A Framework for Certification of Continuous Learning Systems

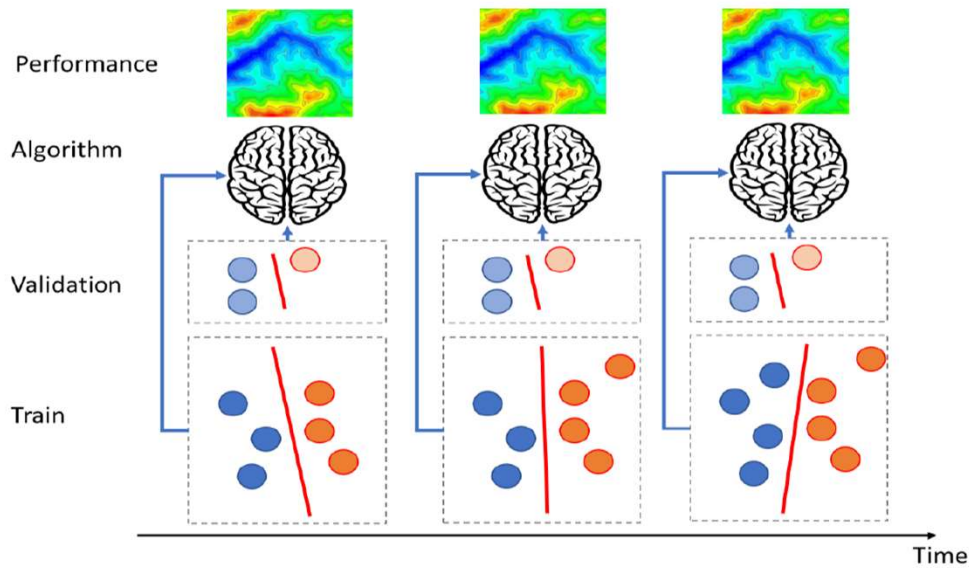
21 September 2022

Presented to: AI4SE & SE4A Workshop

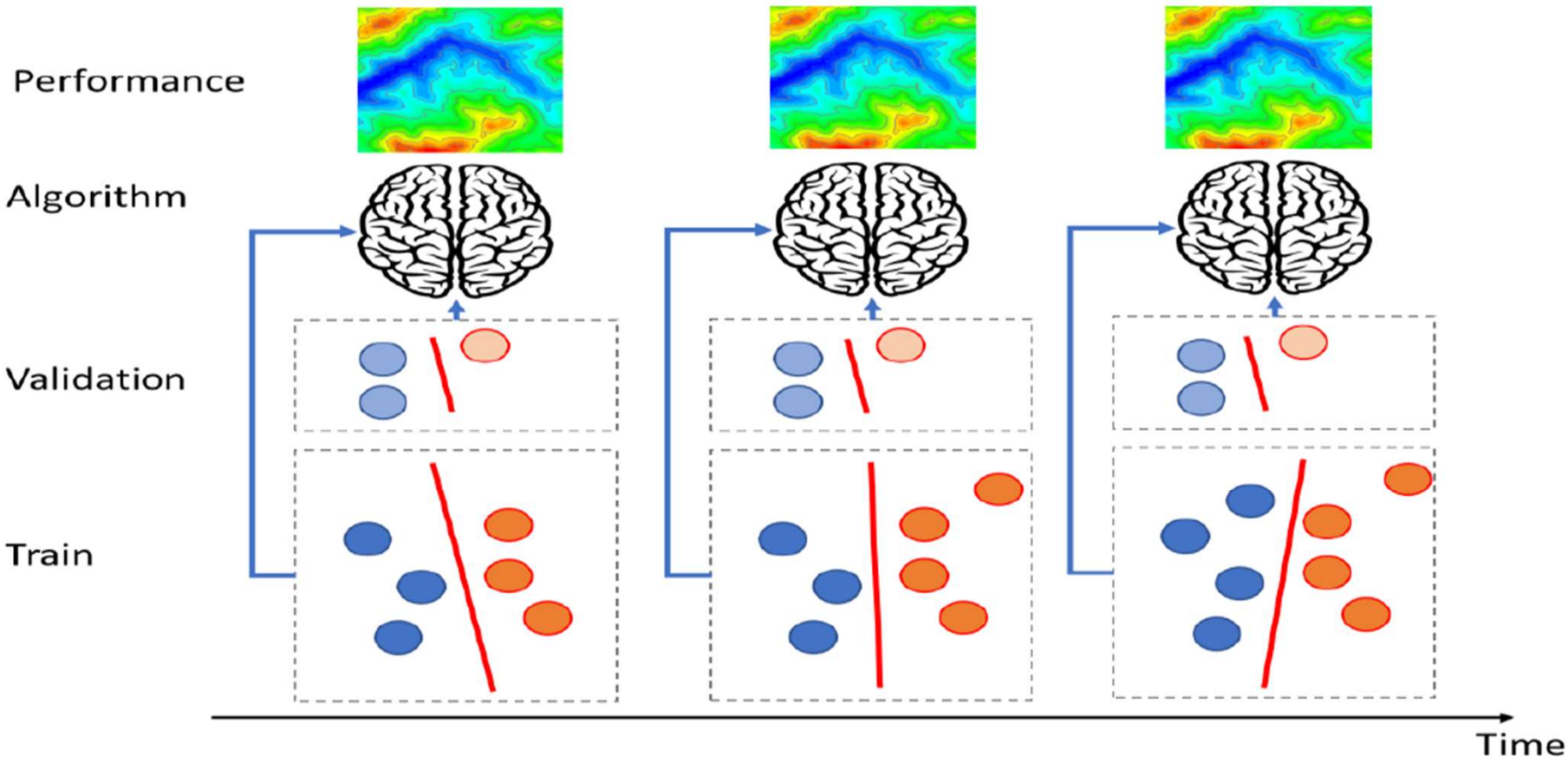
Presented by: Ryan O'Shea

Motivation

- Robotics and Intelligent Systems Engineering Lab (RISE) leverages machine learning to solve difficult problems
- Continual learning (CL) allows systems to adapt to real world conditions, but removes elements of control over the system's actions



