AI Powered ASL Interpreter

AUTHORS

Yasmine Mohammed 22-101174
Laila Khaled 22-101078
Nour Hany 22-101068
Ahmed Sameh 22-101198

AFFILIATIONS

جامعة مصر للمعلوماتية
EGYPT UNIVERSITY OF INFORMATICS

EGYPT UNIVERSITY OF INFORMATICS
FACULTY OF COMPUTING & INFORMATION SCIENCES

01. INTRODUCTION

Over 430 million people worldwide face challenges with hearing loss, which creates barriers to communication. This project aims to bridge the gap by developing an AI-powered Sign Language Interpreter (SLI) that performs real-time translations at the letter and word levels. The project began by evaluating multiple machine learning approaches to identify the most effective solutions, ultimately delivering a system capable of handling real-world scenarios.

02. METHODOLOY

The methodology involved systematically exploring various models for letter-level and word-level translation:

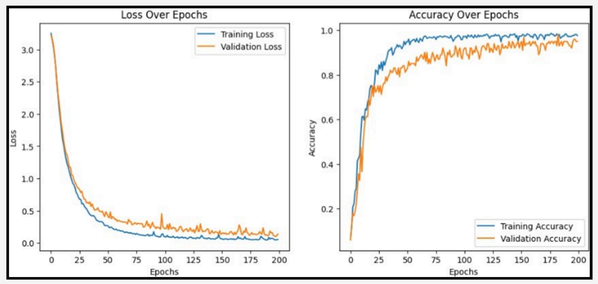
1. Letter Level: Focused on recognizing individual signs corresponding to letters.
2. Word Level: Extended the system to recognize and interpret entire words by analyzing temporal dependencies in gestures.

Throughout the project, the team experimented with several architectures and hyperparameter tuning to identify the most effective solutions.

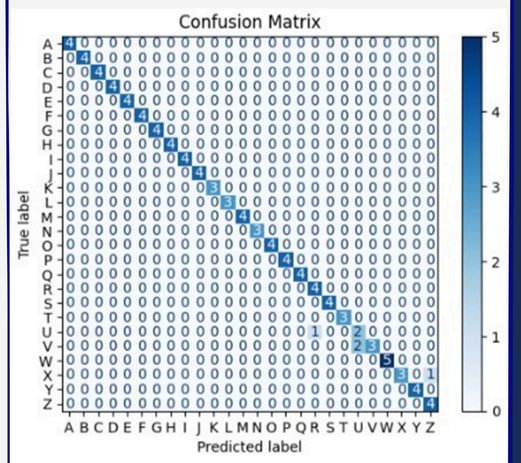
03. LETTER LEVEL ANALYSIS

To interpret individual letters in sign language, the team evaluated two approaches: YOLO and MediaPipe, comparing their performance through accuracy metrics and confusion matrices.

- YOLO: A popular object detection model trained to classify individual letters. It achieved a 69.8% test accuracy after 200 epochs with an ADAM optimizer.
- MediaPipe + FNN: MediaPipe provided real-time hand tracking, feeding data into a Feedforward Neural Network (FNN) for classification. This approach achieved a significantly higher 93.33% test accuracy, leveraging sparse categorical cross-entropy as the loss function..



MediaPipe + FNN's accuracy graph shows steady improvement, stabilizing at 93.33%, reflecting robust and consistent learning.
The loss graph demonstrates a smooth decline with minimal divergence between training and validation curves, indicating strong generalization and no overfitting.



MediaPipe's confusion matrix highlights its ability to distinguish between similar gestures. Misclassifications were minimal, reflecting the robustness of the FNN architecture.

04. WORD LEVEL ANALYSIS

To translate words, the team evaluated multiple models. Transformers showed overfitting due to limited data. Attention-based LSTM improved focus but introduced complexity. A simpler LSTM without attention emerged as the best option, achieving 90% accuracy, handling temporal dependencies effectively, and avoiding overfitting.

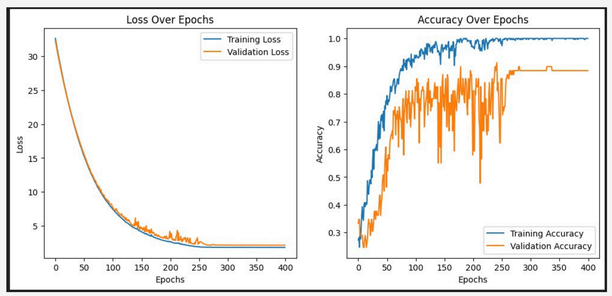
Layer (type)	Output Shape	Param #	Connected to
input_layer_6 (InputLayer)	(None, 30, 33, 3)	0	-
input_layer_7 (InputLayer)	(None, 30, 21, 3)	0	-
input_layer_8 (InputLayer)	(None, 30, 21, 3)	0	-
time_distributed_3 (TimeDistributed)	(None, 30, 99)	0	input_layer_6[0]…
time_distributed_4 (TimeDistributed)	(None, 30, 63)	0	input_layer_7[0]…
time_distributed_5 (TimeDistributed)	(None, 30, 63)	0	input_layer_8[0]…
concatenate_1 (Concatenate)	(None, 30, 225)	0	time_distributed… time_distributed… time_distributed…
bidirectional (Bidirectional)	(None, 30, 256)	362,496	concatenate_1[0]…
batch_normalization (BatchNormalizatio…	(None, 30, 256)	1,024	bidirectional[0]…
dropout_10 (Dropout)	(None, 30, 256)	0	batch_normalizat…
lstm_1 (LSTM)	(None, 30, 64)	82,176	dropout_10[0][0]
dense_9 (Dense)	(None, 30, 128)	8,320	lstm_1[0][0]
dropout_11 (Dropout)	(None, 30, 128)	0	dense_9[0][0]
dense_10 (Dense)	(None, 30, 4)	516	dropout_11[0][0]

Total params: 454,532 (1.73 MB)

Trainable params: 454,020 (1.73 MB)

Non-trainable params: 512 (2.00 KB)

Architecture of the LSTM model used for word-level sign language interpretation, detailing input layers, time-distributed processing, bidirectional LSTM, dense layers, and parameters for efficient temporal gesture recognition.



The LSTM model's accuracy graph shows steady improvement, stabilizing at 90%, demonstrating its ability to effectively learn temporal patterns in word-level gestures.
The loss graph exhibits a smooth decline with minimal divergence between training and validation curves, indicating robust generalization and no overfitting.

05. RESULTS/FINDINGS

The project successfully developed a robust AI-powered Sign Language Interpreter capable of real-time translations at both the letter and word levels.

- Letter-Level Findings: MediaPipe + FNN achieved a test accuracy of 93.33%, significantly outperforming YOLO's 69.8%. The confusion matrix confirmed minimal misclassifications for visually similar letters.
- Word-Level Findings: The LSTM model, without attention mechanisms, achieved a 90% test accuracy, demonstrating its effectiveness in learning sequential word-level gestures while avoiding overfitting.
- Key Insights:
  - MediaPipe's precise landmark extraction directly contributed to its superior performance in letter classification.
  - LSTM's ability to handle temporal dependencies made it ideal for word-level translation.

06. CONCLUSION

This study highlights the potential of AI in sign language interpretation, bridging communication gaps for the Deaf community.

- The two-stage system effectively handled letter- and word-level translations, achieving high accuracy and robust performance.
- Future enhancements include expanding datasets, integrating multimodal inputs , and optimizing real-time deployment for broader accessibility.

These findings establish a foundation for scalable, efficient sign language translation systems.