DELHI TECHNICAL CAMPUS
Greater Noida
Affiliated to GGSIPU and Approved by AICTE & COA

Project Based Learning Report on

Realtime ASL Translator using CNN

BACHELOR OF TECHNOLOGY
In
COMPUTER SCIENCE & ENGINEERING

SUBMITTED BY

CSE IIIrd Year Section - A

Harsh Rathee - 00818002722
Mridul Singh - 02118002722
Naman Chandok - 03018002722
Dhruv - 03518002722

GUIDED BY

Ms. Muntaha Rouf & Ms. Preeti Katiyar
(Teaching Assistant) (Assistant Professor)

Affiliated to GGSIP University, New Delhi
Approved by AICTE & Council of Architecture

Department of Computer Science & Engineering
DELHI TECHNICAL CAMPUS, GREATER NOIDA
(Affiliated with Guru Gobind Singh Indraprastha University)
Session 2024-45 (Even Sem)

Introduction:

Sign languages are vital for communication within the Deaf and Hard-of-Hearing communities. However, their usage is limited due to the communication gap with people unfamiliar with them. This project, **ASL Translator**, bridges that gap using computer vision and machine learning to translate American Sign Language (ASL) gestures into text in real time.

The application utilizes **MediaPipe** for real-time hand tracking, combined with **Keras** to classify ASL signs from hand landmarks. It delivers a responsive web-based interface allowing users to perform signs in front of a camera and see their corresponding letters or words instantly.

Key Objectives:

- Recognize ASL alphabet hand gestures using real-time video feed.
- Use MediaPipe for precise hand landmark extraction.
- Train and use a Keras-based neural network for gesture classification.
- Enable real-time translation of signs to text.

Feasibility Study:

1. Technical Feasibility:

- **Languages:** Python
- **Libraries/Frameworks:**
  - Backend: TensorFlow, OpenCV, Keras, MediaPipe
- **Model:** A CNN trained on ASL alphabet images using Keras.

2. Operational Feasibility:

- Works on standard webcam-equipped systems.
- No additional hardware required.
- Real-time inference with minimal lag.

3. Time Feasibility:

- Model training: ~1 week
- Integration & testing: ~3 days

4. Cost Feasibility:

- All tools are open-source.
- No cloud computing costs involved.

5. Scope Feasibility:

- Demonstrates translation of ASL alphabets.
- Can be extended to detect full words and phrases.
- Future scope includes sentence generation and voice output.

Methodology/Planning of Work:

1. Project Setup:

- Initialized project repo and structured codebase for modularity.

2. Data Acquisition:

- Used a self-generated dataset with webcam images for 26 ASL letters.
- Split into training, validation, and test sets.

3. Feature Engineering:

- MediaPipe used to extract 21 hand landmarks per frame.
- Landmarks normalized before input to the model.

4. Model Training:

- Trained a deep learning model (CNN + Dense layers) on landmarks.
- Used accuracy, confusion matrix for evaluation.

5. Inference Pipeline:

- Input from webcam processed in real time.
- Detected hand pose fed to model for prediction.

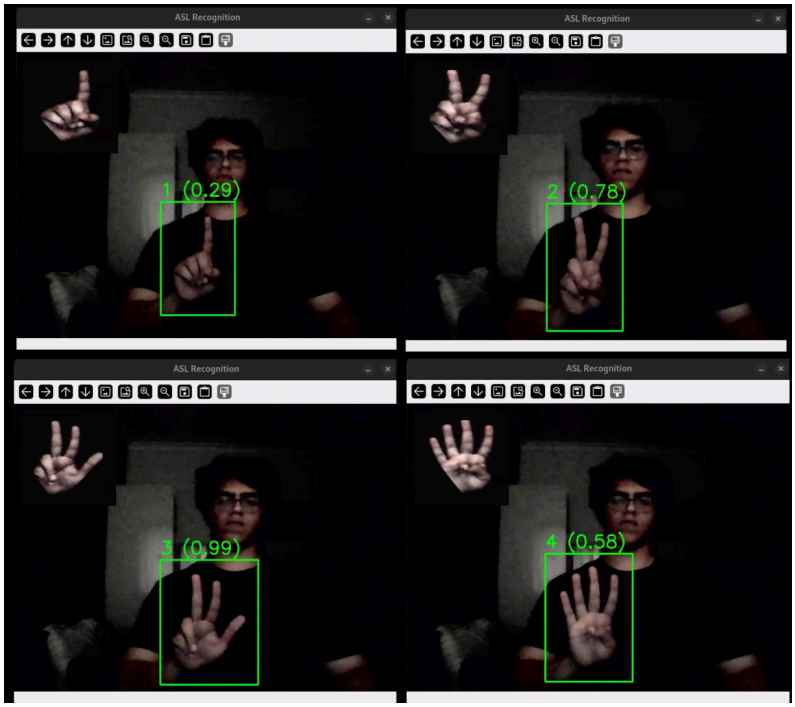
6. Integration and Testing:

