Heterogeneous Machine Learning

DESIGN DOCUMENT

Team Number
ssdec23-02
Client
JR Spidell
Advisers
Diane Rover
Team Members/Roles
Sandro Panchame, Rudolph Nahra, Alek Comstock, Jeffrey Kasper
Team Email
sddec23-02@iastate.edu
Team Website

Revised: April 23, 2023

Executive Summary

Development Standards & Practices Used

List all standard circuit, hardware, software practices used in this project. List all the Engineering standards that apply to this project that were considered.

7001-2021 - IEEE Standard for Transparency of Autonomous Systems.

- Iso/iec/ieee international standard - software and systems engineering - software testing -- part 2: test processes - redline

1532-2002 - IEEE Standard for In-System Configuration of Programmable Devices

Summary of Requirements

List all requirements as bullet points in brief.

- Our system takes, as input, a video feed at 60 frames per second of a human eye
- For each frame, system will use our machine learning model to determine: whether the person is blinking, and, if not blinking, the location of the pupil of the eye in the image.
- Our system must be able to process frames faster than they are received on average, such that minimal storage is needed for incoming frames.
- Our system will be implemented on a single Kria SOM 260 board
- Externally, the neural network will be certified to be safe according to our developed safety criteria using Marabou, a neural network verification tool.

Applicable Courses from Iowa State University Curriculum

List all Iowa State University courses whose contents were applicable to your project.

CPRE308 - Operating Systems

CPRE288 - Embedded Systems

COM S 474/574 - Introduction to Machine Learning (Not taken by any team members)

CPRE381 - Computer Organization

New Skills/Knowledge acquired that was not taught in courses

List all new skills/knowledge that your team acquired which was not part of your Iowa State curriculum in order to complete this project.

Usage of OpenAMP for communication between heterogeneous processing units

Knowledge of convolutional neural networks, how to train them, and various representations to save them in

Usage of Marabou to perform formal analysis on a neural network

Capabilities of Xilinx DPU and necessary input formats

Table of Contents

List of figures/tables/symbols/definitions (This should be the similar to the project plan)

1 Team

1.1 Team Members

Sandro Panchame, Alek Comstock, Jeffrey Kasper, Rudolph Nahra

1.2 Required Skill Sets for Your Project

(if feasible – tie them to the requirements)

- Embedded systems knowledge
- Experience working on a linux operating system
- General knowledge on the python language
- Knowledge of neural networks
- Skills in analyzing neural networks

1.3 Skill Sets covered by the Team

(for each skill, state which team member(s) cover it)

- general/base knowledge of embedded systems: Alek C., Rudolph Nahra, Jeffery K.
- Knowledge of Python Language: Sandro P., Rudolph Nahra, Jeffery K.
- Experience with linux system: Sandro P., Rudolph N., Alek C., Jeffery K.
- Knowledge of neural networks: Rudolph Nahra
- Skills in analyzing neural networks: None

1.4 Project Management Style Adopted by the team

Waterfall Project management: Each week, we meet with our client and he breaks down what has been accomplished each week, and what needs to get done. We are starting with small goals to establish our understanding of the technology, and building up to the final project. We will then find a person best suited for each week's goal and assign it - whether done by self volunteer, best suited for the role, or elimination.

1.5 Initial Project Management Roles

(Enumerate which team member plays what role)

Alek Comstock - Embedded systems design

Jeffery Kasper - Embedded system design

Sandro Panchame - Neural network analysis and optimization

Rudy Nahra - Neural network analysis and optimization

2 Introduction

2.1 Problem Statement

What problem is your project trying to solve? Use non-technical jargon as much as possible.

Our project is concerned with the safety of pilots while piloting aircraft in stressful, demanding situations where humans may encounter problems such as hypoxia or sensory overload that render them unfit to pilot the aircraft. A solution to this problem is to analyze the pilot's eye movements to determine if they are consistent with various bad conditions. Our contribution to this will be a system that takes in video feed of a pilot's eyes, and uses two different machine learning models on the same embedded system to determine the position of the pilot's pupil, whether or not they are blinking, and the type of eye movement that is occurring (fixation, saccade, smooth pursuit). This information could then be used in a broader system to identify unwanted conditions.

2.2 Requirements & Constraints

List all requirements for your project . This includes functional requirements (specification), resource requirements, qualitative aesthetics requirements, economic/market requirements, environmental requirements, UI requirements, and any others relevant to your project. When a requirement is also a quantitative constraint, either separate it into a list of constraints, or annotate at the end of requirement as "**(constraint)**". Other requirements can be a single list or can be broken out into multiple lists based on the category.

Functional Requirements:

- The system takes in a video feed of a human eye and, for each frame, outputs the position of the pupil in the image, whether they are blinking, and the eye movement pattern they are exhibiting
- The system runs on only a Kria SOM KV260 (constraint)
- The system must be able to process each frame of a video feed in 1/60th of a second per frame. **(constraint)**
- In order to meet time constraints and scalability, the system should be able concurrently utilize all available processors.
- The output of the system must have "safe" bounds. The definition of safety will differ with each use case, but we must have verification that the output of the system lies within some predetermined bounds.

The resource requirements:

3 FPGA boards
4 The installation of petalinux on the boards
5 Working, pre-trained Machine Learning Algorithms.

2.3 Engineering Standards

What Engineering standards are likely to apply to your project? Some standards might be built into your requirements (Use 802.11 ac wifi standard) and many others might fall out of design. For each standard listed, also provide a brief justification.

https://ieeexplore.ieee.org/document/9726144

7001-2021 - IEEE Standard for Transparency of Autonomous Systems.

This standard is important for our project because we need to analyze our deep learning model to ensure its output is safe, as our project could be employed in safety-critical applications. The standard will help us measure how safe they are.

https://standards.ieee.org/ieee/29119-2/7498/

29119-2-2021 - Iso/iec/ieee international standard - software and systems engineering - software testing -- part 2: test processes - redline

This standard is useful to use as a form of ensuring we meet many possible problems our project could run into. It ensures that each step along the way, any updates or changes to the project, should meet our testing standards.

https://ieeexplore.ieee.org/document/1176958

1532-2002 - IEEE Standard for In-System Configuration of Programmable Devices

This standard applies to our project because we will utilize an FPGA board. An FPGA board is a type of Programmable Device. We will need to utilize these standards to effectively configure the board.

2.4 Intended Users and Uses

Who benefits from the results of your project? Who cares that it exists? How will they use it? Enumerating as many "use cases" as possible also helps you make sure that your requirements are complete (each use case may give rise to its own set of requirements).

Possible use cases:

- To track the eye movement of an operator of a large vehicle or aircraft and determine the operator's state to determine if they're fit to continue operation. Possible states may include but are not limited to fixation, saccade, and smooth pursuit.
- To aid medical professionals in finding possible eye-related illnesses.
- As a proof-of-concept for other Machine Learning Algorithms to be used on an embedded system, information on this project could help develop other kinds of tools.

Intended users:

- Pilots
- Surgeons
- Clinical Nurse Specialists
- Heavy Machine Operators
- Train Engineers
- Astronauts

3 Project Plan

3.1 Project Management/Tracking Procedures

Which of agile, waterfall or waterfall+agile project management style are you adopting. Justify it with respect to the project goals.

- Waterfall project management. The project goals include two machine learning models that satisfy certain requirements, and a hardware system with the programming and configuration to be able to run both of them at the same time. We will need both of these functions before the goal can be accomplished, so a waterfall style will allow us to ensure these are both working before we try to put them together.

