

(Real-time) American Sign Language Recognition

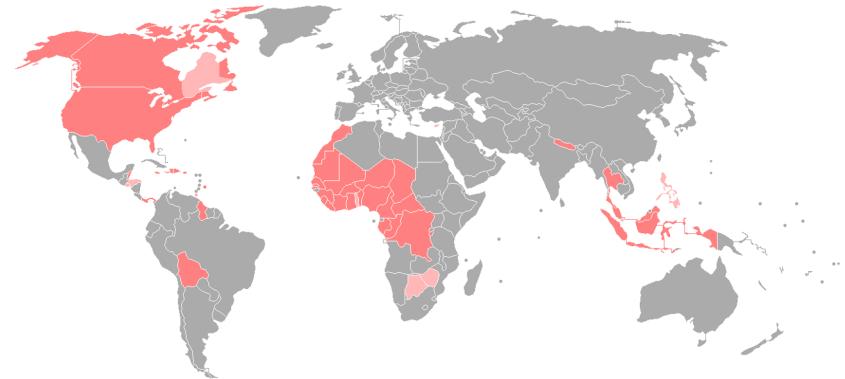
Leveraging Machine Learning

Presented By:
Abdullah Al Maruf

Objectives

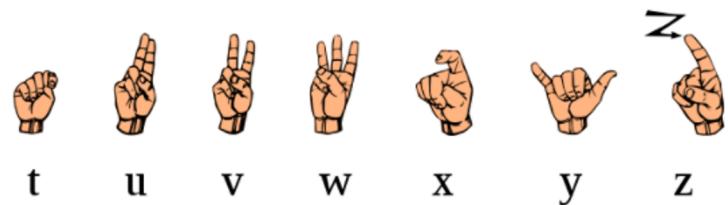
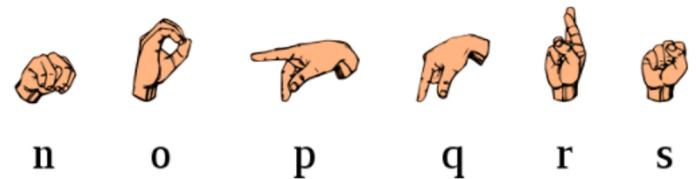
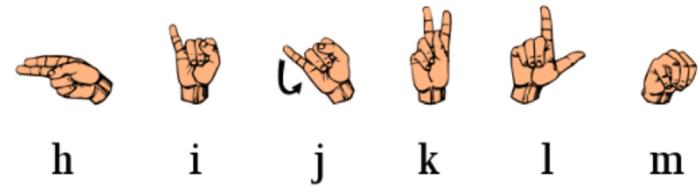
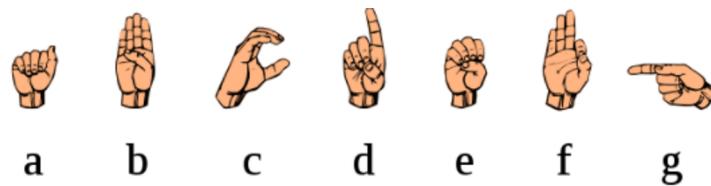
- Problem Description
- Dataset
- Model
- Experiments
- Analysis
- Conclusion/Discussion

American Sign Language



- Natural language with the same linguistic properties as spoken languages
- Grammar differs from English
- Expressed by movements of the hands and face
- Primary language of many North Americans who are deaf and hard of hearing
- Used by approximately 250,000 to 500,000 persons

