Final Report

SAI Learning

Alejandro Adriaensens Martínez

Esteban Sierra Baccio

Danaé Sánchez Gutiérrez

Ana Karen Márquez Escobar

Mariana Amy Martínez Nevárez

May 27, 2024

Monterrey, Nuevo León, México

Abstract

Our proposed project aims to develop a platform for learning Mexican Sign Language (LSM). This platform leverages Artificial Intelligence, specifically Computer Vision, to detect and interpret users' signs. Through Machine Learning methods such as neural networks, a model was trained to classify, identify and show the signs. By employing classification models, it accurately recognizes and classifies signs, enhancing the learning experience. Additionally, we explore the use of Natural Language Generation models to interpret signs into spoken language, further aiding comprehension. The platform is interactive and web-based, offering a user-friendly interface for learners. Through this project, we aim to provide an innovative and accessible tool for LSM education.

Background/Introduction

The ability to communicate through sign language is crucial for individuals with hearing impairments. However, learning sign language can be challenging due to limited resources and accessibility. Our project addresses this need by developing a web-based platform that utilizes cutting-edge AI technologies to facilitate the learning of Mexican Sign Language (LSM). Previous studies have shown that AI and computer vision can effectively interpret sign language, but few platforms have integrated these technologies into an accessible and interactive educational tool. Our project bridges this gap, offering an innovative solution that enhances accessibility and learning efficiency.

Engineering Fundamentals

Our project is grounded in several key engineering principles, including computer vision, machine learning, and web development. We utilize Convolutional Neural Networks (CNNs) for the accurate detection and classification of signs. These networks are trained on a dataset of hand images performing the LSM alphabet, which have been processed to detect 21 key points on the hand using MediaPipe.

Each image generates a 42-parameter vector representing these points, which feeds into our classification model built with scikit-learn. Additionally, our web platform is designed using modern web development frameworks to ensure usability and accessibility.

Prototype Design

The prototype of our LSM learning platform consists of several components: data collection and preprocessing module, the sign classification model, and the web interface. We collected approximately 2500 hand images representing the LSM alphabet. These images were processed using MediaPipe to extract hand landmarks, resulting in a dataset of 42 parameters per image. Our classification model, developed using scikit-learn, is capable of recognizing static signs with high accuracy. The web interface, designed in Figma and implemented using React, along with HTML, CSS, and JavaScript, features a user-friendly layout with sections for user login, main menu, and various learning levels.

Key elements of our design include:

- Data Collection: A script was developed to capture 2500 hand images performing the LSM alphabet. Each image was labeled and stored.
- Data Preprocessing: Images were processed to detect 21 hand landmarks using MediaPipe, resulting in a 42-parameter vector for each image.
- Model Training: Various classification models were tested using scikit-learn, with a focus on neural networks. Simple hyperparameter tuning was performed to optimize model performance.
- Web Development: A user-friendly interface was designed in Figma and implemented using standard web technologies. The interface includes sections for user login, main menu, and learning levels.

Implemented Methodology

Our methodology integrates several key stages: data collection, preprocessing, model training, and web development. We began by collecting a large dataset of hand images performing the LSM alphabet. These images were processed using MediaPipe to detect hand landmarks, resulting in a dataset of 42 parameters per image. We then trained a classification model using scikit-learn, experimenting with different algorithms and performing basic hyperparameter tuning. The model was trained to classify static signs accurately. Concurrently, we designed and developed a web interface to host the learning platform, ensuring it was intuitive and accessible.

Steps Implemented:

1. Data Collection: Approximately 2500 images of hands signing the LSM alphabet were collected.
2. Data Preprocessing: Using MediaPipe, 21 landmarks on each hand were detected, providing a 42-parameter vector per image.
3. Model Training: Different classification models were tested using scikit-learn. The best-performing model was a neural network, which was optimized through hyperparameter tuning.
4. Web Development: The platform was designed in Figma and developed using Django as the framework for the back end and React, HTML, CSS, and JavaScript for the front end, creating a user-friendly interface with clear navigation and multiple learning levels. We used Microsoft SQL Server for the database.

