Kenya Sign Language Finger Spelling Detection

Business Understanding

Overview

Our endeavor involves utilizing cutting-edge computer vision techniques to recognize and comprehend the intricate handshapes and motions involved in fingerspelling. Through real-time implementation, our goal is to create a system that fosters immediate, meaningful communication between individuals with disabilities and their counterparts. We aspire to develop an innovative solution that harnesses modern technology to empower Deaf and Hard of Hearing individuals by enabling instant fingerspelling interpretation. This initiative will enhance the ministry's capacity to cater to individuals with disabilities, particularly those who rely on fingerspelling. It will facilitate more efficient communication between government officials and citizens with disabilities, thereby elevating the quality and accessibility of services provided.

The Problem Statement

Comprehensive communication is a fundamental human right, crucial for social integration and active participation in society. However, individuals with hearing impairments who rely on fingerspelling encounter significant obstacles to effective communication. The prevailing issue lies in the limited accessibility and inclusivity experienced by these individuals. Existing communication solutions fall short in facilitating seamless interaction for those proficient in this visual-gesture language.

In response, the Ministry of Public Services, Gender, Senior Citizens Affairs, and Special Program and Rehabilitative Services Division have initiated this project. Its aim is to develop and implement a real-time fingerspelling detection system. This system will serve as a bridge, enabling individuals with hearing impairments to communicate effectively with government officials, access government services, pursue education and employment opportunities, and ultimately enhance their overall social integration and well-being.

Objectives

The Main Objective

To develop and implement a real-time fingerspelling detection system that empowers individuals with hearing impairments to communicate effectively with the broader community, government officials and service providers, thereby promoting inclusivity, accessibility and social integration.

Specific Objectives

- **Data Collection;** Collect a diverse and comprehensive dataset of fingerspelling gestures performed by individuals with varying signing styles, hand shapes and speeds and merge it with the already existing kaggle dataset.
- **Develop a robust hand detection system;** Create an accurate and robust hand detection system using cvzone and MediaPipe to ensure precise tracking of hand movements during fingerspelling.
- **Implement fingerspelling recognition;** Develop a recognition component that can identify and interpret fingerspelled gestures in real-time using machine learning algorithms.
- **Real-time translation;** Integrate a translation mechanism that converts detected fingerspelled letters into text in real-time, ensuring immediate and accessible communication.
- **User-Friendly Interface;** Design an intuitive user interface that is accessible and user-friendly, allowing individuals with disabilities to interact with the system comfortably
- **Integration into a website;** Develop a website to integrate the system to maximize its reach and impact.

Data Understanding

The dataset, obtained from Kaggle, comprises grayscale images of hand signs representing the American Sign Language alphabet, excluding J and Z due to their involving hand motion. Each image is 28x28 pixels, with pixel values ranging from 0 (black) to 255 (white), allowing for various shades of gray. The dataset closely resembles the MNIST format, with labels and pixel values provided for each case.

In total, there are 27,455 training cases and 7,172 testing cases. Additional images were collected to enhance data diversity and volume. Notably, there's minimal distinction between fingerspelling in American Sign Language (ASL) and Korean Sign Language (KSL), allowing for interchangeability.

Data Preparation:

During the data preparation phase, the dataset images underwent preprocessing techniques to enhance their quality, remove noise, and standardize them for further analysis and model training.

Exploratory Data Analysis (EDA):

The analysis conducted on the data included the following:

Previewing Images: The dataset was visually explored to gain insights into the fingerspelling gestures' characteristics, such as hand shapes, finger positions, and movements.

Univariate Analysis: An analysis of the labels was performed to understand the distribution of individual letters in the dataset and identify any potential class imbalance issues.

Modelling and Evaluation

In this project, three different models were developed for a computer vision task involving fingerspelling recognition. Here's a summary of each model:

Baseline Model:

Architecture: Two hidden layers with 64 and 32 neurons using ReLU activation, and an output layer with 24 neurons using softmax activation.

Training: 100 epochs, batch size of 100, with 20% validation set.

Evaluation: Training accuracy nearly 100%, but validation accuracy drops, indicating potential overfitting. Test accuracy is approximately 69.88%.

First Model:

Architecture: Convolutional layers, MaxPooling layers, Flatten layer, Fully connected layers (Dense) with ReLU and Softmax activation, and Dropout layer for regularization.

Training: 10 epochs, batch size of 100, with 20% validation set.

Evaluation: Test accuracy of approximately 94.06%, suggesting good generalization to unseen data.

Second Model:

Architecture: Conv2D, MaxPooling2D, Flatten, Dense layers with ReLU and Softmax activation, and Dropout for regularization.

Training: 10 epochs, batch size of 100, with 20% validation set.

Evaluation: Test accuracy of approximately 87.31%, indicating good performance on new data.

Third Model:

Architecture: Conv2D, Batch Normalization, MaxPooling2D, Flatten, Dense layers with ReLU, and Dropout for regularization.

Training: 10 epochs, batch size of 128, with 20% validation set. EarlyStopping callback implemented.

Evaluation: Final test accuracy of approximately 94.62%, showing effective learning and good generalization.

In summary, the first and third models demonstrated high effectiveness in learning from the data and generalizing to unseen examples, achieving accuracies around 94%. The second model, while still performing reasonably well, had a slightly lower accuracy of around 87%. The baseline model showed signs of overfitting, with a test accuracy of about 69.88%.

Conclusions

- The analysis of the pixel distribution image indicates that the training data may have important features concentrated in the center, potentially influencing how the models learn.

- The Baseline and Second Models exhibit signs of overfitting, evident in high training accuracy but lower validation accuracy, implying they struggle to generalize to new data.

- Conversely, the First and Third Models demonstrate superior performance with higher test accuracy, suggesting effective learning and good generalization.

- The models vary in complexity, with the Baseline being simplest and the Third Model being the most complex. Performance appears to align with complexity, implying that a more complex model might be beneficial for this task.

- Dropout layers are employed to mitigate overfitting by randomly deactivating neurons during training. Despite this, overfitting persists in some models, indicating a need for additional regularization techniques or adjustments.

Recommendation

- To address overfitting in the Baseline and Second Model, consider implementing techniques like dropout, early stopping, or increasing training data. For models performing well (First and Third Model), regularly monitor their performance on new data to ensure consistent high performance.
- Given the concentration of important features in the center of images, explore preprocessing techniques like center-cropping or attention mechanisms to enhance model performance.
- The Third Model stands out as the optimal choice, showcasing a robust performance with high training accuracy (99.48%) and a commendable test accuracy of approximately 94.62%. This suggests effective learning and strong generalization to unseen data, further supported by the inclusion of various anti-overfitting techniques.

Next steps

- Continuously assess model performance with fresh, unseen data to verify consistent and effective generalization.
- Establish a monitoring system for ongoing performance evaluation, enabling prompt identification and resolution of any potential issues.
- Explore the potential of utilizing pretrained models, which have a track record of success in image classification tasks and have been trained on extensive datasets. Fine-tuning these models for your specific task can serve as a strong initial foundation.