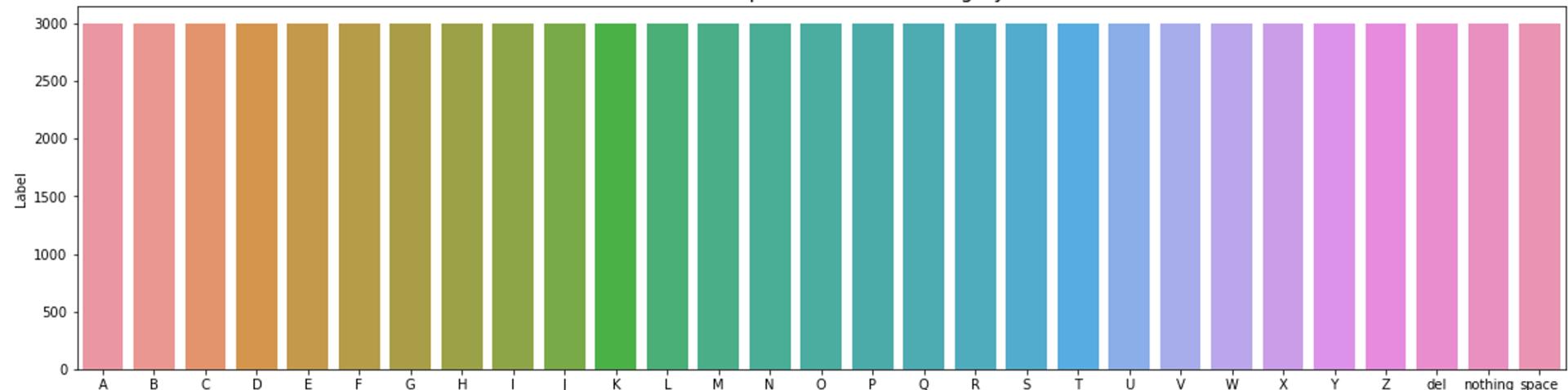
ASL Alphabets



- Similar to English Alphabets
- Multiple alphabets looks similar
 - A,E,M,N,S,T

Dataset

Number of pictures of each category



- Total 29 signs
- 26 alphabets + Delete, Space, and Nothing sign
- Each Sign Contains 3000 images

Dataset

- 90% for training
- 10% for testing
- Training data splitted into training and validation
 - 10% validation data



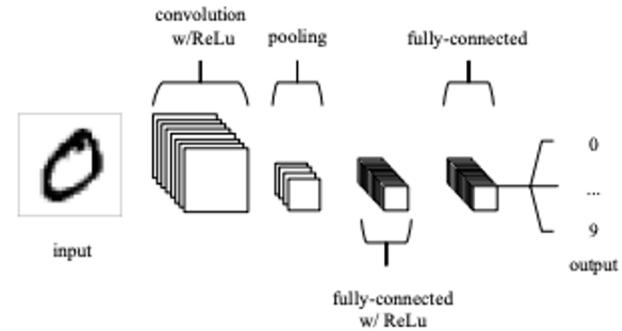
Dataset Shortcomings

- Less varieties
- Almost same background



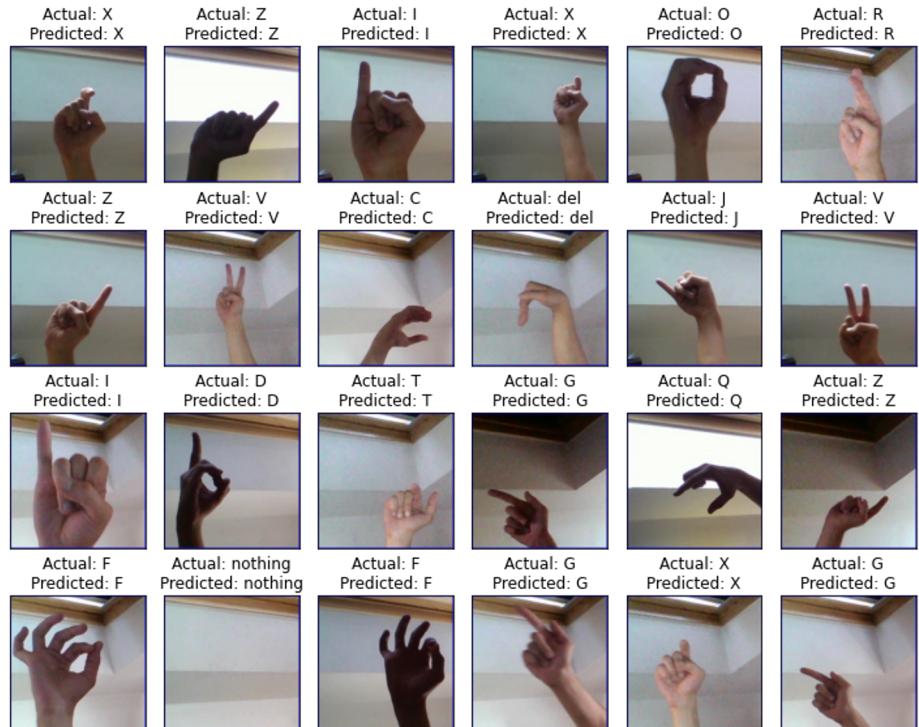
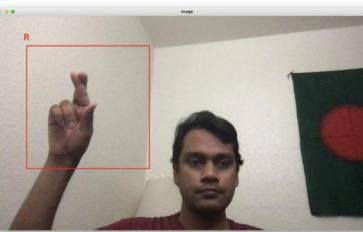
Training Model

