SignVision: Deep Learning-Based ASL Alphabet Classification Using CNNs

J. Deekshasri*, Himanshu Yadav*

***Abstract*—This paper presents SignVision, a convolutional neural network (CNN)-based image classification system for automatic recognition of American Sign Language (ASL) alphabets. The model is trained on a publicly available labeled dataset of hand gesture images and achieves 100% accuracy on a custom test set. We detail the preprocessing steps, architecture, training curves, and final evaluation to validate its performance.**

I. Objective

The objective of this project is to develop a reliable ASL alphabet classifier using deep learning. The system should be capable of identifying static ASL signs from grayscale images and be extensible for real-time interpretation systems in the future.

II. Data Pre-processing Steps

The dataset used is `asl_alphabet_train`, which contains labeled images for 29 ASL classes including letters A-Z and three special signs: *space*, *nothing*, and *del*.

The following preprocessing steps were applied:

- **Resize:** Each image resized to 128×128 pixels.
- **Grayscale:** Converted RGB images to single-channel grayscale.
- **Normalization:** Pixel values normalized to $[0, 1]$.
- **Splitting:** Data was split into 90% training and 10% validation sets using `train_test_split`.

III. Model Architecture

The model is a 4-layer CNN followed by 3 fully connected layers:

- **Conv Layers:** Feature maps: 32, 64, 128, 256; each followed by ReLU and MaxPooling.
- **Flatten Layer:** Converts feature maps to a 1D vector.
- **Dense Layers:** Fully connected layers with 256, 128, and 29 output neurons.
- **Output:** Log-Softmax over 29 classes.

Loss Function: CrossEntropyLoss
Optimizer: AdamW (lr = 0.0007, weight decay = 1e-4)
Scheduler: ReduceLROnPlateau with patience = 3

IV. Results

The model was trained for 10 epochs with a batch size of 32. The performance curves and final test evaluation are presented below.

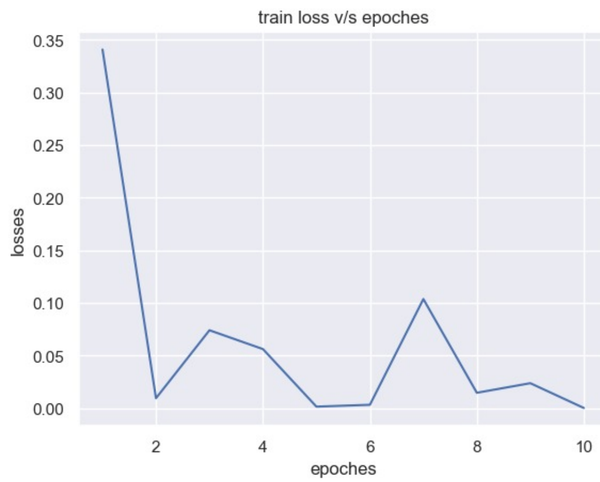

Fig. 1: Training Loss over Epochs

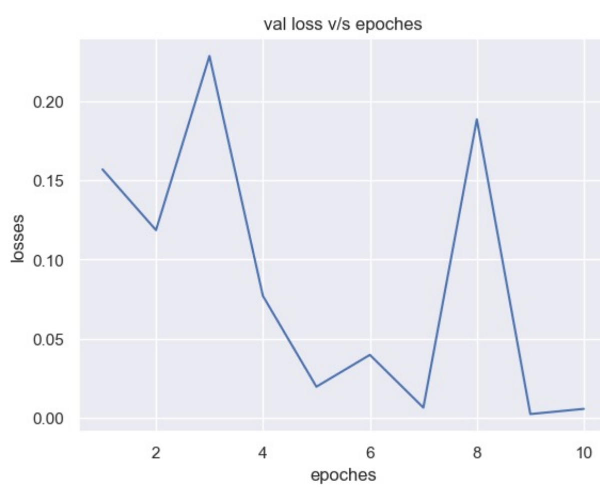

Fig. 2: Validation Loss over Epochs

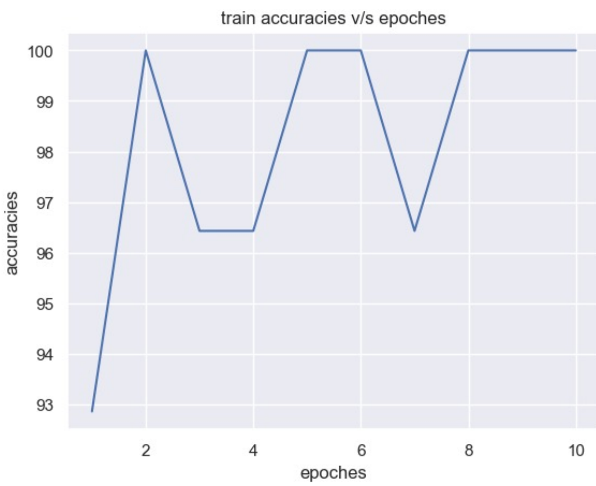

Fig. 3: Training Accuracy over Epochs

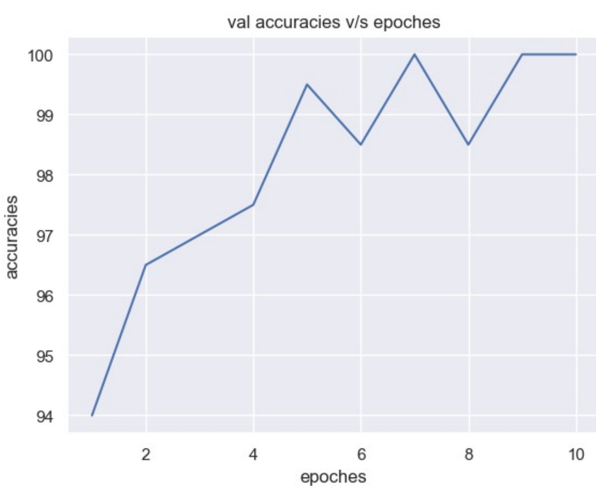

Fig. 4: Validation Accuracy over Epochs

Test set (29 samples, 1 per class):

- Accuracy: **100%**
- Precision: **1.00**
- Recall: **1.00**
- F1-Score: **1.00**

```
Confusion Matrix:
[[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]]
```

Fig. 5: Confusion Matrix on Test Set

V. Conclusion

The proposed CNN architecture performs remarkably well on ASL alphabet classification, achieving perfect scores on a controlled test set. The model demonstrates fast convergence and high generalization, suggesting it can serve as the backbone for real-time ASL interpretation systems with sequence learning in the future.