Deep Learning for Sign Language Recognition

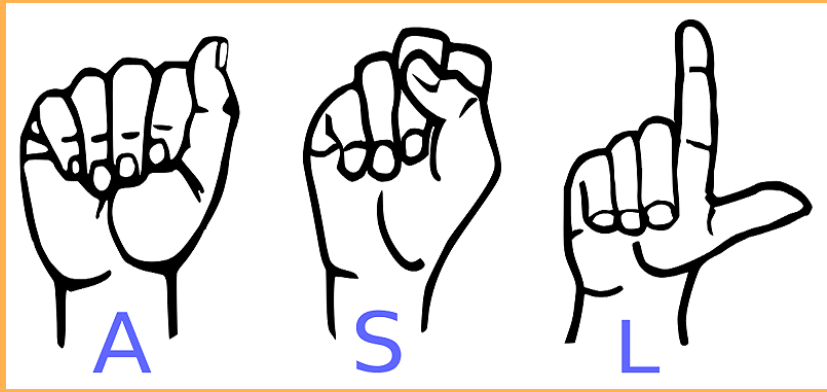
Zimu Su
Metis Data Science Machine learning Bootcamp

Objective and Background

- Use deep learning to translate sign language to text, providing a more convenient communication approach for deaf people.

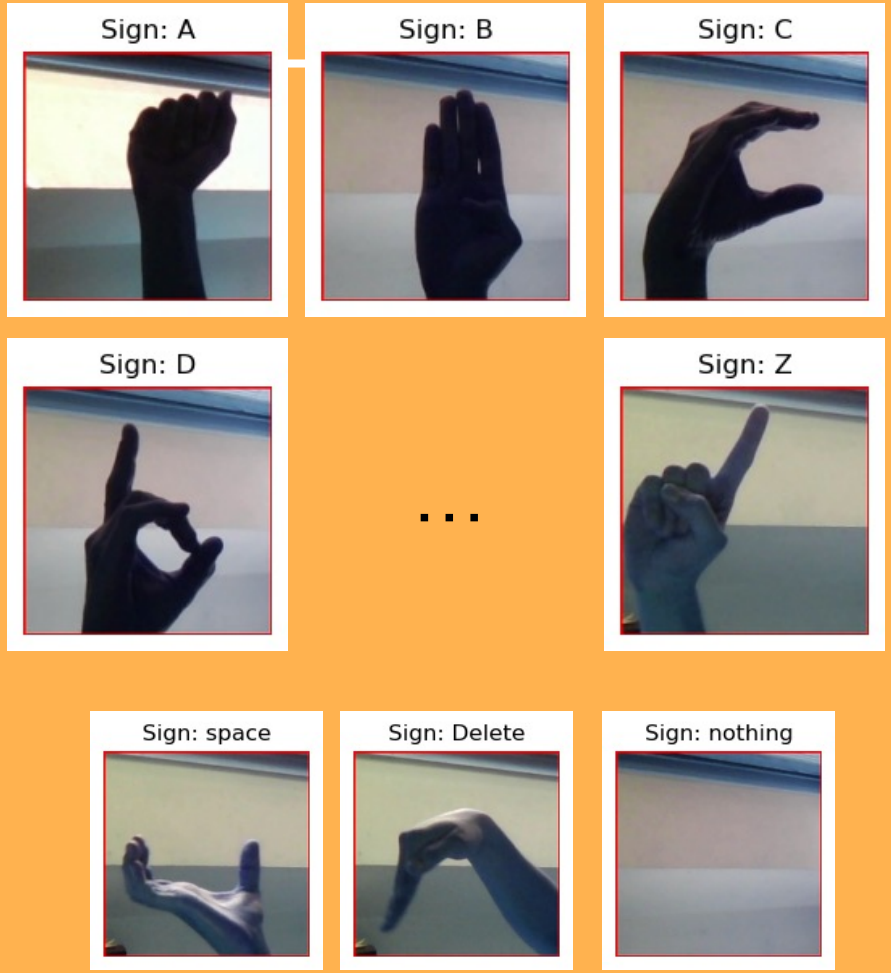
- This project tests on American sign language (ASL).

