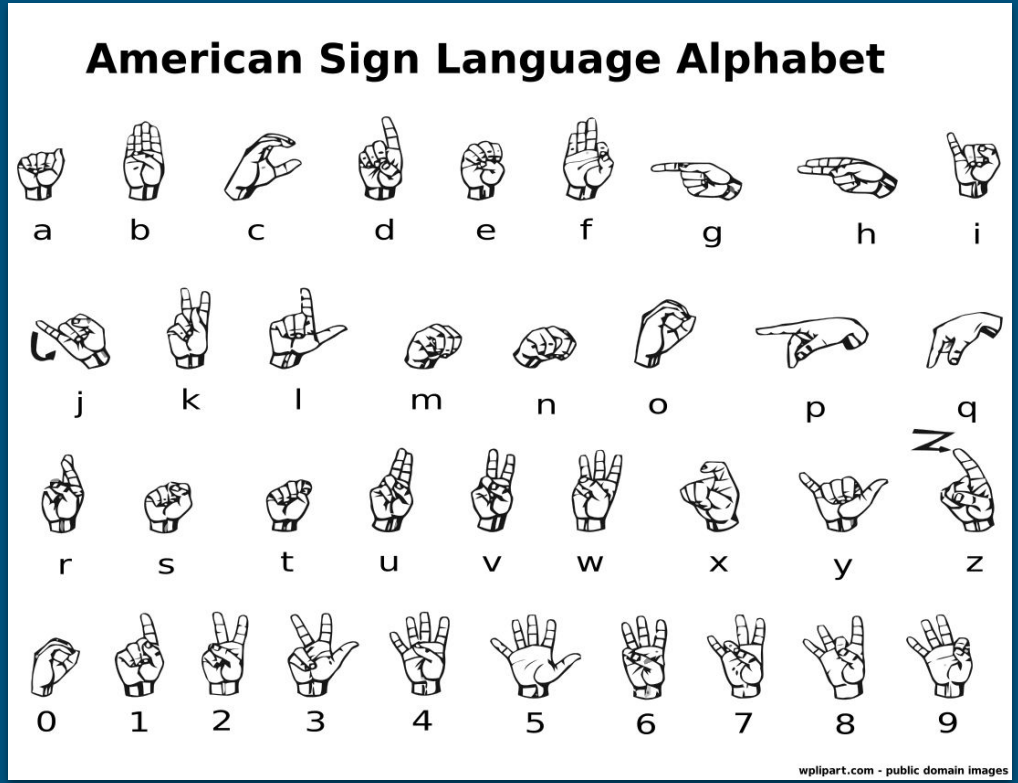
ASL(American Sign Language) Detection Using Convolutional Neural Networks

Presented By: Deep Dand
Assisted By: Shruti Kulkarni
Under Guidance of: Dr.Pablo Rivas

Concept:

- American sign language is method of communicating among the speech and hearing impaired community.
- The project goal is to create American sign language detection model using Convolutional neural network.

What is American Sign Language(ASL)?



- The sign language has same signs for “V” and “2”, “W” and “6”, “O” and “0”.
- Also “j” and “z” need hand gestures.
- Because of these reasons we are not including 5 classes and have 31 classes instead of 36 classes.

