

SIGN LANGUAGE RECOGNITION USING DEEP NEURAL NETWORKS

MANASA KRISHNAN

JOANN RACHEL JACOB

AGENDA

- ▶ Introduction
- ▶ Related Work
- ▶ Datasets
- ▶ Experiments
 - ▶ Baseline Model
 - ▶ Results with American Sign Language(ASL) Dataset
 - ▶ Results with custom dataset
- ▶ Conclusion & Future Work

INTRODUCTION



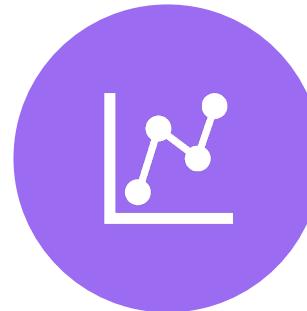
Deaf and mute community will benefit greatly through sign language.



An efficient translation system that recognizes sign language using deep neural networks.



Image and video datasets come with problems such as occlusion, improper background detection, intensity of light etc. However, CNNs prove to be effective in creating a robust model with a high level of resistance to noise.



Our pipeline consists of dataset generation, pre-processing, modeling and prediction of signs.

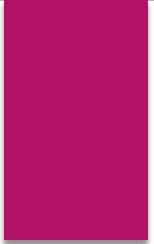
RELATED WORK

- ▶ Previous models - SVM, basic CNNs, InceptionNet and ResNet variants.
- ▶ VGG16, ResNet50 and InceptionV3 improved the overall accuracy of the model for this problem domain.
- ▶ Image pre-processing techniques like HOG, data augmentation by padding with black pixels, segmentation etc. were used.
- ▶ Loss functions - Categorical cross entropy loss and squared hinged loss.
- ▶ Optimizers - Adam, SGD, RMSProp.
- ▶ They all acquired an accuracy of more than 90 per cent, but in general CNNs tend to work better than the other models.

DATASETS

- ▶ **ASL Dataset** - Includes 29 classes with signs for 26 alphabets, space, delete and nothing. Size – 1.11GB (<https://www.kaggle.com/datasets/grassknotted/asl-alphabet>)
- ▶ **Sign Language MNIST Dataset** - Includes 24 classes with 24 alphabets. Size: 105.8 MB (<https://www.kaggle.com/datasets/datamunge/sign-language-mnist>)
- ▶ **Custom Dataset** - 1.7 GB. Includes 38 classes - 26 alphabets, 9 digits and 3 phrases. Size – 1.7GB
(<https://www.kaggle.com/datasets/joannracheljacob/american-sign-language-dataset>)

ASL DATASET



OUR CUSTOM ASL DATASET



EXPERIMENTS – Baseline Model

- ▶ Two baseline models.
- ▶ Simple CNN using ASL Alphabet dataset – 6 convolutional layers with input size 64*64 and batch size 64.
- ▶ Achieved accuracy of 95.54 percent in 5 epochs.
- ▶ Simple CNN using Sign language MNIST dataset – 3 convolutional layers with input size 28*28 and batch size 128.
- ▶ Both used ReLU activation, Categorical cross entropy loss and Adam optimizer.
- ▶ Achieved accuracy of 95.5 per cent in 50 epochs.

