HALO: High Autonomous Low-SWaP Operations

Team Members:

- Sloan Hatter shatter2022@my.fit.edu
- Blake Gisclair bgisclair2022@my.fit.edu

Faculty Advisor: Dr. Ryan T White rwhite@fit.edu

Client: Dr. Ryan T White rwhite@fit.edu

Progress Matrix of Milestone 1

Task	Completion %	Sloan	To Do
Literature Review	100%	Research	None
Load ViT onto Raspberry Pi	95%	Interface with hardware	Write custom scripts
Requirement Document	100%	Write 100%	None
Design Document	100%	Write 100%	None
Test Plan	100%	Write 100%	None

Discussion of Accomplished Tasks for Milestone 1:

- **Literature Review**
  - The goal of the literature review was to gain more knowledge and understanding of the project's direction.
- **Load ViT onto Raspberry Pi**
  - The goal of loading a vision transformer (ViT) onto a Raspberry Pi was to be able to procure baseline metrics to use as a starting point and for future testing. A neural network was able to be loaded onto a Raspberry Pi AI HAT+. Unable to visualize model inference due to the necessity for custom post-processing scripts.
- **Requirement Document**
  - The goal of the Requirement Document is to outline the requirements of the project and how those requirements are going to be achieved. The document outlines external requirements, including system, hardware, and software interfaces, functional requirements, including real-time object detection, low SWaP deployability, and binary quantization, and non-functional requirements, including performance requirements.
- **Design Document**
  - The goal of the Design Document is to outline the design of the system in a way that shows how the system will be able to satisfy the required features and behaviors as outlined in the Requirement Document. The document outlines the project's purpose, scope, system design and architecture, and algorithmic design, along with an entity relationship diagram and a data description.
- **Test Plan**

    - The goal of the Test Plan is to outline verification features of the system that show that the system is accomplishing the required features and behaviors as outlined in the Requirement Document and Design Document. The document outlines testing strategies, including test levels and model testing, test cases, proper test environments, potential risks and mitigation strategies, and success criteria.

Discussion of Contribution to Milestone 1:

- **Sloan Hatter:** Tasks contributed to this milestone include development of the format of the Requirement Document, Design Document, and Test Plan, as well as all content therein. Tasks also included performing a literature review on relevant papers pertaining to the subject of the project, with topics including vision transformers, neural network weight representations, and vision transformers being deployed on smaller computer hardware, such as a Raspberry Pi. A neural network model was also loaded onto a Raspberry Pi AI Hat+.

Task Matrix for Milestone 2:

Task	**Sloan**
Literature Review for 4-bit Representation	100%
4-bit Representation	100%
Recoup Accuracy Losses	100%

Discussion of Planned Tasks for Milestone 2:

- Literature Review for 4-bit Representation
    - A second literature review will be carried out on methods used to achieve a 4-bit representation for a neural network. There are substantial methods, open-source packages, and toolkits available that support 4-bit quantization or a quantization of less than 8. The goal is to research available methods and packages and choose the best one for HALO.
- 4-bit Representation
    - Once the optimal method and package have been chosen, they will be applied to the neural network in order to gain a 4-bit representation. Though there are packages available to shrink bitwise representations, achieving a 4-bit representation can be slightly more involved, as the packages available might not get HALO down to 4-bit on their own due to hardware restrictions. In this case, the hardware will need to be switched from the Raspberry Pi to one of the available Jetsons.
- Recoup Accuracy Losses
    - In order to recoup losses in accuracy, two potential methods can be used: FP32 training or Post-Training Quantization (PTQ). FP32 training uses a 32-bit precision for neural network parameters and computations, ensuring numerical stability while providing the highest baseline accuracy. PTQ, on the other hand, allows a model to be quantized after full-precision training, often with additional calibration or fine-tuning to minimize degradation. The decision on which method to use will be made once a 4-bit representation has been achieved.

Date of Meetings:

- 09/05/25
- 09/12/25
- 09/26/25

Client Feedback on Milestone 1:

See Faculty Advisor Feedback below.

Faculty Advisor Feedback on Milestone 1:

- Literature Review: Student reviewed recent literature on vision transformers, including the original paper and SwinV1-2, 1.58-bit transformers, and quantizing neural networks, in addition to computer vision work in astronautics for unknown spacecraft component detection and its use in seeding guidance and navigation algorithms in current use.
- Load ViT onto Raspberry Pi: Student successfully demonstrated a ViT running on Pi + AI-HAT module for object detection. It's a good step. Student used pretrained weights in her demo, so next she needs to learn to load her own trained weights into the model and to import models she builds with various architectures.
- Requirement Document: Completed
- Design Document: Completed
- Test Plan: Completed

Faculty Advisor Signature: _____ Date: 29 Sept 2025