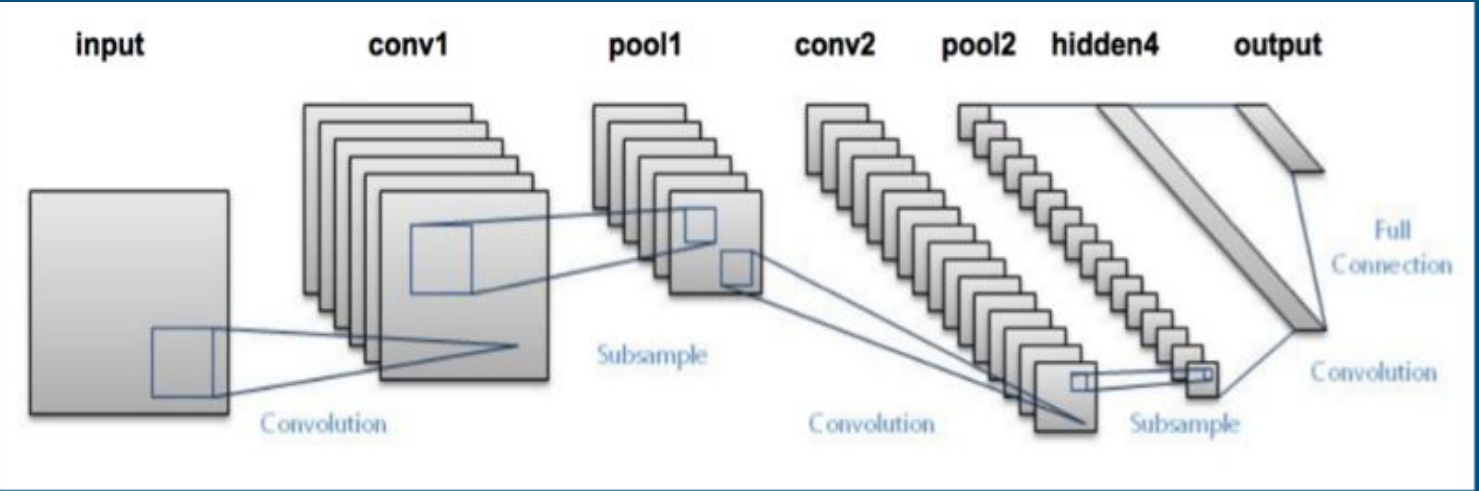
Existing Systems:

Author	Algorithm	Process	Result
Nachanau. M	SIFT	Preprocessing the Color images - 16x16 size	100% accuracy in test data
Brandon Garcia and Sigberto Alarcon Viesca	SVM	Utilized pretrained GoogLeNet architecture and resized images to 256x256	Letters a-y = 72% accuracy Letters a-k = 74% accuracy Letters a-e = 97% accuracy

About the Data:

- We use a dataset of segmented images captured with a depth-sensor camera for different subjects [1].
- We have data of five different subjects(From S01-S05).
- Each subject has 6,200 images .
- Images are divided into 31 different classes (From C01-C31).
- Currently, images of subject S01, S02, S03 are used for training and subject S04 and S05 are used for testing.

CNN Architecture:



CNN Architecture:

- Number of Layers - 2
- Layer 1
  - 32 filters
  - Relu activation
  - Maxpool filter size- 1
- Layer 2
  - 64 Filters
  - Relu activation
  - Maxpool filter size- 2
- Fully Connected Layer -
  - 512-1024 units.
  - Softmax activation
- Dropout layer - configured to drop results less than 0.8
- Number of Classes - 31
- IMG_SIZE = 256x256
- Learning rate - 1e3

Results:

Filter		Maxpool filter		Units in Fully Connected layer	Epochs	Test Accuracy	Val Accuracy
L1	L2	L1	L2				
3	3	1	2	512	5	0.04	0.03
5	5	1	1	512	3	NA	NA
5	5	1	1	756	3	NA	NA
5	5	1	2	512	3	0.03	0.03
5	5	1	2	512	5	0.03	0.03
3	3	3	3	1024	5	0.03	0.03
3	5	3	5	1024	5	0.037	0.032
5	3	5	3	1024	5	0.098	0.095
5	5	5	5	1024	5	0.9	0.9

Next experiments

Future Enhancements:

- Since the images we are dealing with are 256x256, adding more convnet layers will improve feature extraction and result in the better accuracy of the model.
- Also, once the experiments complete with existing config, the next config will have more epochs with architecture that gives best result with 5 epochs.

Acknowledgments

- I would like to thank Dr Pablo for the guidance in this project and Shruti for assisting in the research for CNN architecture.

References

[1] Byeongkeun Kang, Subarna Tripathi, and Truong Q Nguyen. Real-time sign language fingerspelling recognition using convolutional neural networks from depth map. In Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on, pages 136–140. IEEE, 2015.

Questions?

Thank You!