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. The goal of this project is to detect and accurately translate the letters in ASL.

1.2 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 the ASL. Generally speaking, a person in the US can understand spoken English in the UK but this is not the case with sign language.

It is interesting to note that there are more than 200 sign languages that are used across the world. However, 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. Perhaps the major limitation is classifying each sign from the sheer corpus of signs in ASL. However, since our focus is classifying the alphabets alone, our range is limited to 24 alphabets since we exclude the alphabets J & Z since these require motion.

2 Literature Survey

2.1 Population Statistics

A combined study in [2] 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.

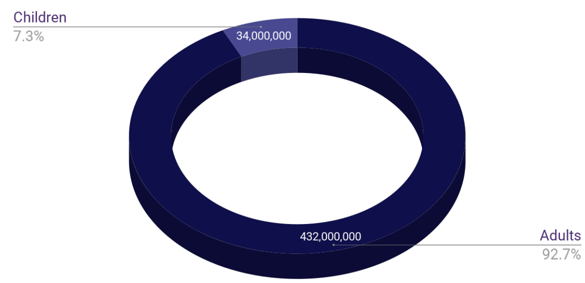

Figure 1: Distribution of the population

The Survey of Income and Program Participation (SIPP) collects data for the US population. From the research of Ross E. Mitchell [2] 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," 2021. `https://www.csd.org/`, Last accessed on 2021-05-31.

[2] 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

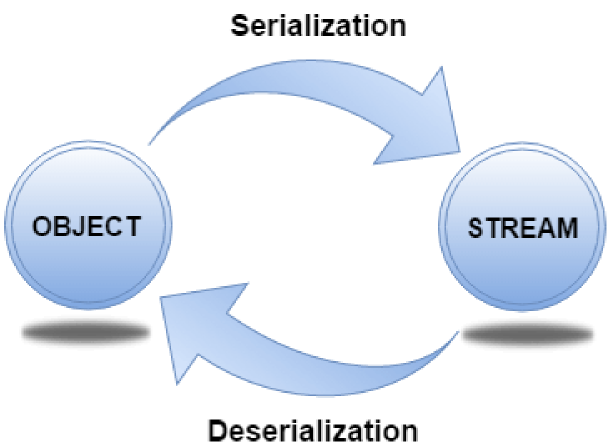

Figure 2: Process of serialization & deserialization

Model	Size	Accuracy	Paramet
MobileNetV2	14 MB	0.713	3,538,98
InceptionV3	92 MB	0.779	23,851,7
Xception	88 MB	0.790	22,910,480
InceptionResNetV2	215 MB	0.803	55,873,736