Adaptive Resilience

Throughout the project, we faced several challenges, including the collection of a sufficiently large and diverse dataset, the optimization of our classification model. To address these challenges, we employed strategies such as iterative testing and refinement of our model using grid search. These efforts ensured that we could adapt to obstacles and maintain progress towards our project goals.

Specific challenges and solutions:

- Data Diversity: We initially struggled to capture a diverse set of hand images. To overcome this, we expanded our collection efforts and employed data augmentation techniques.
- Model Optimization: Finding the optimal model required extensive testing. We performed iterative testing and hyperparameter tuning to improve model accuracy.
- Web Interface Usability: Ensuring an intuitive user experience was crucial. We conducted user testing and incorporated feedback to refine the interface design.

Engineering Innovation

Our project stands out due to its innovative use of AI to facilitate LSM learning. By integrating computer vision and natural language generation, we have created a tool that not only recognizes signs but also translates them into spoken language. This dual functionality enhances the learning experience and provides a more comprehensive educational tool. Additionally, our use of web-based technologies ensures that the platform is widely accessible, promoting inclusivity in digital learning environments.

However, this project represents only the beginning. To achieve its full potential, our platform requires expansion through larger datasets, more robust models, and powerful computational resources. Such advancements will necessitate additional funding, but they promise to break further barriers in language learning and accessibility. With further investment, our platform could support dynamic sign recognition, real-time translation, and integration with other educational tools.

Financial Feasibility

The financial feasibility of our project is supported by a detailed analysis of costs and benefits. Initial expenses include the development of the AI models, data collection, and web development. Operational costs are minimal due to the use of open-source tools and cloud-based hosting services. The potential benefits, including the widespread adoption of our platform and its impact on LSM education, justify these costs. We have also considered potential financial risks, such as the need for ongoing maintenance and updates, and have proposed mitigation strategies to ensure long-term sustainability.

Cost Analysis:

- Initial Development: Costs for data collection, model training, and web development.
- Operational Costs: Minimal, due to the use of open-source tools and cloud-based hosting.
- Projected Benefits: Widespread adoption and impact on LSM education.
- Financial Risks: Need for ongoing maintenance and updates; mitigated through strategic planning and potential funding sources.

Outcome Presentation

The outcome of our project is a functional prototype of the LSM learning platform. This prototype meets our initial expectations, demonstrating accurate sign detection and classification capabilities. The final product is a testament to our successful integration of AI and web technologies to create an accessible and innovative educational tool.

Prototype Features:

- Accurate Sign Detection: High accuracy in classifying static signs.
- User-Friendly Interface: Positive feedback on usability and design.
- Educational Impact: Effective tool for LSM education.

Results Analysis

The results of our project indicate significant progress in the use of AI for LSM learning. Our classification model demonstrates high accuracy for static signs, and the web interface provides a user-friendly learning environment. However, there are areas for improvement, such as expanding the dataset to include dynamic signs and refining the model for better performance. These findings highlight both the strengths and limitations of our approach, providing valuable insights for future development.

Strengths:

- High Accuracy: Effective classification of static signs.
- User Interface: Intuitive and accessible design.
- Educational Value: Positive impact on LSM learning.

Areas for Improvement:

- Dynamic Signs: Need to include dynamic sign recognition.
- Model Robustness: Further refinement for enhanced performance.
- Computational Resources: Requirement for more powerful computational resources to support advanced features.

References

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools. Recuperado de https://opencv.org/

Lugaresi, C., Tang, J., Nash, H., McGuire, M., Kossler, D., Garrido, P., & Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. Recuperado de https://arxiv.org/abs/1906.08172

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. Recuperado de https://scikit-learn.org/

Secretaría de Educación Pública. (2023). Diccionario de lengua de señas mexicana. Recuperado de https://educacionespecial.sep.gob.mx/storage/recursos/2023/05/xzrfl019nV-4Diccionario_lengua_%20Senas.pdf