What will your group use to track progress throughout the course of this and the next semester. This could include Git, Github, Trello, Slack or any other tools helpful in project management.

- We have a Google Slides presentation that is shared with our client where we list progress, challenges, milestones, and future plans. We have a discord server and a telegram channel where resources are shared with each other. We will be using Git for version control.

3.2 Task Decomposition

In order to solve the problem at hand, it helps to decompose it into multiple tasks and subtasks and to understand interdependence among tasks. This step might be useful even if you adopt agile methodology. If you are agile, you can also provide a linear progression of completed requirements aligned with your sprints for the entire project.

The team is divided into two teams, those that work on the boards and those that work on the files provided by the client and the ML algorithm (represented by white and black bullet points respectively).

- set up Kria SOM KV260:
  - Get Kria SOM KV260 boards firmware updated
  - Get PetaLinux running on Kria SOM KV260
- develop communication between inter-processes:
  - Write pseudo code for communication between processors

        - Using OpenAMP package write code for processor communication
    - Using Vitis write code to optimize Deep-Learning Processing units
    - set up Ultra 96 board:
        - Get Kria Ultra96 boards firmware updated
        - Get PetaLinux running on Ultra96
    - develop communication between inter-processes:
        - Write pseudo code for communication between processors for Ultra96
        - Using OpenAMP package write code for processor communication
    - Using Vitis write code to optimize Deep-Learning Processing units

- Set up the Software Environment on Ubuntu 20.04 systems
    - Setup a virtual Machine to Host the OS or boot with the OS
    - install appropriate packages for the project
- Analyze the files a previous team worked on
    - Analyze the programs and file interactions
- Learn and use Marabou, to analyze the ML algorithm, and to find areas of improvement
    - Install Marabou and test out sample code
    - Use Marabou on ML algorithm
- Verify that the neural network does not have unsafe output
    - Develop criteria to evaluate safety of the neural network
    - Evaluate the safety of the neural network
- Modify the parameters and/or dataset of the neural network and reevaluate until we meet safety criteria
    - Receive dataset by mail
    - Clean dataset and transform it to be like the one we are currently using
- Get new large dataset and train the neural network on it

3.3 Project Proposed Milestones, Metrics, and Evaluation Criteria

What are some key milestones in your proposed project? It may be helpful to develop these milestones for each task and subtask from 2.2. How do you measure progress on a given task? These metrics, preferably quantifiable, should be developed for each task. The milestones should be stated in terms of these metrics: Machine learning algorithm XYZ will classify with 80% accuracy; the pattern recognition logic on FPGA will recognize a pattern every 1 ms (at 1K patterns/sec throughput). ML accuracy target might go up to 90% from 80%.

In an agile development process, these milestones can be refined with successive iterations/sprints (perhaps a subset of your requirements applicable to those sprint).

- Understand the current state of the software and hardware design
- Setup the software and hardware environments

- Message passing process/algorithm (RPU to APU and APU to RPU)
- Determine safety criteria for machine learning model
- Modify machine learning model to fit safety criteria
- get pattern recognition logic onto board
- the pattern recognition logic for eye detection on FPGA will recognize a pattern every 17 ms (at 60 frames/sec throughput)

3.4 Project Timeline/Schedule

• A realistic, well-planned schedule is an essential component of every well-planned project

• Most scheduling errors occur as the result of either not properly identifying all of the necessary activities (tasks and/or subtasks) or not properly estimating the amount of effort required to correctly complete the activity

• A detailed schedule is needed as a part of the plan:

– Start with a Gantt chart showing the tasks (that you developed in 2.2) and associated subtasks versus the proposed project calendar. The Gantt chart shall be referenced and summarized in the text.

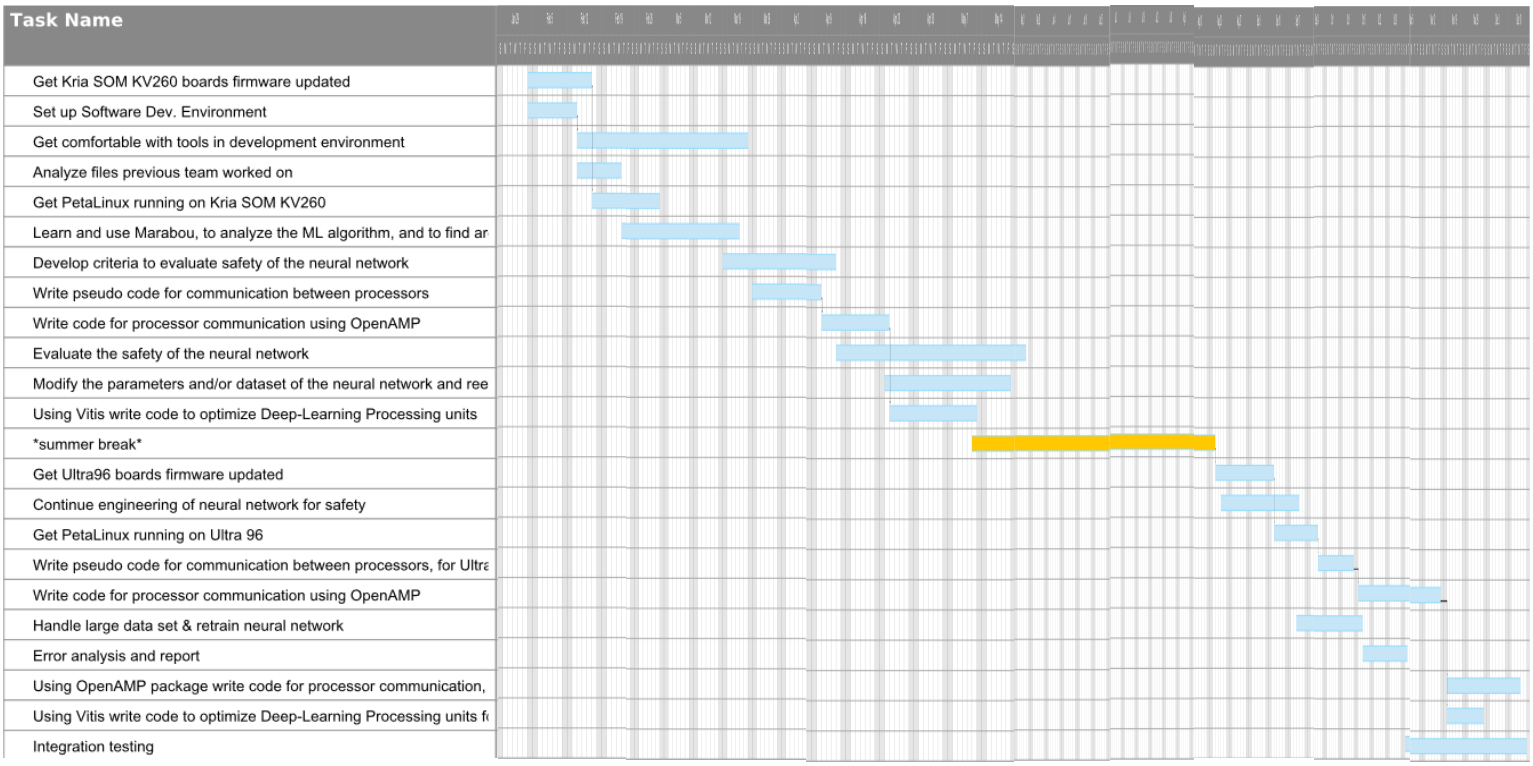
– Annotate the Gantt chart with when each project deliverable will be delivered

• Project schedule/Gantt chart can be adapted to Agile or Waterfall development model. For agile, a sprint schedule with specific technical milestones/requirements/targets will work.