Table 1: Transfer learning model statistics

Figure 3: ASL alphabets from our dataset

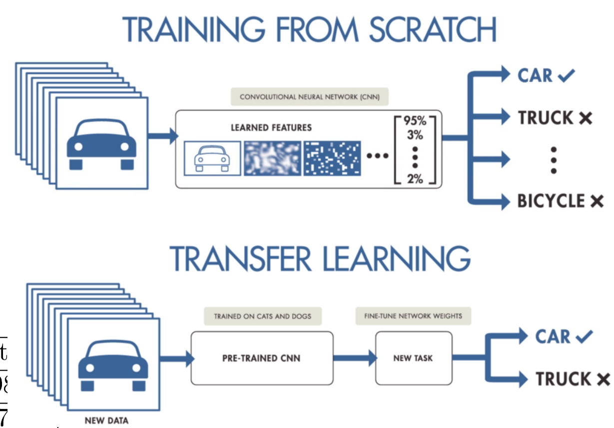

Figure 4: Training from scratch vs. transfer learning

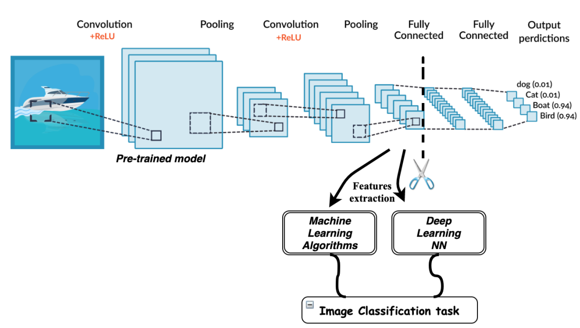

Figure 5: Transfer learning process

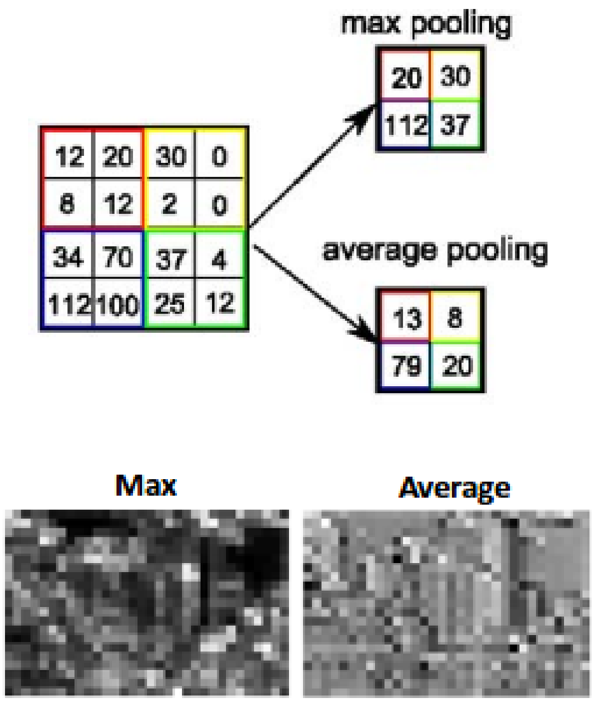

Figure 6: Max pooling vs. Average pooling

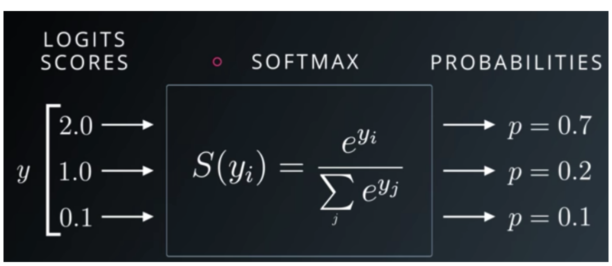

Figure 7: Softmax function pushing high scores close to 1 and low scores close to 0

Model	Size	Accuracy	Paramet
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

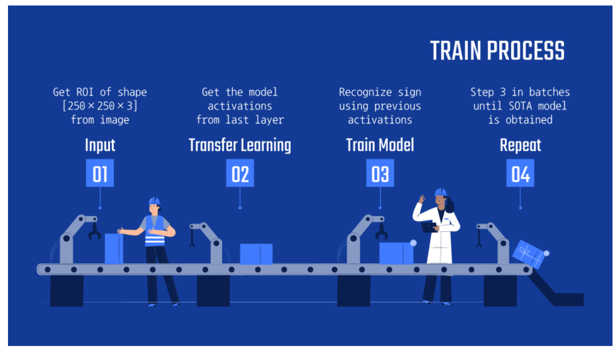

Figure 8: Process of training the model

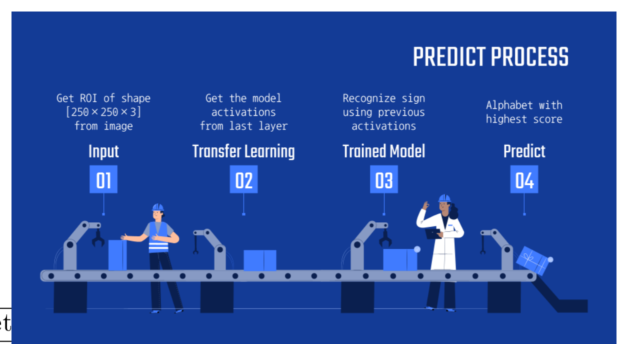

Figure 9: Process of testing the model

ID	Layer (Type)	Number of Parameters
1	dense_1 (Dense)	*dependent*
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)	*dependent*

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