Group 9 Final Project
Sign Language Gesture Recognition Wit
Different Light Source

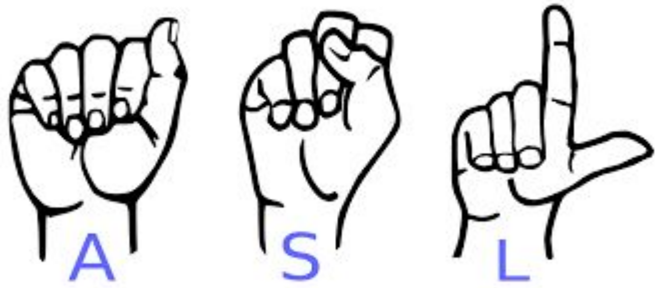
Howard Chi, Yu Cheng, Zhaowei Yu

Background

- The American Sign Language (ASL)
- ASL is a widely used language in US and Canada.

Problem:

- The ASL could not be translated using normal method.



Background

Solution:

- Image recognition and CNN for the translation.

Enhancement:

- Preprocess the image to make the system more robust under different environments.

Literature Survey

[4] used CNN and RNN model to train the ASL data independently.
By the result, the accuracy of the CNN model is greater than 90%, and the accuracy of the RNN model is approximately 55%.

[5] used various classifier to do the recognition. The recognition rates of ANN is 75%, deep ANN is 84%, and CNN is 91%.

Based on the two aboved paper, we know that CNN could be the most powerful method for sign language recognition of the mentioned neural networks.

Literature Survey

[1] shows us the power of the CNN model!

- The dataset contains 27455 images (28*28) with 26 different hand gestures made by different people.

- The accuracy this paper get is 99.3%.

Therefore, we decide to use CNN as our model.

Table 1 Architecture of CNN model

Layer (type)	Output shape	Param #
conv2d_1 (Conv2D)	(None, 26, 26, 32)	320
batch_normalization_1 (Batch)	(None, 26, 26, 32)	128
conv2d_2 (Conv2D)	(None, 24, 24, 32)	9248
batch_normalization_2 (Batch)	(None, 24, 24, 32)	128
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 32)	0
conv2d_3 (Conv2D)	(None, 10, 10, 64)	18,496
batch_normalization_3 (Batch)	(None, 10, 10, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 5, 5, 64)	0
conv2d_4 (Conv2D)	(None, 3, 3, 128)	73,856
max_pooling2d_3 (MaxPooling2D)	(None, 1, 1, 128)	0
flatten_1 (Flatten)	(None, 128)	0
dense_1 (Dense)	(None, 128)	16,512
dense_2 (Dense)	(None, 25)	3225

Total params: 122,169
Trainable params: 121,913
Non-trainable params: 256

Is ML suitable for ASL recognition?

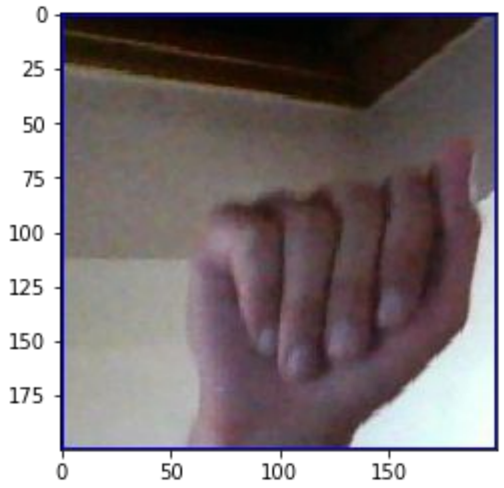
Of course! it does!!

Dataset

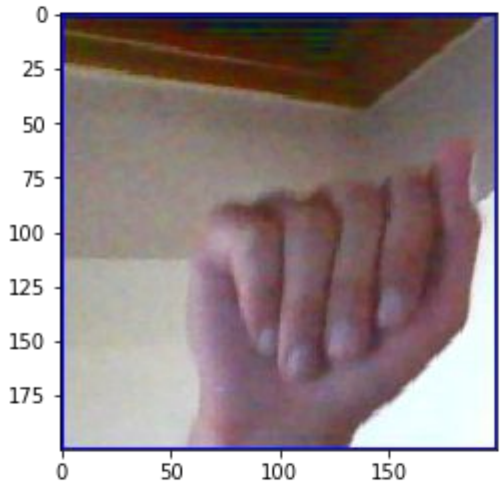
https://www.kaggle.com/grassknoted/asl-alphabet

- 80% for training and 20% images for testing.
- Each image is 200*200 pixels, and has 3 channels for RGB.
- Images are preprocessed before feeding into the model.