Simple Chart:

A	B	C	D	E	F
1st half Semester 1	2nd half Semester 1	Summer Break	1st half Semester 2		2nd half Semester 2
Ideation					
	Creation		Creation		
				Testing	
					Evaluation/refinement

Detailed Chart:


Task Name
Get Kria SOM KV260 boards firmware updated
Set up Software Dev. Environment
Get comfortable with tools in development environment
Analyze files previous team worked on
Get PetaLinux running on Kria SOM KV260
Learn and use Marabou, to analyze the ML algorithm, and to find ar
Develop criteria to evaluate safety of the neural network
Write pseudo code for communication between processors
Write code for processor communication using OpenAMP
Evaluate the safety of the neural network
Modify the parameters and/or dataset of the neural network and ree
Using Vitis write code to optimize Deep-Learning Processing units
summer break
Get Ultra96 boards firmware updated
Continue engineering of neural network for safety
Get PetaLinux running on Ultra 96
Write pseudo code for communication between processors, for Ultra
Write code for processor communication using OpenAMP
Handle large data set & retrain neural network
Error analysis and report
Using OpenAMP package write code for processor communication,
Using Vitis write code to optimize Deep-Learning Processing units f
Integration testing

3.5 Risks And Risk Management/Mitigation

Consider for each task what risks exist (certain performance target may not be met; certain tool may not work as expected) and assign an educated guess of probability for that risk. For any risk factor with a probability exceeding 0.5, develop a risk mitigation plan. Can you eliminate that task and add another task or set of tasks that might cost more? Can you buy something off-the-shelf from the market to achieve that functionality? Can you try an alternative tool, technology, algorithm, or board?

- Set up the Software Environment on Ubuntu 20.04 systems
    - Issues with how the environment may be set up.
    - Issues on location of the OS. (boot from drive vs. virtual machine)
    - Overall Risk Factor: 0.1
- Analyze the files a previous team worked on
    - Resource allocation/availability on a user's machine. There may not be enough RAM available to run certain programs or software.
    - Overall Risk Factor: 0.1
- Learn and use Marabou, to analyze the ML algorithm, and to find areas of improvement
    - Marabou is new to the client, there will be issues where they may not be able to assist.
    - Overall Risk Factor: 0.5
    - Mitigation Plan: Actively contact maintainers of Marabou, for general advice, and troubleshooting issues.
- Develop criteria to evaluate safety of the neural network
    - The current criteria is tied to the effectiveness of the Neural Network. The criteria may change due to unforeseen behavior
    - Overall Risk Factor: 0.4
- Evaluate the safety of the neural network
    - Neural Network may fail the evaluation
    - Overall Risk Factor: 1.0
    - Mitigation Plan: Perform the below task
- Modify the parameters and/or dataset of the neural network and reevaluate until we meet safety criteria
    - The ML algorithm not reach the required criteria
    - Overall Risk Factor: 0.3
- Get new large dataset and train the neural network on it
    - Time to train the algorithm. We may run into issues where we can't train it on the entire dataset. If we try to train it with an entire dataset, something could happen in the middle of training (i.e a power outage). If we train in chunks, there may be chunks leftover by the end of the year.
    - Overall Risk Factor: 0.4
- Avoid damaging the boards
    - the board is an open piece of components, so avoiding any static shock to avoid damage is crucial

    - Overall Risk Factor: 0.1
- Making sure Baremetal works on the board
    - Our entire project relies on baremetal working so we can set up our code on it. As it is designed with these boards in mind, the odds of it failing is low
    - Overall Risk Factor: 0.1
- Out of box firmware does not allow PetaLinux
    - The board comes with a factory firmware setup. If there are bugs with this firmware we may need to upgrade or downgrade the firmware.
    - Overall Risk Factor: 0.5
    - Mitigation Plan: Find older or newer firmware that is known to work with PetaLinux
- Computation time on DPUs is more than 17ms
    - If the computation time on one DPU is too long we will need to utilize more than one DPU to complete the computation.
    - Overall Risk Factor: 0.8
    - Mitigation Plan: We have budgeted the ability to purchase DPU fabric expansion cards. We will need to setup and program these cards.

3.6 Personnel Effort Requirements

Include a detailed estimate in the form of a table accompanied by a textual reference and explanation. This estimate shall be done on a task-by-task basis and should be the projected effort in the total number of person-hours required to perform the task.

inTask Name	Duration (Hours)	Start	Finish	Task Details
Get Kria SOM KV260 boards firmware updated	8.5	02/04/23	02/16/23	acquire boards, set up and flashed firmware
Set up Software Dev. Environment	6	02/04/23	02/13/23	Install Ubuntu linux and setup multiple tools including Vitis, balena etcher, Juypiter notebook, etc.
Get comfortable with tools in development environment	25	02/14/23	03/24/23	Utilize the tools in order to become comfortable in the environment.
Analyze files previous team worked on	6	02/14/23	02/22/23	Get familiar with the existing codebase that our client has
Get PetaLinux running on Kria SOM KV260	9.5	02/17/23	03/03/23	Flash PetaLinux onto bootable SD cards and

				fixed boot issues.
Learn and use Marabou, to analyze the ML algorithm, and to find areas of improvement	18	02/23/23	03/22/23	Examine Marabou documentation and experiment with different constraints
Develop criteria to evaluate safety of the neural network	18	03/19/23	04/13/23	Give a mathematical definition of a “safe” neural network that can be expressed in the Marabou framework
Write pseudo code for communication between processors	11	03/26/23	04/10/23	2 members responsible for board development write inter-process communication and resource allocation outline
Write code for processor communication using OpenAMP	9	04/11/23	04/24/23	write actual code using software designed for Machine learning resource management on select boards
Evaluate the safety of the neural network	25	04/14/23	05/25/23	Use Marabou to evaluate whether the neural network meets our safety criteria, and continue to reevaluate as we modify the network
Modify the parameters and/or dataset of the neural network and reevaluate until we meet safety criteria	18	04/24/23	05/19/23	Continue modifying the network while it does not meet the safety criteria
Using Vitis write code to optimize Deep-Learning Processing units	12	04/25/23	05/12/23	Use Vitis, a program to optimize Machine learning algorithms to make the machine learning algorithms run faster on the specific Kria board

summer break	72d	05/12/23	08/21/23	
Get Ultra96 boards firmware updated	12	08/22/23	09/11/23	acquire boards, set up and flashed firmware
Continue engineering of neural network for safety	18	08/24/23	09/20/23	described above
Get PetaLinux running on Ultra 96	10	09/12/23	09/27/23	Flash PetaLinux onto bootable SD cards and fixed boot issues.
Write pseudo code for communication between processors, for Ultra 96	10	09/28/23	10/13/23	2 members responsible for board development write inter-process communication and resource allocation outline
Write code for processor communication using OpenAMP	21	10/16/23	11/17/23	write actual code using software designed for Machine learning resource management on select boards
Using OpenAMP package write code for processor communication, for Ultra 96	15	11/20/23	12/13/23	write code using software designed for Machine learning resource management on Ultra96 board
Using Vitis write code to optimize Deep-Learning Processing units for Ultra96	12	11/20/23	12/01/23	Use Vitis to make the machine learning algorithms run faster on the specific Kria board

3.7 Other Resource Requirements

Identify the other resources aside from financial (such as parts and materials) required to complete the project.