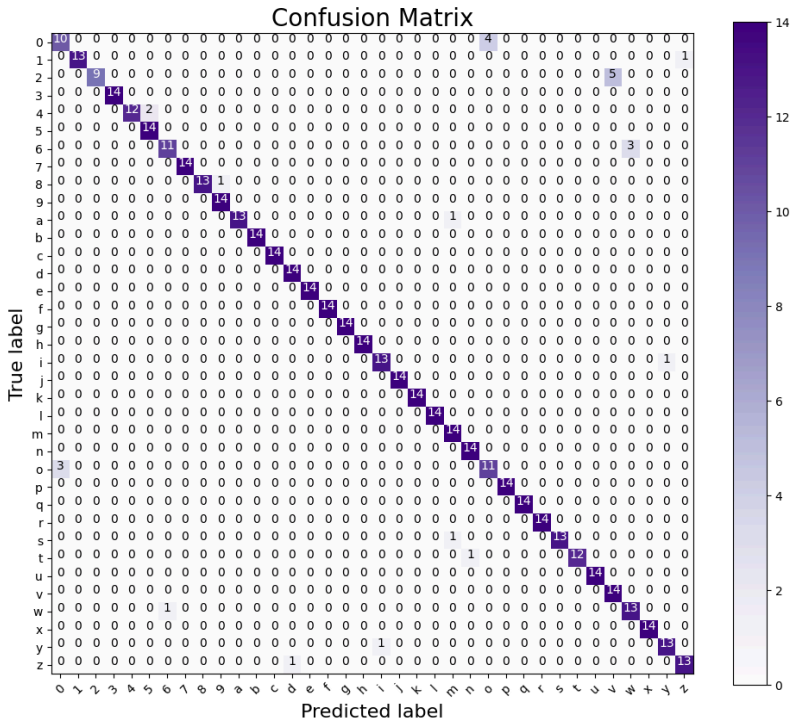
- Validated real-time performance across different lighting/backgrounds.
- Tested alphabet recognition speed and robustness.

System Architecture/Flow:

- **asl_dataset**
  - Kaggle dataset with various images for every character in ASL
- **main.py**
  - ASL Dataset is loaded and split into test and train data
  - Images are preprocessed to be sent to model
  - CNN Model is trained on the dataset
  - Confusion Matrix is plotted based on model results
  - Model is saved to local .keras file
- **cam.py**
  - Model is loaded from local .keras file
  - Input is taken from webcam using OpenCV
  - Input is preprocessed
  - Input sent to CNN Model for evaluation

Deployment Screenshots:





Conclusion:

The Realtime ASL Translator project effectively demonstrates how deep learning and computer vision can bridge communication gaps for the Deaf and Hard-of-Hearing community. By leveraging MediaPipe for hand tracking and a CNN-based classifier for gesture recognition, the system provides a functional prototype that translates ASL alphabet signs into text in real time. The project was implemented using open-source tools, making it accessible and cost-effective. While it currently supports only alphabets, it lays the groundwork for more comprehensive systems capable of word, sentence, and even voice translation.

Overall, the project served as a comprehensive learning experience, laying a strong foundation for future machine learning projects and enhancing our readiness for real-world challenges.

Key Takeaways:

- Feasibility and effectiveness of using open-source tools for impactful social applications.
- CNNs are powerful tools for gesture recognition when paired with effective feature extraction techniques like hand landmarks.
- MediaPipe provides robust real-time tracking capabilities that are essential for human-computer interaction applications.
- Well-structured modular codebases make future scalability (e.g., to full sentences or audio output) achievable.
- The PBL approach helped bridge the gap between theory and practice, improving our problem-solving, collaboration, and debugging skills.