Synchronic Automatic Sign Language Recognition Project

Zidong Xu(zx2507), Caiwu Chen(cc4786)

Overview

This project aims to build a synchronic automatic sign language recognition system to facilitate communication between people with hearing challenges or people who experience deafness. Building upon the midterm project, we enhanced the robustness and generality of the model while contributing to improved human interaction.

The system processes real-time video input, extracting human gestures featuring Convolutional Neural Networks (CNNs) and modeling temporal dynamics using either Long Short-Term Memory networks (LSTMs) or Vision Transformers (ViTs). To reduce environmental noise, tools like MediaPipe Hands will be used to extract hand landmarks as inputs, improving model focus and accuracy.

The system will be trained on the Kaggle Sign Language MNIST Dataset. The real-time sign language recognition system will need permission to access the user's computer camera to fetch information regarding your gestures.

We will adopt the following metrics to evaluate the effectiveness and robustness of our sign language recognition system:
Accuracy: Measures the overall proportion of correct predictions among all attempted classifications.
Precision: The ratio of correctly predicted positive observations to total predicted positives.
Recall: The ratio of correctly predicted positives to all actual positive samples.
F1-score: Harmonic mean of Precision and Recall, balancing false positives and false negatives.
Latency: Measures the time the system takes to process a frame and return a prediction. This is crucial for assessing real-time usability, particularly under resource-constrained environments.

We will build upon the midterm project previously investigated by Zidong Xu, aiming to enhance the robustness and generality of the model while contributing to improved human interaction.(xuefeng16513/6156_Midterm_paper)

Dataset

1. How2Sign Dataset
2. Word-Level American Sign Language (WLASL)

Research Questions

1. How can sign language be identified from the noisy background?
2. How can we detect the changing of sign language among a series of images, i.e. video?
3. How can the sign language translator be embedded into a personal computer?
4. How can the robustness of an automatic sign language recognition model be improved for accurate translation?

Value to User Community

The project contributes to the development of building a bridge among the hearing-impaired community, sign language users, and the broader society by enabling seamless, real-time communication through advanced sign language recognition technology. The system is part of an accessible tool for doing real-time translation for sign-language users. Due to the flexibility and easily-embedded feature of this project, it can also be a tool of self-learning. Our unique value lies in that our model supports dynamic sign language phrase/sentence recognition, not just letter or static image recognition. At the same time, due to the combination of CNN model and LSTM time series modeling, the model is suitable for low-cost deployment, which makes promotion possible.

Demo

We will demonstrate:

- Real-time ASL recognition from live webcam input
- On-screen translation of signed gestures into English text
- Effectiveness of hand landmark preprocessing using MediaPipe (maybe)

The demo will showcase the system's performance under various lighting conditions and backgrounds to validate robustness.

Delivery

Github: https://github.com/xuefeng16513/6156_Final_Project. Currently having model, dataset. Will delete it afterward.

Reference

1. S. Mhatre, S. Joshi and H. B. Kulkarni, "Sign Language Detection using LSTM," *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*, Bhopal, India, 2022, pp. 1-6, doi: 10.1109/CCET56606.2022.10080705. Sign Language Detection using LSTM | IEEE Conference Publication | IEEE Xplore
2. Nimisha, KP, and Agnes Jacob. "A brief review of the recent trends in Sign language recognition." *2020 International Conference on Communication and Signal Processing (ICCSP)*, July 2021, pp. 186–190, https://doi.org/10.1109/iccsp48568.2020.9182351.
3. Shin, Jungpil, et al. "Korean sign language recognition using transformer-based deep neural network." *Applied Sciences*, vol. 13, no. 5, 27 Feb. 2023, p. 3029, https://doi.org/10.3390/app13053029.
4. D. Li, C. R. Opazo, X. Yu, and H. Li, "Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison," *CoRR*, vol. abs/1910.11006, 2019. [Online]. Available: http://arxiv.org/abs/1910.11006.
5. A. C. Duarte, S. Palaskar, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giró-i-Nieto, "How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language," *CoRR*, vol. abs/2008.08143, 2020. [Online]. Available: https://arxiv.org/abs/2008.08143.
6. A. Tayade and A. Halder, "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning," May 2021. DOI: 10.13140/RG.2.2.32364.03203. Available: https://www.researchgate.net/publication/369945035_Real-time_Vernacular_Sign_Language_Recognition_using_MediaPipe_and_Machine_Learning