- Additional RAM in order to run Marabou and analyze the ML algorithms with it.

Additional DPU chips may be needed if the ones that came with the board aren't powerful enough to keep up with a video stream

4 Design

4.1 Design Context

4.1.1 Broader Context

Describe the broader context in which your design problem is situated. What communities are you designing for? What communities are affected by your design? What societal needs does your project address?

*Our project is designed for anyone interested in monitoring eye movements, behaviors, or potential abnormalities. Our main focus is the medical and flight industry. It will help monitor eye movement and condition to evaluate health of the person non-invasively and quickly for situations that require fast response time.

List relevant considerations related to your project in each of the following areas:

Area	Description	Examples
Public health, safety, and welfare	How does your project affect the general well-being of various stakeholder groups? These groups may be direct users or may be indirectly affected (e.g., solution is implemented in their communities)	Increasing/reducing exposure to pollutants and other harmful substances, increasing/reducing safety risks, increasing/reducing job opportunities
Global, cultural, and social	How well does your project reflect the values, practices, and aims of the cultural groups it affects? Groups may include but are not limited to specific communities, nations, professions, workplaces, and ethnic cultures.	Development or operation of the solution would violate a profession's code of ethics, implementation of the solution would require an undesired change in community practices
Environmental	What environmental impact might your project have? This can include indirect effects, such as deforestation or unsustainable practices related to materials manufacture or procurement.	Increasing/decreasing energy usage from nonrenewable sources, increasing/decreasing usage/production of non-recyclable materials
Economic	What economic impact might your project have? This can include the financial viability of your product within your team or company, cost to consumers, or broader economic effects on communities, markets, nations, and other groups.	Product needs to remain affordable for target users, product creates or diminishes opportunities for economic advancement, high development cost creates risk for organization

4.1.2 User Needs

List each of your user groups. For each user group, list a needs statement in the form of:

User group needs (a way to) do something (i.e., a task to accomplish, a practice to implement, a way to be) because some insight or detail about the user group.

- Pilots need to monitor their eyes for potentially dangerous conditions because they may be in a state where they are unfit to fly the aircraft - then autopilot can take over once the condition is detected.
- Doctors need a way to cheaply and non-invasively detect irregularities in a patient's eyes because rapid detection of eye diseases may improve patient outcomes.

4.1.3 Prior Work/Solutions

Include relevant background/literature review for the project

– If similar products exist in the market, describe what has already been done

– If you are following previous work, cite that and discuss the **advantages/shortcomings**

– Note that while you are not expected to "compete" with other existing products / research groups, you should be able to differentiate your project from what is available. Thus, provide a list of pros and cons of your target solution compared to all other related products/systems.

Detail any similar products or research done on this topic previously. Please cite your sources and include them in your references. All figures must be captioned and referenced in your text.

While we are not seeking to compete with other products on the market, and our project is a bit more specific towards eye movement, there exists embedded systems that make use of Machine Learning. For instance, Radio City Music Hall used one to bar attorneys involved in a legal case against them from their shows(Lawyer Removed from Radio City Music Hall After Facial Recognition Flagged Her As Opposing Counsel - FindLaw.). Wal-mart is using one to catch thieves (Walmart's using AI-enhanced computers and cameras to stop theft | Mashable). China is implementing something similar for its social credit system (China's 'Social Credit' System - Artificial Intelligence +).

What it boils down to is this: getting Machine Learning models working effectively on embedded systems. With Marabou helping us develop better models.

4.1.4 Technical Complexity

Provide evidence that your project is of sufficient technical complexity. Use the following metric or argue for one of your own. Justify your statements (e.g., list the components/subsystems and describe the applicable scientific, mathematical, or engineering principles)

1. The design consists of multiple components/subsystems that each utilize distinct scientific, mathematical, or engineering principles
    a. Analyze the safety of a neural network using mathematical concepts - constrained optimization, linear algebra.
2. The problem scope contains multiple challenging requirements that match or exceed current solutions or industry standards.
    a. Optimizing a Neural Network to run efficiently on an embedded system.
    b. Designing an algorithm to effectively utilize all of the features of an embedded system. This includes processors, real-time processors, deep-learning processing units (DPUs), and buses to memory.
    c. Determining the safety of a neural network.

4.2 Design Exploration

4.2.1 Design Decisions

List key design decisions (at least three) that you have made or will need to make in relation to your proposed solution. These can include, but are not limited to, materials, subsystems, physical components, sensors/chips/devices, physical layout, features, etc.

- We decided to use real-time processors as our controller processor; it will manage the other processors and assign video frames to work while managing the memory as well
- We decided to get additional DPU cores as backup processing power in case the machine learning algorithms were far more incentives than anticipated
- We decided to use state machines to represent our processors, as it makes identifying which processors are busy and available a lot more streamlined.
- Safety criteria will be made; We will eventually have a long list of things that we decide are “unsafe” for our neural network to do in our design context. One example we have is that an output is unsafe; it is outside the bounds of possible true values in our training data.

4.2.2 Ideation

For one design decision, describe how you ideated or identified potential options (e.g., lotus blossom technique). List at least five options that you considered.

The project has been pretty linear, there are decisions that will have to be made that we haven’t gotten to yet. Possible future decisions could be:

- Determine criteria for what is safe in the system.
    - Create a bounding box, do not go outside box
    - Everything inside the bounding box of camera feed is free game.
    - Marabou can (and will) be used to improve model
    - training the model on specific dataset to improve accuracy.
- Determine how many (APU and DPU) processing units we will use.
    - The time it takes to process the images, and classify the movements will determine how we go about this.
    - Times we were working with were assumed, may not be the case.

4.2.3 Decision-Making and Trade-Off

Demonstrate the process you used to identify the pros and cons or trade-offs between each of your ideated options. You may wish you include a weighted decision matrix or other relevant tool. Describe the option you chose and why you chose it.

- There weren't many other options for determining safety criteria. Without the bounding box, the model would behave very unpredictably. We need our model to be as accurate as possible.

- The number of processing units was largely predetermined; part of our challenge was to get the algorithms working on the specified board, which came with its own processing units. There is the opportunity to add one or 2 more, but it's not necessary until we determine we don't have enough processing power.