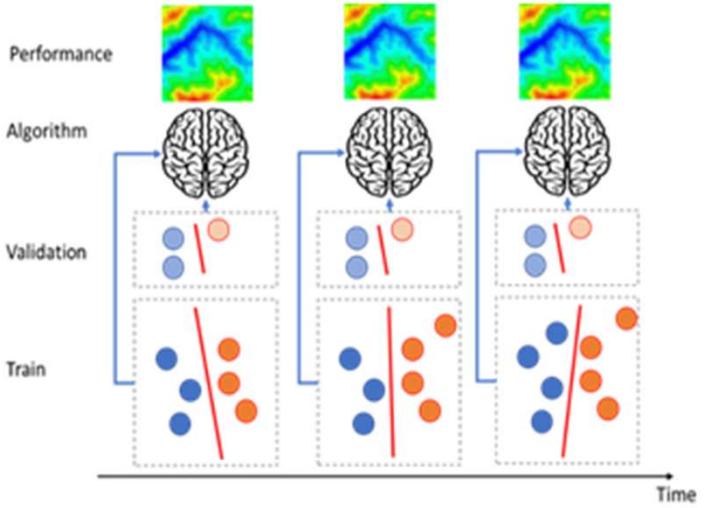
Background



Objective & Need

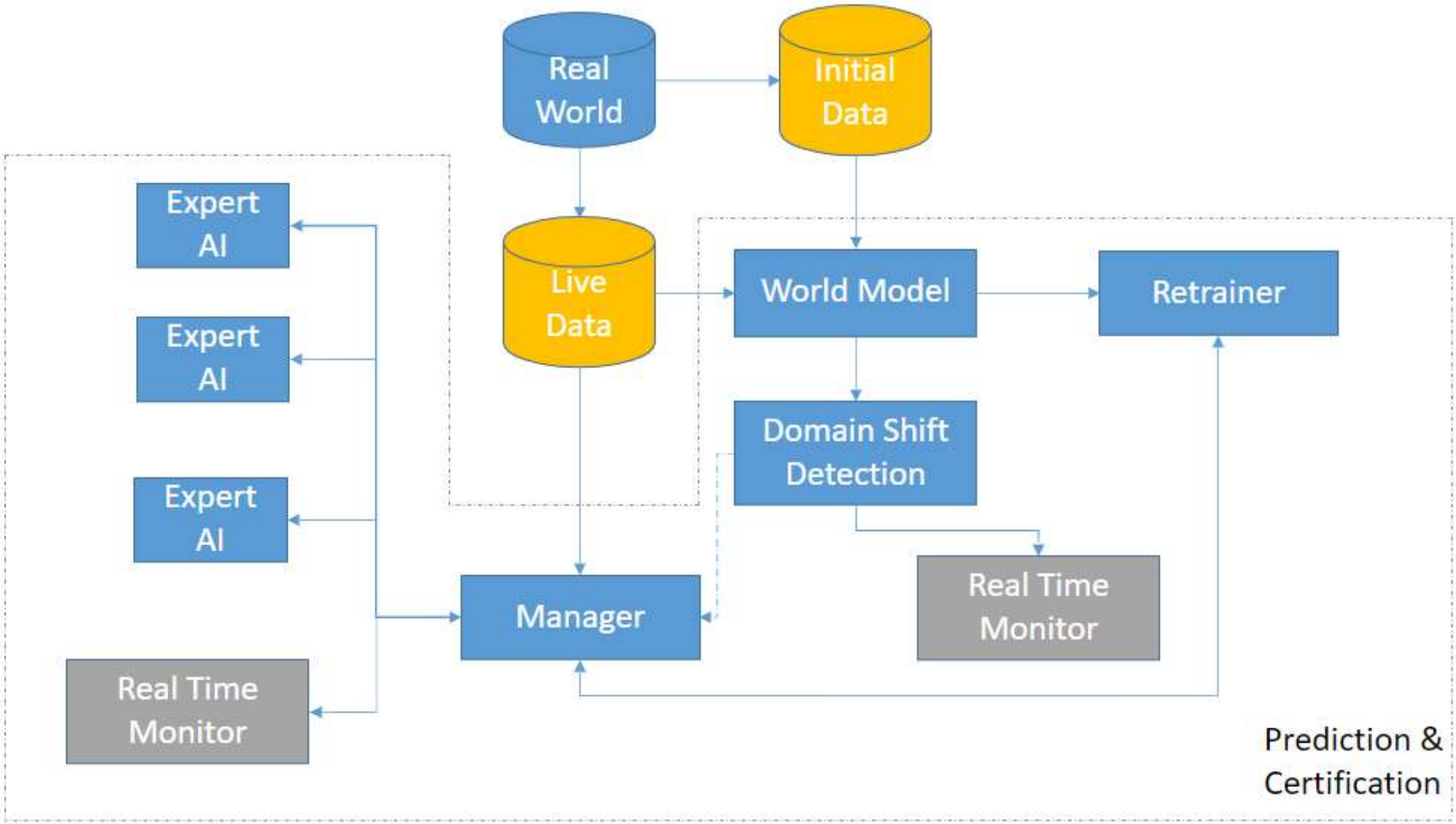
Objective: Produce a prototype capable of evaluating and monitoring a continuously learning system to enable acquisition of lifelong learning artificial intelligence.

- Develop new methods to certify, monitor, and control continual learning systems focusing on current R&D products with possibility to transition
- Enable the safe transition of continual learning systems through advancing T&E techniques
- Create a set of validating experiments to support prototype evaluation process
- Produce a prototype capable of evaluating and monitoring a continuously learning system to enable acquisition of lifelong learning artificial intelligence



A continual learning system over three points in time showing underlying distribution and decision boundaries changing

