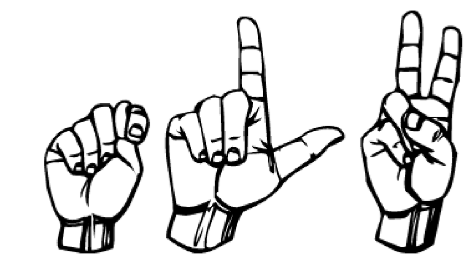
TransLingualVisionary



E6 Kavish Purani, Neeraj Ramesh, Sandra Serbu
18-500 Capstone Design, Spring 2024
Electrical and Computer Engineering Department
Carnegie Mellon University

Product Pitch

For many deaf or hard of hearing (HOH) individuals, sign language is a faster and more efficient way to communicate than written text. Written text can be inaccessible to deaf individuals due to its speed and the difficulty of learning a written language without it's phonetic component.

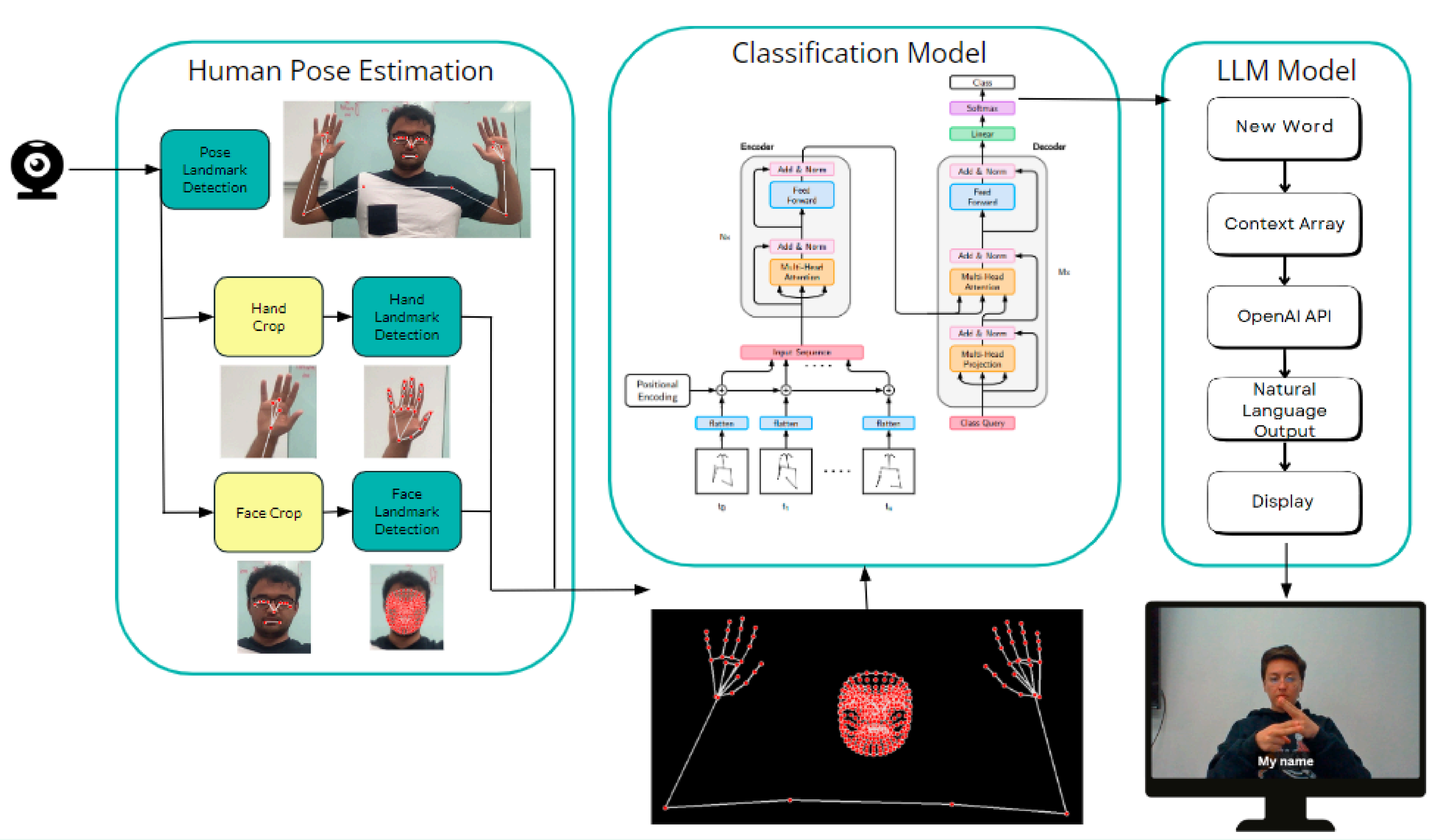
Our project aims to break down this communication barrier by developing an automated speech-to-text system to **translate ASL to written English in real-time.** We plan to do this using computer vision and machine learning to instantly transcribe a live video feed into written English text. The output text will then be displayed, enabling efficient two-way communication.

