

American Sign Language (ASL) Fingerspelling Detection

Shoaib Mohammed

01 June 2021

Abstract

Approximately 70 million people around the world are deaf-mute. While translation services have become easily accessible for about 100 languages, sign language is still an area that has not been explored. Our goal is to detect & translate the letters of ASL in real-time.

Keywords — ASL, fingerspelling, Convolutional Neural Network (CNN), transfer learning, computer vision

1 Introduction

1.1 Background

According to the Communication Service for the Deaf (CSD) [1], there are 360 million deaf people worldwide. Another report by the World Health Organization (WHO) [2] bumps up the number to 466 million people suffering from disabling hearing loss. Future projections estimate 630 million people by 2030 and over 900 million people by 2050. But even in this age of technology and communication, we are yet to see a universal translation system that helps bridge the gap between people that can and cannot speak. ASL is a natural language meaning it was not created and was spread by the people who employ the signs by the movement of hands, facial expressions, and body posture. The goal of this project is to detect and accurately translate the letters in ASL.

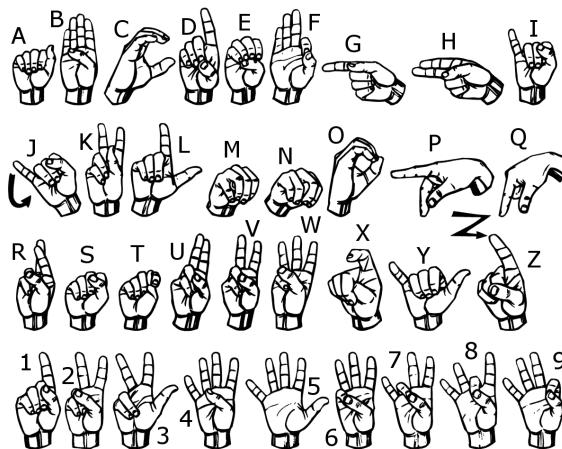


Figure 1: Alphabets in ASL

1.2 Geographical Distribution

The true count for the number of sign languages is still unknown given the vast majority, however, *Ethnologue* [3] lists this number to be 137 [4]. Given the vast majority, ASL is still the most popular sign language and is being widely used around the globe. In addition to being the primary source of communication for a sign language in the United States, ASL is being used throughout most of the provinces in Canada [5].



Figure 2: ASL being used around the world

Variations of ASL are also being used worldwide. Sign language similar to ASL is being used throughout Africa in places such as Nigeria, Ghana, Guyana, Central African Republic, Jamaica, Zimbabwe, and Kenya [6].

1.3 Scope & Limitations

The bottom line is that there is no universal sign language. For instance, the British Sign Language (BSL) differs by a great margin from ASL, this is clear when comparing Figure 1 with Figure 3. Generally speaking, a person in the US can understand spoken English in the UK but this is not the case with sign language. Even though there is a multitude of sign languages being used across the world, if an accurate model was developed to recognize a sign language, in our case, ASL, the same methodology could be applied to recognize other sign languages.

One big limitation for classification is that most signs require motion and are not static. Even more so as there are signs which are formed by the same motion but the repetition or the number of times a motion is repeated differs. Things get even more challenging as

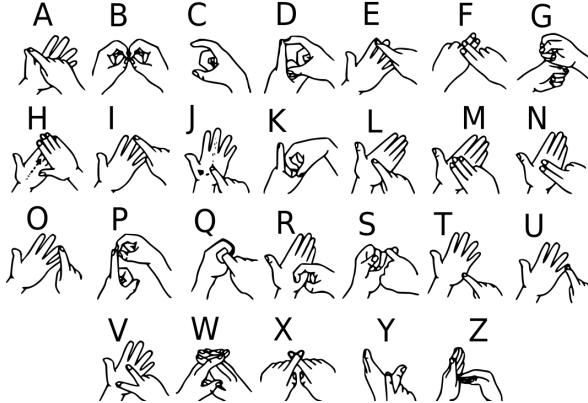


Figure 3: Alphabets in BSL

facial expressions are important in sign languages which is akin to the vocal tone of a person's voice when speaking. Two signs may be exactly the same visually but the face gesture of the signer makes them different.