Framework Overview

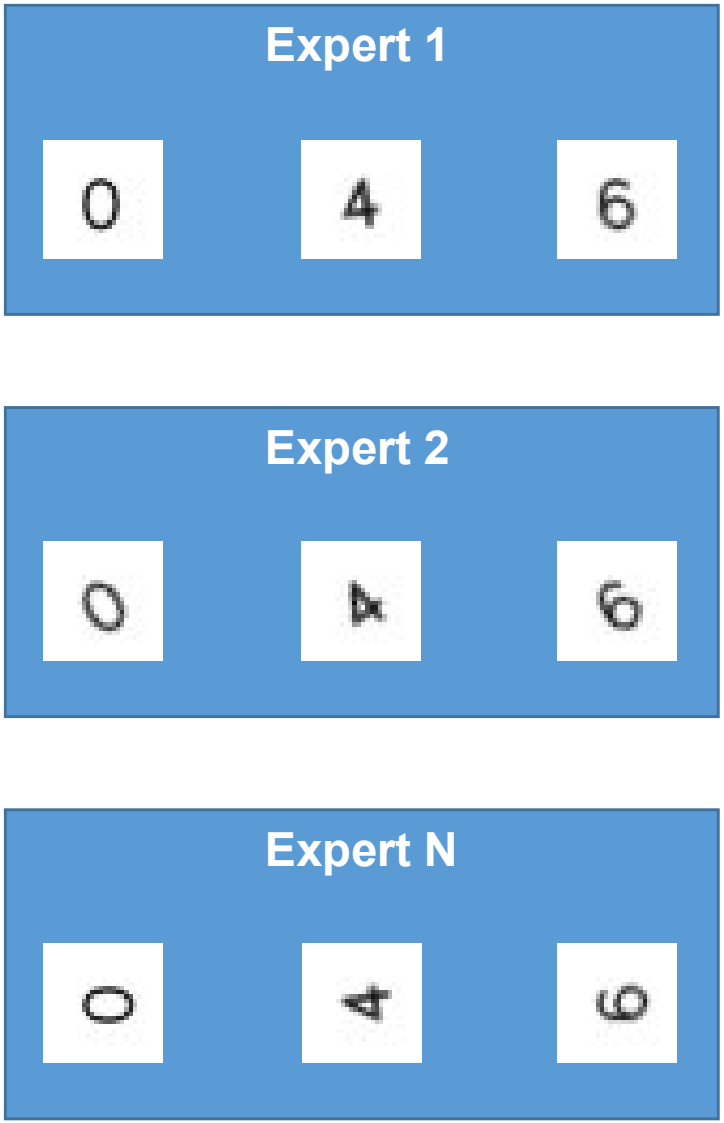


The Dataset

- Synthetic rotated numbers for digits 0-9
  - 28x28 grayscale images
- Used rotation of the image to simulate a change in the domain of the data
- Generated for a specific range of angles
  - i.e., 0-20, 30-90, 80-120
- All classes generated with equal representation but angle was randomly selected from the range

Expert AI

- Simple convolutional neural network (CNN) classifier
- Trained to be an expert for a specific domain (angle range)
- Different experts for different angle ranges to maintain performance
  - i.e, expert 1 trained to classify numbers rotated between 0 and 40 degrees while expert 2 was trained to classify numbers rotated between 40 and 70 degrees



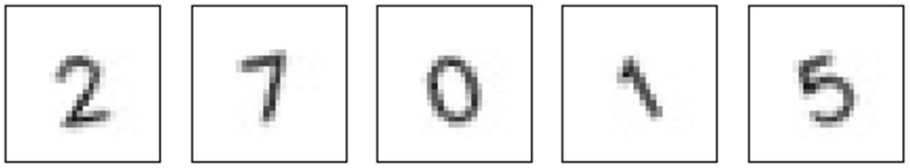
Expert AI Metrics

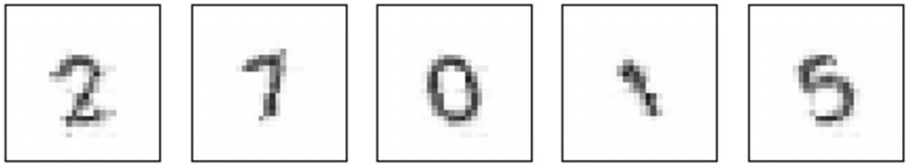
- Confidence
  - Top two SoftMax difference
  - Entropy
  - Variation ratio
- Competence
  - Training data K Nearest Neighbor
    - Using Earth Mover's Distance as distance metric
    - 0 or 1 depending on whether a neighbor can be found withing a threshold distance
  - Domain shift monitor domain competence
    - How similar is incoming data to what the expert was trained on

Domain Shift Monitor (DSM)

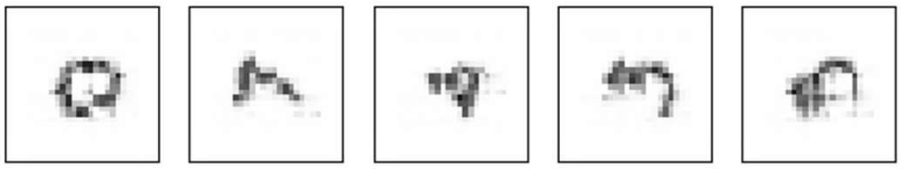
- Each expert AI has an associated DSM trained on the exact same data
- Autoencoder trained to reconstruct input data
- High reconstruction loss corresponds to an image being “out of domain”
- Provides a measure of domain competence for a given expert AI

In Domain (0-30 Degree Rotation)

