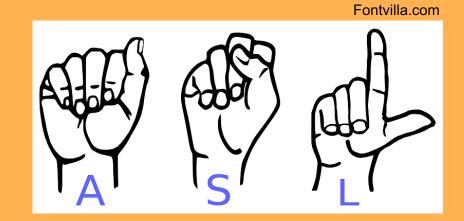


Objective and Background

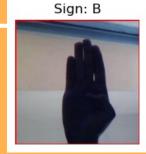
- Use deep learning to translate sign language to text, providing a more convenient communication approach for deaf people.
- This project tests on American sign language (ASL).



Dataset

- Source: https://www.kaggle.com/datasets/grassknot ed/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.

Sign: A



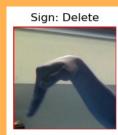














>>> Workflow

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 Opency

Training:
Custom 3 layers CNN
Transfer modeling
(MobileNetV2, EfficientNetB0)

Custom CNN layer setup (1 channel)

conv2d batch normalization max_pooling2d dropout conv2d_1 batch_normalization_1 max_pooling2d_1 dropout_1 conv2d 2 batch normalization 2 max_pooling2d_2 dropout 2 global_average_pooling2d flatten dense dense 1

(Conv2D) (Batch Normalization) (MaxPooling2D) (Dropout) (Conv2D) (Batch Normalization) (MaxPooling2D) (Dropout) (Conv2D) (Batch Normalization) (MaxPooling2D) (Dropout) (Global Average Pooling 2D) (Flatten) (Dense) (Dense)

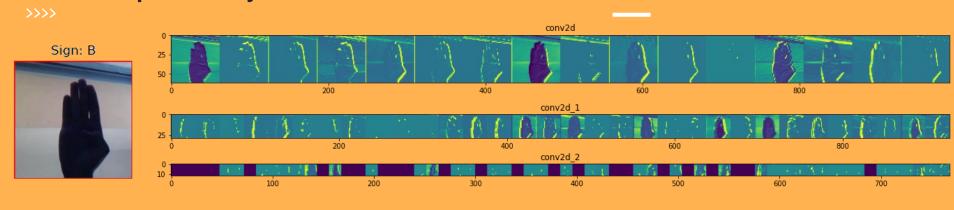


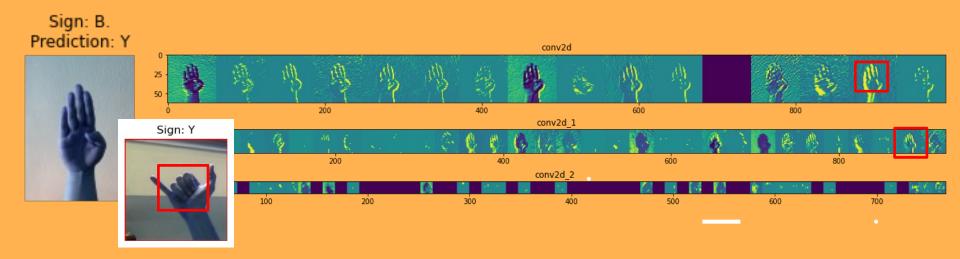
	Training accuracy	Validation accuracy
Custom CNN	0.9306	0.9690
MobileNetV2	0.8455	0.7731
EfficientNetB0	0.9859	0.9922

Custom CNN Test results

Custom CNN Test results											
>>>>	Sign: A. Prediction: A	Sign: B. Prediction: B	Sign: C. Prediction: C	Sign: D. Prediction: D	Sign: E. Prediction: E	Sign: F. Prediction: F	Sign: G. Prediction: G	Sign: H. Prediction: H	Sign: I. Prediction: I	Sign: J. Prediction: J	
Test image	446					11-		1	Jan .		
provided by datasets			1	2		40	3			10	
	Size I	Gira M	Size N	Sign: O.	Sign: P.	Sign: Q.	Sign: R.	Sign: S.	Sign: T.	Sign: U.	
Sign: K. Prediction: K	Sign: L. Prediction: L	Sign: M. Prediction: M	Sign: N. Prediction: N	Prediction: O	Prediction: P	Prediction: Q	Prediction: R	Prediction: S	Prediction: T	Prediction: U	
M					7			44	-		
	Sign: A.	Sign: B.	Sign: C.	Sign: D.	Sign: E.	Sign: F.	Sign: G.	Sign: H.	Sign: I.	Sign: J.	
	Prediction: Y	Prediction: Y	Prediction: C	Prediction: D	Prediction: I	Prediction: F	Prediction: P	Prediction: P	Prediction: I	Prediction: P	
Test image	100	di		4	de		-	-	1-	- 2	
provided by myself	9				1	10	The same		115	100	
,											
Sign: K. Prediction: K	Sign: L. Prediction: F	Sign: M. Prediction: J	Sign: O. Prediction: Z	Sign: P. Prediction: P	Sign: Q. Prediction: Q	Sign: R. Prediction: X	Sign: S. Prediction: X	Sign: T. Prediction: X	Sign: U. Prediction: X	Sign: V. Prediction: V	
Trediction: K					riediction. Q	Trediction: X					
V	1.	1	0	7	450	1	-23	24			
						45	1		4		

Feature map in CNN layers





>>>>

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!