This product would benefit any hearing impaired individual who prefers communicating via ASL and may need to interact with hearing non-ASL users in virtual environments. Described below is a **local preliminary implementation** of the described product.

System Architecture

Our architecture is composed of three main parts, being our Human Pose Estimation Model, Classification Model, and LLM Model. We beign with our camera input, which sends a frame to our Human Pose Estimation model. The pose estimation model provides landmarks of points for the users face, hands, and general pose. This gets formatted and input into the classification model, which takes in the skeletal data and outputs the corresponding word through an encoder-decoder transformer. The outputted words are then compiled and sent into the LLM component, which is an API call to OpenAI's ChatGPT4, which utilizes an engineered prompt to output the fully translated sentence.

System Design Flow Chart



Conclusion & Additional Information

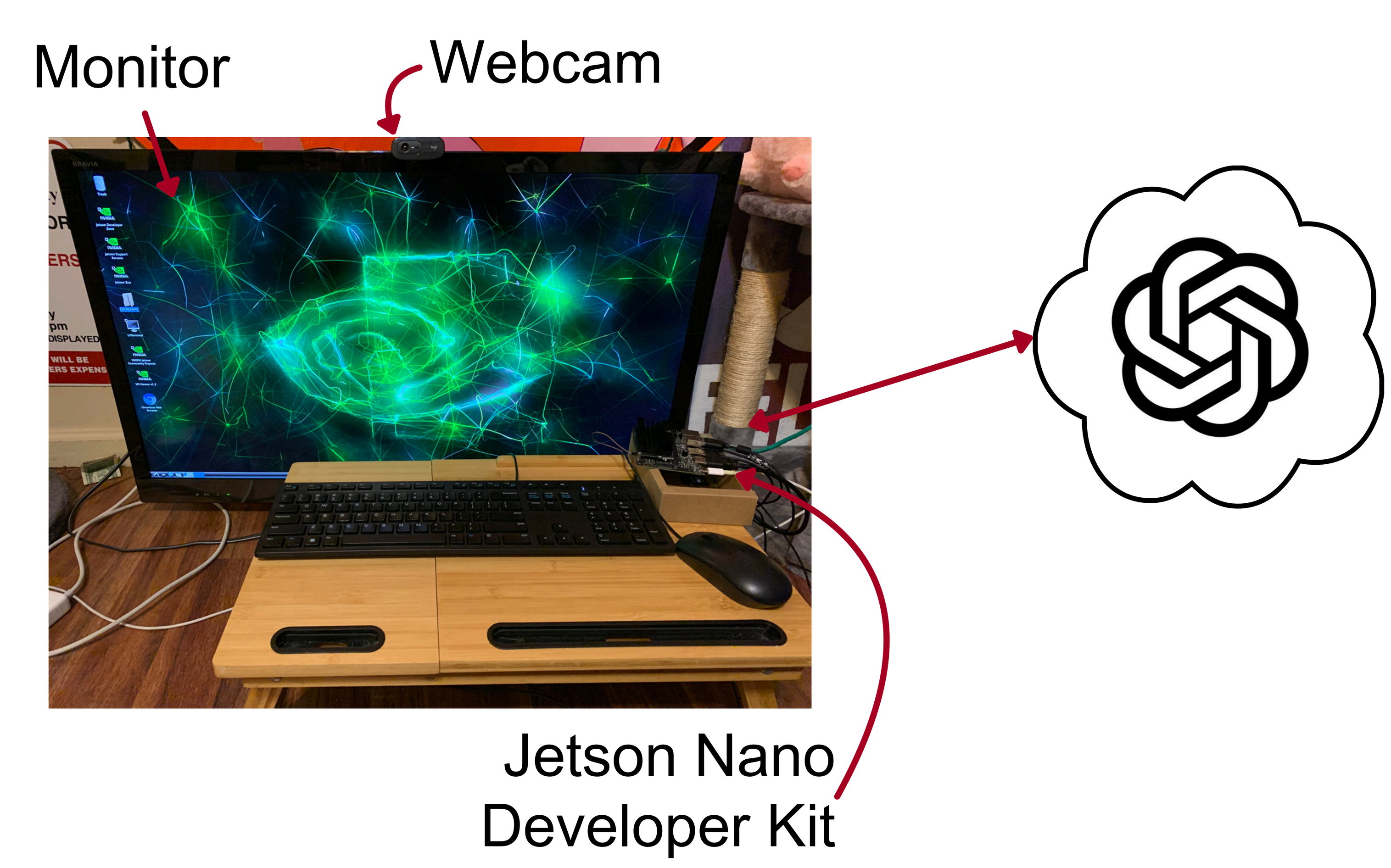
More information at http://www.ece.cmu.edu/~ece500/projects/S24-teamE6

Due to time and technical limitations, our final system is less extensive than we'd originally hoped we could make it. Despite the group's background knowledge, successfully training and implementing a modified ML classification model posed more of a challenge than originally expected. If we could continue to work on this project, we would prioritize incorporating MediaPipe's face mesh landmarks into our classification algorithm to better inform the predicted speech of the user through their facial expression in tandem with their signs.