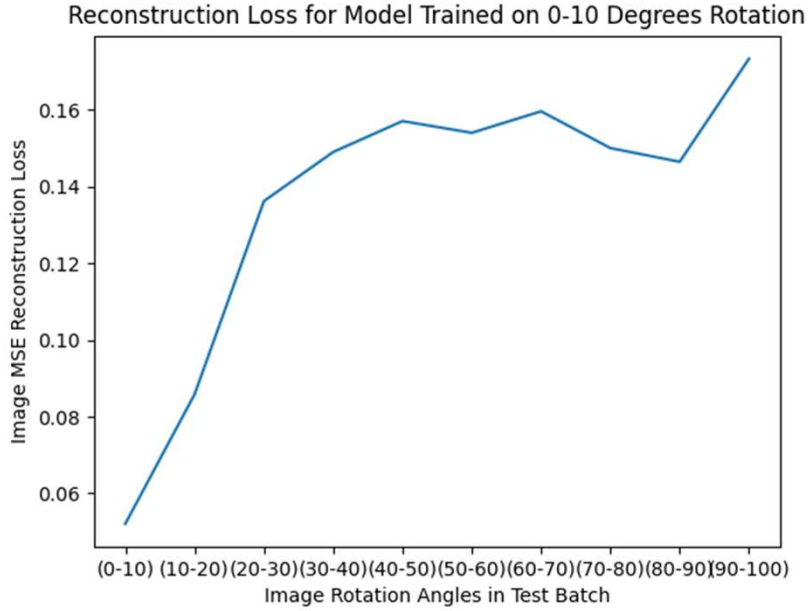


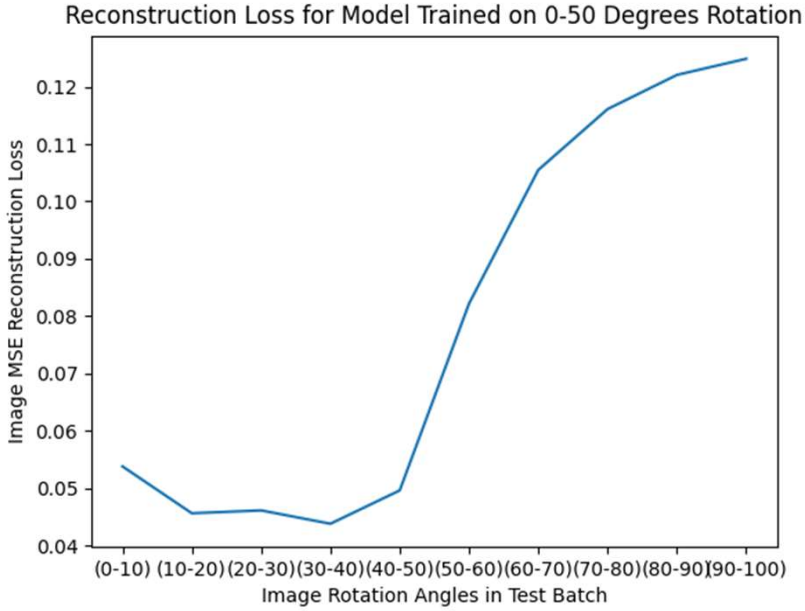


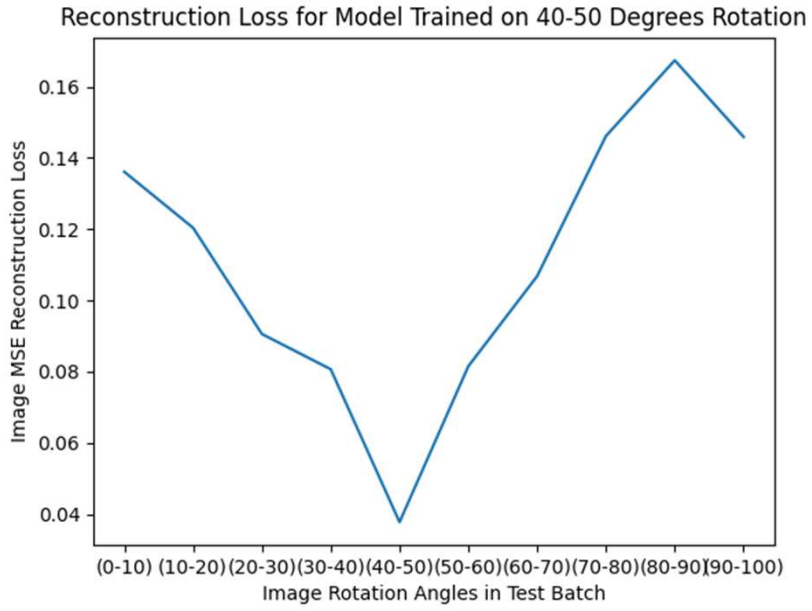
Out of Domain (80-100 Degree Rotation)

