ASL to Text

Katie Bruett[1], Anthony Caivano[1], and Kimberly Ngai[1]
{klb270, apc87, kn246}@scarletmail.rutgers.edu
Advisor: Kristin Dana[1]
[1] *Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey*

Introduction

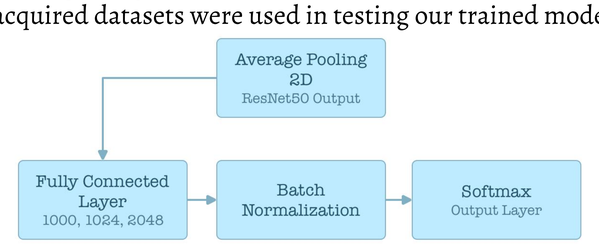
This project aims to help close the gap between the hearing and deaf communities through computer vision and deep learning. The program has been trained to pinpoint and differentiate between letters in the American Sign Language alphabet. It will then translate the signs to an easy-to-read text for the user to see what the signer is saying. This is done through a Raspberry Pi with a camera attachment, which is low-cost compared to previous work. Training and testing of the program was done through images and video collected from the Internet, previous research databases, and in person. The goal of this project is to make it easier for English speakers to understand any ASL they may encounter, and as a proof of concept that can later be expanded to more complicated signs.

Previous Work

- Last year's Capstone had a group that used the Leap Motion device for ASL recognition
- Previous research utilized the Microsoft Kinect to capture images and depth of ASL signs
- Expensive depth cameras have also been used for ASL recognition

Methodology

- **Convolutional Neural Network (CNN):** ResNet50 was chosen. It is a 50 layer residual neural network, pre-trained with ImageNet images. Pre-trained output weights were used in fine-tuning our network for image classification of letters
  - Addition of fully connected, batch normalization, and softmax layers for letter recognition
- **Training:** Over 100,000 static images of 24 letters of the alphabet. Images were pre-processed to create more variation. Therefore, features learned by the network would be lighting and pose independent
- **Testing:** Images both similar and different to that in our acquired datasets were used in testing our trained model



- After many alterations in fine-tuning our network, we were able to achieve almost 50% total accuracy for all 24 letters
- Smaller tests examining only a section of letters rather than all 24 yielded higher results
  - A-K gave over 93% testing accuracy (Fig. 3)
- M, N, and O are the main issue letters lowering accuracy in larger test batches
- There is still evidence of overfitting in our model as can be seen in Fig. 5, by the high validation and testing accuracies and low testing accuracy

Figure 1: Example Input of ASL letter "A"

```
[9.0561748e-01 1.8752752e-04 5.4413895e-03 2.2881843e-04 2.5585326e-03
 4.1687209e-04 2.2031943e-04 7.2387367e-04 2.0510899e-03 3.3048338e-05
 7.1134395e-03 6.5687753e-04 7.2267425e-04 2.1235552e-03 7.5420627e-05
 1.1124815e-03 7.3352517e-06 1.8587114e-02 2.4447393e-02 6.0902988e-05
 1.8148111e-04 4.3640661e-04 4.4352558e-04 2.6552342e-02]
I think the letter is A with 90.56174755096436 % confidence.
Actual letter:  A .
```

Figure 2: Example Output of accurately identified ASL letter "A"

Challenges

- We were unable to achieve training on the two letters (J and Z) of the ASL alphabet that had insufficient data to train with
- Use of ResNet50 does not allow independent usage of the Raspberry Pi
  - MobileNets were not able to identify the letters of the alphabet accurately enough
- Accuracy of fine-tuned ResNet50 model for 24 different classes of hands
  - Model often mixed up similar looking letters, e.g. "M" and "N"
- Large amount of dataset letters were front-facing images, making discrepancies between letters like "E" and "O" very hard to capture

Acknowledgments

Thank you to our advisor, Kristin Dana, for her input and help throughout the duration of our capstone design

Results

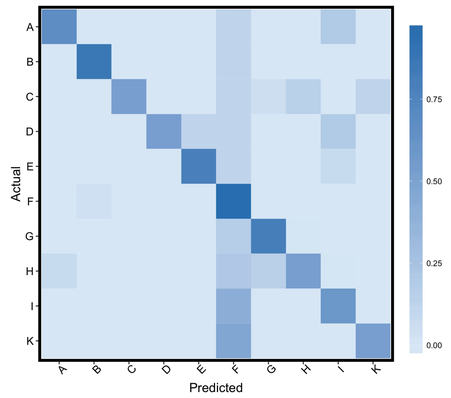


Figure 3: Confusion Matrix for letters A through K. Every letter tested with 75 images

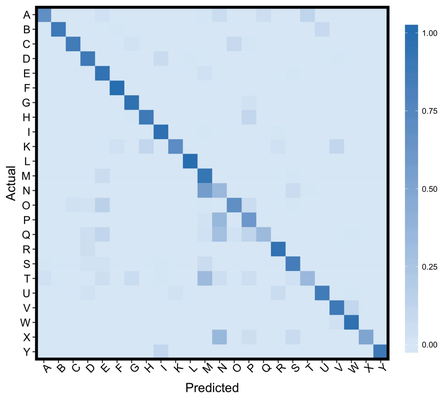


Figure 4: Confusion Matrix for all letters, sans J and Z. Every letter tested with 75 images

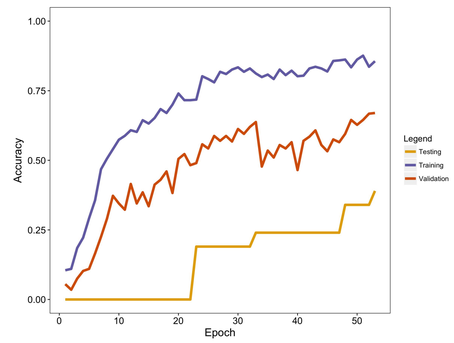


Figure 5: Line Graph of Training, Validation, and Testing Accuracies over 50 epochs of all letters in ASL alphabet, sans J and Z

Discussion

- There is a decent level of accuracy for determining individual letters of the ASL alphabet, but it can be improved
- Better results may have been achieved via a self-made CNN and database rather than a pre-trained network and online databases
  - Time restrictions were a large factor in the decisions made
- Model's accuracy is dependent on the images used
  - Images with controlled lighting and background are easily identified
  - Images with poor resolution and quality are often classified incorrectly

References

Sign language and static gesture recognition using sklearn. https://github.com/mon95/Sign- Language-and-Static-gesture-recognition-using-sklearn.

N. Pugeault and R. Bowden. Spelling it out: Real-time asl fingerspelling recognition. 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision, 2011.