American Sign Language Alphabet Classification Using Deep Convolutional Neural Networks

Aswaajieet J M*, Abhishek Kumar†
*230236, BSBE
Email: aswaajiiet23@iitk.ac.in
†230047, BSBE
Email: abhishekp23@iitk.ac.in

***Abstract*—This work presents a deep learning approach for classifying American Sign Language (ASL) alphabets using convolutional neural networks (CNNs). We use a publicly available ASL image dataset and train a CNN from scratch to achieve high accuracy in recognizing 29 classes of hand-signed letters. The model achieves competitive accuracy and shows strong generalization capability across a held-out validation set.**

I. Objective

The objective of this project is to build an image classification model that accurately recognizes hand gestures representing ASL alphabets using a convolutional neural network. The goal is to automate gesture recognition to support communication accessibility for hearing-impaired communities.

II. Data Pre-processing Steps

We use the ASL Alphabet dataset, which contains labeled RGB images for 29 alphabet categories (A–Z, except J and Z, which require motion, plus "nothing", "space", and "del"). The dataset was extracted from a zip archive and loaded using `torchvision.datasets.ImageFolder`. The following preprocessing steps were applied:

- Images resized to 64×64 pixels.
- Converted to tensors and normalized to $[-1, 1]$ using mean 0.5 and std 0.5.
- Dataset split into 80% training and 20% validation sets.

III. Model Architecture

The CNN model was implemented using PyTorch and consists of:

- **Convolutional Block 1:** 32 filters, 3×3 kernel, ReLU, MaxPooling
- **Convolutional Block 2:** 64 filters, 3×3 kernel, ReLU, MaxPooling
- **Convolutional Block 3:** 128 filters, 3×3 kernel, ReLU, MaxPooling
- **Fully Connected Layers:** Flatten $\rightarrow$ Dense (512 units, ReLU) $\rightarrow$ Output layer with 29 units

The model uses cross-entropy loss and is optimized with Adam optimizer with a learning rate of 0.001. Training was done for 10 epochs with batch size 64.

IV. Results

Model evaluation was conducted on the validation split using several metrics:

- **Accuracy:** 0.9942
- **Precision (Macro):** 0.9945
- **Recall (Macro):** 0.9941
- **F1 Score (Macro):** 0.9942

The classification report and confusion matrix indicate very strong performance across all classes, with minimal misclassification.

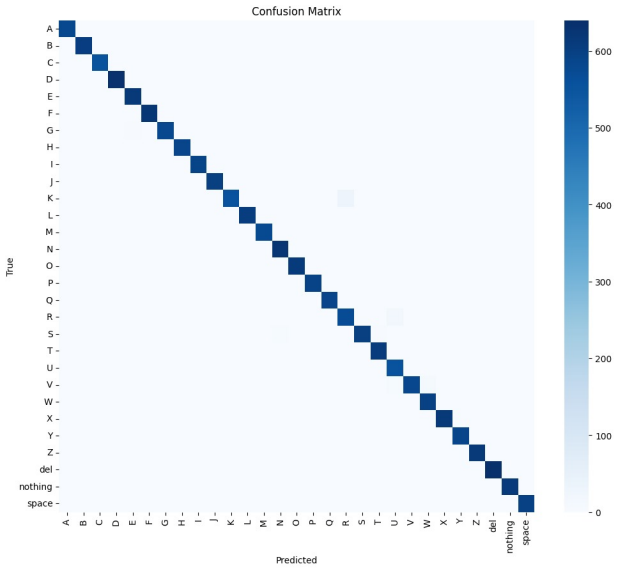

Fig. 1. Confusion matrix for ASL alphabet classification.

V. Conclusion

The proposed CNN model demonstrates high performance in ASL alphabet classification. Future improvements may include data augmentation, use of transfer learning (e.g., ResNet, EfficientNet), and deployment on mobile devices. The system provides a promising foundation for real-time ASL interpretation tools.