Perhaps the major limitation is classifying each sign from the sheer corpus of signs in ASL. A dictionary on ASL contains illustrations for more than 1,600 signs [7]. However, our focus is classifying the alphabets alone which limits our range to 24 alphabets since we exclude the alphabets J & Z because these require motion.

2 Literature Survey

2.1 Population Statistics

A combined study in [8] estimates there were more than 250,000 deaf people and as many as 500,000 people who used ASL in 1972. Over the years, this number has been on the rise.

The Survey of Income and Program Participation (SIPP) collects data for the US population. From the research of Ross E. Mitchell [8] in the year 2006, fewer than 1 in 20 Americans or 10,000,000 people suffered from hard of hearing and close to 1,000,000 were classified as functionally deaf.

References

- [1] C. Soukup, "Communication service for the deaf," 1975. <https://www.csd.org/>, Last accessed on 2021-05-31.

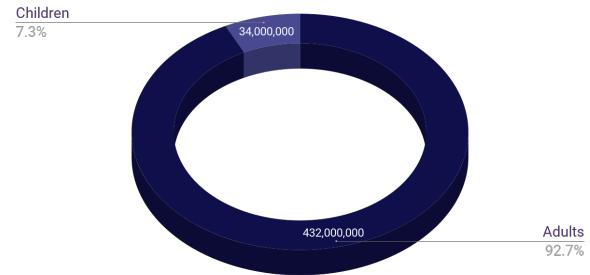


Figure 4: Distribution of the population

- [2] R. E. C. David W Dowdy, "World health organization. office of library and health literature services," 1988. <https://www.who.int/>, Last accessed on 2021-05-31.
- [3] S. International, "Ethnologue languages of the world," 2000. <https://www.ethnologue.com/>, Last accessed on 2021-05-31.
- [4] J. Fenlon and E. Wilkinson, "Sign languages in the world," *Sociolinguistics and Deaf communities*, pp. 5–28, 2015.
- [5] K. Al-Fityani and C. Padden, "Sign language geography in the arab world," *Sign languages: A Cambridge survey*, vol. 20, 2010.
- [6] V. Nyst, "Sign languages in west africa," *Sign languages*, pp. 405–432, 2010.
- [7] R. A. Tennant, M. Gluszak, and M. G. Brown, *The American sign language handshape dictionary*. Gallaudet University Press, 1998.
- [8] R. E. Mitchell, T. A. Young, B. Bachelda, and M. A. Karchmer, "How many people use asl in the united states? why estimates need updating," *Sign Language Studies*, vol. 6, no. 3, pp. 306–335, 2006.

List of Figures

1	Alphabets in ASL	1
2	ASL being used around the world	1
3	Alphabets in BSL	2
4	Distribution of the population	2
5	Process of serialization & deserialization	4
6	ASL alphabets from our dataset	4

7	Training from scratch vs. transfer learning	4
8	Transfer learning process	4
9	Max pooling vs. Average pooling	5
10	Softmax function pushing high scores close to 1 and low scores close to 0	5
11	Process of training the model	5
12	Process of testing the model	5

List of Tables

1	Transfer learning model statistics	6
2	Trained model statistics using varying transfer learning models	6
3	Model architecture and number of parameters in each layer	7
4	Shape of each layer in trained model for different transfer learning models	8

