CSMSS
CHH. SHAHU COLLEGE OF ENGINEERING,
CHHATRAPATI SAMBHAJINAGAR, MAHARASHTRA - 431011

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND

DATA SCIENCE ENGINEERING

A

PROJECT TITLE ON

"Indian Sign Language to Text/Speech Translation"

Under the Guidance of

Dr. S.R. Zanwar.

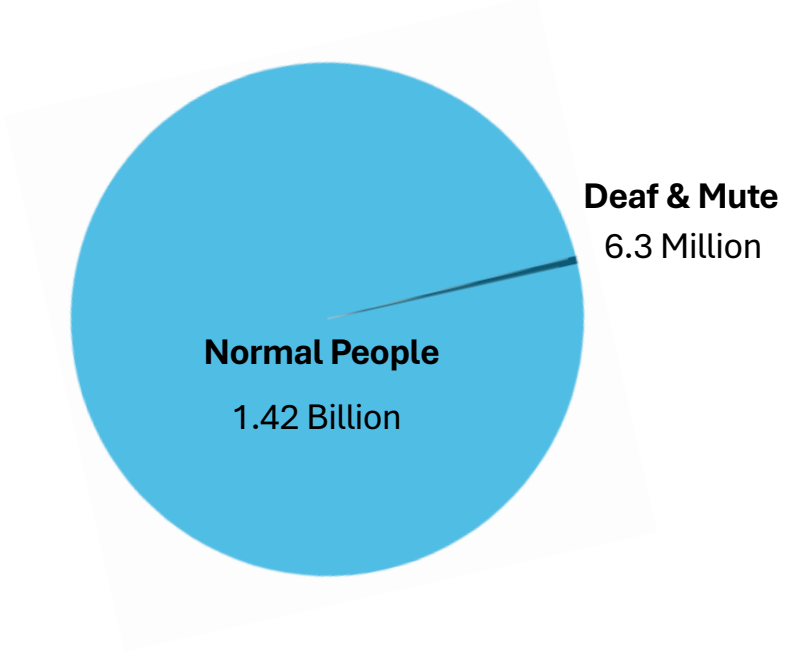
Submitted by

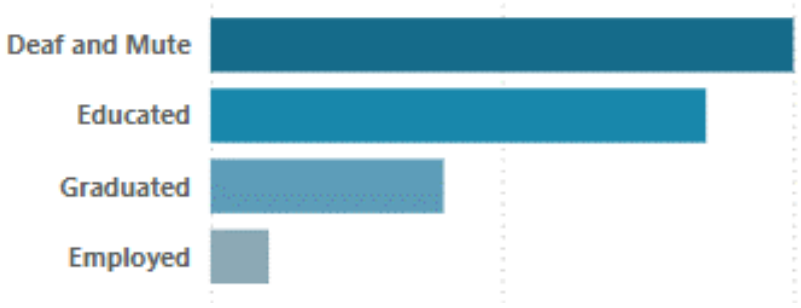
Swaraj Solanke(AI4072)
Parth Dhone (AI4073)
Pradyumna Digraskar (AI4075)

INDEX

INTRODUCTION

- Indian Sign Language is a predominant sign language since the only disability D&M people have been communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language.
- Communication is a vital aspect of human interaction, enabling individuals to express their thoughts, emotions, and needs. However, for individuals with speech or hearing impairments, traditional communication methods can be challenging.
- Indian Sign Language (ISL) is widely used by the deaf and hard-of-hearing communities in India as a means of communication.





