

TransLingualVisionary

Team E6

Kavish Purani, Neeraj Ramesh, Sandra Serbu

Recap - Use Case

Problem:

- Difficult for deaf or hard of hearing (HOH) individuals to participate in live digital environments (online meetings, live streams, etc.)
- Lack of widespread understanding of American Sign Language (ASL)
- Often require assistance from translators to communicate

Solution:

- A real-time ASL speech to English text translator on a user friendly web application

Recap – Our Solution

TransLingualVisionary (or TLV) is an ASL-to-Text translator that includes:

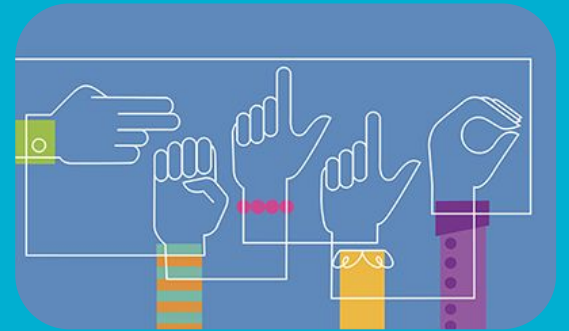
- Live translation of ASL to text
- Accelerated FPGA pre- and post- image processing
- User-friendly web app to visualize processed ASL input and text output

TLV will allow ASL users to:

- Quickly communicate to non-ASL users
- Document their speech in a simple and efficient manner

ECE Focus Areas:

- Software Systems
- Hardware Systems



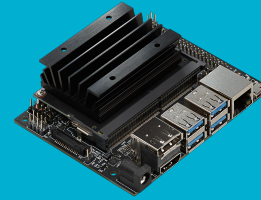
Recap - Design Requirements

<u>Requirement</u>	<u>Metric</u>
Recognize when a user is signing	~95% sign recognition rate
Correctly identify ASL words	Recognize 2000 words at ~80% accuracy
Correctly interpret ASL semantics	Translate identified clusters of words into full english sentences with a BLEU score of ~40%
Classification Distance	Recognize and retain accuracy of the classification model up to 4-5 feet away from the camera.
Text Accessibility	Display and collect the ASL Speech in an accessible user format that can be easily found and read.
Overall Latency ~ real time	Present visual feed and translation on web UI within ~3 seconds

Solution Approach

Increase accessibility to online spaces

- Increased community involvement via increasing range of communication
- Reduce digital isolation