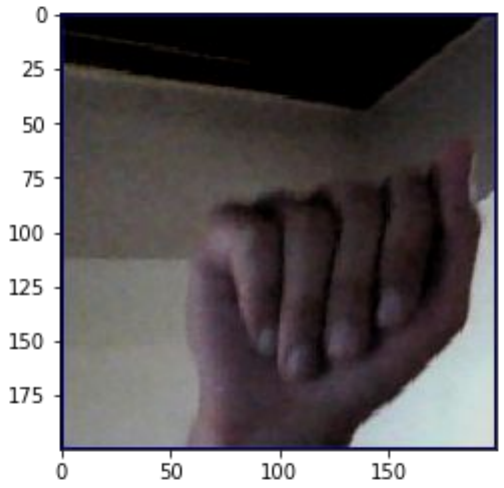
Image Processing



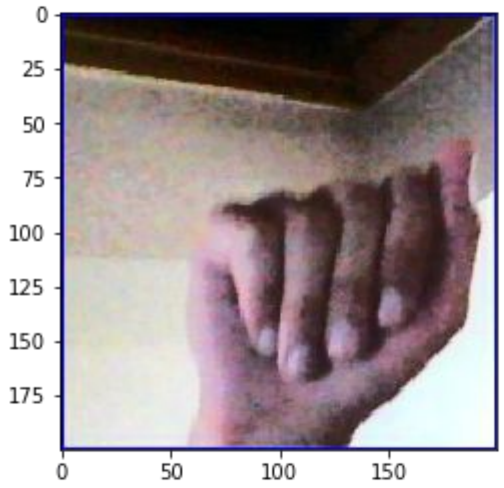
Original Image



Turn bright

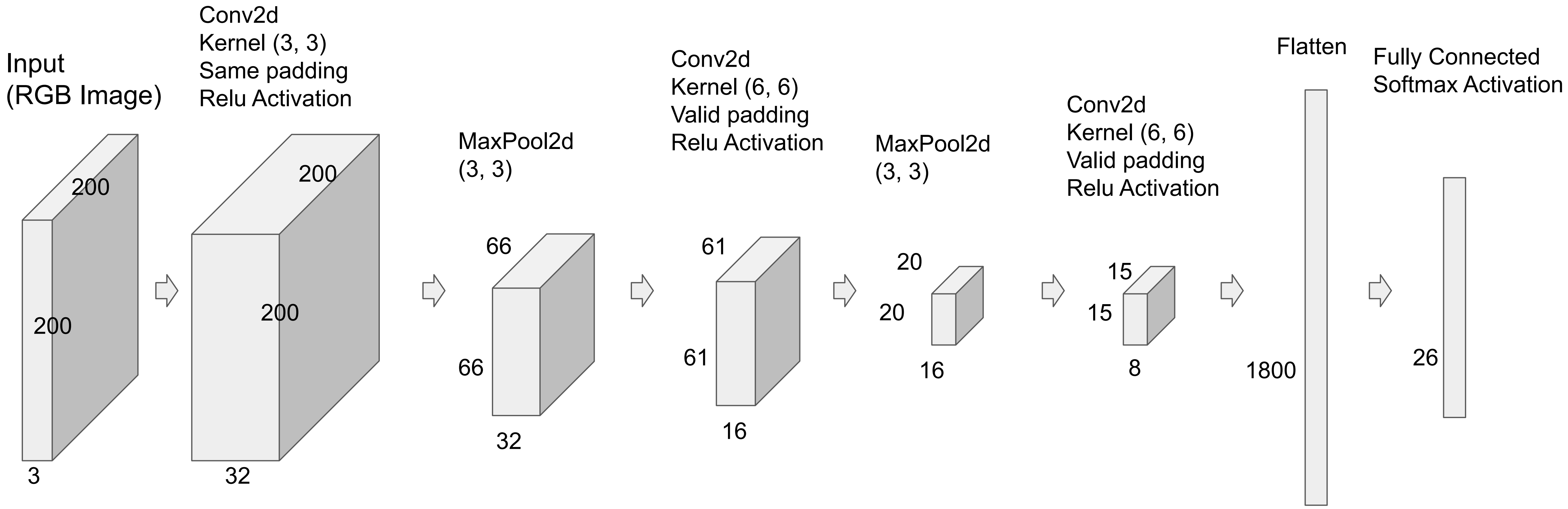


Turn dark



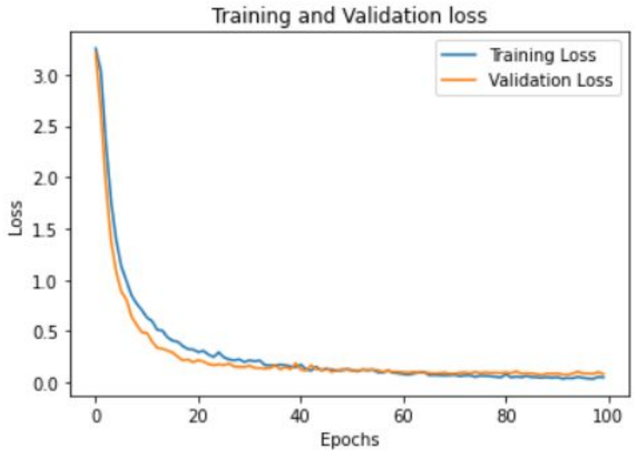
Equalization

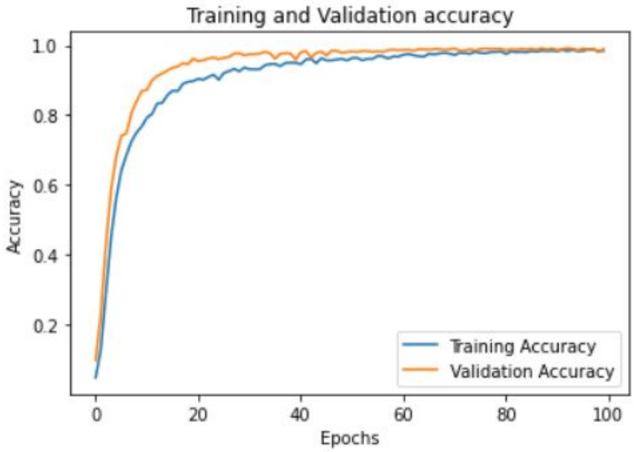
Model


Input
(RGB Image)
200
200
3
Conv2d
Kernel (3, 3)
Same padding
Relu Activation
200
200
32
MaxPool2d
(3, 3)
66
66
32
Conv2d
Kernel (6, 6)
Valid padding
Relu Activation
61
61
16
MaxPool2d
(3, 3)
20
20
16
Conv2d
Kernel (6, 6)
Valid padding
Relu Activation
15
15
8
Flatten
1800
Fully Connected
Softmax Activation
26