LITERATURE SURVEY

Sr. No.	Title of paper	Author	Year	Methodology	Findings
[1]	Real-Time Sign Language Detection Using TensorFlow, OpenCV, and Python	Prashant Verma	2024	Discusses a system for real- time detection of sign language using TensorFlow and OpenCV.	Focuses on Indian Sign Language (ISL) and leverages machine learning for gesture-to-text conversion. Uses computer vision techniques to avoid wearables and emphasizes GPU-accelerated training for efficient detection.
[2]	Sign Language Recognition Using Deep Learning	Harini Priya	2023	Explores using deep learning models like CNNs and LSTMs for ISL recognition.	Focuses on real-time applications and highlights the challenges in interpreting sequential gestures. The model integrates temporal features for better gesture-to- text translation accuracy
[3]	Conversion of Sign Language Into Text	Shalini Gupta,	2023	Proposes a vision-based sign language recognition system using gesture recognition.	The paper emphasizes low-cost solutions for ISL translation into text and explores gesture classification using machine learning models and image processing techniques
[4]	Sign Language Recognition Based on Computer Vision	Divya Sharma	2023	Investigates the use of computer vision and artificial neural networks.	Incorporates preprocessing techniques to enhance gesture detection and highlights the use of non- manual features like facial expressions in recognition tasks.
[5]	Sign Language Conversion to Text and Speech	Sneha Rao	2024	Develops a bilingual system for ISL, converting gestures to text and speech in regional languages.	Uses sequence-to- sequence deep learning models with attention mechanisms, showcasing high accuracy in aligning gestures with text and speech outputs.

Limitation of previous model

1. Unidirectional flow which in turn only make one way communication for normal people.

2. Hardware requirements

3. Less accuracy

4. No sentence Generation

5. No multi-lingo support

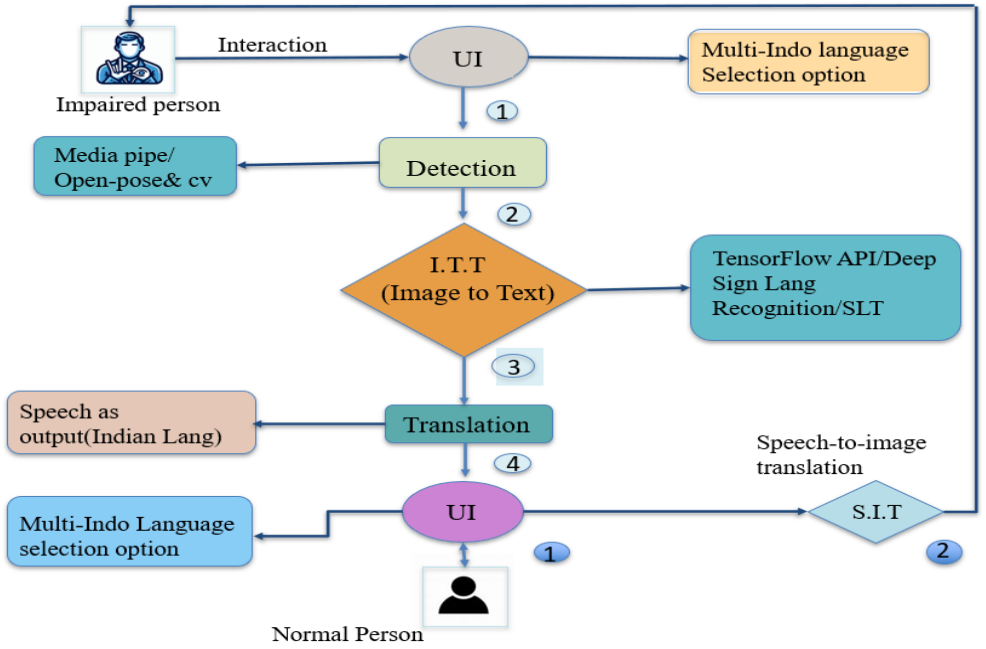
6. High infrastructure cost.

