**krusty krab and big data**

Author Supervisor

MHD Khaled Maen Assoc. Prof Dr. Amelia Ritahani Ismail

mk.maen93@gmail.com amelia@iium.edu.my

Abstract

Communication is an essential part of our life. Unfortunately, some of us were born in various types of disability such as deaf, since hearing impaired people cannot listen they loosed the ability to learn how to speak so they developed a new communication way to interact with other people by using distinct hand gestures, which was not enough to overcome this issue since most of hearing people do not understand sign language, even now with all technologies and tools it is remain a challenging problem to solve. For the mentioned reason, the intention of this paper is to improve the ordinary model to translate the sign language gestures into a voice. In that, Deep learning is remarkably serviceable for this mission, first by detecting a hand in a video frame using Convolutional Neural Network (CNN) algorithm followed by recognizing the letter and state the matching sound. The accuracy achieved for a hand gesture detection using CNN model is more than 90% for all the proposed signs.

# 1. Introduction

Communication is a process of sending and receiving data among individuals. People communicate with each other’s by too many ways yet the best one is by talking. Numerous researchers trusted that the significance of communication is like the importance of breathing. Indeed, communication facilitates the spread of knowledge and structures connections between individuals.

Deep learning added an immense lift to the already rapidly developing field of computer vision. Moreover, a lot of new utilization of computer vision techniques have been presented and they are currently ending up some portion of our regular day to day existence.

Alongside with the intensity of the present computers, there are now various algorithms that were developed to empower computers to perform tasks such as object tracking and pattern recognition.

In this paper, the attention will be on hand gestures detection and make an interpretation of them into voice.

# 2. Related Work

In spoken languages words being pronounced by vocal cords which produce sound wave can be heard, however for deaf people this sounds are not heard [1], so sign language is based on the visual tools to communicate and receive information.

American Sign Language (ASL) is completely different from English. It has the basic features of a language such as the letters as shown in Figure 1.

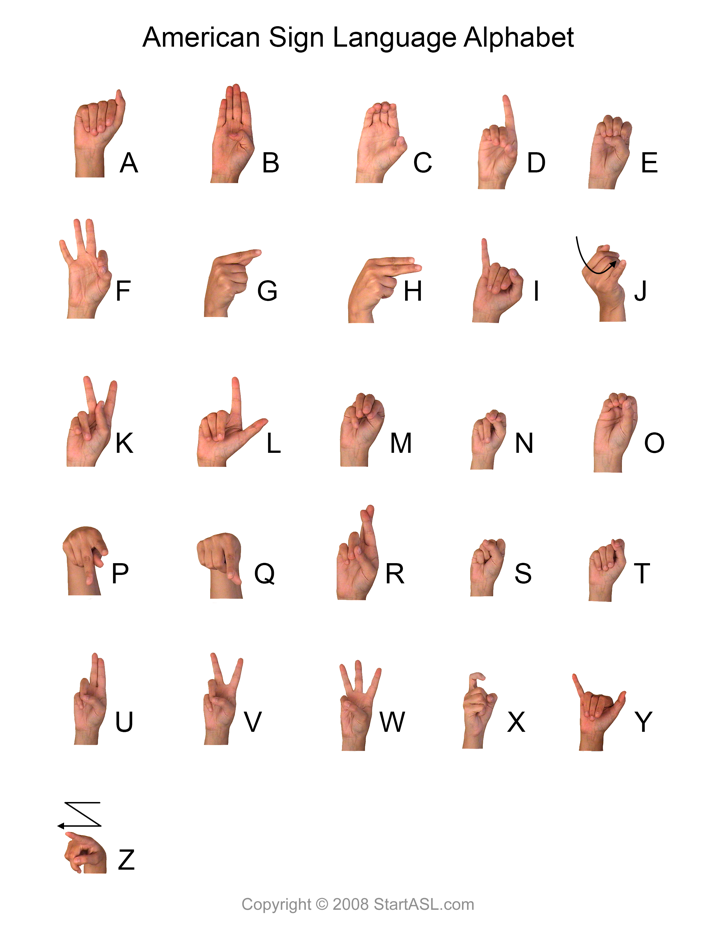


Figure 1- American Sign Language

Following the same path many previous attempts has been conducted to find a solution for this problem, marvelous work has been done, but most of them was just a prove of concept were the authors concentrate on a few signs to be detected.

Starting with Bao, Maqueda, del Blanco, & Garc´ıa

They proposed a Deep convolutional neural network algorithm for hand-gesture recognition for 7 different gestures without hand localisation, since the hands only occupy about 10% of the image. They used a combination of 9 convolution layers, 3 fully connected layers, interlaced with Rectified Linear Unit (ReLU) and dropout layers. Alongside this architecture the apply some image processing techniques to have sufficient computation efficiency and memory requirement. According to the paper the accuracy achieved was 97.1% in the images with simple backgrounds and 85.3% in the images with complex backgrounds [2].

The second work conducted by Pyo, Ji, You, & Kuc they introduced a depth-based hand data with convolution neural networks (CNNs).The hand gesture dataset has roughly 6,000 RGB-D images in each of 12 labels. In all, there are approximately 60,000 training images, 1S,000 validation images, and 12,000 training images [3].

