

Project Proposal

Zhanqin Ai, Dinglun Zhang

November 15, 2023

Project Title: CV for American Sign Language

Topic summary

American Sign Language (ASL) is a comprehensive natural language that is the primary sign language of the deaf community in the United States. Our project focuses on building a deep learning model to recognize and interpret ASL from images. The goal is to leverage the data set available on Kaggle, which is made up of images representing 29 categories that correspond to 26 letters of the alphabet, a space, deleted gestures and nothing. The dataset contains 3,000 images for each category, all captured as 200x200 RGB images. The structure of the dataset is divided into training subsets and test subsets, with special emphasis on model training without test sets to encourage the use of real-world images.

The method includes several suggested steps: necessary pre-processing of the dataset images, a literature survey of existing work in the field to understand the network architectures used, experiments on different architectures and hyperparameters, and including the results of testing the model on an external image set due to the small size of the test set provided.

In addition, the project could be expanded to develop an application capable of capturing video and providing real-time transcription, which could be further used to generate speech, combined with features such as Google Text to Speech.

The success of this project will depend on the accuracy of the model in category recognition and its functionality in real-time applications. The potential impact of this project is enormous, providing the deaf community and those who wish to learn ASL with a tool to aid communication.

Introduction

In a world that increasingly relies on visual communication, the ability to accurately interpret American Sign Language (ASL) through technological means stands as a significant milestone in bridging communication gaps. The Deaf and Hard of Hearing (DHH) community, which extensively uses ASL, can immensely benefit from a solution that allows for seamless translation and understanding across different mediums, including digital platforms.

The significance of developing an accurate ASL recognition model extends beyond the DHH

community. It promises enhanced accessibility, offering real-time translation services that facilitate clearer communication for all, especially in educational, social, and professional settings. A solution to this problem would also be a testament to the inclusivity efforts, demonstrating how technology can cater to diverse communication needs.

In industry, companies like Zoom or Skype could integrate this model to provide real-time ASL interpretation during video calls, effectively broadening their user base and providing inclusive services that cater to the DHH community. Furthermore, in academic research, such advancements can fuel further studies in linguistics and computational language models, enriching the theoretical understanding of sign languages and their integration into modern communication tools.

From an applied perspective, the project stands to enhance the fields of computer vision and machine learning, pushing the boundaries of what is achievable with current deep learning architectures. It involves leveraging convolutional neural networks (CNNs), which have shown exceptional proficiency in image recognition tasks. These networks, through their capability to discern intricate patterns and gestures in ASL, can transform the algorithmic approach to real-time language translation.

This proposal will delve into the theoretical underpinnings of CNNs, the applied mechanisms of image processing, and the algorithmic intricacies of model training and validation. Our objective is to forge a path that leads to a robust, real-time ASL interpretation model, setting a new benchmark for accessibility technology and industry standards.

Related Work

The problem of American Sign Language (ASL) recognition using deep learning has been tackled by various researchers. Kshitij Bantupalli and Ying Xie at the College of Computing and Software Engineering approached ASL recognition by employing machine learning and computer vision techniques, contributing to the body of knowledge in robotics and intelligent systems[1].

In the realm of ASL alphabet recognition, significant progress has been reported. Researchers have proposed models that recognize the ASL alphabet from RGB images, focusing on the pre-processing of images and employing a squeeze net architecture suitable for mobile devices[2]. This indicates a trend towards developing more accessible and portable ASL recognition systems.

Further research conducted a comprehensive review of automated sign language recognition based on machine and deep learning methods. This study spanned publications from 2014 to 2021, suggesting that despite advancements, there is still a need for improved conceptual classification within current methods[3].

Another study focused on the development of a sign language translation system based on deep learning, underscoring the importance of sign language as the primary communication tool for people with hearing and language impairments[4].

Comparatively, our proposed project builds on these foundational works with the aim to refine the accuracy and real-time processing capabilities of ASL recognition models. While the cited

studies have set the stage for mobile and efficient ASL recognition and provided valuable reviews of existing techniques, our work intends to address some of the limitations they have identified. Specifically, we plan to explore advanced neural network architectures and training techniques to enhance the model's ability to generalize from training data to real-world applications.

Our project is poised to contribute to the existing body of work by potentially introducing innovative approaches that could lead to more nuanced and sophisticated ASL recognition models, thereby expanding the tools available for facilitating communication for the Deaf and Hard of Hearing community.

Dataset description

[ASL Alphabet \(kaggle.com\)](https://www.kaggle.com/datasets/assafshalev/asl-alphabet)

[GitHub - grassknotted/Unvoiced: American Sign Language to Speech Application.](https://github.com/grassknotted/Unvoiced: American Sign Language to Speech Application)

The dataset is 1.02 GB. The dataset has 87,000 images of 29 labels, representing the 26 letters of the English alphabet and three additional signs: SPACE, DELETE and NOTHING. The images are in RGB format and have a resolution of 200 by 200 pixels.

Estimated Compute Needs

- Our primary compute resources will be personal workstations equipped with high-performance GPUs and CPUs.
 - GPU: Nvidia Geforce RTX 2060
 - CPU: AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz
 - RAM: 16.0 GB
- We anticipate needing a GPU with at least 4 GB of memory and a CPU with at least 8 GB of RAM to handle the large dataset and the complex network architectures.
- In addition to our personal workstations, we plan to use online compute resources with GPU acceleration, such as Kaggle or Google CoLab, as a backup.

Likely Outcome and Expected Results

- We expect our final output measure to show a high accuracy (above 90%) in classifying the ASL images into the correct categories.
- We also expect our model to generalize well to external images that are not part of the dataset, such as real-world images captured by a camera or a webcam.
- The most likely reasons for project failure are:
 - The dataset is imbalanced or noisy, which could affect the model's performance and

robustness.

- The network architectures are not suitable or optimal for the task, which could lead to overfitting or underfitting.
- The hyperparameters are not tuned properly, which could affect the model's convergence and efficiency.
- We will be able to conclude that we have successfully applied deep learning techniques to recognize and interpret ASL from images, and that we have contributed to the field of computer vision and sign language recognition.

Primary References and Codebase

- [1] Bantupalli, K., & Xie, Y. (2019). American Sign Language Recognition Using Machine Learning and Computer Vision. Kennesaw State University.
- [2] Kasukurthi, N., Rokad, B., Bidani, S., & Dennisan, A. (2019). American Sign Language Alphabet Recognition using Deep Learning. arXiv.
- [3] Al-Qurishi, M., Khalid, T., & Souissi, R. (2021). Research of a Sign Language Translation System Based on Deep Learning. IEEE Access, 9, 126917-126951.
- [4] He, S. (2019). Deep Learning for Sign Language Recognition: Current Techniques. In 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM) (pp. [Page numbers not provided]). Dublin, Ireland: IEEE.