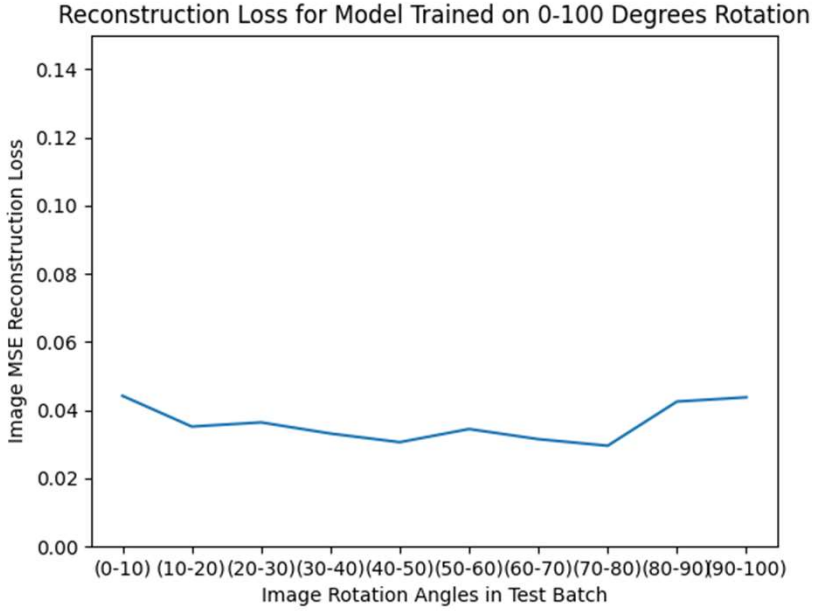


Domain Shift Monitor (DSM)



