TRANSLATING SIGN LANGUAGE TO SPEECH

Group ID: PCSE 25-65
Guide: Dr Parita Jain

SDG Mapping :
SDG 4: Quality Education
SDG 3: Good Health and Well being
SDG 9: Industry, Innovation, and Infrastructure
SDG 10: Reduced Inequalities

Research Paper: Real-Time American Sign Language(ASL) Translation using Deep Learning
https://ieeexplore.ieee.org/document/10986685

Team Members

Vikas Kumar
Member

Tanya Sharma
Member

Srishti Upadhyay
Leader

Project Abstraction :

Overview :

- This project bridges the communication gap between sign language users and non-signers by developing a real-time system to translate gestures into spoken English, using CNN, OpenCV, and YOLO.

key Objectives :

- Translate sign language gestures into speech for seamless communication.
- Ensure high accuracy and real-time performance.
- Provide a robust system for diverse conditions, promoting inclusivity.

Approach :

- The system uses CNN for gesture recognition, OpenCV for video processing, and YOLO for object detection. Recognized gestures are converted to text and then to speech via a text-to-speech module, ensuring reliable real-world application.

Project Goals and Objectives

Design a system to bridge the communication gap between sign language users and non-signers. The system should translate sign language gestures into spoken English in real time, utilizing technologies like Convolutional Neural Networks (CNN), OpenCV, and YOLO. It must ensure high accuracy in gesture recognition and contextual precision in translations. The solution should prioritize inclusivity and accessibility, enabling hearing-impaired individuals to communicate effectively in social and professional settings. Emphasize creating an assistive device that integrates advanced technologies to foster mutual understanding and enhance accessibility, promoting seamless interaction between diverse groups in various environments.

Project Goals and Objectives :

Strategy Development

- Develop a framework using CNN for gesture recognition and OpenCV for image preprocessing
- Model optimization,Speech synthesis integration

Analysis Phase

- Identify the problem,Study existing solutions,Technology assessment and Dataset selection.
- Address the communication barrier between sign language users and non-signers.

Implementation Plan

- Gather and preprocess data, ensuring consistency and quality for training the model.
- gesture recognition with real-time video input.

Sustainable Development Goals (SDGs)

SDG 4: Quality Education

- By enabling communication for the deaf and hard-of-hearing, project can help provide inclusive education opportunities, reduce barriers, and foster better learning environments.

SDG 3: Good Health and Well-being

- Improved communication can lead to better mental health and social well-being for those who might otherwise feel isolated due to communication barriers.

SDG 9: Industry, Innovation, and Infrastructure

- Developing an innovative technological solution aligns with fostering sustainable innovation and building resilient infrastructure.

SDG 10: Reduced Inequalities

- The project promotes social inclusion by bridging communication gaps, thus reducing inequalities faced by individuals with hearing or speech impairments.

Methodologies:

Data Acquisition:

- The ASL Alphabet Dataset from Kaggle, containing 87,000 images representing 29 sign language gestures (26 letters of the alphabet and 3 special signs), was chosen for training the model.
- The dataset captures static gestures that serve as the foundation for sign language recognition, offering a broad representation of individual hand signs.

Data Preprocessing:

- Image Resizing: All images were resized to a uniform dimension of 224×224 pixels to ensure consistency across the dataset and optimize input for CNN models.
- Normalization: Pixel values were normalized to a range of 0-1 by dividing by 255, improving the model’s training speed and stability

Testing and Refinement:

- The system was rigorously tested using multiple sign language inputs to assess its accuracy and performance in diverse settings.
- Feedback from testing was used to fine-tune the model, address misclassification issues, and improve the real-time processing speed.

Methodologies:

Model Development:

The system used a Convolutional Neural Network (CNN) with 3×3 filters for feature extraction, max-pooling layers for dimensionality reduction, and fully connected layers for classification. The model was trained on 80% of the dataset with Adam optimizer (learning rate of 0.001) and categorical cross-entropy loss for 10 epochs, using early stopping to prevent overfitting. Model performance was evaluated using training and validation accuracy, and a confusion matrix helped identify areas for improvement.

System Integrations:

OpenCV was integrated for real-time gesture recognition, processing video input frame-by-frame with the CNN model. The recognized gestures were converted to spoken English via a text-to-speech module, enabling smooth communication. The system was tested under various conditions to ensure robustness and accuracy in real-world scenarios.

Model Design:

Input Layer:

The system takes images resized to 224×224 pixels as input, ensuring uniformity and compatibility with pre-trained models like ResNet or VGG.

Convolutional Layers:

These layers use 3×3 filters to extract spatial features from the input images. ReLU activation functions are applied to introduce non-linearity, enabling the model to learn complex patterns.

Pooling Layers:

Max-pooling to reduce dimensionality while retaining key features.

