

Name: Joey Hussain

Project Name: Computer Vision to Translate Sign Language

1. Introduction

There are over 500 thousand people who live in the United States in which their native language is that of American Sign Language. The issue however is that people who use ASL as there only way to communicate struggle with holding conversations with people who do not possess the understanding of Sign Language. This project I plan to provide an overview of data from alphabetical signals in sign language, including data preparation, exploratory data analysis, modeling, and evaluation to build a computer vision model to translate sign language with high accuracy.

2. Dataset:

The ASL dataset is available on Kaggle and consists of thousands of images in .jpg format for each alphabetical letter in American Sign Language

3. Data Preparation:

The first step in data mining on the ASL dataset is to prepare the data. This involves cleaning and transforming the data to ensure that it is in a format that can be easily analyzed. Once the dataset has been downloaded, it is important to check for any wrong images being labeled so that are data is clean and consistent for analysis.

4. Exploratory Data Analysis:

After the data has been prepared, the next step is to perform exploratory data analysis (EDA). EDA involves visualizing and summarizing the data to gain insights into the relationships between the variables. For the ASL dataset, I plan to see the distributions between each class of images, seeing the number of pixels in the image, and lastly viewing a sample of the images in each class. This is so that I can help identify any patterns or relationships that may exist in the data and guide the selection of appropriate modeling techniques.

5. Modeling

The next step is to select a modeling technique. I plan to use three algorithms for the ASL dataset, including k-means, convolutional neural networks, and support vector machines. The goal of modeling here is to predict what alphabetical letter is shown sign language based upon hand shape. The dataset I plan will use ten different letters: A-J. The modeling process involves splitting the data into training and testing sets, training the model on the training set, and then evaluating the model's performance on the testing set.

6. Evaluation

Model evaluation is an important step in data mining on the ASL dataset. This involves calculating performance metrics, such as accuracy, precision, recall, and F1 score, to determine how well the model performs. The metrics will depend on the modeling technique used, but in general, a higher accuracy indicates a better-performing model for ASL.