Additional Resources

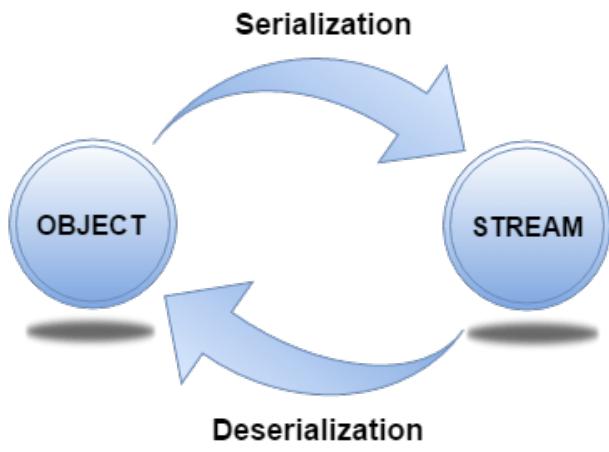


Figure 5: Process of serialization & deserialization



Figure 6: ASL alphabets from our dataset

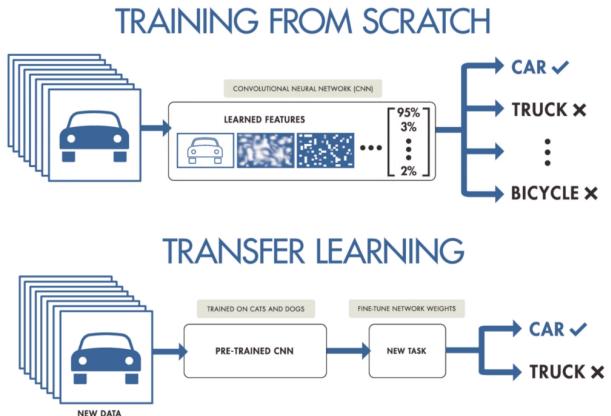


Figure 7: Training from scratch vs. transfer learning

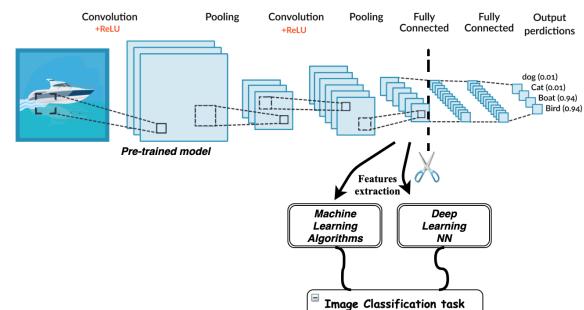


Figure 8: Transfer learning process

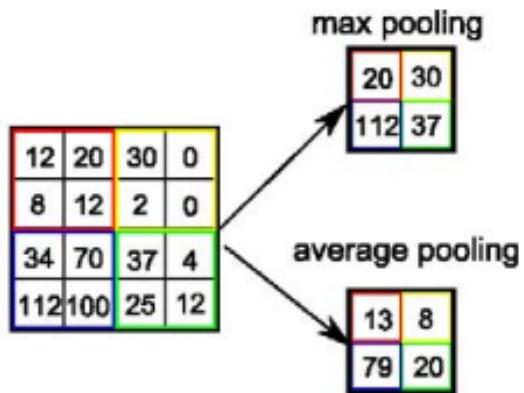


Figure 11: Process of training the model

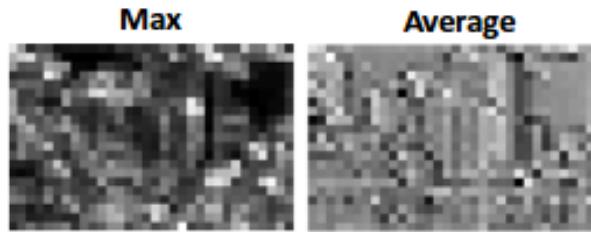


Figure 9: Max pooling vs. Average pooling

$$\begin{array}{c}
 \text{LOGITS SCORES} \\
 \circ \quad \text{SOFTMAX} \\
 \text{PROBABILITIES}
 \end{array}
 \quad
 \begin{aligned}
 y &\left[\begin{array}{l} 2.0 \\ 1.0 \\ 0.1 \end{array} \right] \longrightarrow S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \longrightarrow p = 0.7 \\
 &\longrightarrow p = 0.2 \\
 &\longrightarrow p = 0.1
 \end{aligned}$$

Figure 10: Softmax function pushing high scores close to 1 and low scores close to 0