Camera

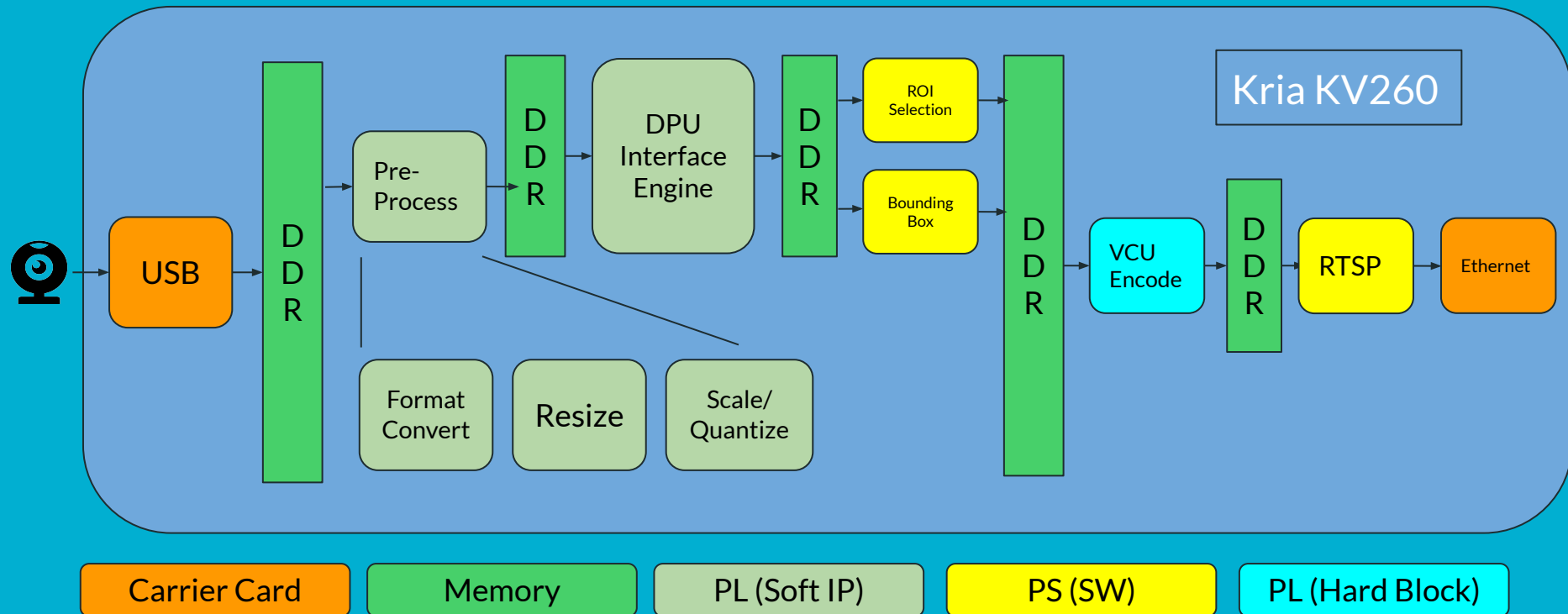
FPGA

Jetson

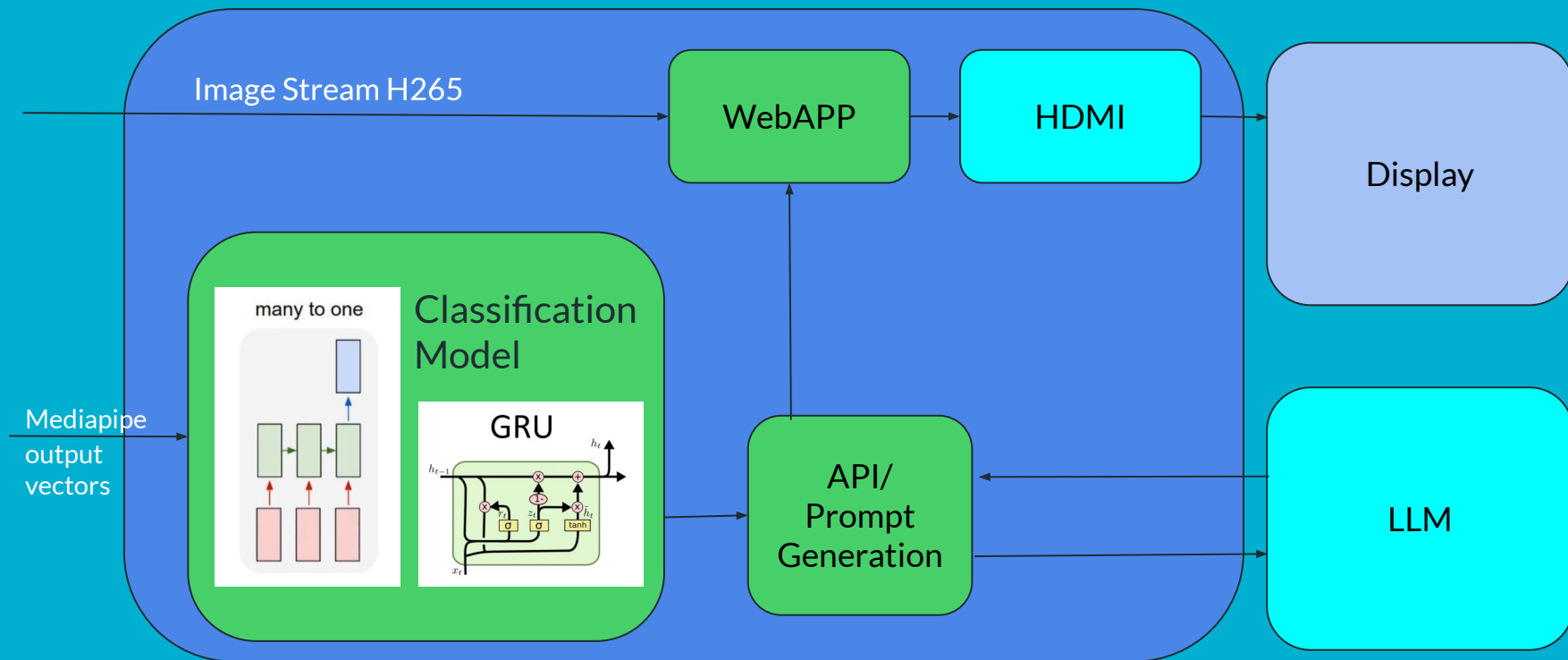
Web App



Block Diagram - MediaPipe on FPGA



Block Diagram - Jetson



Implementation Plan

Logitech C920S (pre-owned)

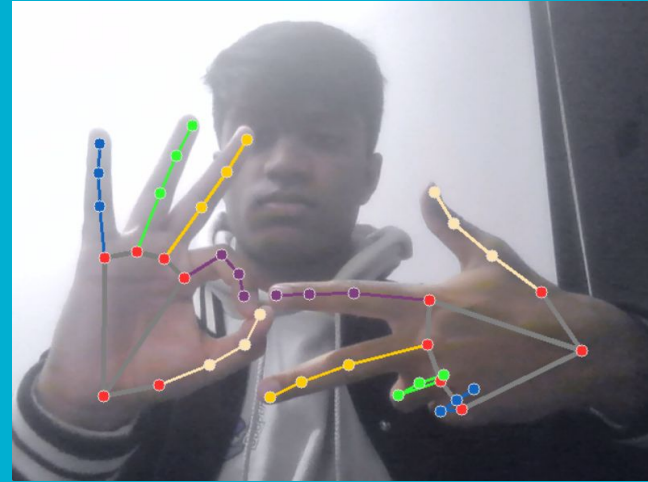
- USB Interface

Kria KV260 (pre-owned)

- PL (Soft IP) will be developed ourselves using Vitis AI
 - Will rely on pre-trained MediaPipe model and Vitis model zoo
- PS (SW) will be built using open source libraries

Jetson Nano (from inventory)

- Self-train RNN model using WLASL dataset
- Personalized GUI to view overlaid text and video
- Use ChatGPT4 API to do prompt generation



Testing & Verification

Latency	<ul style="list-style-type: none">• Unit component latency testing<ul style="list-style-type: none">◦ Create benchmarking set and measure latency via inter-component timestamps• Overall inference latency ~3 seconds
Accuracy and Correctness	<ul style="list-style-type: none">• Calculate validation accuracy of RNN and LLM<ul style="list-style-type: none">◦ RNN Tuning Target: 85% accuracy◦ Prompted LLM Target: BLEU Score ~45%• Calculate inference accuracy of RNN and LLM<ul style="list-style-type: none">◦ RNN Target: 80% accuracy◦ Prompted LLM Target: BLEU Score ~40%• Method correctness of FPGA accelerated operations<ul style="list-style-type: none">◦ Accuracy comparison with CPU generated results of ~100%
Usability and Accessibility	<ul style="list-style-type: none">• System Usability Scale score of over 85%

Risks & Unknowns

- Porting Human Pose Estimation on FPGA
 - Could significantly improve latency and space flexibility for RNN
 - Test FPGA vs Jetson HPE compute time
 - Test feasibility early to pivot to Jetson in case of failure
- Using Mediapipe's Hand Pose Estimation
 - Captures intricacies of hand position relative to image space, but not relative to person
 - Randomly modify the size of training videos to train for variance in distance
 - Test accuracy over variant distances
- GRU RNN
 - Light model that captures short term temporal data; might not capture patterns between frames
 - Test alongside LSTM model to check for validation accuracy and latency

Division of Labour

	Kavish	Neeraj	Sandra
FPGA MediaPipe Implementation	✗		
RNN Model	✗	✗	✗
Prompt Generation & LLM		✗	
Web Application			✗
Testing & Integration	✗	✗	✗

Gantt Chart

Sandra, Kavish, Neeraj

Kavish

Sandra, Neeraj