UNIQUENESS OF WORK

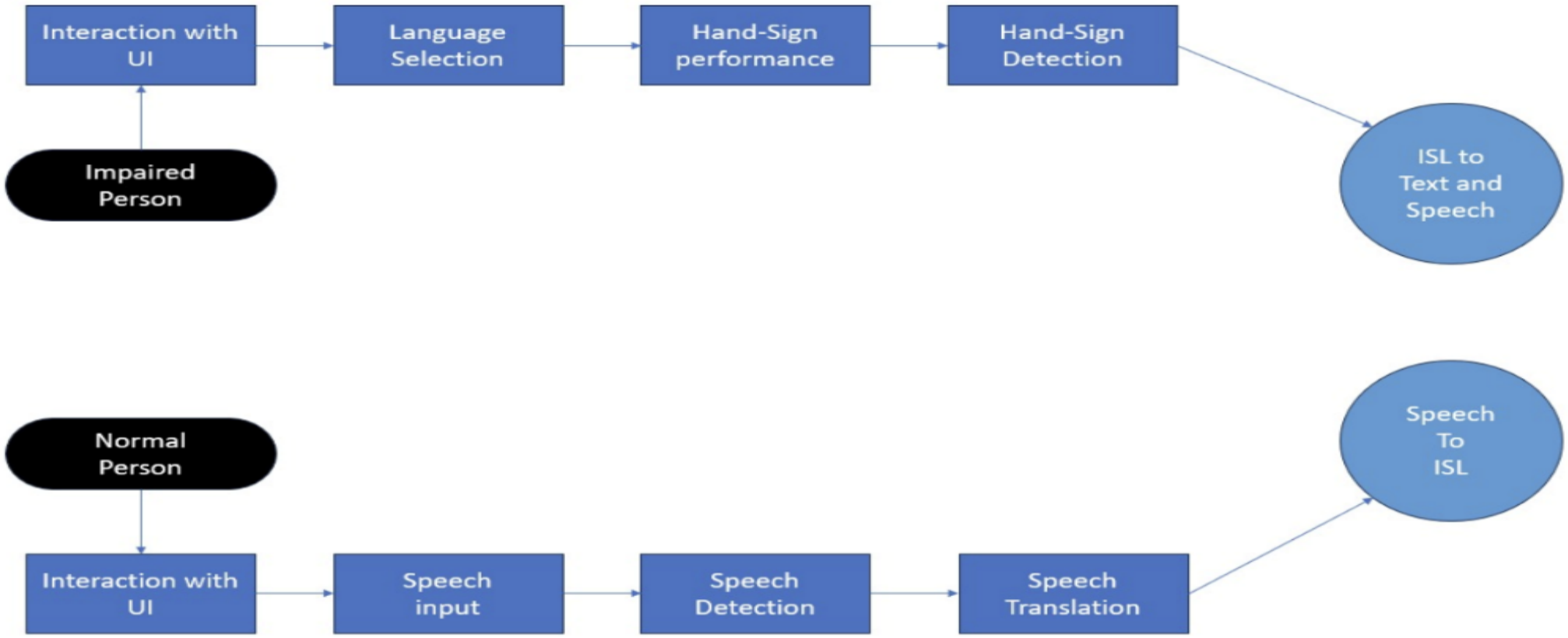
- Our solution provides an **Bi Directional flow** where not only the Impaired person can use it but also the Normal person can use it too.
- Also our solution provides **low infrastructure cost** , which can be used on already available devices.
- It also provides **Multi Lingo Support** that means user with different native backgrounds can use our System efficiently.
- User do not have to do any complex tasks to use our system and it is very **easy to use**.
- Rather than the cost of the internet there will be **no other extra cost** required to use the system.
- It is a **Real-Time System** that means users will get instant output of there actions and don’t have to wait for times.

PROPOSED METHODOLOGY

- Our solution provide **Two-Way Communication** to bridge the gap between normal and Impaired people.
- user has to interact with our website in order to make the communication effectively.
- Our method provides **Multi-Language Support** so ,that people with different native background use it.
- For the Impaired people's, they have to show the sign that sign is **detected in real time** and later will be converted into text and speech for normal people.
- For the normal person, has to speak in his native language system will convert that it into **sequence of image** for Impaired people.