Fontvilla.com



Dataset

- Source: https://www.kaggle.com/datasets/grassknoted/asl-alphabet
- 29 Classes: letters A-Z, SPACE, DELETE and Nothing (Nothing for no sign)
- 87,000 images in total. 3000 images for each class.



Workflow and tool

Keras

Generate batches of image data with keras.imagedatagenerator

Validation using test image from dataset and custom images (photo by my own)

Get image array with Opencv, Numpy

Training:
Custom 3 layers CNN (1 channel, in google colab)
MobileNetV2, EfficientNetB0 (3 channels, embedded in keras)

OpenCV

Custom CNN layer setup (1 channel)

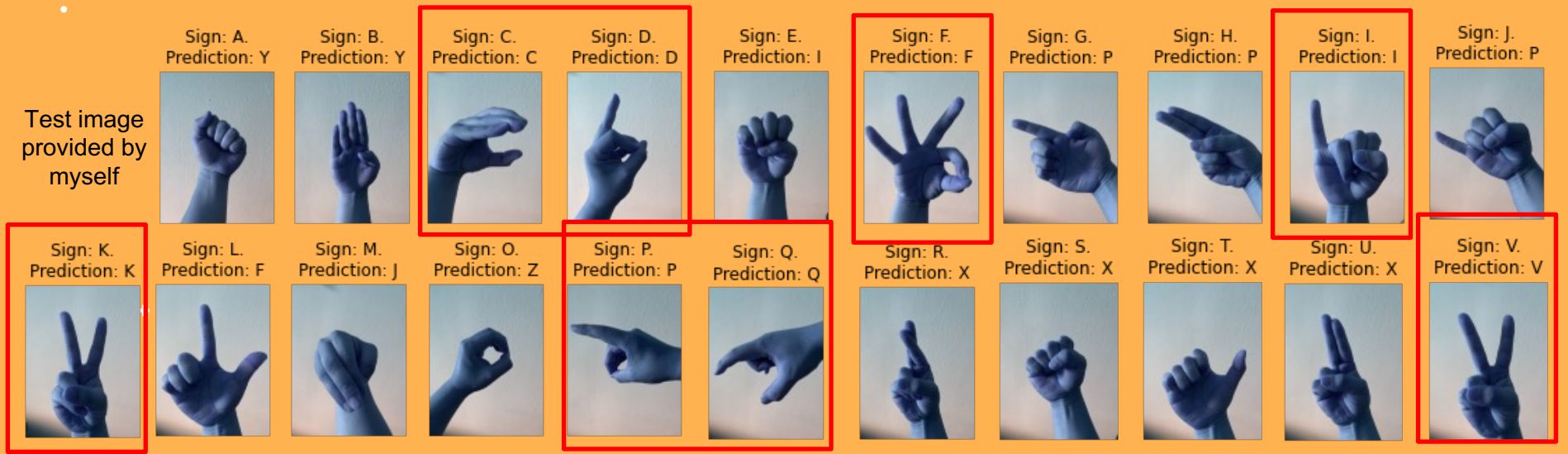
conv2d	(Conv2D)
batch_normalization	(Batch Normalization)
max_pooling2d	(MaxPooling2D)
dropout	(Dropout)
conv2d_1	(Conv2D)
batch_normalization_1	(Batch Normalization)
max_pooling2d_1	(MaxPooling2D)
dropout_1	(Dropout)
conv2d_2	(Conv2D)
batch_normalization_2	(Batch Normalization)
max_pooling2d_2	(MaxPooling2D)
dropout_2	(Dropout)
global_average_pooling2d	(Global Average Pooling 2D)
flatten	(Flatten)
dense	(Dense)
dense_1	(Dense)

	Training accuracy	Validation accuracy
Custom CNN	0.9306	0.9690
MobileNetV2	0.8455	0.7731
EfficientNetB0	0.9859	0.9922