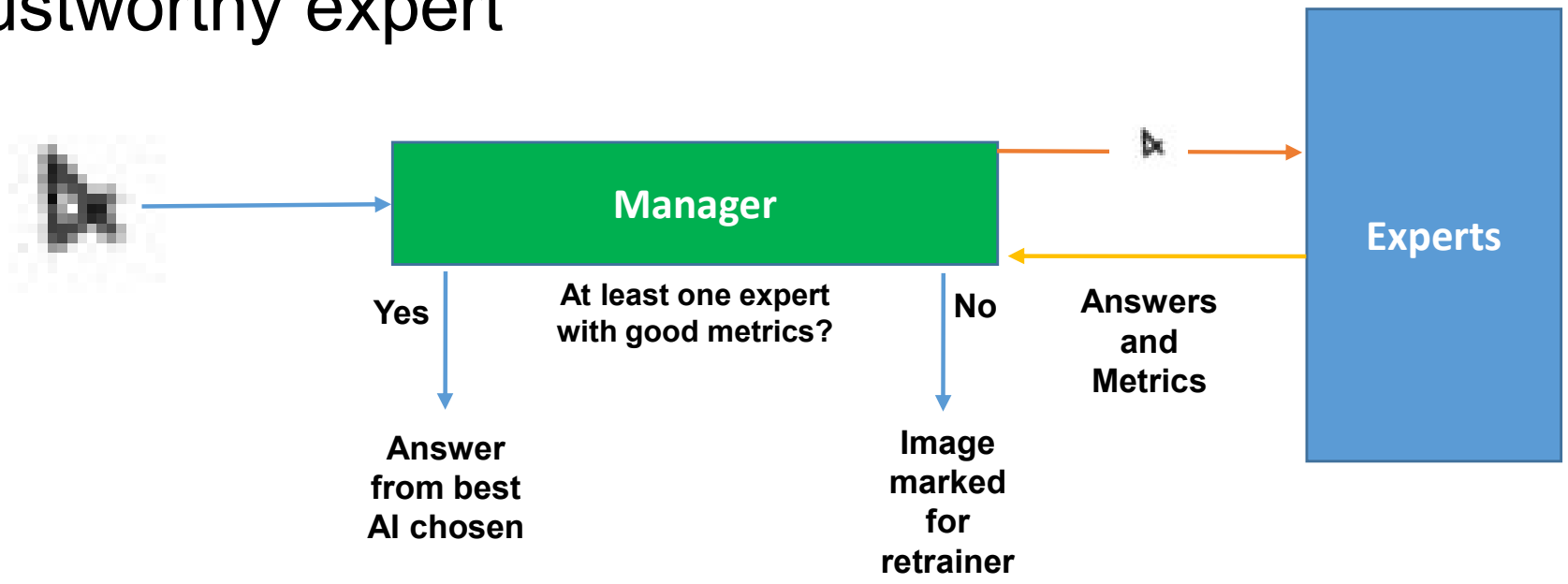






Manager AI

- Passes incoming data to each expert AI and selects the best output
  - Expert AI returns an answer and a vector of confidences and competences
- Selects best answer provided by the experts
- If returned confidence and competence metrics are all poor, then the data is stored later for retraining
- Feed forward neural network trained to choose the most trustworthy expert