EXPERIMENTS – Using ASL Dataset

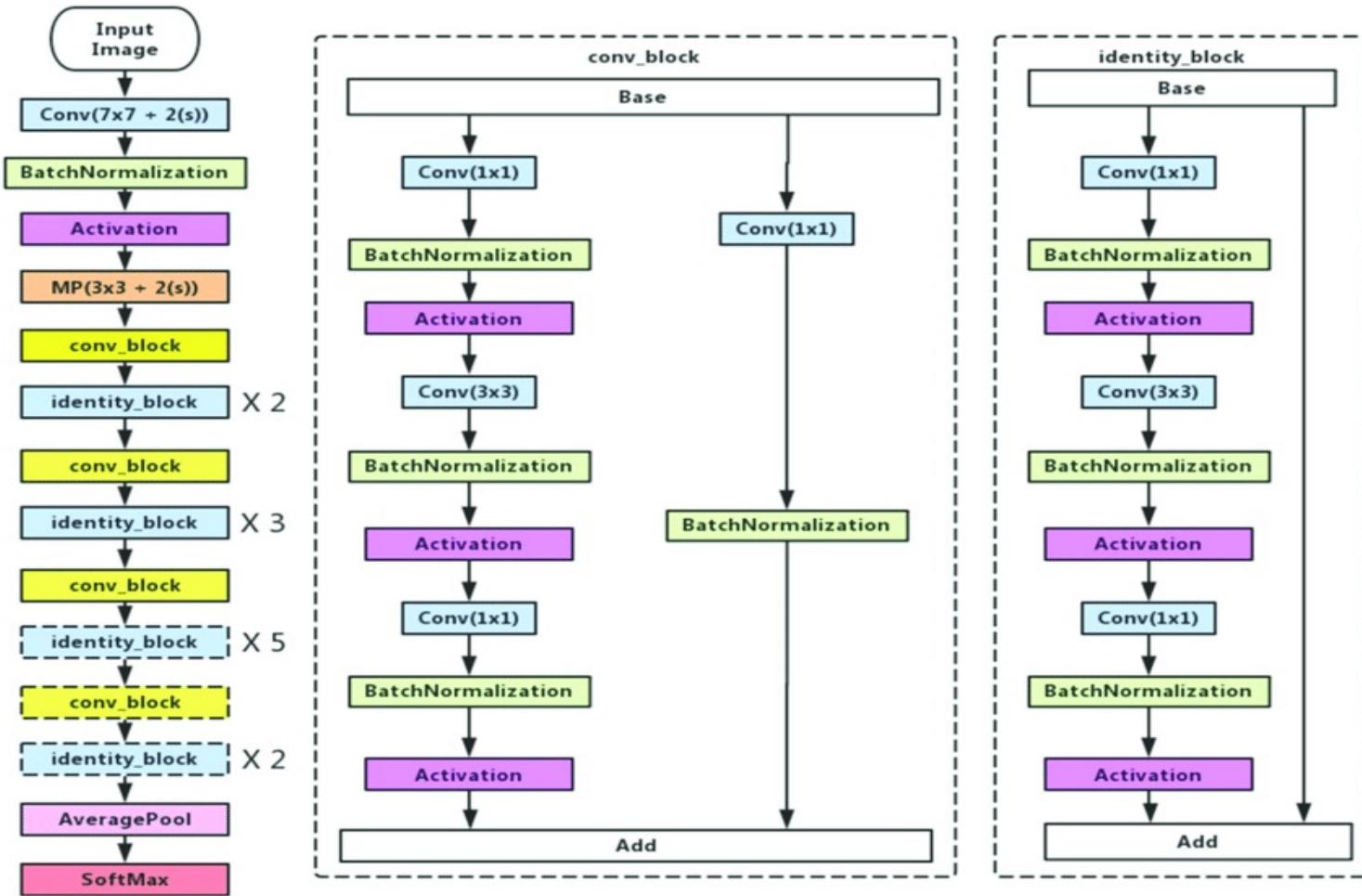
VGG16

- ▶ Input Image Size - 48 * 48
- ▶ Weight Initialization - Keras pretrained
- ▶ Cross Entropy Loss with Adam Optimizer
- ▶ Evaluation metrics - Accuracy, classification report to know the average accuracy and the confusion matrix.
- ▶ Epochs - 10
- ▶ Accuracy - 93%

ResNet50 (FastAI)

- ▶ Input Image Size - 200 * 200
- ▶ Train-Validation Split - 80-20 of the training data
- ▶ Batch Size - 16 using dataloader
- ▶ Cross Entropy Loss
- ▶ Metrics - Accuracy
- ▶ Epochs - 5
- ▶ Accuracy - 99.6%