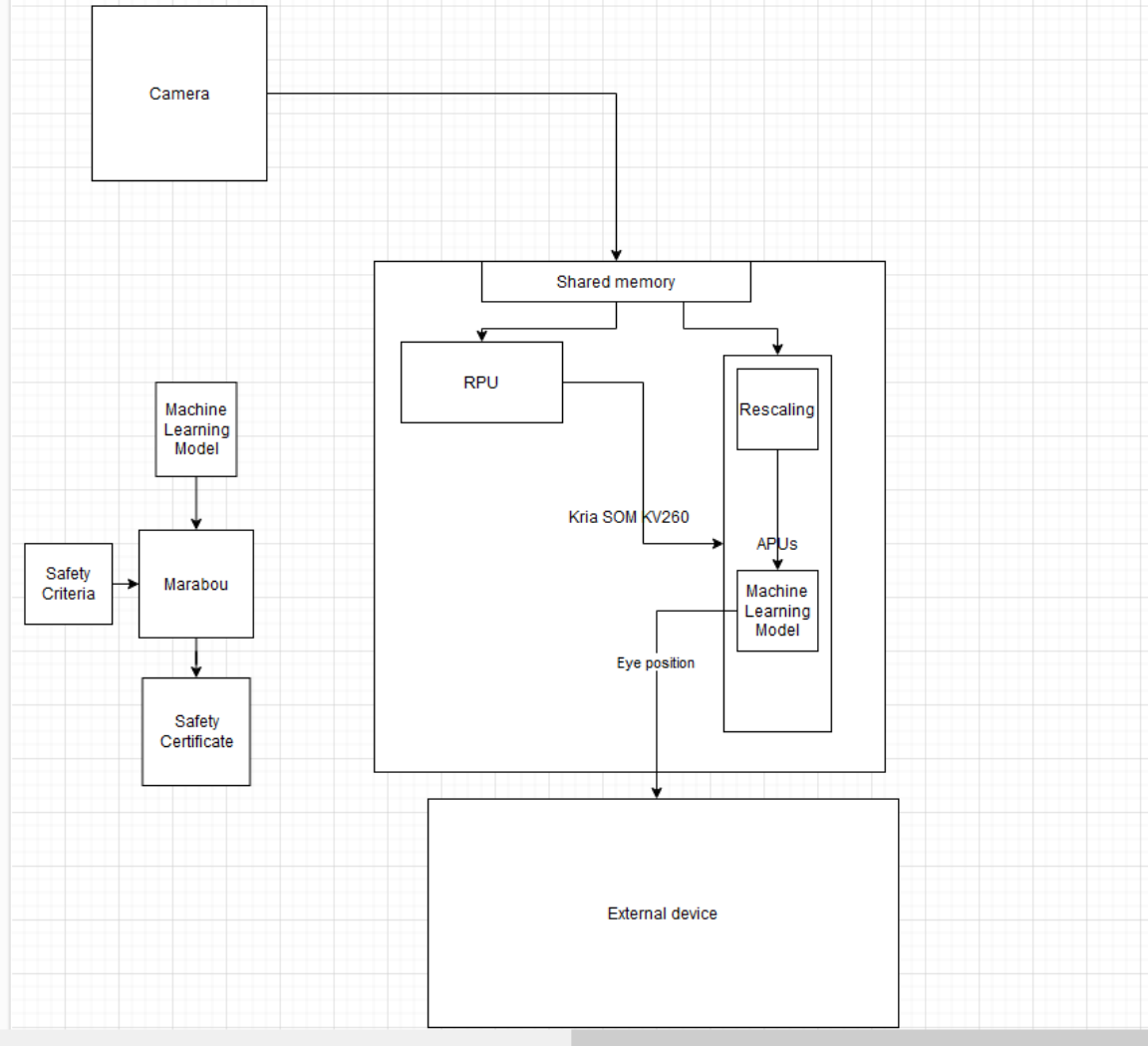
- While there may exist other ways to reach our goals, there are reasons as to why we haven't explored those avenues. This is mostly due to time constraints, with how much time we put into research and getting things to work on our current plan, we may not have time to explore other options.

4.3 Proposed Design

Discuss what you have done so far – what have you tried/implemented/tested?

4.3.1 Design Visual and Description

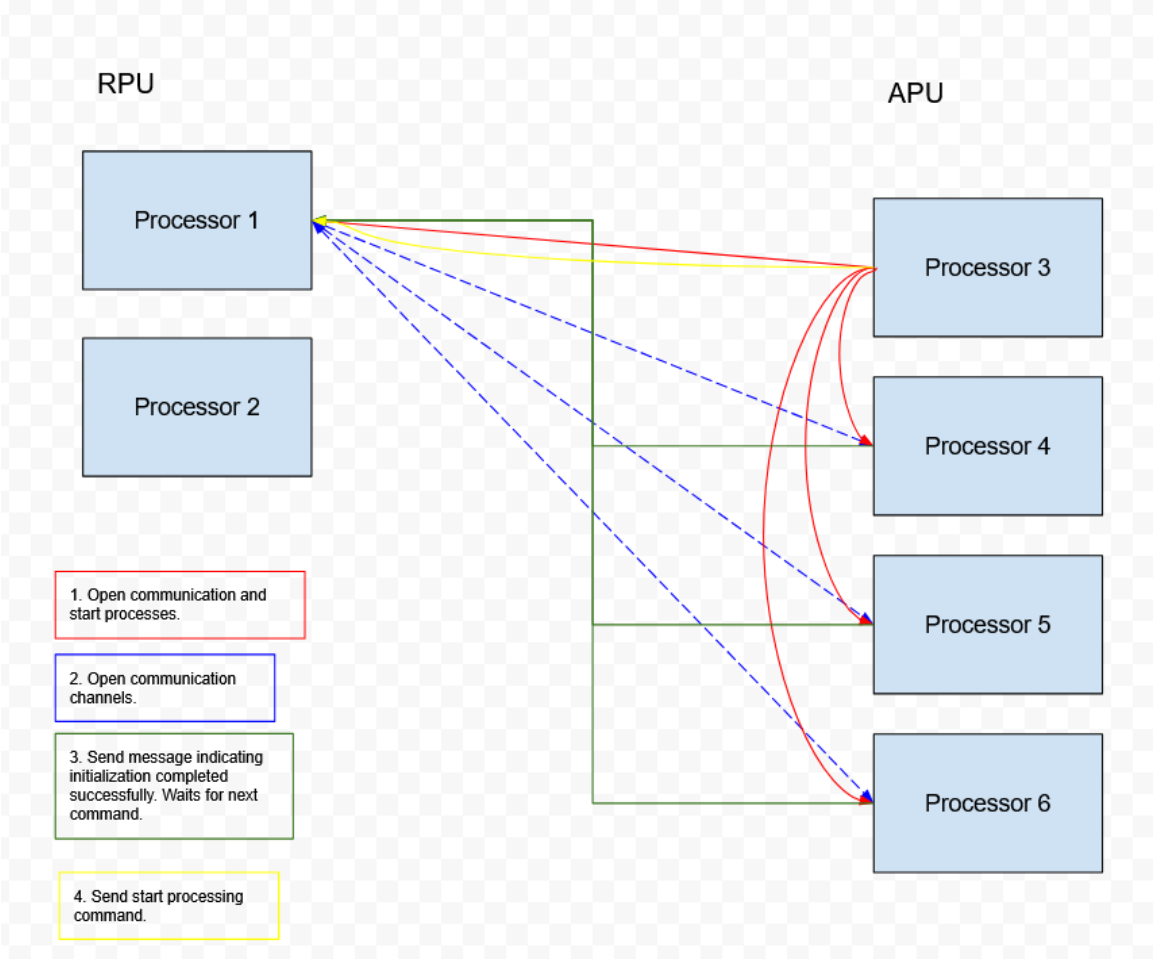
Include a visual depiction of your current design. Different visual types may be relevant to different types of projects. You may include: a block diagram of individual components or subsystems and their interconnections, a circuit diagram, a sketch of physical components and their operation, etc.



Describe your current design, referencing the visual. This design description should be in sufficient detail that another team of engineers can look through it and implement it.

