computer vision with applications spanning autonomous vehicles, healthcare, and security systems. The development of deep learning models, particularly Convolutional Neural Networks (CNNs), has significantly improved classification accuracy. This report discusses an image classification project using a CNN architecture on the Hagrid 150k dataset, exploring existing research and external findings to contextualize and validate the approach.

This project aims to compare custom classification CNN architecture with standard models used for most research applications in this field and utilize a mobile app-based survey to find the perfect balance between complexity and accuracy for gesture recognition.

Related Research and Work

<https://openaccess.thecvf.com/content/WACV2024/html/Kapitanov_HaGRID_--_HAnd_Gesture_Recognition_Image_Dataset_WACV_2024_paper.html>

<https://ieeexplore.ieee.org/abstract/document/10543590/>

<https://ieeexplore.ieee.org/abstract/document/10484421/>

Dataset Overview

<https://www.kaggle.com/datasets/innominate817/hagrid-classification-512p-no-gesture-150k>



The Hagrid 150k dataset is a large-scale collection of hand gesture images designed for classification tasks. This dataset includes 150,000 images with varied lighting conditions, hand postures, and backgrounds, offering a robust benchmark for assessing model performance in practical settings. The diversity in the data presents challenges typical of real-world applications, such as noise and occlusion.

It has 19 distinct classes with multiple people from varying surroundings in order to adapt to any environment.

Methodology:

The existing models based off the Hagrid dataset involve training well know image classification models, Hence this project aims to experiment with alternative model architectures :

* Simple CNN involving few layers
* Complex CNN involving multiple layers

For comparison we have also trained a ResNET50 model on the same dataset to compare our architecture and finding the perfect balance between complexity and accuracy.

The general pipeline for our custom models is as follows:

* **Data Preprocessing**: Images were normalized and resized to 256x256 pixels for uniform input. Data augmentation (flipping, rotation, zoom) was applied to enhance generalization.
* **Model Architecture**: A custom CNN architecture comprises of:
  + **Convolutional Layers**: For feature extraction, followed by ReLU activation.
  + **Pooling Layers**: Max-pooling was employed for down-sampling.
  + **Fully Connected Layers**: For classification, concluding with a softmax activation.
* **Training and Optimization**: The model was trained using categorical cross-entropy as the loss function and Adam as the optimizer, with an initial learning rate of 0.001. Early stopping and dropout regularization were employed to prevent overfitting.