Semantic Segmentation Optimization

Project Planning

By: sddec25-01

Problem Statement

Problem

- People with disabilities face risks from undetected medical issues. Traditional methods lack real-time monitoring.
- Using eye movement tracking with semantic segmentation can detect warning signs and automatically reposition users to prevent incidents, improving safety needs.

Client

- Volunteered to help individuals with cerebral palsy.
- Create an assistive wheelchair technology.

Team

- Update the system to increase throughput.

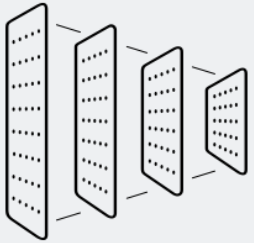
Project Overview

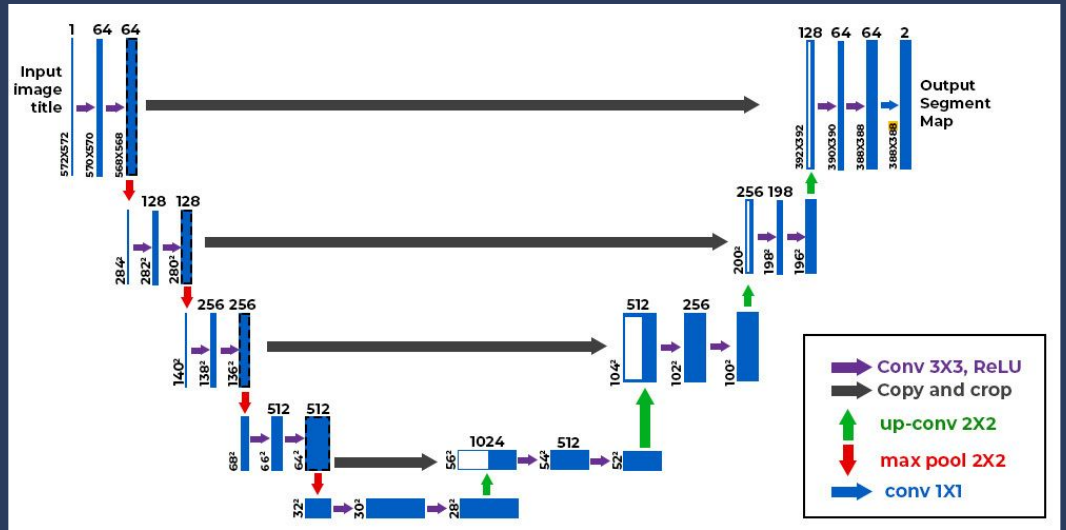
Components/Software:

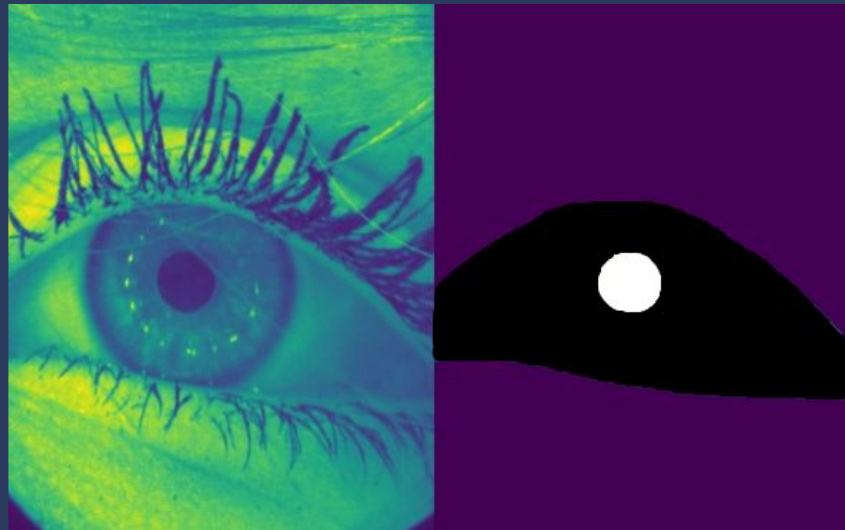
- Kria Board Kv260
- Vivado
- Xilinx
- Vitis-AI
- Pytorch
- ONNX & ONNX-Runtime

U-net Semantic Segmentation

- Deep CNN
- Contracting Encoder
- Expanding Decoder







Each “forward step” applies a relu function to the output of a repeated convolutional layer application over input channels.

U-net Semantic Segmentation cont.