Architecture Diagram



Workflow Diagram

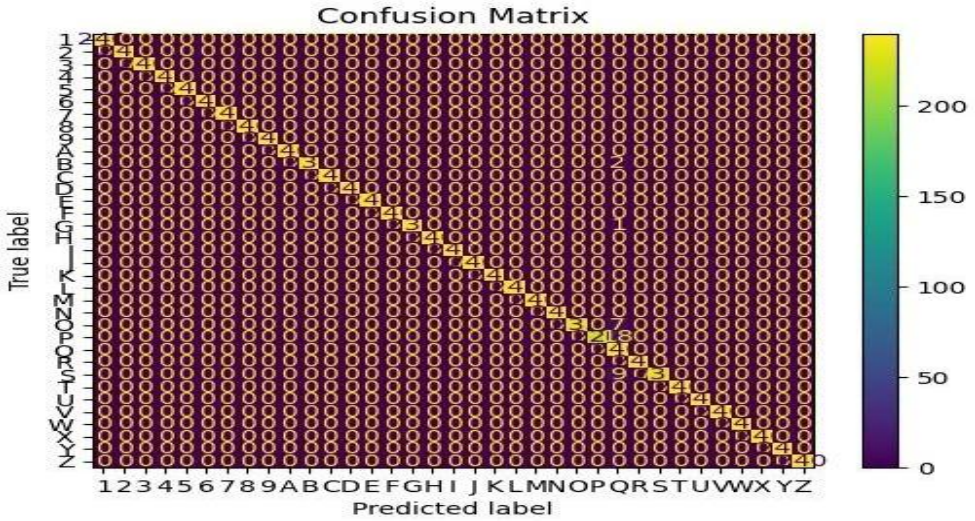
TECHNOLOGY STACK

- **Frontend Technology:**
  HTML, CSS, JavaScript

- **Backend Technology:**
  1. **Python:** programming language for creating backend.
  2. **flask:** web framework for building web applications.

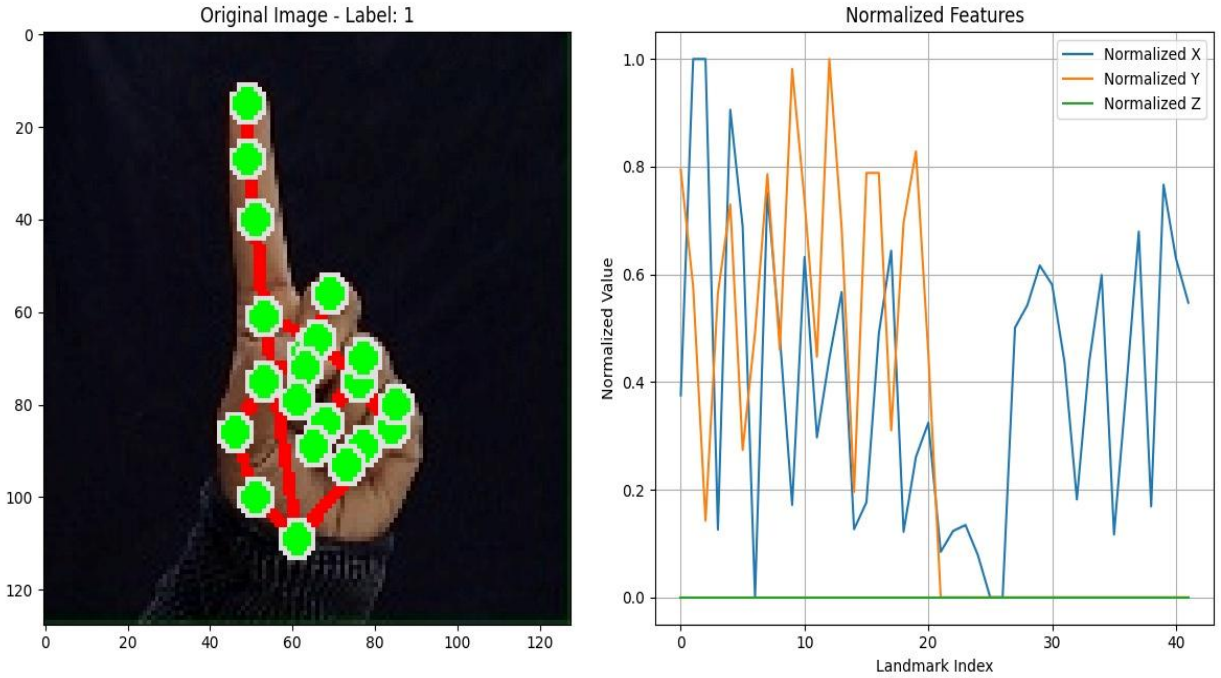
- **Libraries/framework:**
  1.**MediaPipe**: open-source framework for hand sign detection.
  2. **gTTS** : it is library for converting text-to-speech
  3.**speech_recognition** : is used to convert speech to text
  4.**deep_translator**: it is used for translating between two languages.

