

# American Sign Language interaction with Robot

Jeremy Hale

**Abstract**—There are roughly half a million users of American Sign Language (ASL) [1]. With the growing presence of companion robots in the world, it is necessary for the Deaf to be able to communicate with Robots using ASL.

We present a method for recognizing the hand gestures used in ASL to represent the latin alphabet using neural networks.

**Index Terms**—Robot, IEEEtran, Udacity, L<sup>A</sup>T<sub>E</sub>X, deep learning.

## 1 INTRODUCTION

WHILE detailed research into the history of communication between deaf persons and robots was not conducted, the need for this communication to exist is readily apparent. The data science website, Kaggle, has a dataset for mnist using ASL [2], however, a new dataset was created to fulfill the requirements of the project. Furthermore, the Kaggle dataset is in tabular form as opposed to image form and not ready to be used with convolutional neural networks.

The dataset includes the entire latin alphabet except for letters "j" and "z" which require motion. A neural network that reconigzes motion gestures is beyond the scope of the project.

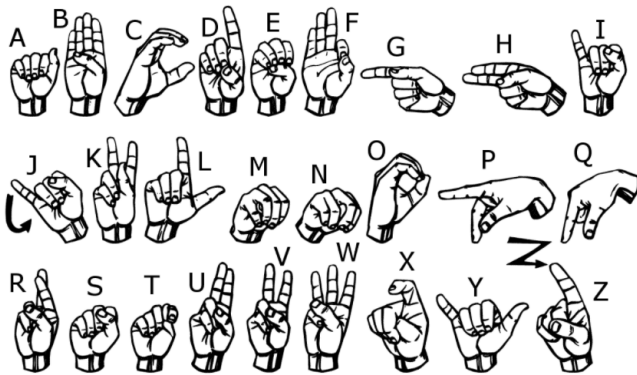


Fig. 1. American Sign Language. [2]

## 2 BACKGROUND / FORMULATION

The dataset was classified using a LeNet [3] network which is available as a standard network in the nVidia DIGITS application. The network is a simple 2 block network with each block consisting of a convolutional neural network (CNN) followed by max pool layer. The CNNs have kernel size 5 and stride 1, while the max pool layers have kernel size 2 and stride 2. The CNN blocks connect to a fully-connected layer with 500 units followed by a fully-connected with the number of units equal to the number of classes.

The number of epochs was set to 30 while the learning rate was the default 0.01.

In general, the network with the fewest parameters that yields sufficient accuracy should be chosen. In this case, the LeNet is the simplest model available and was able to perform satisfactorily.

## 3 DATA ACQUISITION

The nVidia Jetson TX2 was used for acquiring a dataset of 11,000 images. The udacity-modification of Peter Morgan's code [4] was programmed on to the TX2 and then used to capture images of the author's hand in various ASL poses. Approximately 400 images of each letter gesture were capture for the dataset.

The images were captured using the builtin camera on the TX2 against a white wall. The author attempted to keep their body and face outside of the field of view and to rotate their hand as well as move their hand around the field of view to generate a variety of camera angles.

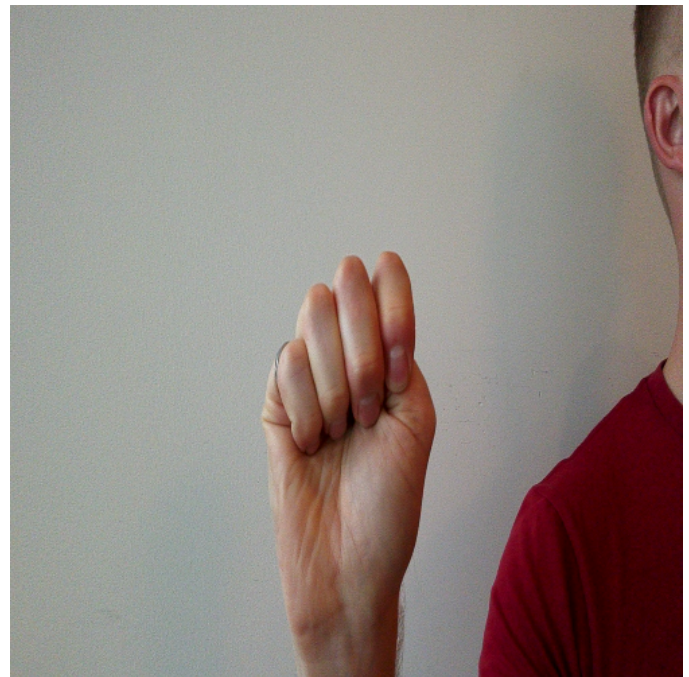


Fig. 2. Letter "M".

The images were then resized to 28 x 28 using Python code from [5] to work with the expected input size for

LeNet. Using nVidia DIGITS, the images were imported into a dataset of 8272 training images and 2755 validation images. They were imported as grayscale images.

## 4 RESULTS

The LeNet network achieved greater than 99 percent accuracy on the validation data after only 5 epochs.

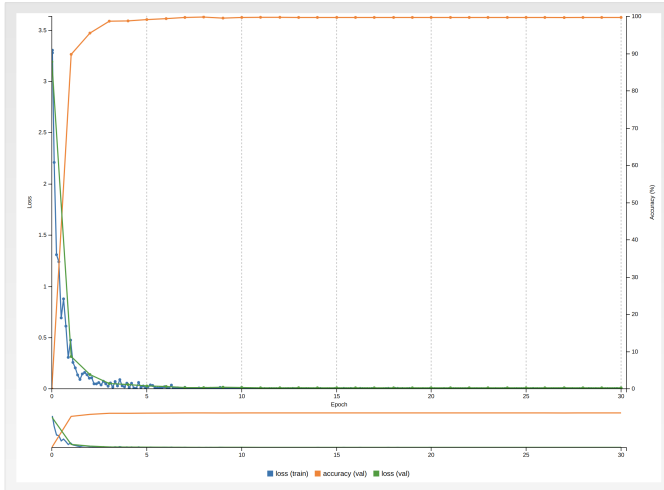


Fig. 3. Results.

## 5 DISCUSSION

The excellent results are likely due to the limitations of the dataset. The dataset most likely has similarities between images in the same class that are not related to the hand gesture being made. For example, the author's red shirt is visible in most the the label "M" images but not other labels. It's possible the network was able to "cheat" by learning to recognize the presence of the shirt instead of the position of the hand.

## 6 CONCLUSION / FUTURE WORK

The work presented here shows that it is indeed feasible to have a robot recognize the ASL alphabet and that a simple CNN-based network can be employed accurately.

For a real-world application, the network would need to be improved to not only recognize the letters "J" and "Z" which include motion, but also to include an entire vocabulary of ASL words, many of which include motion.

## REFERENCES

- [1] Wikipedia contributors, "American sign language — Wikipedia, the free encyclopedia." [https://en.wikipedia.org/w/index.php?title=American\\_Sign\\_Language&oldid=859986021](https://en.wikipedia.org/w/index.php?title=American_Sign_Language&oldid=859986021), 2018. [Online; accessed 8-October-2018].
- [2] tecperson, "Sign language mnist." <https://www.kaggle.com/datamunge/sign-language-mnist>, 2018. [Online; accessed 8-October-2018].
- [3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, pp. 2278–2324, 1998.
- [4] P. Moran, "Csi cameras on the tx2 (the easy way)." <http://petermoran.org/csi-cameras-on-tx2/>, 2017.
- [5] B. K. Ir, "Resize thousands of images with python." <https://dzone.com/articles/resize-thousands-images-python>, 2014.