Encoder

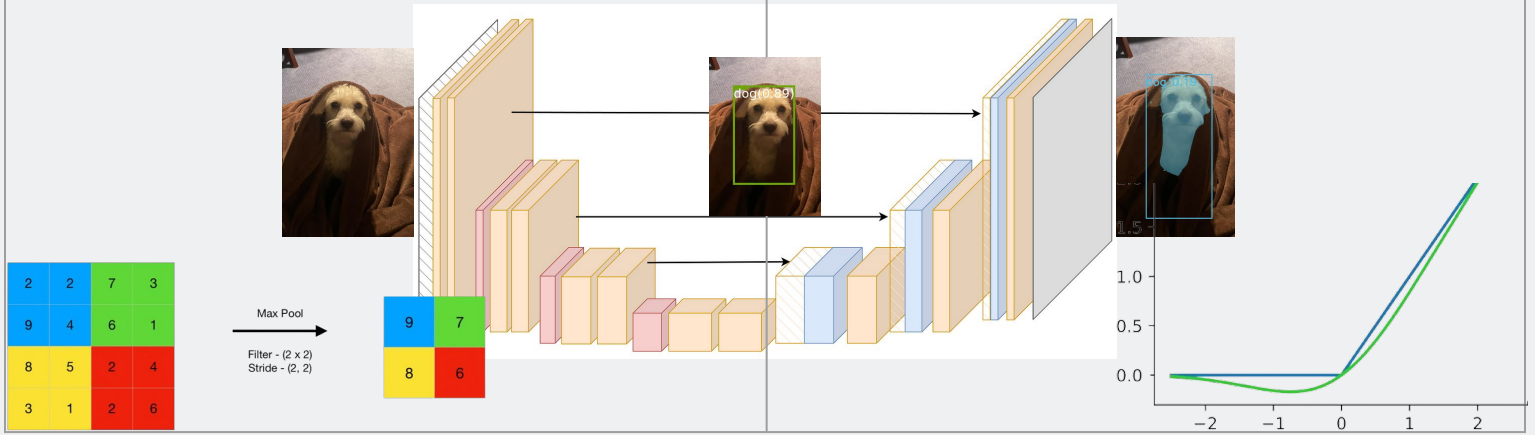
- Downsampling (i.e. 2x2 Max Pool) compensated by the doubles # channels
- Transmit to across to decoder

Includes spatial info

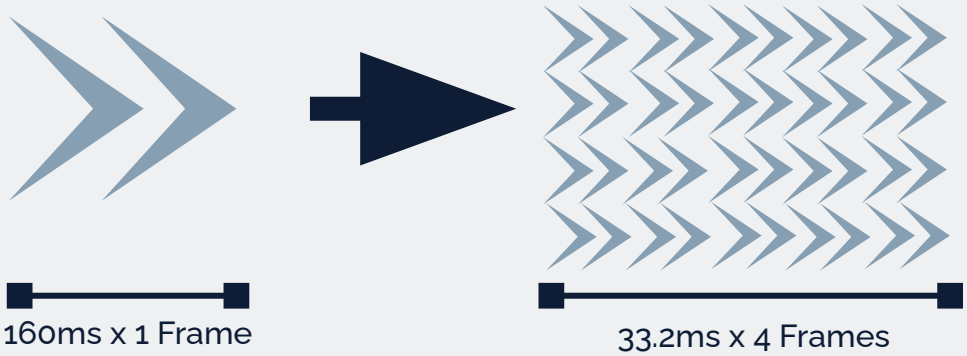
Decoder

- Receive from encoder
- Upsampling (i.e. 2x2 Convolution) decreases the amount of channels