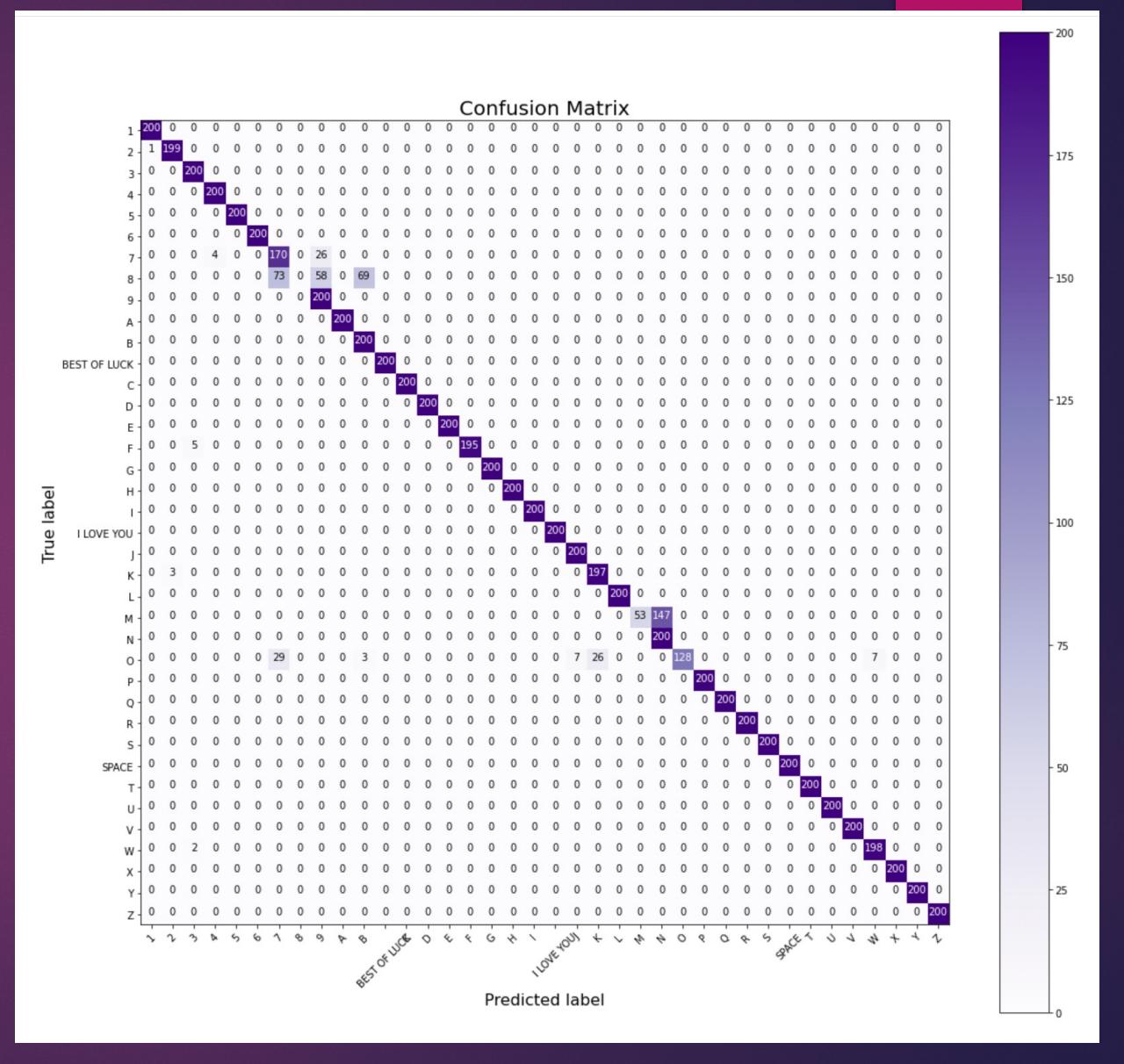
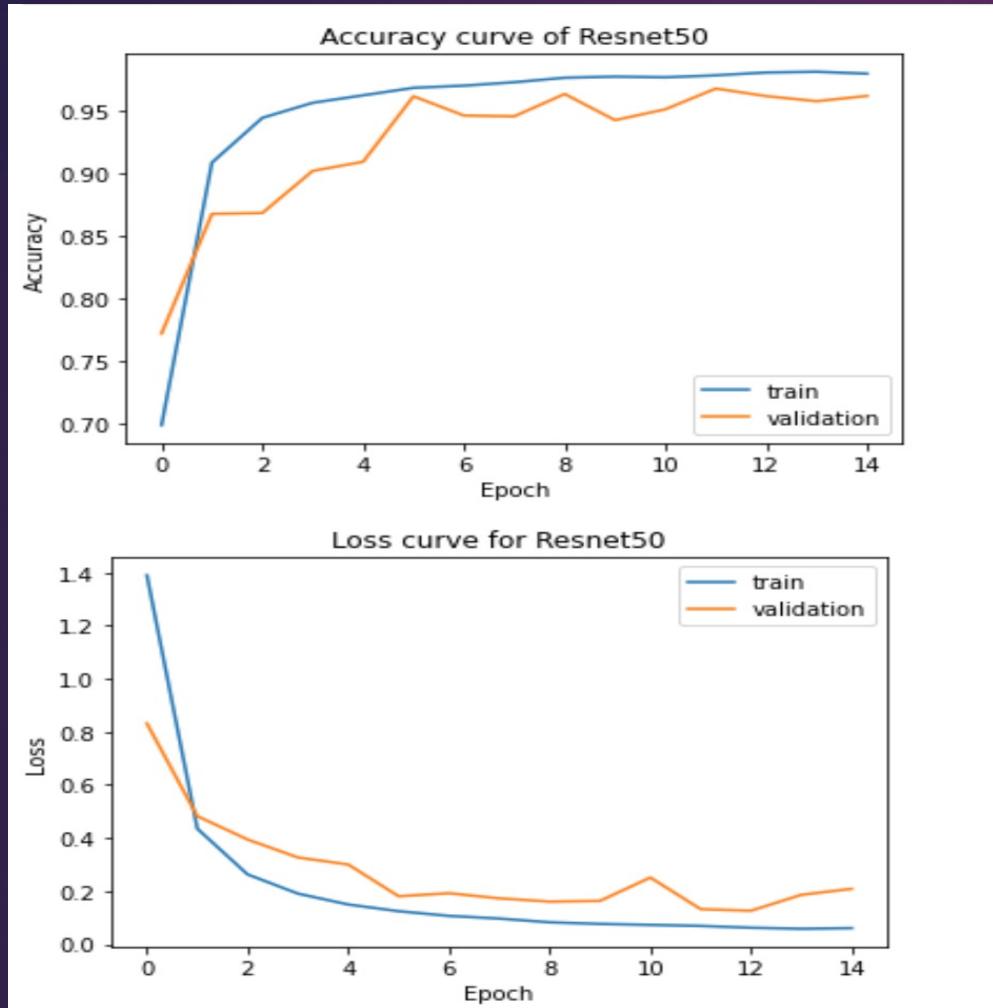
RESNET50 MODEL ARCHITECTURE



FINAL MODEL SPECIFICATIONS

- ▶ Train validation test data split - 60-20-20
- ▶ ResNet50 with pretrained weights.
- ▶ Added Dropout(0.2) and fully connected layer for 38 classes
- ▶ Image input size - 128*128
- ▶ Batch size - 32
- ▶ Epochs – 15
- ▶ Categorical Cross Entropy Loss and Adam optimizer
- ▶ Accuracy – 93.75 %

RESULTS



CONCLUSION & FUTURE WORK

Among the models tried, ResNet50 gave best results. The project proposes an optimal approach to recognize 38 signs.



Handling phrases and symbols with motions.



Better pre-processing techniques.



Including NLP with sign language translation in real-time video.



Implementing other efficient models.

THANK YOU
QUESTIONS?