RESULT AND ANALYSIS



Accuracy=TP+TN / FP+FNTP+TN

This image shows a **confusion matrix**, which is a visualization used in classification tasks to evaluate the performance of a machine learning model. It represents the frequency of true labels versus predicted labels.



The tracking points on the sign image are displayed in the above figure. The **84-tracking points** per hand , system was used to track the hand gesture using Media Point to identify the data points.

Model	Accuracy	Loss
CNN	78.8%	0.269
Ada Boost	80%	0.22
Random Forest	88%	0.15

Model comparison Table

Cross-Entropy Loss (Log Loss): It measures the difference between the predicted probabilities and the true binary labels.

Accuracy :is a performance metric used to measure how often a model correctly predicts the target labels.

Cross-Entropy Loss (Log Loss): It measures the difference between the predicted probabilities and the true binary labels.

Formula:

$$L = -\frac{1}{N}\left[\sum_{j=1}^{N}\left[t_j \log(p_j) + (1 - t_j)\log(1 - p_j)\right]\right]$$

Accuracy :is a performance metric used to measure how often a model correctly predicts the target labels.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Dataset Description

The dataset consists of 35 distinct classes, encompassing 9 numerical digits (1 through 9) and 26 alphabetic characters (A through Z). In total, the dataset contains 210000 images, with an equal distribution of 6000 samples per class. Each image is formatted with dimensions of 128×128 pixels, ensuring consistency in size across the dataset. The images are in RGB color format, comprising three channels to capture a full spectrum of colors.

Dataset Description

Dataset Information :- The dataset contains images of Indian Hand sings by alphabetical Order.

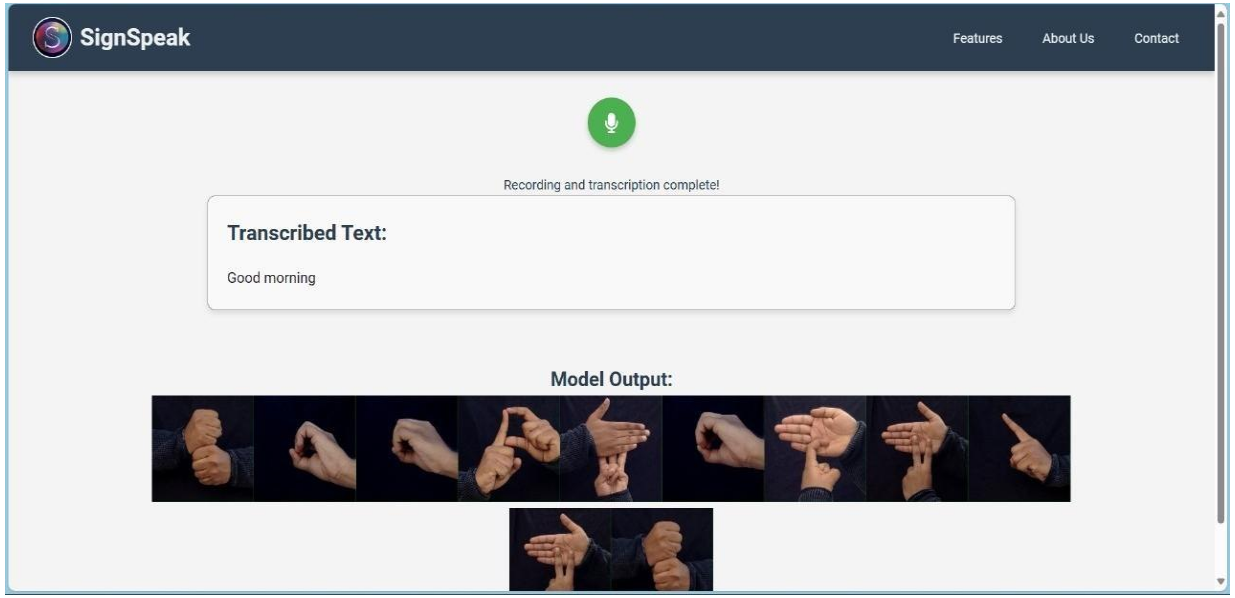
Dataset Size :- Dataset contains 210000 images, with an equal distribution of 6000 samples.