Our system begins with input via bus from an external camera (camera not included in our project) which occupies space in the shared memory of the Kria SOM KV260 board. Each frame of video feed of the camera is sent to our system separately, as soon as they are generated. The RPU on the Kria board makes note of each frame's location in memory, and assigns one of four APUs to rescale and run inference on the image. The communication

between the RPU and APUs is described in the diagram below.



The task that an APU must run for each frame of the video is as follows: rescale the image, then use the rescaled image as the input to a neural network. The output of the neural network is then sent via bus to an external device.

4.3.2 Functionality

Describe how your design is intended to operate in its user and/or real-world context. This description can be supplemented by a visual, such as a timeline, storyboard, or sketch.

How well does the current design satisfy functional and non-functional requirements?

The design is intended to be placed in an aircraft to process a video feed of a pilot's eye, and should be giving its output to another system that is capable of interpreting and acting on the output of our system. One possible scenario: A pilot flying an aircraft has just entered a very stressful situation. They enter a state of anxiety, where they are unable to think clearly about their actions. The state of anxiety causes their eyes to move frantically. They have a camera watching their eyes, sending its video feed to our system. Our system processes the video feed to determine the position of the pupil of the eye and the type of eye movement

that is occurring. This information is given to another system, which determines that the eye movement is not normal, so this system turns off the pilot controls and turns on autopilot.

The current design will help us meet the functional requirement of speed by having an adjustable number of cores to process the image. If our current board is too slow, we can instead use our same design on another board with more processors to meet timing requirements.

The fact that we are using Marabou to verify the safety of our neural network allows us to formally certify that our neural network will have safe output bounds, meaning the broader system will not turn off the pilot controls when nothing is wrong.

4.3.3 Areas of Concern and Development

Based on your current design, what are your primary concerns for delivering a product/system that addresses requirements and meets user and client needs?

What are your immediate plans for developing the solution to address those concerns? What questions do you have for clients, TAs, and faculty advisers?

- Even with multiple cores we are not sure we could meet an acceptable processing time. We won't know until we test everything out.
    - Solution: implement and test everything
- Although we have an idea of what is 'safe,' the model may end up 'mostly safe' if not 'completely safe.'
    - Solution: Marabou will be used to improve performance of the model, but given the context of the situation...Since won't necessarily be used in safety critical system, 'mostly safe' could be acceptable.

4.4 Technology Considerations

Highlight the strengths, weaknesses, and trade-offs made in technology available.

Discuss possible solutions and design alternatives

- Marabou (Neural Network verification tool)
    - has a high RAM requirement in terms of verifying neural networks. What could be considered small for a neural network may be a heavy workload for the tool. We have run into issues with crashing when running the tool with effectively 30GB of RAM.
    - A possible solution may be to quantize the model.
- Kria SOM board
    - Well supported, has a lot of documentation, lower power requirement
    - Six Cores
        - 2 real time processors
        - 4 computational processors
    - Outdated firmware originally on boards, but has been solved

- OpenAMP
    - Allows for inter-processor communication at and during runtime
    - Not sure how the framework views the board and how interaction works entirely

4.5 Design Analysis

– Did your proposed design from 4.3 work? Why or why not?

- We have not reached that stage yet

– What are your observations, thoughts, and ideas to modify or iterate over the design?

- We have not reached that stage yet

4.6 Design Plan

Describe a design plan with respect to use-cases within the context of requirements, modules in your design (dependency/concurrency of modules through a module diagram, interfaces, architectural overview), module constraints tied to requirements.

Our project has two major components, each of which will have independent development:

For the board system:

- set up openAMP IDE for communication between processors
- set up openAMP and resource allocation on the Kria SOM board
- confirm inter-process communications (IPC) success via output file
- keep IPC overhead to minimal runtime for more overhead for Machine Learning algorithm
- Use Vitis to get algorithms onto the processors
- Ensure the board can process each frame such that when the next image is ready to process the previous image's results have been reported.
- add more processors until either board is maxed or the goal of keeping up with 60FPS is achieved.

For the machine learning model:

- Choose structure of neural network
- Reformat training data
- Train model with quantization
- Develop criteria to judge safety of model
- Encode criteria into mathematical constraints to use with the Marabou tool
- Evaluate model using Marabou
- Modify model (more training, different structure, different preprocessing of inputs) repeatedly until model passes Marabou safety tests

5 Testing

Testing is an **extremely** important component of most projects, whether it involves a circuit, a process, power system, or software.

The testing plan should connect the requirements and the design to the adopting test strategy and instruments. In this overarching introduction, given an overview of the testing strategy. Emphasize any unique challenges to testing for your system/design.

5.1 Unit Testing

What units are being tested? How? Tools?

- The neural network is tested by separating our training data into two parts: training data, ,and testing data. After training, the model is evaluated based on its performance on the testing data.
- We will be performing a verification of the model using Marabou, a formal methods tool for neural network verification. Using this tool, we will formally ensure that our neural network has a safe output range. Marabou allows us to perform queries on the network to see whether or not some output can be reached given some constraints on the input, which allows us to develop safety criteria then use Marabou to determine whether or not the network satisfies those safety criteria.
- The runtime of the Machine Learning algorithm will be tested on the cores themselves; We will be testing it with the prototypes of the neural networks. We will be utilizing the Clock function in C to note the start and end time of the algorithms and will compare to our end goal of 40 ms per frame.
- We will test the message passing design by having our RPU processor core send a ordered list of messages, one at a time, to an APU, having it confirm it received our message, modify it (in this case, by adding 100) and sending the modified message back to the RPU, and having it confirm it received it.

5.2 Interface Testing

What are the interfaces in your design? Discuss how the composition of two or more units (interfaces) are being tested. Tools?

- There will be two interfaces in our design. One interface will be a camera in which a user would position in front of their eye. This will be tested with an initialization step that will identify if we are receiving images.
- The second interface will be within the linux kernel. When the application is started the process will allow the user to interact with the application through the terminal. This interface will allow the user to see the current states of processors that are controlled by the application as well as provide the ability to stop the application. Issues with this interface will be logged into an error file that will be created on startup.

5.3 Integration Testing

What are the critical integration paths in your design? Justification for criticality may come from your requirements. How will they be tested? Tools?

- There will be 2 main critical integration paths. The first will be integrating out Inter-Process Communication protocol onto the board. openAMP also allows us to set up our own communication processes with its APIs. We will model the initial communication with a handshake protocol, and using our master core, we will dump any and all outputs into a file for later review. The other critical integration will be getting the algorithms onto the processors. Vitis will assist by allowing us to program the Deep-learning processing units with the model and will confirm success via output file. We will need to compare the results of the model inference on the board with the results generated on our computers.

5.4 System Testing

Describe system level testing strategy. What set of unit tests, interface tests, and integration tests suffice for system level testing? This should be closely tied to the requirements. Tools?

- The system test will involve a video feed fed into memory at 60 frames per second. The outputs will be monitored, which will include the result of the algorithm running which are as follows; if the eye is open, where the center of the eye is, and the movement of the eye. The other output will be a dump file which will include how fast the processors running the machine learning algorithm took and any results/failures of inter-process communications.

5.5 Regression Testing

How are you ensuring that any new additions do not break the old functionality? What implemented critical features do you need to ensure they do not break? Is it driven by requirements? Tools?

- To ensure additions and removal of functionality does not break the overall functionality of the system we will be segmenting the system. Some tools that will allow us to do this are the use of OpenAMP. Utilizing OpenAMP allows us to start a process and give it an algorithm similar to multi-threading; however, we will have the ability to communicate between processors which will allow us to send them jobs to complete and have them send the results to another processor without returning out of the function.