# 3. Methodology

The main structure of our project is knowing students’ food preferences and how they will in the situation of opening a new restaurant. That's why we search for a suitable dataset for our projects.

## A picture containing text Description automatically generated

## 3.1. Business understating

The main target of Krusty Krab is university students, so in order to make a success restaurant is to have a menu that satisfy the taste of majority students.

## 3.2. Data set

The data set is a collection of a 61 attributes include information on food choices, nutrition, preferences, childhood favorites, and other information from college students (GPA, Gender, breakfast, calories chicken, calories day, calories scone, coffee, comfort food, comfort food reasons, comfort food reasons coded, cook, comfort food reasons coded, cuisine, diet current, diet current coded, drink, eating changes, eating changes coded, eating changes coded1, eating out, employment, ethnic food, exercise, father education, father profession, fav cuisine, fav cuisine coded, fav food, food childhood, fries, fruit day, grade level, Greek food, healthy feeling, healthy meal, ideal diet, ideal diet coded, income, Indian food, Italian food, life rewarding, marital status, meals dinner friend, mother education, mother profession, nutritional check, on off campus, parents cook, pay meal out, Persian food, self-perception weight, soup, sports, Thai food, tortilla calories, turkey calories, type sports, veggies day, vitamins, waffle calories, weight)[4].

A screenshot of a cell phone

Description automatically generated

Fig. 1 Visualization of some data set attributes

# 3.3. Data preprocessing

In order to use the data set many operation has to be done, starting with data encoding by transfer nominal variables into numerical variables then filling the null column values with proper value in our case we filled the null values with the median of the column that belongs to, after that non related column should be dropped which was 13 columns in this case.

# 3.4. Algorithm

Since our problem is unsupervised machine learning clustering K-Means tend to be suitable for the task. K-Means is one of the most popular clustering algorithms, it stores K centroids used to define the clusters [5]. A point is considered to be in a particular cluster if it is closer to that cluster's centroid than any other centroid. K-Means finds the best centroids by alternating between (1) assigning data points to clusters based on the current centroids (2) choosing centroids (points which are the centre of a cluster) based on the current assignment of data points to clusters and that can be done after many iterations as shown in Fig.2.

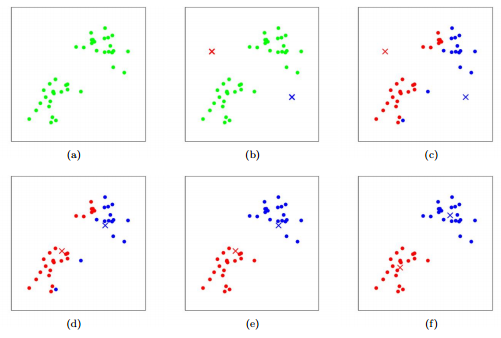


Fig. 2 K-Means Visualization

# 3.4. Evaluation

Unlike supervised learning where there is a ground truth for model performance evaluation, clustering algorithms including K-Means don’t have a solid evaluation metric that can be used to evaluate the model outcome. However, K-Means require k as input and there is no right answer in terms of clusters’ number, some methods are useful to give an intuition about k:

# 3.4.1. Elbow Method

Elbow method helps to find out what a good k number of clusters based on the sum of squared distance (SSE) as shown in Fig.4.

A close up of a map

Description automatically generated

Fig. 3 Elbow Method

In this case the target is to have a big cluster of student so the selected value of k is 3, it might not be the optimal value but it should be acceptable for this kind of tasks.

# 3.4.2. Silhouette Analysis

Silhouette analysis can be used to determine the degree of separation between clusters. For each sample it computes (1) the average distance from all data point in the same cluster. (2) the average distance from all data points in the closest cluster. (3) the coefficient.



Fig. 4 Silhouette Analysis

# 4. Results

After 13587 steps the model being able to detect and identify each and every sign of the 29 signs including the empty one with classification loss 0.143 and localization loss 0.0327.

**5. Feature works**

One of the potential improvements is to collect data for complete words and build a model which translate the voice into signs to make it two-ways connections.

**6. References**

[1] K. Ishizaka and J. L. Flanagan, “Synthesis of Voiced Sounds From a Two‐Mass Model of the Vocal Cords,” *Bell Syst. Tech. J.*, 1972.

[2] P. Bao, A. I. Maqueda, C. R. Del-Blanco, and N. Garciá, “Tiny hand gesture recognition without localization via a deep convolutional network,” *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 251–257, 2017.

[3] J. Pyo, S. Ji, S. You, and T. Kuc, “Depth-based hand gesture recognition using convolutional neural networks,” in *2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI 2016*, 2016.

[4] “Food choices | Kaggle.” [Online]. Available: https://www.kaggle.com/borapajo/food-choices.

[5] “K-Means.” [Online]. Available: http://stanford.edu/~cpiech/cs221/handouts/kmeans.html.

[6] “TensorFlow | Object\_detection.” [Online]. Available: https://github.com/tensorflow/models/tree/master/research/object\_detection.

[7] “MobileNet-SSD | Architecture.” [Online]. Available: https://hey-yahei.cn/2018/08/08/MobileNets-SSD/index.html.

[8] A. G. Howard *et al.*, “MobileNets,” *arXiv Prepr. arXiv1704.04861*, 2017.

[9] W. Liu *et al.*, “SSD: Single shot multibox detector,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016.