- Convolutional Neural Network (CNN)
- Transfer learning with MobileNetV2 base model
 - Pre-trained on the ImageNet dataset
 - with millions of images and
 - 1000 classes of web images.
- Adaptive Moment Estimation (Adam) Optimization
- ‘categorical_crossentropy’ loss function
- 5 Epochs with early stopping



Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	-

Result



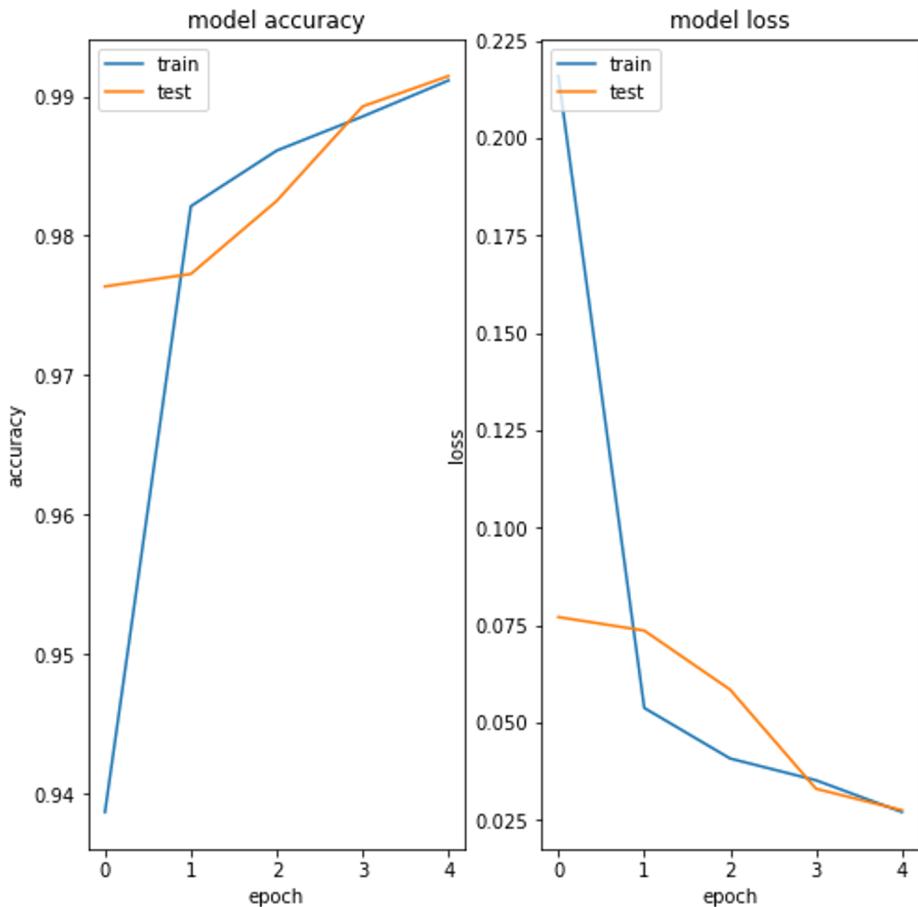
Test Set

Realtime UI

Analysis

- Training vs Validation accuracy and loss
- 0.991 validation accuracy and 0.027 validation loss.
- On Test set:
 - 0.994 accuracy and 0.0227 loss.