System Description

Our system runs on a Jetson Nano with three main sub-systems as described in the system architecture: human pose estimation, sign language classification, and LLM language construction. The pose estimation is achieved through Mediapipe's Holistic Model, from which we obtain 54 landmarks (21 per hand and 12 pose) that are normalized to input into the classification model. The classification model - SPOTER - runs the normalized landmarks through a encoder-decoder based transformer to output a sign classification. Finally, the outputs are queried through OpenAI's GPT4 API to fix the english language sentence structure.

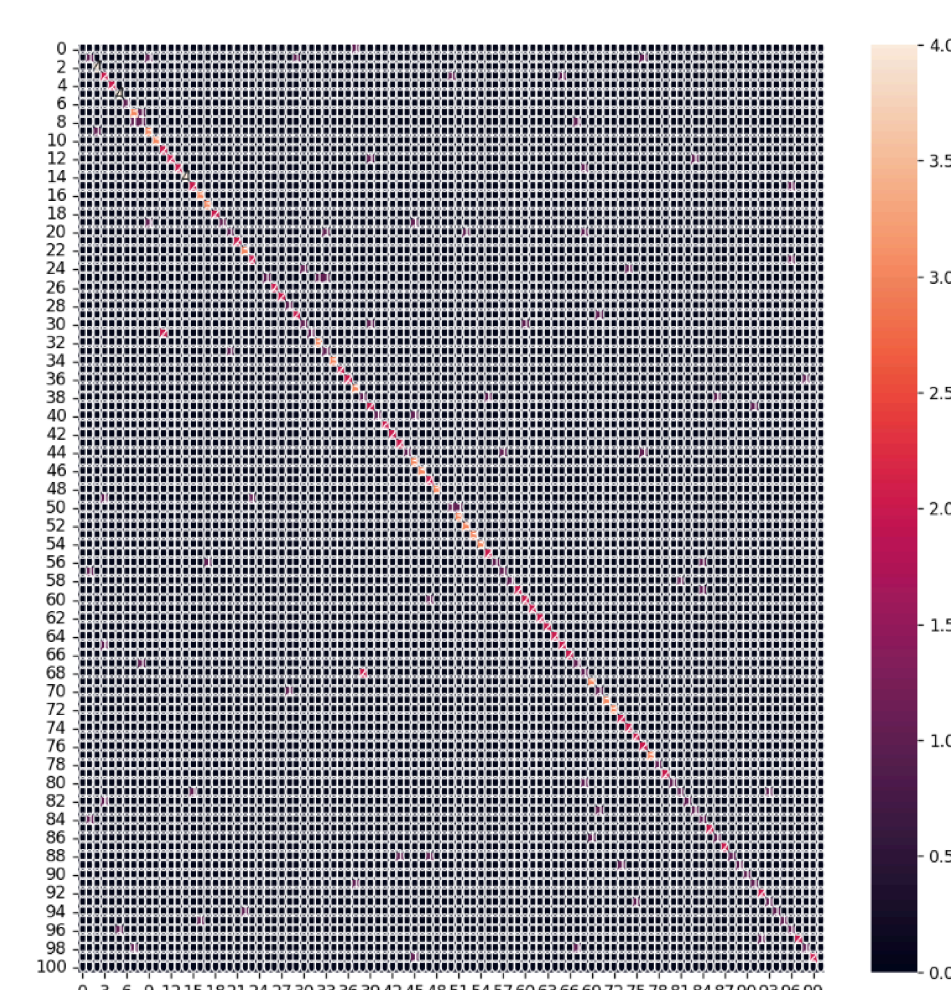
Product Implementation



System Evaluation

Results and Confusion Matrix

Metric	Goal	Results
Recognition Rate	~95%	~98%
Word Classification Accuracy	Training ~95% Validation ~85%	Training ~ 97.82% Validation ~ 72.44%
Inference Accuracy	~80%	~55%
Overall Latency	~ 3 seconds	~ 2.2 seconds
Unit Latency	HPE: ~600 ms Classification: ~800ms	HPE: ~65 ms Classification: ~12 ms



FPGA vs Jetson Tradeoffs

HPE	Jetson	FPGA	
Pros	- More CPU processing power - Models on device limits communication latency	- Opportunity to optimize models - Splitting models between devices to prevent excessive compute cost	
Cons	- More models running Jetson could decrease compute speed - Tighter space restriction due to multiple models on device	- Higher performance dependent on quantizability of models - Communicating between devices would increase latency	
FPS	~ 17 fps	Unaccelerated	Accelerated
		~3 fps	~25 fps*

Electrical & Computer ENGINEERING
Carnegie Mellon