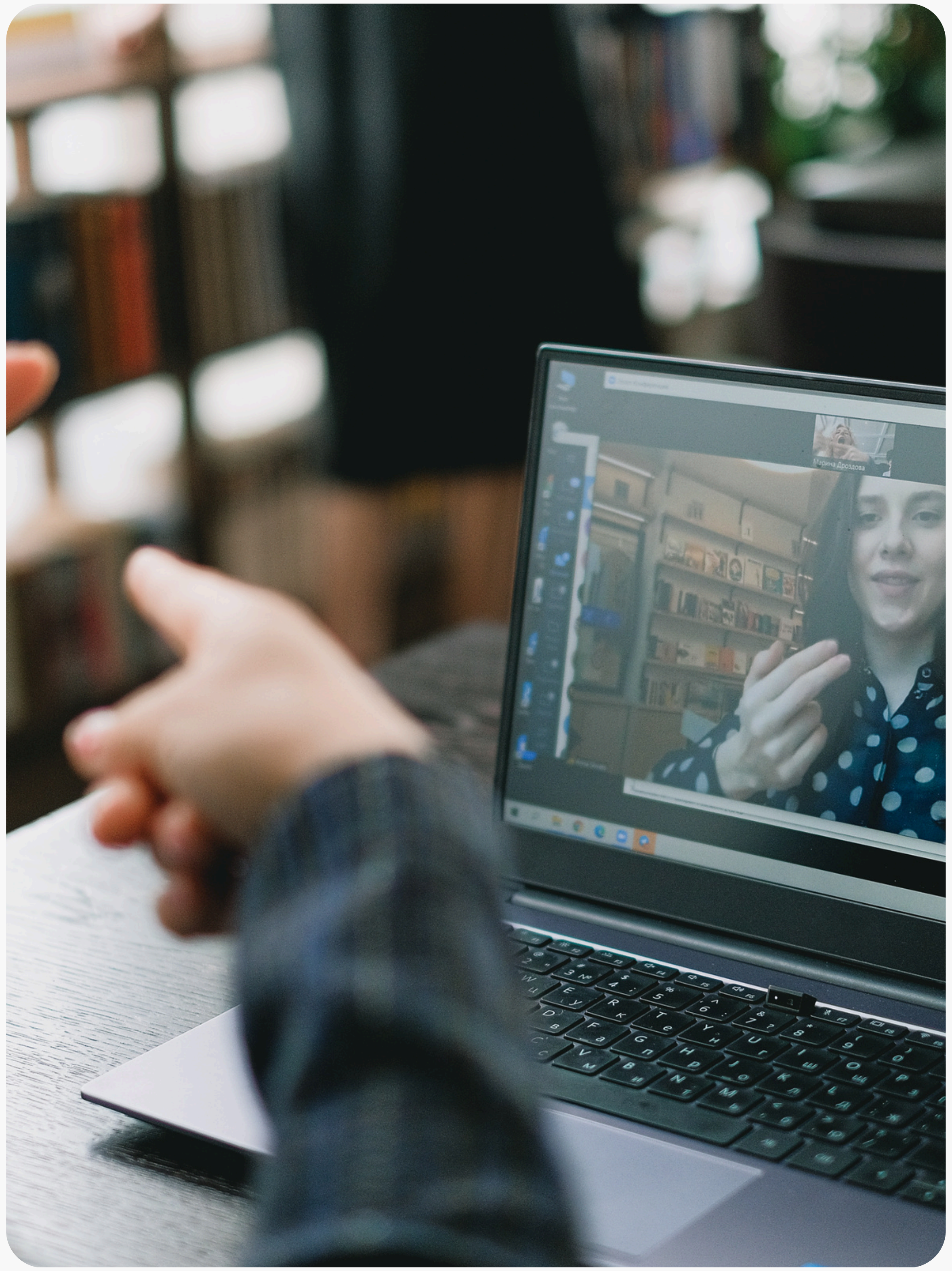
Fully Connected Layers:

Dense layers for combining features and classification.

Output Layer:

Softmax activation to classify gestures based on the highest probability.

Expected Outcomes



High Accuracy in Gesture Recognition:

- The CNN model will effectively identify sign language gestures with high training and validation accuracy.

Real-Time Performance:

- The system will process video input and classify gestures in real-time, ensuring smooth communication without significant delays.

Seamless Text-to-Speech Conversion:

- Recognized gestures will be accurately mapped to spoken English, promoting effective communication between sign language users and non-signers.

System Robustness:

- The system will perform reliably across different lighting and background conditions, demonstrating its practicality for real-world applications.

Key Project Deliverables

Real-Time Gesture Recognition System :

- A functional system that uses CNN, OpenCV, and YOLO to accurately recognize and classify sign language gestures in real time.

Text-to-Speech Integration :

- A module that converts recognized gestures into spoken English, providing clear and natural speech output.

Robust and Tested System

- A thoroughly tested system capable of handling diverse environmental conditions such as varying lighting and backgrounds.

User-Friendly Interface

- An intuitive and accessible interface that enables seamless interaction for both sign language users and non-signers.

Conclusion

The "Translating Sign Language to Speech" project developed a real-time system that converts sign language gestures into spoken English using CNN, OpenCV, YOLO, and text-to-speech technology. The system successfully bridges the communication gap between sign language users and non-signers, promoting inclusivity. While it performed well, further improvements can be made in recognizing dynamic gestures, handling varying conditions, and supporting multilingual outputs. This project advances assistive technologies and lays the groundwork for future developments in sign language translation, enhancing communication for individuals with hearing impairments.

Thank you