Exploratory Data Analysis

3.6.1 WLASL Dataset

In WLASL, as the dataset was a collection of videos from sign language educational website and YouTube, the videos were recorded in different background and under different settings. Majority of the words were presented with at least two sign variations due to the existence of dialect in sign language. As presented in table 3, each word consists of at most 7 videos recorded with 3 to 7 different signers. Even though there are less than 3 videos difference between each class, each video stands a high proportion in contributing to the specific class, this dataset is still considered imbalance.

ID	Word	Number of videos	Number of signers
01	Africa	6	6
02	After	7	7
03	Again	6	6
04	Ago	5	4
05	Allday	5	5
06	And	5	5
07	Bacon	5	5
08	Baseball	5	5
09	Because	6	3
10	Beginning	5	1
11	Benefit	5	5
12	Birthday	5	5
13	Camera	5	5
14	Can	5	4
15	card	5	5

Table 3: Number of videos and signers for each word in WLASL dataset.

As shown in figure 11, the frame count of the videos range between 25 to 150. Thus, the total number of frames to be extracted for method 1, MediaPipe Keypoints was 150 frames. Majority of the videos have 30 frames per second, with a combination of different resolutions. Hence, for method 2 which uses frame sequences, the resolution of all videos were hold constant at 100*100 for model 1 and 80*80 for model 2.

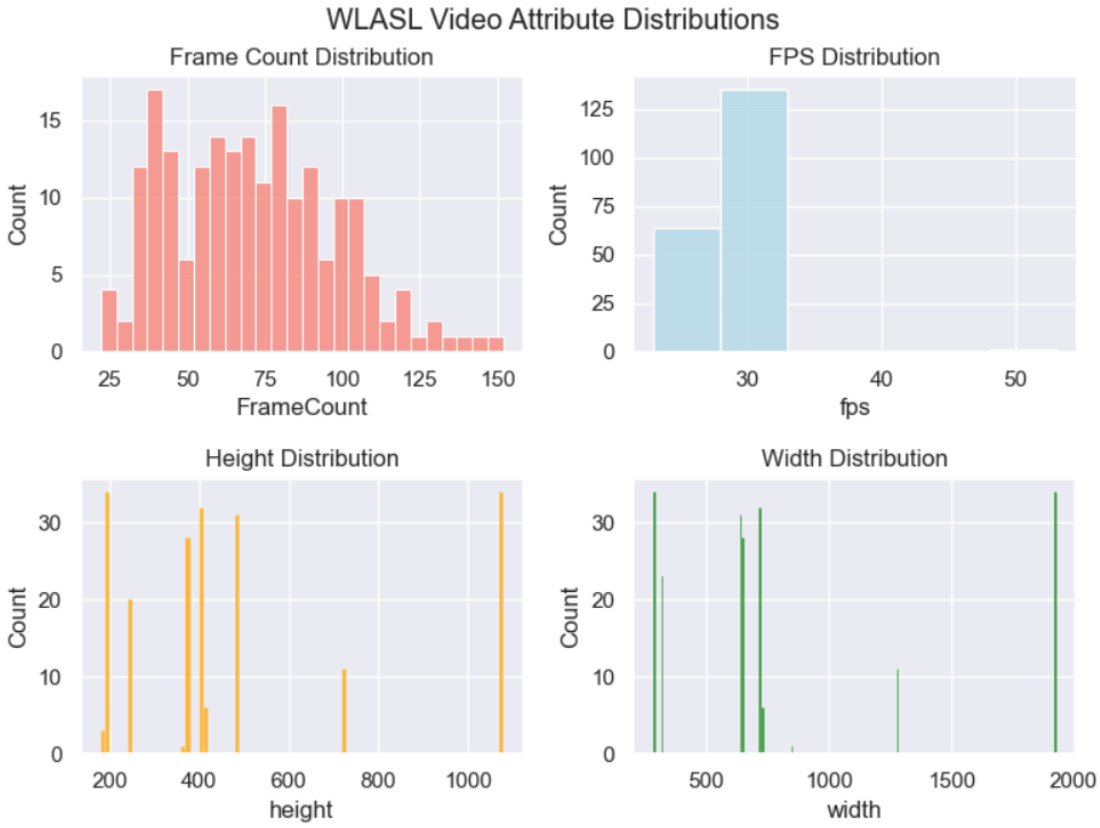

Figure 1 Video Attribute Distribution of WLASL Dataset

It was also noted some of the words consists of more than one type of sign variations. As noted in figure 12, the movement and position of the hands are different between these signers for the same word 'add'. The signer at the top row raised his right hand from bottom to top while holding his left hand still. On the other hand, the signer at the bottom row joined both hands together from each side. Due to the small amount of data available, for each word, only videos with the same sign were used. Thus, the selection of the 15 words subset used in this study were also depending on the number of videos available.