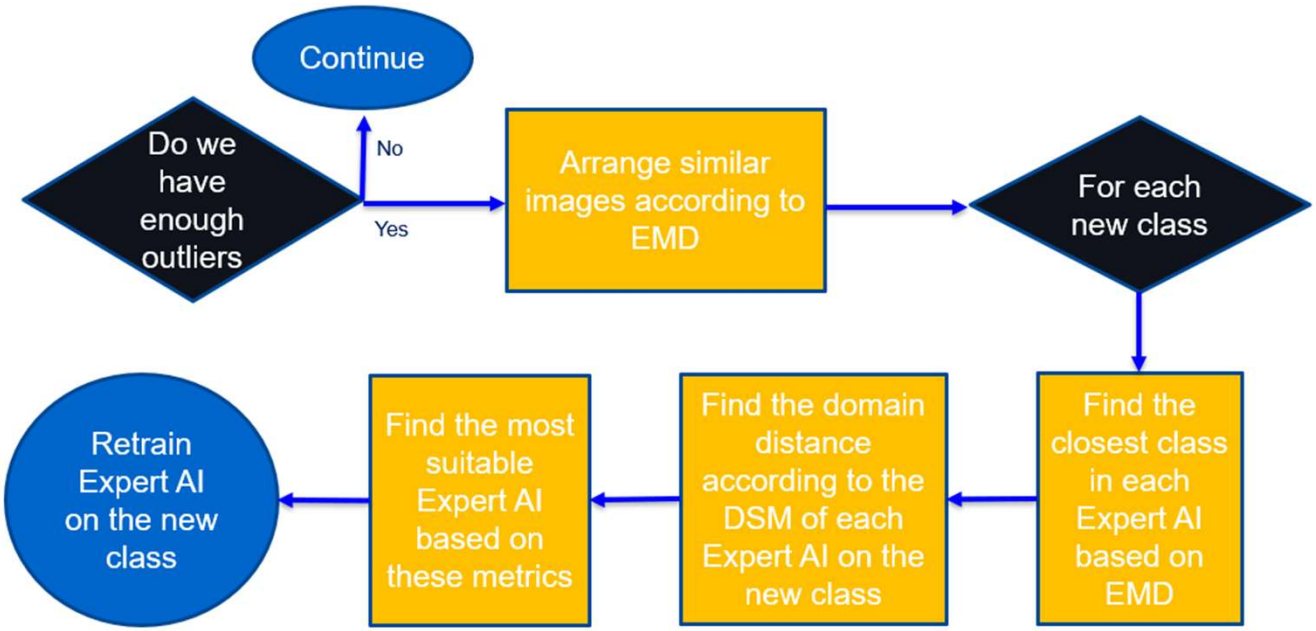


Real Time Monitor

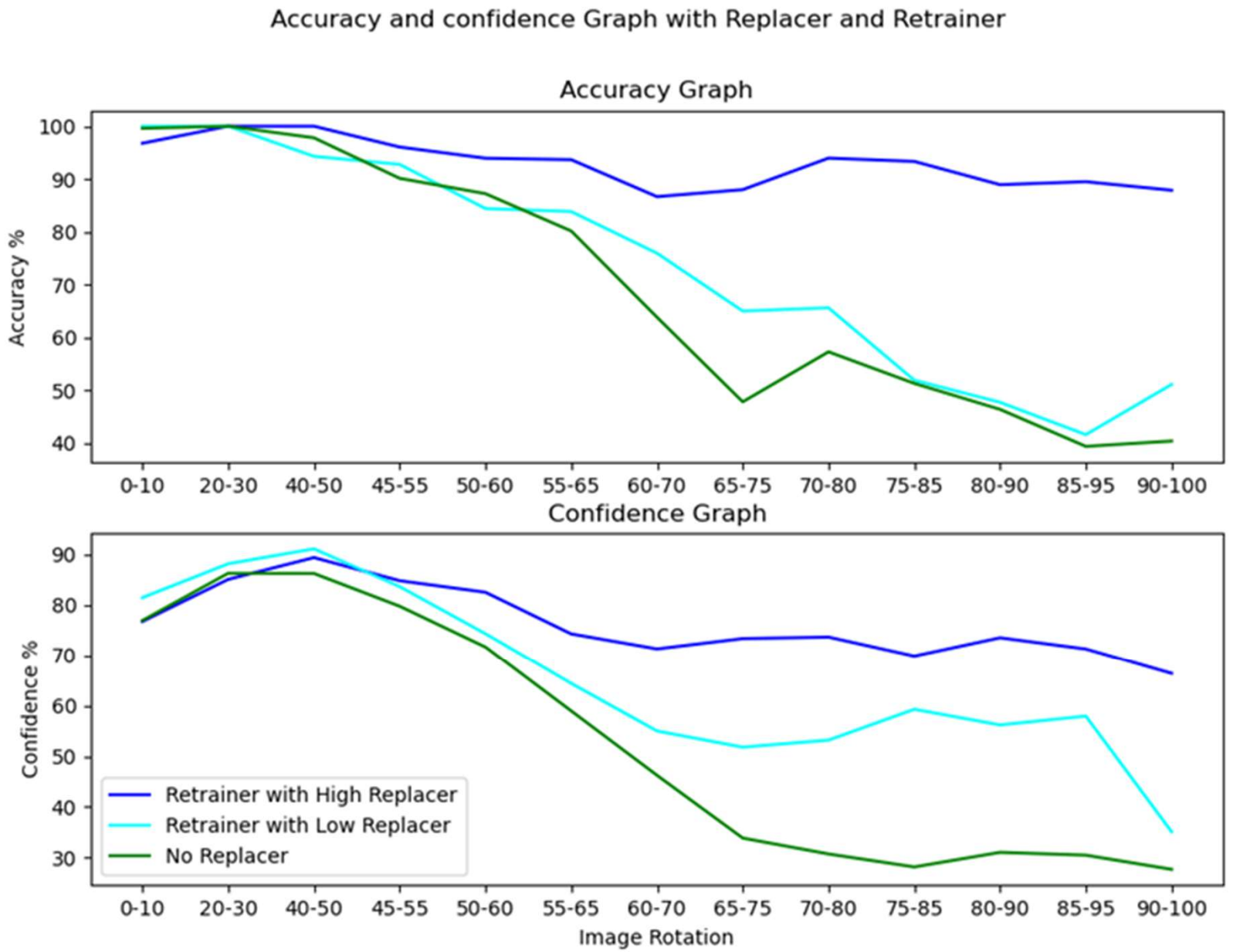
- Continuously monitors overall system performance and collects outliers from the manager
- Collects confidently classified data and uses it to replace existing training data
  - Allows the dataset to shift over time while still maintaining some of its originality
  - Level of replacement determined by an experimentally tuned adaptability variable
- Raises a warning and kills the system if performance degrades to a point that cannot be recovered from

Retrainer

- Selects data for retraining and selects AI for retraining and/or dataset for replacement
- Active and non-active components
- Outliers are stored in a Hashmap for clustering
- When enough outliers have been collected, they are used to retrain expert AI's



Retrainer



Future Work

- More complex datasets and tasks
  - Will the framework scale to more complex images?
  - Can the framework be transitioned to non-classification tasks?
  - Will the framework work well with distorted images?
- Immediate use case: Panoramic Asset Tracking of Real-time Information for the Ouija Tabletop (PATRIOT)
  - Complex and dynamic environment
  - Varying lighting and weather conditions
  - More complex task
    - Pose estimation

Thank you!

If you have any questions, please reach out to my lab at riselab@us.navy.mil