EE 541 Computational Introduction to Deep Learning
Project Report

Abid Hassan, Benjamin Fein-Ashley

1 Introduction

In this work, we utilize several deep learning algorithms to classify American Sign Language (ASL) alphabet letters from images, as well as real-time classification using live video. The dataset used for this project is the ASL Alphabet dataset, which contains images of hand signs representing each letter of the ASL alphabet. The goal is to build a model that can accurately classify these hand signs into their corresponding letters. We examine ResNet18 and DenseNet121 architectures, and compare their performance in terms of accuracy on a test dataset provided separately from the training dataset. The training data set contains 87,000 images which are 200x200 pixels. There are 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. The dataset is provided on kaggle

2 Methods

2.1 Data Preprocessing

The dataset is preprocessed by resizing the images to a uniform size of 200x200 pixels and normalizing the pixel values to be between 0 and 1. Data augmentation techniques such as rotation, flipping, and zooming are applied to increase the diversity of the training data and improve the model's generalization ability.

2.2 Model Architecture

We implement several deep learning architectures, including ResNet18, ResNet34, ResNet50, and DenseNet121. These architectures are chosen for their proven performance in image classification tasks. The models are trained using the Adam optimizer with a learning rate of 0.01 and a batch size of 32. The loss function used is categorical cross-entropy.

2.3 Training and Evaluation

The models are trained on the training dataset for 20 epochs. The provided training data is randomly split in an 80 to 20 ratio of training to validation data. The model at the epoch with best validation accuracy is saved as the final model to be evaluated. The performance of each model is evaluated on a separate test dataset, and metrics such as accuracy, precision, recall are calculated.

3   Results

The results of the experiments are summarized in Table 1.

Table 1: Model Performance on Test Dataset

Model	Test Accuracy	Precision	Recall
ResNet18	100%	–	–
DenseNet121	99%	–	–

The models were also tested on live video input, achieving real-time classification

4   Literature Review

Principal Component Analysis (PCA) combined with SVM classification achieved around 94% accuracy, demonstrating a baseline for evaluating more complex deep learning architectures [1].

A study utilizing multiple deep learning models, including ResNet-50 and EfficientNet, showed significant success using transfer learning techniques, achieving accuracies up to 99.98% with ResNet-50 [1].

MobileNetV2 was employed as a lightweight architecture, achieving 98.77% accuracy, suitable for real-time classification on resource-constrained devices [2].

SqueezeNet was another compact neural network architecture used for ASL recognition on mobile platforms, achieving a recognition accuracy of 83.29% [?].

A convolutional neural network combined with Mediapipe hand landmark detection demonstrated real-time performance with an accuracy of 99.95% [?].

A DeepCNN model combined with extensive data augmentation techniques was proposed, achieving 99.67% accuracy [3].

A novel hardware-based approach used neuromorphic sensors coupled with an artificial neural network implemented on an FPGA, achieving 79.58% accuracy [4]. Although the accuracy of this method is lower than others, it represents a more practical system for real time application.

Table 2: Accuracy Results from the Literature Review

Model	Accuracy (%)
ResNet-50 (Transfer Learning) [1]	99.98
MobileNetV2 [2]	98.77
DeepCNN [3]	99.67
Neuromorphic Approach [4]	79.58
PCA + SVM [1]	94.00

References

[1] Mdpi, "American sign language alphabet recognition using deep learning and transfer learning techniques," *Sensors*, vol. 23, no. 18, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7970

[2] Techscience, "A lightweight mobilenetv2 for american sign language recognition," *Intelligent Automation & Soft Computing*, vol. 35, no. 3, pp. 493–508, 2022. [Online]. Available: https://www.techscience.com/iasc/v35n3/49360

[3] NCBI, "American sign language recognition using a deep convolutional neural network with data augmentation," *Journal of Medical Internet Research*, vol. 24, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9078784/

[4] Mdpi, "Neuromorphic vision-based recognition of american sign language letters using an fpga," *Sensors*, vol. 17, no. 10, p. 2176, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/10/2176