Dataset Shape :- 210000 Samples and 35 Classes.

Image Size :- 128*128*3 per Image

Training Size :- The 70% (i.e. 147000 images) has been used for training

Testing Size :- The 30% (i.e. 63000 images) has been used for training.



First module

Second module

LIMITATIONS

- **Limited Dataset Availability**
- **Complex Gesture Recognition**
- **Variability in Signing Styles**
- **Dependencies on Infrastructure.**
- **Real-Time Challenges.**

FUTURE SCOPE

- **Integration with Educational Platforms**
- **Facial and posture Recognition.**
- **Real-Time ISL Avatar Animation**
- **Offline Accessibility**
- **Integration with Video conferencing platform such as(Zoom , google meet)**

References

1. Ming Jin Cheok, Zaid Omar, and Mohamed Hisham Jaward. A review of hand gesture and sign language recognition techniques. International Journal of Ma- chine Learning and Cybernetics, 10(1):131–153, 2019.

2. Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. Sign language recognition using convolutional neural networks. In European conference on computer vision, pages 572–578. Springer, 2014.

3. Ankita Wadhawan and Parteek Kumar. Deep learning-based sign language recognition system for static signs. Neural computing and applications, 32(12):7957–7968, 2020.

4. By great learning team ,“Real-Time Object Detection Using TensorFlow”, december 25 ,2021 https://www.mygreatlearning.com/blog/object-detection-usingtensorflow/

5. Jeffrey Dean, minute 0:47 / 2:17 from YouTube clip “TensorFlow: Open source machine learning”. Google. 2015. Archived from the original on November 11, 2021. "It is machine learning software being used for various kinds of perceptual and language understanding tasks.

6. Li, Y.; Chen, X.; Zhang, X.; Wang, K.; Wang, Z.J. A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data. IEEE Trans. Biomed. Eng. 2012, 59, 2695–2704. [Google Scholar] [PubMed]

References

7. Ding, L.; Martinez, A.M. Modelling and recognition of the linguistic components in American sign language. Image Vis. Comput. 2009, 27, 1826–1844. [Google Scholar] [CrossRef] [PubMed]
8. Trigueiros, P.; Ribeiro, F.; Reis, L.P. Vision-based Portuguese sign language recognition system. Adv. Intell. Syst. 2014, 275, 605–617. [Google Scholar]
9. Fang, G.; Gao, W.; Zhao, D. Large vocabulary sign language recognition based on fuzzy decision trees. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 2004, 34, 305–314. [Google Scholar] [CrossRef]
10. Doliotis, P.; Athitsos, V.; Kosmopoulos, D.; Perantonis, S. Hand Shape and 3D Pose Estimation Using Depth Data From a Single Cluttered Frame.

THANK YOU