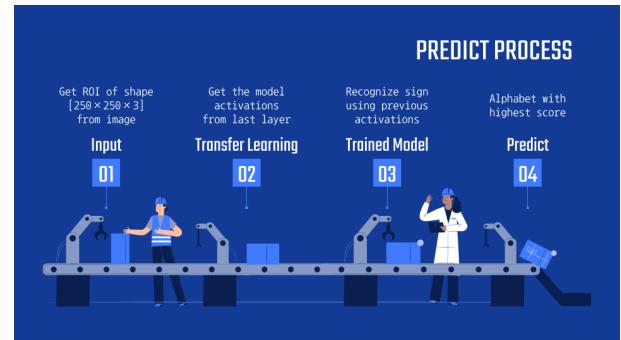


Figure 12: Process of testing the model

Model	Size	Accuracy	Parameters
MobileNetV2	14 MB	0.713	3,538,984
InceptionV3	92 MB	0.779	23,851,784
Xception	88 MB	0.790	22,910,480
InceptionResNetV2	215 MB	0.803	55,873,736

Table 1: Transfer learning model statistics

Model	Size	Accuracy	Parameters
MobileNetV2	91 MB	0.918	7,934,872
InceptionV3	99 MB	0.924	8,598,424
Xception	100 MB	0.908	8,721,304
InceptionResNetV2	93 MB	0.848	8,074,136

Table 2: Trained model statistics using varying transfer learning models

ID	Layer (Type)	Number of Parameters
1	dense_1 (Dense)	<i>dependent</i>
2	dense_2 (Dense)	524,800
3	dense_3 (Dense)	131,328
4	dense_4 (Dense)	32,896
5	up_sampling2d_1 (UpSampling2D)	0
6	conv2d_5 (Conv2D)	131,136
7	depthwise_conv2d_1 (DepthwiseConv2D)	1,088
8	up_sampling2d_2 (UpSampling2D)	0
9	depthwise_conv2d_2 (DepthwiseConv2D)	1,088
10	conv2d_6 (Conv2D)	65,600
11	dense_5 (Dense)	8,320
12	dense_6 (Dense)	33,024
13	dense_7 (Dense)	131,584
14	dense_8 (Dense)	525,312
15	dense_9 (Dense)	2,099,200
16	dense_10 (Dense)	2,098,176
17	dense_11 (Dense)	524,800
18	dense_12 (Dense)	131,328
19	dense_13 (Dense)	32,896
20	max_pooling2d_1 (MaxPooling2D)	0
21	flatten_1 (Flatten)	0
22	dense_14 (Dense)	<i>dependent</i>

Table 3: Model architecture and number of parameters in each layer

ID	Layer (Type)	Number of Parameters
1	(None, 8, 8, 1024)	(None, 6, 6, 1024)
2	(None, 8, 8, 512)	(None, 6, 6, 512)
3	(None, 8, 8, 256)	(None, 6, 6, 256)
4	(None, 8, 8, 128)	(None, 6, 6, 128)
5	(None, 16, 16, 128)	(None, 12, 12, 128)
6	(None, 13, 13, 64)	(None, 9, 9, 64)
7	(None, 10, 10, 64)	(None, 6, 6, 64)
8	(None, 20, 20, 64)	(None, 12, 12, 64)
9	(None, 17, 17, 64)	(None, 9, 9, 64)
10	(None, 14, 14, 64)	(None, 6, 6, 64)
11	(None, 14, 14, 128)	(None, 6, 6, 128)
12	(None, 14, 14, 256)	(None, 6, 6, 256)
13	(None, 14, 14, 512)	(None, 6, 6, 512)
14	(None, 14, 14, 1024)	(None, 6, 6, 1024)
15	(None, 14, 14, 2048)	(None, 6, 6, 2048)
16	(None, 14, 14, 1024)	(None, 6, 6, 1024)
17	(None, 14, 14, 512)	(None, 6, 6, 512)
18	(None, 14, 14, 256)	(None, 6, 6, 256)
19	(None, 14, 14, 128)	(None, 6, 6, 128)
20	(None, 7, 7, 128)	(None, 3, 3, 128)
21	(None, 6272)	(None, 1152)
22	(None, 24)	

Table 4: Shape of each layer in trained model for different transfer learning models