Custom CNN Test results

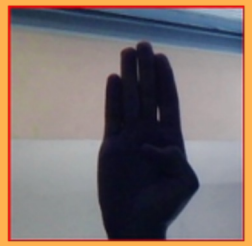
Test image provided by datasets

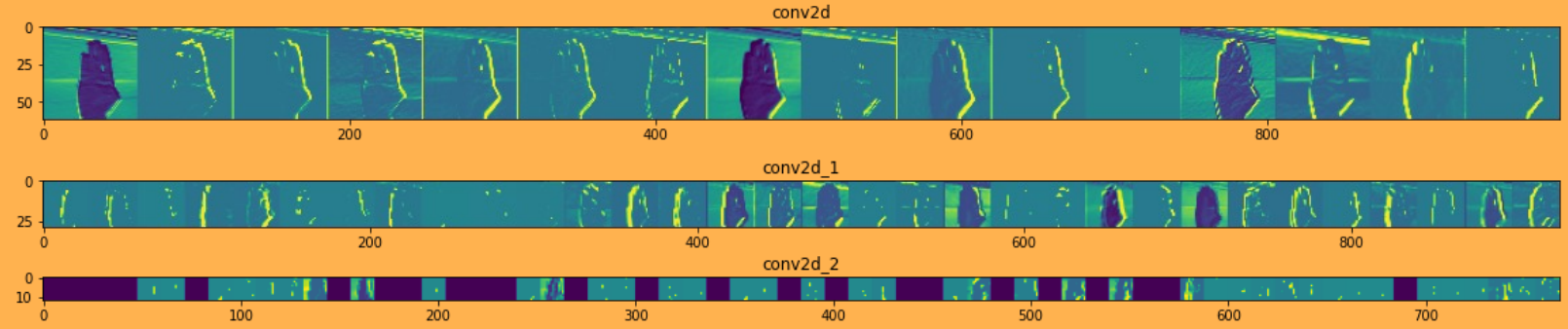
Test image provided by myself



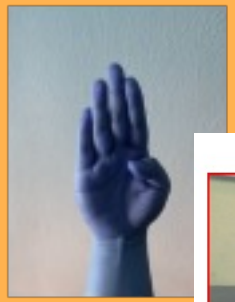
Feature map in CNN layers

Sign: B

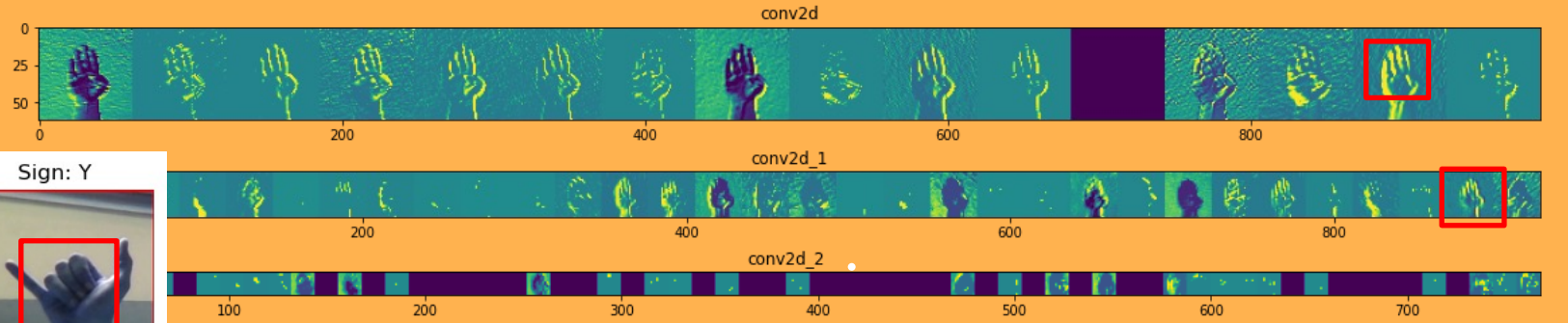




Sign: B.
Prediction: Y



Sign: Y



Take away points

- 3 layer CNN model or transfer modeling EfficientNetB0 is adequate for ASL translation (hand gesture recognition).
- Training dataset should contain more details of hand or fingers and diversify the image background. Unified images could generate bias.

Thank you for the attention!