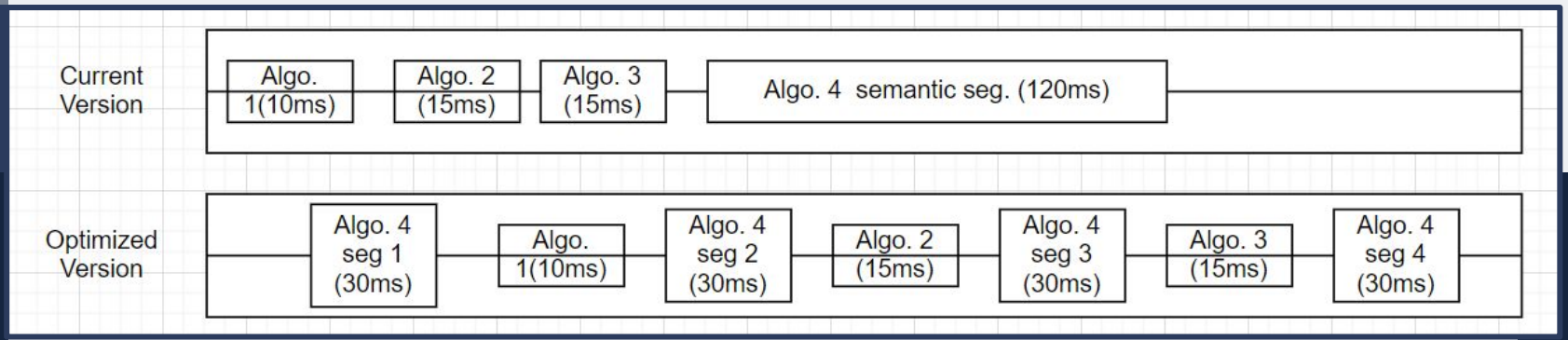
Includes semantic Info



Objective

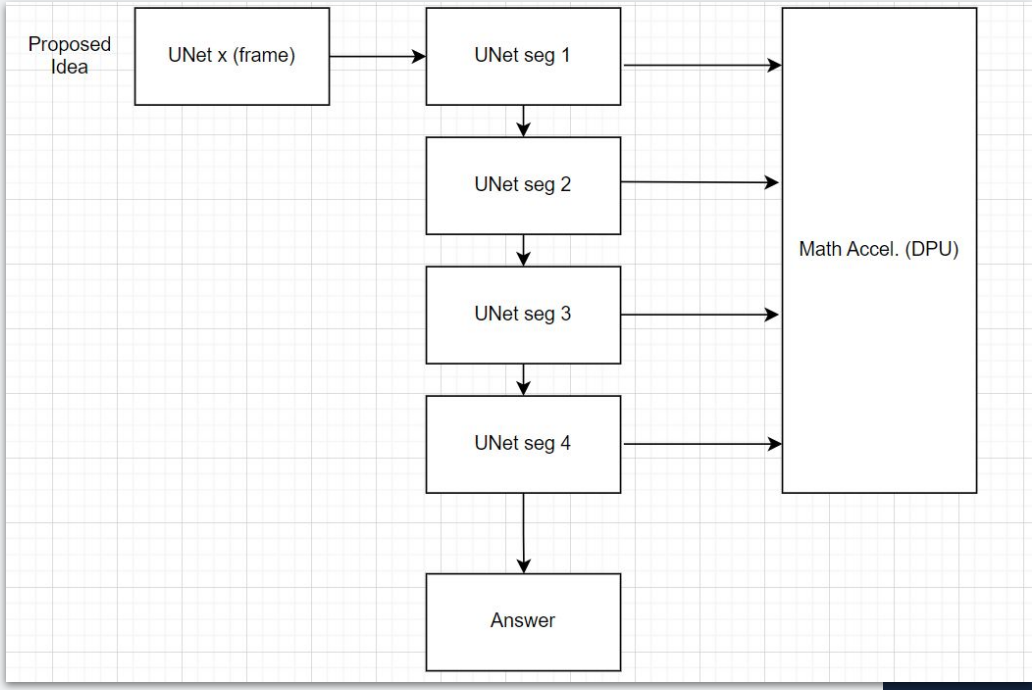


Increase throughput through pipelining U-net algorithm over 4 cores and across MPU.



Task Decomposition

Role1 - Tyler	Dividing out the algorithm into 4 equal parts, code it in a way to have 4 segments to produce the same result and accuracy as prior design.	Skillset: equal math distribute, reasonable coding capability.
Role2 - Aidan	Implement algorithm into current codebase, (already has 4 pipelines) Understand how each thread works on a part of the equation.	Skillset: C++ expertise (OOP)
Role3 - Conner	Works with OS Docker environment to optimize all of the activities that need to happen. Splitting up the Onnx for loading into the MPU for Multicore Loading and Layering.	Skillset: OS Knowledge, comfort with scheduler (what is done when), Docker.
Role4 - Joey	Understand Hardware (Kira board) inside and out to give other members accurate benchmarks of what the board can handle.	Skillset: Hardware focused, intensive data sheet research.



