Real-Time Italian Sign Language Recognition with Deep Learning

Project Work in Deep Learning (91274)

Abdullah Naveed
MS - Artificial Intelligence (9063)
Matricola: 0001129140

February 24, 2025

Abstract

This report presents a deep learning approach for real-time Italian Sign Language (LIS) recognition. Starting from a dataset of sign language images with a black background, data augmentation was performed by extracting hand contours and compositing them on multiple backgrounds. A convolutional neural network (CNN) was then trained using the processed dataset. This report details the data preprocessing, network architecture, training and validation procedures, and demonstrates the final real-time recognition system using a webcam.

1 Introduction

Sign language recognition is a challenging problem due to variations in hand shapes, lighting conditions, and backgrounds. In this project, the original dataset [1] consisting of sign language images with a single (black) background was augmented by isolating the hand region and compositing it onto various background colors. The goal was to improve the robustness and generalization of the CNN classifier. The final system provides real-time recognition through a webcam interface.

2 Methodology

The project workflow consists of the following steps:

1. **Data Preprocessing and Augmentation:** Hand regions were isolated using contour detection, and the segmented hands were composited onto a set of fixed and randomly generated background colors. This augmentation increased the diversity of the dataset.

2. **CNN Model Training:** A deep CNN was built with five convolutional blocks followed by fully connected layers. The network was trained using data augmentation on the training set and evaluated on a validation set.

3. **Real-Time Testing:** The trained model was integrated into a real-time recognition demo using a webcam, where a 224x224 ROI is captured every few seconds and processed for prediction.

3 Data Preprocessing and Augmentation

The original dataset consisted of images with a uniform black background. To improve generalization, a preprocessing script was developed that:

- Recursively searches for image files in the dataset directory.
- Isolates the hand by finding the largest contour in a binarized image.
- Composites the segmented hand onto multiple backgrounds (a set of fixed colors and several random colors).

This process significantly increased the variability of the training data, resulting in better model performance.

4 CNN Architecture

The CNN model is composed of five convolutional blocks and a classifier. Each convolutional block consists of a convolution layer, batch normalization, ReLU activation, and max pooling. The classifier comprises a fully connected layer (flattening the features), a hidden layer with dropout, and an output layer that produces logits for 22 classes.

4.1 Architecture Diagram

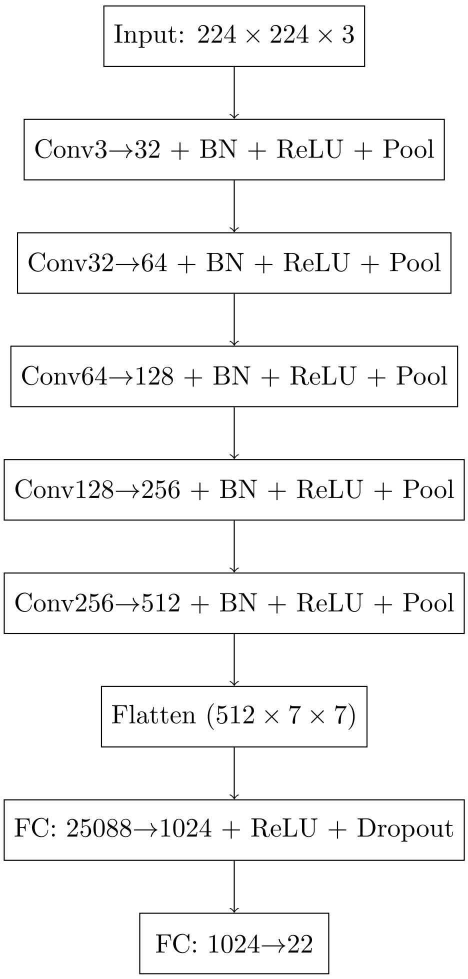

Figure 1: CNN architecture used for Italian Sign Language recognition.

5 Training Setup and Results

The dataset was split into training (80%) and validation (20%) sets. Data augmentation (random horizontal flips, rotations, and color jittering) was applied to the training data to improve robustness. The model was trained using the Adam optimizer and cross-entropy loss for 5 epochs. The following table summarizes the training and validation metrics:

Table 1: Training and Validation Metrics

Epoch	Train Loss	Train Acc	Val Loss	Val Acc
1	3.0615	8.81%	2.2291	27.58%
2	2.0019	29.02%	1.1167	64.19%
3	1.5083	45.14%	0.7354	76.55%
4	1.2465	54.73%	0.6438	80.63%
5	1.0672	61.36%	0.4874	86.03%

The best model, achieving a validation accuracy of 86.03%, was saved and subsequently deployed in a real-time recognition system using a webcam. In the demo, every 3 seconds a 224×224 region is captured from the center of the webcam feed and processed by the network to predict the corresponding sign language letter.

6 Conclusion

In this project, data augmentation by compositing hand contours onto multiple backgrounds significantly improved the performance of the CNN. The final model achieved a validation accuracy of 86.03% over 22 classes. The integration of the model into a real-time recognition demo demonstrates its practical applicability for Italian Sign Language recognition. Future work may include exploring more advanced architectures or additional augmentation techniques to further boost performance.

References

[1] R. G. et al., "Real-time Italian Sign Language Recognition with Deep Learning," *CEUR-WS Proceedings*, vol. 3078, paper 17, `https://ceur-ws.org/Vol-3078/paper-17.pdf`.