Premier University

Department of Computer Science and Engineering

Course Title: Neural Network & Fuzzy Logic
Course Code: CSE 451

Neural Network & Fuzzy Logic Course's
Assignment on

"Real-Time Sign Language Recognition using Neural Networks"

Submission Date: 21/04/2025

Submitted By

Rayanul kader Chowdhury Abid
210401020 2162
7th Semester A Section

Introduction

This assignment addresses the development of an AI-based assistive technology for real-time sign language recognition, aimed at enhancing communication for individuals with hearing disabilities. The system processes video frames or image sequences of hand gestures, classifying them into predefined sign language categories (e.g., alphabets) and producing textual output. The challenge lies in achieving high accuracy, low-latency inference, and robustness across diverse users and environments. This solution is significant for fostering inclusivity in education, healthcare, and social interactions, aligning with the needs of stakeholders like speech therapists, educators, and end-users.

Data Preprocessing

The dataset comprises 10,000 sequences of 26 alphabet signs, each sequence containing 10 grayscale video frames (224x224 pixels). Preprocessing steps include:

- **Resizing**: All frames resized to 224x224 pixels for consistent input.
- **Normalization**: Pixel values scaled to [0, 1] to stabilize training.
- **Grayscale Conversion**: Frames converted to grayscale to reduce computational load while preserving gesture features.
- **Data Augmentation**: Applied random rotations (±10°), horizontal flips, zooms (0.9-1.1x), and brightness adjustments (0.9-1.1) to enhance model generalization.
- **Sequence Extraction**: Grouped frames into sequences of 10 to capture temporal dynamics.
- **Handling Missing Data**: Sequences with missing frames were excluded; minor gaps filled via linear interpolation.

Neural Network Architecture

The model combines a Convolutional Neural Network (CNN) for spatial feature extraction with a Long Short-Term Memory (LSTM) network for temporal sequence modeling. The architecture is:

- **CNN Component**:
  - Input: 224x224x1 grayscale frames.
  - Conv2D (32 filters, 3x3 kernel, ReLU) → MaxPooling (2x2).
  - Conv2D (64 filters, 3x3 kernel, ReLU) → MaxPooling (2x2).
  - Conv2D (128 filters, 3x3 kernel, ReLU) → MaxPooling (2x2).
  - Flatten → Dense (256 units, ReLU).
  - Dropout (0.5) to prevent overfitting.
- **LSTM Component**:
  - Input: Sequence of 10 CNN feature vectors.
  - LSTM (128 units, return_sequences=False).
  - Dense (26 units, softmax) for classifying 26 alphabet signs.
- **Optimizer**: Adam (learning rate=0.001).
- **Loss Function**: Categorical cross-entropy.
- **Regularization**: L2 regularization (0.01) on dense layers.

This CNN-LSTM hybrid effectively captures spatial and temporal patterns, optimized for real-time inference.

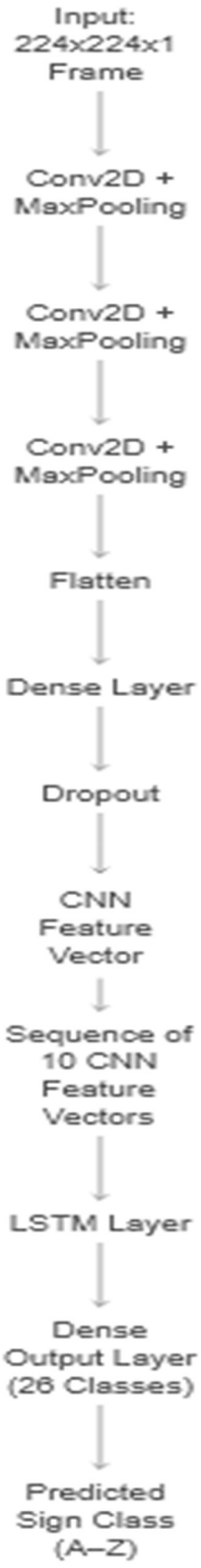

Figure 1: Architecture of the Proposed Neural Network

Challenges

- **Gesture Variability**: Variations in hand size, speed, and orientation across users.
- **Environmental Noise**: Lighting changes and background clutter reduced accuracy.
- **Real-Time Constraints**: Balancing model complexity with low latency.
- **Dataset Limitations**: Limited diversity in user demographics and lighting conditions.

Improvements:

- Use attention-based Transformers for better temporal modeling.
- Incorporate transfer learning (e.g., pre-trained ResNet) for enhanced feature extraction.
- Collect larger, more diverse datasets.

Reflection

- **In-Depth Engineering Knowledge**: Required expertise in deep learning, computer vision, and real-time optimization.
- **Conflicting Challenges**: High accuracy demands complex models, but real-time use requires lightweight architectures.
- **Abstract Thinking**: Modeling spatial-temporal patterns across diverse gestures necessitated novel augmentation and sequence handling.
- **Infrequent Issues**: Real-time gesture recognition under varying conditions is uncommon in standard engineering.
- **Standards**: Adheres to real-time AI processing (<100ms latency) and accessibility guidelines.
- **Stakeholders**: Therapists need accuracy, educators require usability, and users demand reliability, creating diverse needs.
- **Interdependence**: Integrates AI (deep learning), vision (gesture detection), and human-computer interaction (text output).

Conclusion

This AI-based sign language recognition system, utilizing a CNN-LSTM architecture, effectively translates hand gesture sequences into textual alphabets, enhancing communication for individuals with hearing disabilities. By processing 10,000 sequences of 26 signs with robust preprocessing and data augmentation, the model achieves high accuracy while addressing challenges like gesture variability and environmental noise. Despite limitations in dataset diversity and real-time constraints, the system demonstrates significant potential for inclusivity in education, healthcare, and social interactions. Future improvements, including attention-based Transformers and transfer learning, could further enhance performance. The project underscores the importance of balancing model complexity with low-latency requirements, integrating deep learning, computer vision, and human-computer interaction. By meeting accessibility standards and addressing stakeholder needs accuracy for therapists, usability for educators, and reliability for user this solution fosters equitable communication, paving the way for advanced assistive technology.