5.6 Acceptance Testing

How will you demonstrate that the design requirements, both functional and non-functional are being met? How would you involve your client in the acceptance testing?

- To demonstrate the functional requirements of the project, it is required that the Machine Learning algorithm functions within what is considered 'safe' (it functions as it was intended to with little to no errors), and that it can run on the Kria SOM KV260 boards. It is also required that the algorithms run at average under 40 ms per frame. A non-functional requirement would, perhaps, be placing the boards in a shell to protect it. The client is involved by offering his insights on how we may want to go about testing the functionality, and when he is informed on progress and any stop gap.

5.7 Security Testing (if applicable)

- This is not applicable to the project because it is concerned with finding the center of a pupil. If, say, identifying information of the people (name, age, race, etc.) involved were stored somewhere, there would be a concern for security.

5.8 Results

What are the results of your testing? How do they ensure compliance with the requirements? Include figures and tables to explain your testing process better. A summary narrative concluding that your design is as intended is useful.

- We are not at this stage yet.

6 Implementation

Describe any (preliminary) implementation plan for the next semester for your proposed design in 3.3. If your project has inseparable activities between design and implementation, you can list them either in the Design section or this section.

- Optimize the neural network and utilize verification tools to prove the model is still effective. We will repeatedly make new iterations of the neural network to ensure it fits our safety criteria.
- Utilize OpenAMP to design an algorithm to take in real time events (video frames) and facilitate computational cores to run the neural network on the events

7 Professionalism

This discussion is with respect to the paper titled "Contextualizing Professionalism in Capstone Projects Using the IDEALS Professional Responsibility Assessment", *International Journal of Engineering Education* Vol. 28, No. 2, pp. 416–424, 2012

7.1 Areas of Responsibility

Pick one of IEEE, ACM, or SE code of ethics. Add a column to Table 1 from the paper corresponding to the society-specific code of ethics selected above. State how it addresses each of the areas of seven professional responsibilities in the table. Briefly describe each entry added to the table in your own words. How does the IEEE, ACM, or SE code of ethics differ from the NSPE version for each area?

Area of Responsibility	**IEEE Code of Ethics (verbatim)**	**Description (our own words)**	**How does it differ from NSPE?**
Work Competence	Maintain and improve our technical competence and to undertake technological tasks for others only if qualified by training or experience, or after full disclosure of pertinent limitations;	The IEEE Code of ethics tells us to be competent in our work by not performing work we are not qualified for, or at least our employer knows we are not qualified.	The IEEE Code would allow us to perform work we aren't qualified for - as long as everyone is aware we aren't qualified - and of course without breaking other parts of the code. The NSPE simply says not to

			perform such work
Financial Responsibility	Avoid real or perceived conflicts of interest whenever possible, and to disclose them to affected parties when they do exist; Reject bribery in all its forms	IEEE Code tells us to be responsible with our employer's financial choice to hire us by avoiding conflicts of interest, so we don't act against our employer's best interest.	The NSPE code highlights this as being more honest with your coworkers and clients, whereas IEEE is more focused on 'things' of value or rejecting bribes.
Communication Honesty	Be honest and realistic in stating claims or estimates based on available data; to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others;	Be truthful to clients and employers in regards to engineering issues, such as safety or feasibility. Do not hide information that would otherwise change a person's stance on an issue.	The IEEE code of ethics focuses on the relationship between the client and the employer. The NSPE code focuses on both interoffice communication and client-employer relationship. The standards that engineers should uphold are the same.
Health, Safety, Well-Being	Avoid injuring others, their property, reputation, or employment by false or malicious action;	Avoid putting yourself and others in precarious situations.	NSPE highlights situations as the focus for safety, IEEE focuses on the self and avoiding taking the

			direct action.
Property Ownership	Avoid injuring others, their property, reputation, or employment by false or malicious action;	Do not hurt others or other people's property, especially with intent.	IEEE highlights a focus on reputation of others; NSPE focuses on the client and more one-on-one interaction
Sustainability	to accept responsibility in making decisions consistent with the safety, health, and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;	The IEEE code tells us to be aware of and disclose any possible harm to the environment, and to be responsible in decisions that might harm it.	IEEE Code of ethics shows concern with more of a public focus, NSPE focuses on sustainable development rather than safety and welfare.
Social Responsibility	to accept responsibility in making decisions consistent with the safety, health, and welfare of the public, and to disclose promptly factors that might endanger the public or the environment;	Engineers must only design products that will benefit society and humankind.	NSPE has a global or macro view on things. It focuses on producing products that benefit society, while the IEEE is concerned with safety after production.

7.2 Project Specific Professional Responsibility Areas

For each of the professional responsibility area in Table 1, discuss whether it applies in your project's professional context. Why yes or why not? How well is your team performing (High, Medium, Low, N/A) in each of the seven areas of professional responsibility, again in the context of your project. Justify.

Area of Responsibility	**Application to our project**	**Team Performance (Justify)**
Work Competence	Yes. As it is possible that our work may be used in critical applications, it is important that we are competent enough to understand how it works and what its limitations are.	High. While we are not fully competent (we are still learning), we are fully transparent about this fact so there are no questions about our competence.
Financial Responsibility	Yes, there are expensive boards given to us by our client that we are working on and there is always potential for damage or loss.	High. We are keeping costs to a minimum, and are being respectful to the boards our client sent us; we are using EMI equipment to reduce any potential damage. The boards are stored in the TLA with access only being available with a password only the group knows.
Communication Honesty	This is very applicable to our project. Safety issues may present themselves as the project becomes more mature. These issues will need to be reported so that other groups who utilize our design will understand the risks even if those risks are mitigated.	Moderate. We are currently creating weekly documents to outline struggles and pending issues. As our project continues to mature we will need to continue this practice and document potential risks that are discovered.
Health, Safety, Well-Being	Yes. The development of computer vision techniques and applications (what we are doing) has many	Medium. In meetings, we do take this into account when discussing the

	potential health consequences ranging from better medical treatment to more deadly warfare.	software and its potential applications.
Property Ownership	Not really. Our work is primarily software, and has no chance of damaging anyone's property.	(High)Our project is a proof of concept to run two ML algorithms on a six-core processor, no physical interaction with the world.
Sustainability	No real application to our project. Tangentially, the production of the kria boards may have some environmental impact, but we have no control over that.	(Low)In terms of environmental sustainability, there is not much to consider with our project.
Social Responsibility	Yes. Similarly to the Health, Safety, Well-Being section, the development of computer vision techniques and applications (what we are doing) has many potential societal consequences ranging from better medical treatment to more deadly warfare.	Medium. In our meetings, we discuss potential applications of our work so we are aware of the potential societal consequences.

7.3 Most Applicable Professional Responsibility Area

Work competence is important for our project because of the potential concerns of safety that might arise if we do not have sufficient understanding of our project. We are working with neural networks, which are notoriously difficult to interpret and can produce unpredictable results. Because our work has the potential to be employed in critical applications, such as the medical field, it is important that we understand the limitations and potential of our work such that it does not underperform in critical applications. During the development of our project, we demonstrate work competence by keeping each other informed of the details of other team members' work through weekly presentations on what team members did or learned. This way, each team member has a complete understanding of the project so they can assess what it is and isn't capable of.
A specific example of this is how we presented on the process of training a neural network that was done by a previous senior design group before us. We took the time to understand this piece of code and improve our competence so we have a full understanding of how the technology works so we can properly assess its limitations.