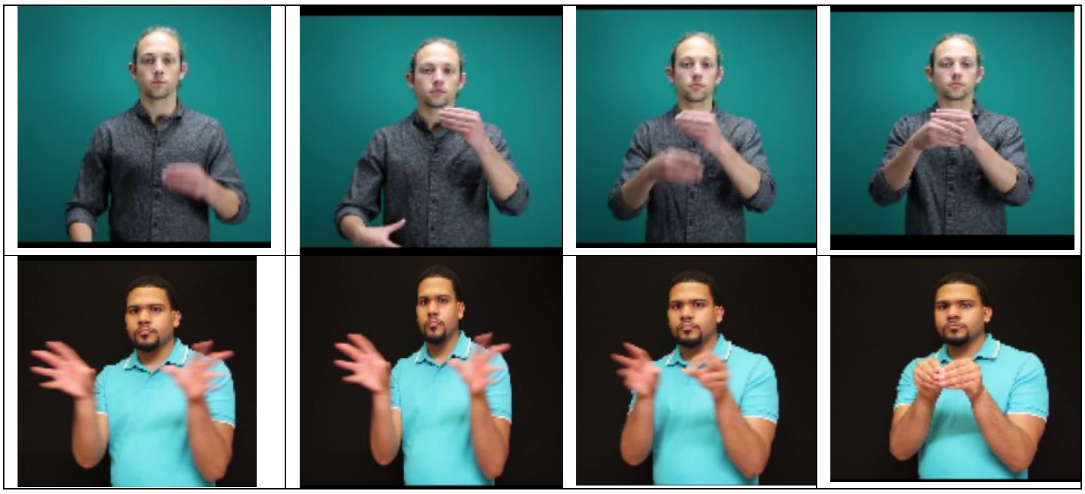

Figure 2 Different sequence of hand sign position for the word 'add'.

3.6.2 LSA64 Dataset

In LSA64 dataset, each word covers only one type of sign performed by 10 different subjects with 5 repetitions. Hence, the dataset consists of 50 videos. All videos have a resolution of 1080*1920 and fps 60 (Rochetti et al., 2016). As shown in figure 13, the frame count of the videos range between 60 to 180. Thus, the total number of frames to be extracted for method 1, MediaPipe Keypoints was 190 frames. Similar to WLASL dataset, for method 2 which uses frame sequences, the resolution of all videos were hold constant at 100*100 for model 1 and 80*80 for model 2.

Table 3.4: 15 Words from the LSA64 dataset used for this study.

ID	Word	ID	Word
01	Accept	09	Born
02	Appear	10	Breakfast
03	Argentina	11	Bright
04	Away	12	Buy
05	Barbecue	13	Call
06	Bathe	14	Candy
07	Birthday	15	Catch
08	Bitter		

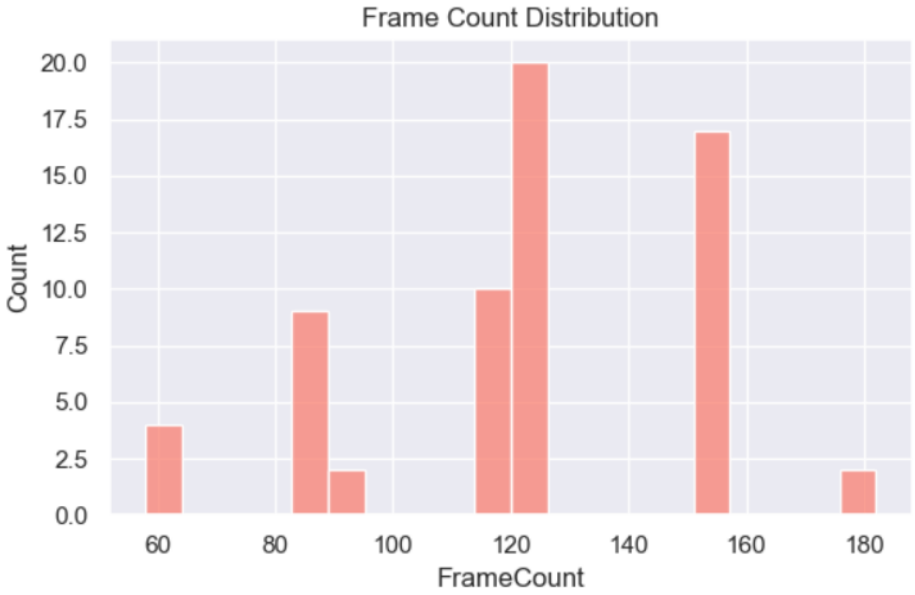

Figure 3 Frame Count Distribution of videos of LSA64 dataset.