Applying Swin Architecture to Diverse Sign Language Datasets

Annaliese Watson, Chin-Chien Lin, Drs. Yulia Kumar, Kuan Huang & J. Jenny Li (faculty advisors)

Department of Computer science & Technology, Kean University, Union, NJ

ABSTRACT

In the era of AI, understanding non-verbal communication is paramount. This research utilizes AI Swin Architecture to bridge communication gaps in sign language communities, comparing classification of American (ASL) and Taiwanese Sign Languages (TSL).

INTRODUCTION

Incorporating ASL into digital platforms enhances accessibility. This project focuses on a website with ASL features, including a fingerspelling dictionary and a practice section. Unique to this study is the development of a custom ASL dataset. The practice section employs the Swin Architecture for data classification. Leveraging AI and Deep Learning technologies, particularly the Hierarchical Vision Transformer using Shifted Windows (Swin) model, the project achieves 100% accuracy in ASL dataset classification. Custom datasets of ASL and TSL including both alphabet letters and short phrases (Fig. 1.) were created especially for this project. The prototype of ASL Educational App was then developed (Fig.2.) and currently being tested.

MATERIALS and METHODS

This research utilizes the Swin Transformer model, introduced by Liu et al. from Microsoft Research [1, 2, 3], for sign language recognition, focusing on ASL and TSL. The methodology is structured into five main phases: describing diverse datasets, used for the study, understanding of Swin Architecture, model training, results evaluation, and app development.

Each phase is crucial for the Sign Language recognition and translation.

electronics

Project Datasets

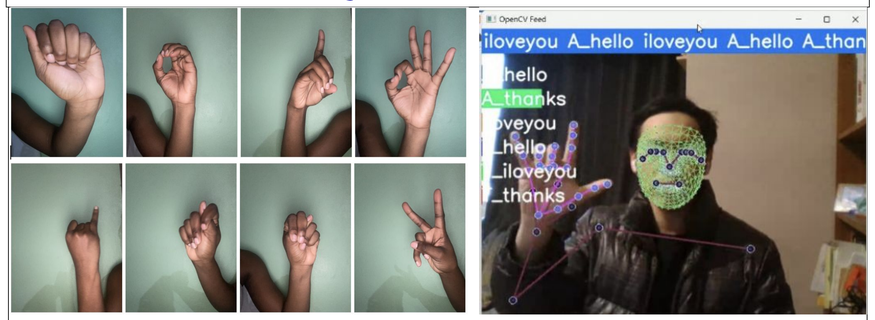

Figure 1. Snapshot of the project dataset: Alphabet Letters (left) vs Short phrases (right) in ASL and TSL.

ASL Educational App

Features on the website utilize an ASL dataset for fingerspelling, incorporating elements such as a dictionary and quiz matching.

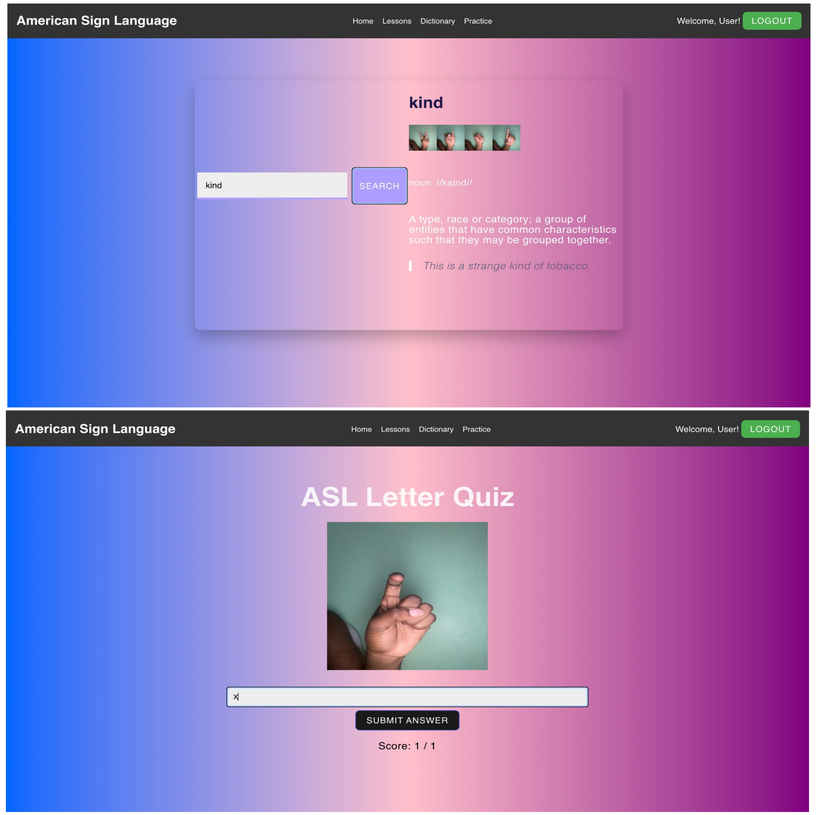

Figure 2. Snapshot of the ASL Educational app

Current Result and Future Work

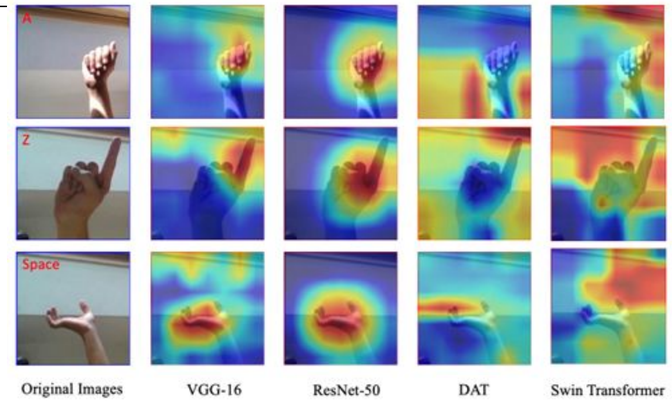

Figure 3. The CAMs of the Swin Transformer in comparison with other DL models

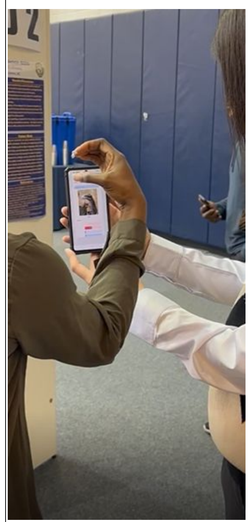

Figure 4. The early app prototype under test

Various AI models were trained on the ASL and TSL datasets, achieving up to 100% of accuracy. Future work includes exploring other sign languages, refining the website's features and developing our own bias mitigation strategies for Transformers like Swin. The associated paper is under consideration by Electronics Q2 journal [4] and goal of the researchers is to publish current results of this work and continue it further in 2024-2025.

Acknowledgments

This research was funded by the NSF, grants awards 1834620 and 2137791, and Kean University's Students Partnering with Faculty 2023 Summer Research Program (SPF).

References