Tracked Metrics

- Throughput
- Accuracy
- Resource Utilization

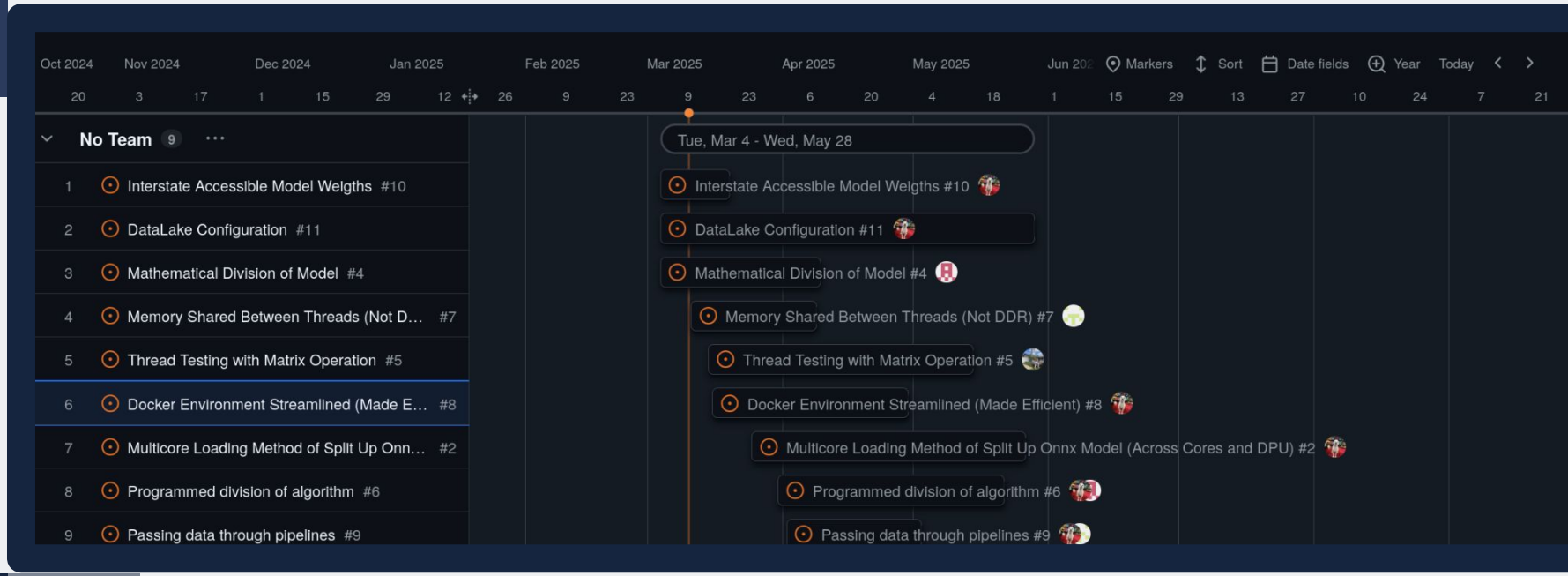
Milestones

- Mathematical division of the Algorithm
- Loading of Split Algorithm weights onto MPU
- Pipelined Implementation of the Semantic Segmentation algorithm across the 4 developed threads.
- Increased Throughput over multiple frames

Project Management Style

Waterfall & Agile

Gantt Chart


Oct 2024
Nov 2024
Dec 2024
Jan 2025
Feb 2025
Mar 2025
Apr 2025
May 2025
Markers
Sort
Date fields
Year
Today
No Team 9
Tue, Mar 4 - Wed, May 28
1 Interstate Accessible Model Weigths #10
2 DataLake Configuration #11
3 Mathematical Division of Model #4
4 Memory Shared Between Threads (Not D… #7
5 Thread Testing with Matrix Operation #5
6 Docker Environment Streamlined (Made E… #8
7 Multicore Loading Method of Split Up Onn… #2
8 Programmed division of algorithm #6
9 Passing data through pipelines #9
Interstate Accessible Model Weigths #10
DataLake Configuration #11
Mathematical Division of Model #4
Memory Shared Between Threads (Not DDR) #7
Thread Testing with Matrix Operation #5
Docker Environment Streamlined (Made Efficient) #8
Multicore Loading Method of Split Up Onnx Model (Across Cores and DPU) #2
Programmed division of algorithm #6
Passing data through pipelines #9

Communication Methods

- Telegram - main form of communication with client and prior years team members
- Phone Messaging apps
- Discord

Risks & Mitigation

- Completion Delays
  - As most of the required parts of the project need to be done serially, we work and assist each other as a team.
- Damage to Hardware
  - Keep hardware in secure location within safe container away from environmental contaminants.
- Data Security
  - United States distributed data store (s3-compatible)
  - Git-based Source and Data Version Control

Conclusion

Problem Solved: Real-time monitoring for individuals with disabilities using eye tracking.

Outcome: Improve safety and throughput with pipelined U-Net on MPU.

Next Steps: Optimize Performance and and explore scalability for broader use.

Thank You :)

Questions ?