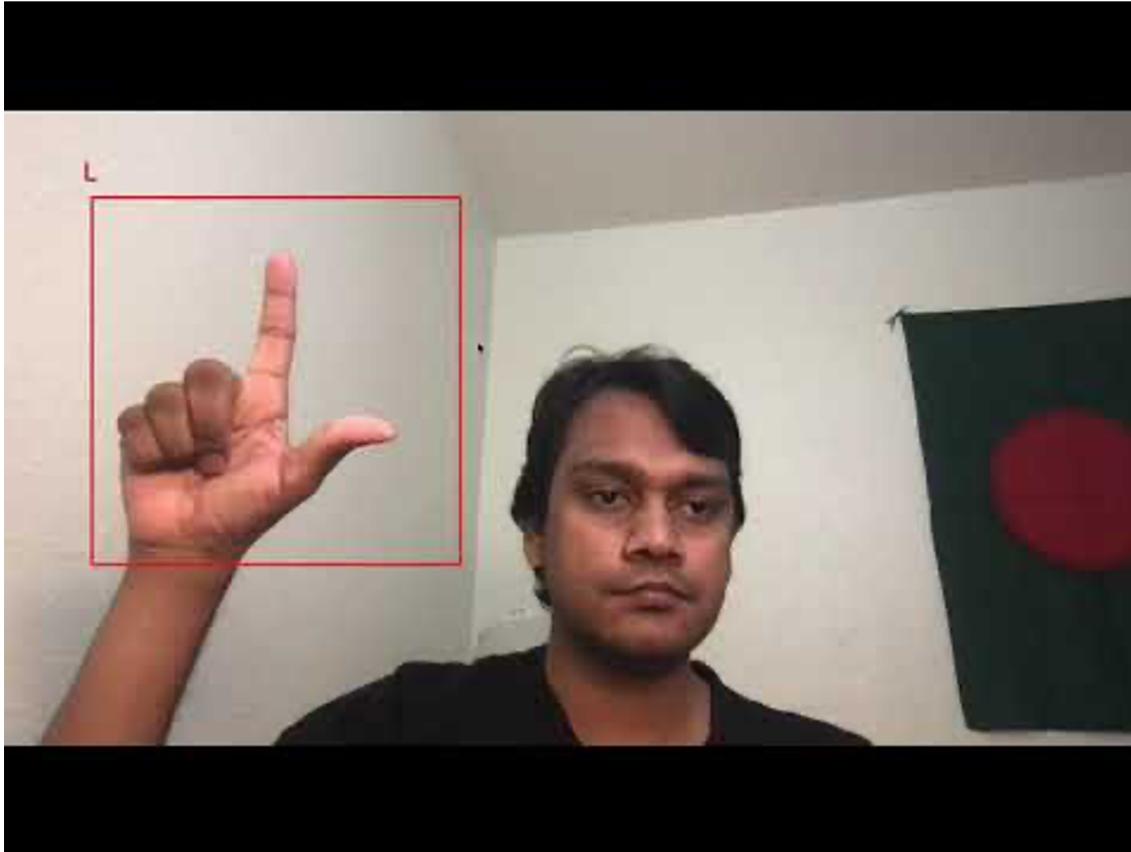
EPOCHS	TRAINING LOSS	VALIDATION LOSS	TRAINING ACCURACY	VALIDATION ACCURACY
1	0.219	0.082	0.935	0.973
2	0.055	0.043	0.981	0.984
3	0.039	0.054	0.986	0.981
4	0.035	0.053	0.988	0.983
5	0.029	0.050	0.990	0.984



Confusion Matrix

Normalized Confusion Matrix

Demo! Let's spell BAYLOR!



Conclusion and Future work

- The real-time application had difficulties to identify few letters
- A larger dataset with varieties of images will improve real time prediction
- Implementation of real-time application for mobile devices

Thanks