8 Closing Material

8.1 Discussion

Discuss the main results of your project – for a product, discuss if the requirements are met, for experiments oriented project – what are the results of the experiment, if you were validating a hypothesis – did it work?

- We have not completed this part yet. This will be completed late in the Fall semester of 2023.

8.2 Conclusion

Summarize the work you have done so far. Briefly reiterate your goals. Then, reiterate the best plan of action (or solution) to achieving your goals. What constrained you from achieving these goals (if something did)? What could be done differently in a future design/implementation iteration to achieve these goals?

On the hardware we have had success installing and interacting with the PetaLinux operating system. In order to run this operating system we needed to update the firmware. We have begun writing code to utilize functionality provided by the OpenAMP framework. This code will be used to run the neural network in a system similar to multi-threading. This system will need to finish fast enough so that when a new frame is provided a processor is able to begin running the neural network against the frame.

On the Machine Learning side, we have structured, trained, and saved a model such that we are able to analyze it with Marabou. We have outlines of the safety criteria we will develop to test the model with. We hope to repeatedly make new iterations of our model (changing structure, using more training data, or using different preprocessing of inputs) until our model passes our safety criteria according to Marabou.

8.3 References

List technical references and related work / market survey references. Do professional citation style (ex. IEEE).

[1]

J. Grese, C. S. Păsăreanu, and Erfan Pakdamanian, "Formal Analysis of a Neural Network Predictor inShared-Control Autonomous Driving," *AIAA Scitech 2021 Forum*, Jan. 2021, doi: https://doi.org/10.2514/6.2021-1580.

[2]

"Kria KV260 and PetaLinux 2022.1: Part 01-Getting Started," *Hackster.io*. https://www.hackster.io/mohammad-hosseinabady2/kria-kv260-and-petalinux-2022-1-part-01-getting-started-ed5a25 (accessed Apr. 24, 2023).

[3]

G. Katz *et al.*, “Con sis te n t * Comple te * W ell Docu m e n te d * Easy to R e us e * The Marabou Framework for Verification and Analysis of Deep Neural Networks.” Accessed: Apr. 24, 2023. [Online]. Available: https://aisafety.stanford.edu/marabou/MarabouCAV2019.pdf

8.4 Appendices

Any additional information that would be helpful to the evaluation of your design document.

If you have any large graphs, tables, or similar data that does not directly pertain to the problem but helps support it, include it here. This would also be a good area to include hardware/software manuals used. May include CAD files, circuit schematics, layout etc,. PCB testing issues etc., Software bugs etc.

8.4.1 Team Contract

Team Members:

1) _____Sandro Panchame_____ ____ 2) ____Alek Comstock_____

3) _____Jeffery Kasper_____ 4) ___Rudolph Nahra_____

Team Procedures

1. Day, time, and location (face-to-face or virtual) for regular team meetings:

Sundays, 5pm for meeting with the client. We will meet before or after class if something needs to be done in person

2. Preferred method of communication updates, reminders, issues, and scheduling (e.g., e-mail, phone, app, face-to-face):

Discord, Zoom, Telegram, and face-to-face

3. Decision-making policy (e.g., consensus, majority vote):

Consensus; everyone must agree.

4. Procedures for record keeping (i.e., who will keep meeting minutes, how will minutes be shared/archived):

Individual records will be recorded by individuals, team meeting minutes will be kept by the current team leader, a role which will rotate each week (described below)

Participation Expectations

1. Expected individual attendance, punctuality, and participation at all team meetings:

Everyone should attend pre-determined meetings except for emergencies or accepted absences. If unexcused absence, they will need to explain why they are absent and will be responsible for catching themselves up.

2. Expected level of responsibility for fulfilling team assignments, timelines, and deadlines:

Goals should be completed at the end of each week, except when the user has shown effort towards the goal without completion. If member has repeated weeks without reaching goal and showing effort, the other members should confront them and ask what they need to help complete their goals

3. Expected level of communication with other team members:

Team members should be reachable and provide responses in a timely manner, within a 48 hour window.

4. Expected level of commitment to team decisions and tasks:

Members should voice concerns and opinions. No such thing as a dumb question.

Leadership

1. Leadership roles for each team member (e.g., team organization, client interaction,

individual component design, testing, etc.):

Sandro: Creation of documents
Rudolph: Communications with client
Jeffery: Communications with Advisor
Leader: Responsible for scheduling and setting up virtual meetings, and for keeping track of other team member's progress on top of their own. Each week, a new team leader is decided in such a way each person will have an equal number of Team Leader weeks.

2. Strategies for supporting and guiding the work of all team members:

Setting time aside to meet with team members to resolve issues and solve problems.
Meet with the client for "head-banging" issues.

3. Strategies for recognizing the contributions of all team members:

Discuss the project and talk about contributions. Each week, each of us discusses our contributions and difficulties in a meeting with the whole team and the client. Go out to eat sometimes.

Collaboration and Inclusion

1. Describe the skills, expertise, and unique perspectives each team member brings to the

team.

Alek: Embedded systems Knowledge, OS knowledge as a TA for Intro to

Operating Systems. Comfortable in C and Java, likes to break down tasks into bite-size goals
Sandro: Knowledge of python, getting things done in advance (panicky otherwise), good file/document organization
Rudolph: Knowledge of hardware and computer organization, linux, python, and experience with machine learning as a research assistant.
Jeffery Kasper: Embedded systems knowledge. Expertise in Linux-based operating systems and kernels. Proficient in C language programming. Comfortable scripting in Bash.

2. Strategies for encouraging and support contributions and ideas from all team members:

We encourage all questions and ideas to be asked and shared, no matter how small or simple you may feel it might be. We will discuss and celebrate contributions during weekly meetings.

3. Procedures for identifying and resolving collaboration or inclusion issues (e.g., how will

a team member inform the team that the team environment is obstructing their

opportunity or ability to contribute?)

If a member has a problem, it should be shared, ideally in one of our communication channels, so it can be evaluated and come up with alternatives.

Goal-Setting, Planning, and Execution

1. Team goals for this semester:

Get a machine learning algorithm working on a Kria SOM KV260 board. Manage memory and data transfer efficiently and effectively.

2. Strategies for planning and assigning individual and team work:

Allow everyone to discuss what tasks they would like to be responsible for and assign them based on preferences.

3. Strategies for keeping on task:

Gentle reminders. Possible posts on Discord. We will refocus and set goals every week during our meetings.

Consequences for Not Adhering to Team Contract

1. How will you handle infractions of any of the obligations of this team contract?

Speak with the team member, either one-on-one or with the entire team.

2. What will your team do if the infractions continue?

Ask for advice from TA and/or advisor for how to solve the issue. If nothing works, possibly contact the professor.

****

a) I participated in formulating the standards, roles, and procedures as stated in this contract.

b) I understand that I am obligated to abide by these terms and conditions.

c) I understand that if I do not abide by these terms and conditions, I will suffer the consequences as stated in this contract.

1) _____Sandro Panchame_____ DATE ____4/23/23_____

2) _____Alek Comstock_____ DATE _____4/23/23_____

3) _____Jeffery Kasper_____ DATE _____4/23/23_____

4) _____Rudolph Nahra _____ DATE ___4/23/23____