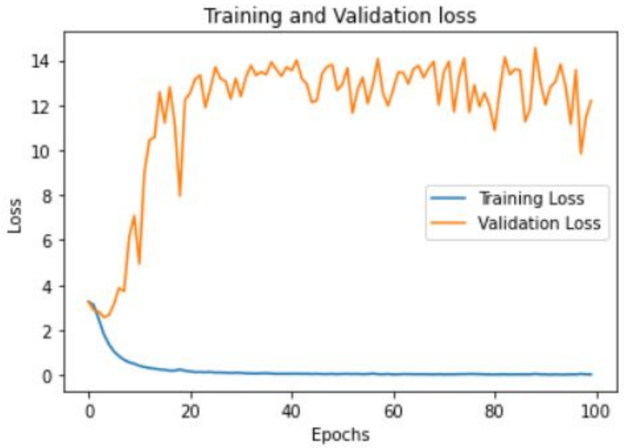
Results

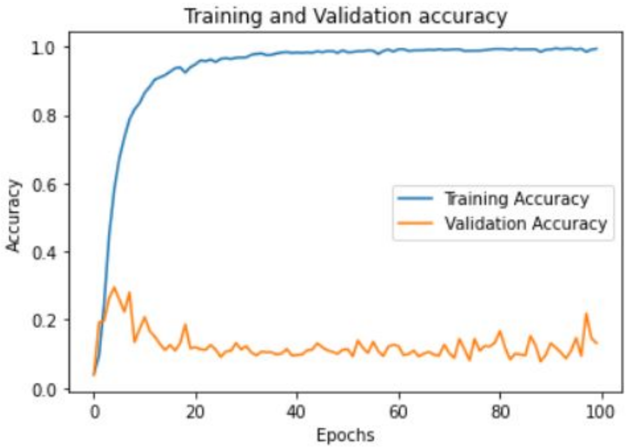
1. All data have appropriate intensity level





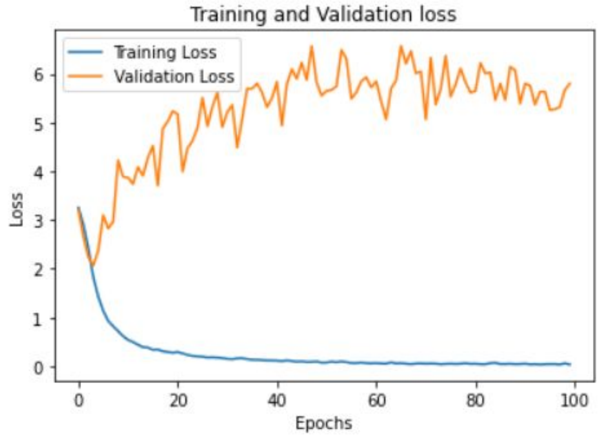
2. Validation data has been darkened

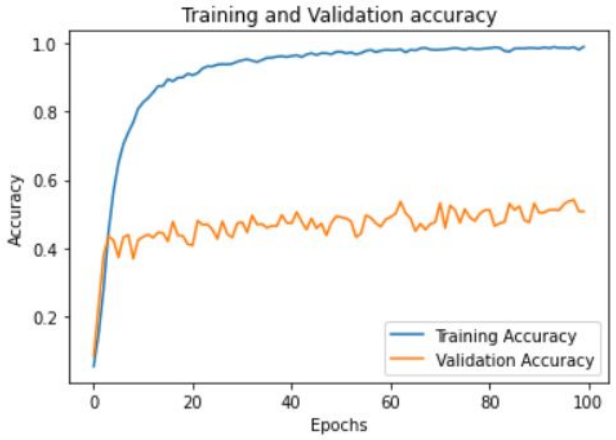




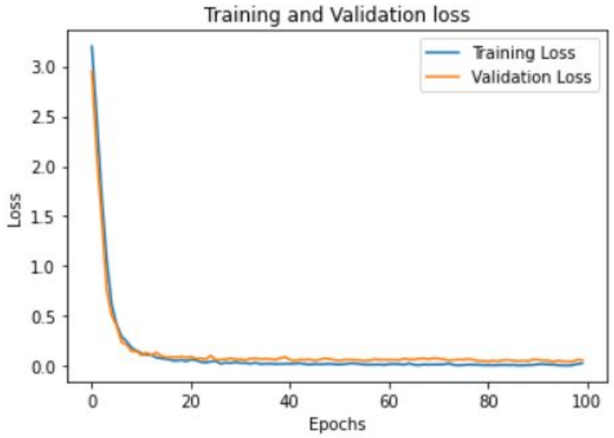
Results

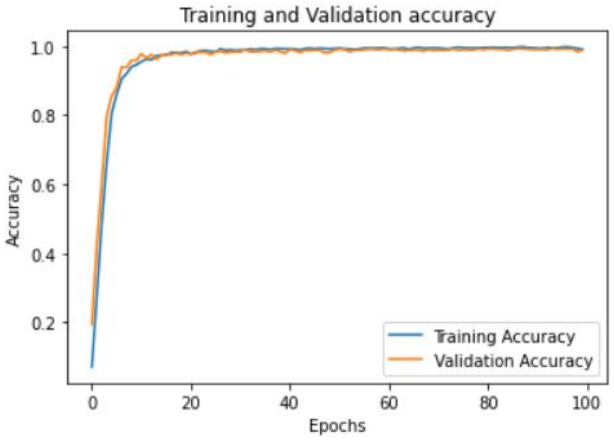
3. Apply histogram equalization to the darkened validation data





4. Apply histogram equalization to all the data





Conclusions

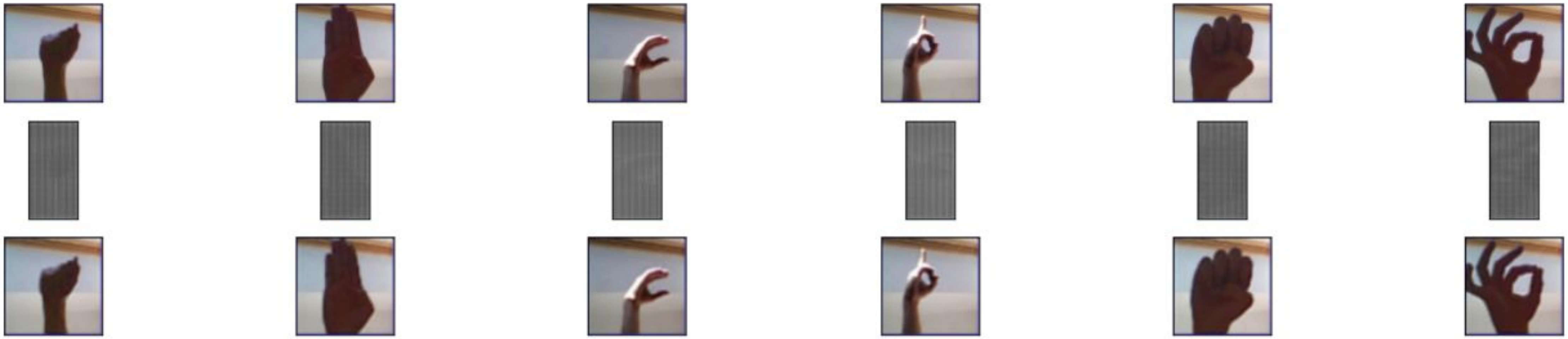
Lighting condition	Original	Darkened	Histogram equalization (Validation)	Histogram equalization (All)
Training accuracy	98%	99.39%	99.04%	99.17%
Validation accuracy	99.04%	13.14%	50.71%	**98.85%**
Training Loss	0.0923	0.0218	0.0346	0.0297
Validation Loss	0.0829	12.20	5.7989	0.0608

<u>We expect that if the input images can be histogram equalized before fed into the model, then the model can distinguish the letters under any environment!!</u>

Further work

- Test on new data

- Feature extraction using convolutional auto encoder



References

CNN Gesture Recognition:

[1]Goswami T., Javaji S.R. (2021) CNN Model for American Sign Language Recognition. In: Kumar A., Mozar S. (eds) ICCCE 2020. Lecture Notes in Electrical Engineering, vol 698. Springer, Singapore.

[2]Garcia, Brandon, and Sigberto Alarcon Viesca. "Real-time American sign language recognition with convolutional neural networks." *Convolutional Neural Networks for Visual Recognition* 2 (2016): 225-232.

[3]M. J. Hossein and M. Sabbir Ejaz, "Recognition of Bengali Sign Language using Novel Deep Convolutional Neural Network," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 2020, pp. 1-5, doi: 10.1109/STI50764.2020.9350418.

[4]K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[5]G. A. Rao, K. Syamala, P. V. V. Kishore and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 2018, pp. 194-197, doi: 10.1109/SPACES.2018.8316344.

References

Histogram Equalization:

[6]W. Okado, T. Goto, S. Hirano and M. Sakurai, "Fast and high-quality regional histogram equalization," *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, 2013, pp. 445-446, doi: 10.1109/GCCE.2013.6664884.

HSI:

[7]K. Yoshinari, K. Murahira, Y. Hoshi and A. Taguchi, "Color image enhancement in improved HSI color space," 2013 International Symposium on Intelligent Signal Processing and Communication Systems, 2013, pp. 429-434, doi: 10.1109/ISPACS.2013.6704588.