Object Detection For Robotic Activity Support

Nathaniel Burley, Sean Kallaher, James Irwin, Gabriel de la Cruz,
Matthew E. Taylor, and Diane Cook

College of Electrical Engineering and Computer Science, Washington State University

College of Arts and Sciences, Washington State University

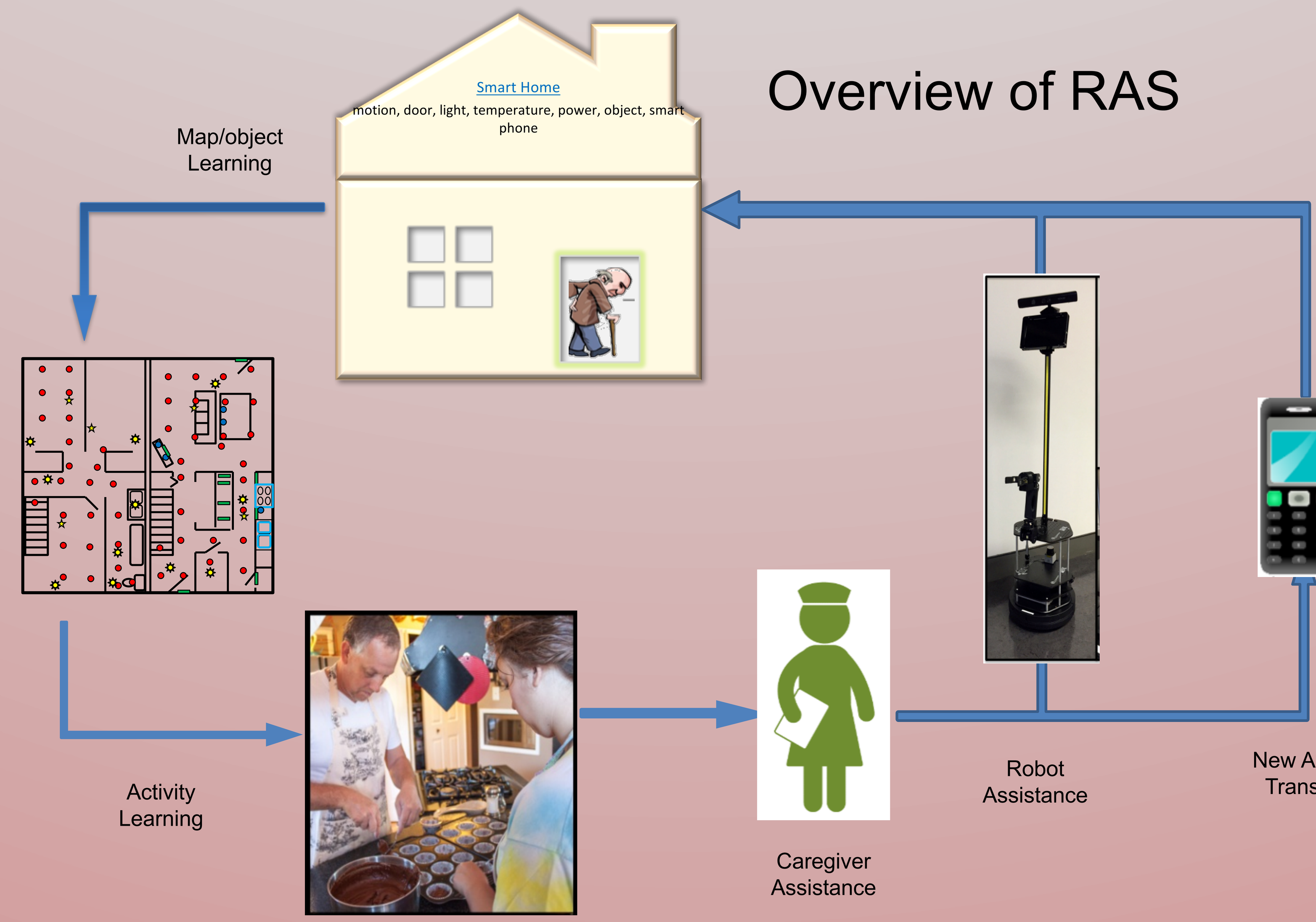
Introduction

- As our population continues to age, the number of nurses and doctors needed to monitor the health of the elderly will continue to decline. The longer an elderly or disabled person can safely live independently, the greater their quality of life will be, and the less of an emotional and financial burden they will be on their family.
- Robotics represents a rapidly growing sector of gerontechnology, a field that merges gerontology with technology. As a technological tool with a physical presence, robots can provide personal assistance. They can aid in performing simple chores, retrieve critical items, and interact with residents throughout a physical environment to provide reminders on how and when to complete activities of daily living.
- This research project aims to utilize neural networks and simultaneous localization and mapping (SLAM) algorithms to program a robot equipped with a 3D camera, robotic arm, and Digital Memory Notebook to map its surroundings, and autonomously locate and retrieve objects.
- Our contribution to this project is adding computer vision to the robot, specifically object detection. Object detection is an exciting new field in computer vision allows computers to accurately recognize objects images. We trained a neural network as a proof of concept that can recognize 3 objects.
- Object detection is a subset of the computer vision and image processing fields that deals with identifying instances of objects belonging to a general class. Examples could include cars, faces...or hot dogs.
- The long term goal for this project is to train a network that can recognize many common objects, and add code to allow the robot to place recognized objects on its map of the environment. This will allow the robot to, when prompted, either retrieve objects with its onboard arm, or provide a reminder of the location of the object on a map of the environment. The robot will also interface with the sensors in the environment and the Digital Memory Notebook.

Methods

- Initially, as a test of the platform, we attempted to train it to recognize just a black Expo marker. After figuring out the nuances of the platform, we attempted to identify, as a proof of concept, a toothbrush, pill bottle, and an Expo marker. We chose these objects for visibility, size, (the robotic arm on the Turtlebot can only grab and hold small, thin objects) and commonality- the aforementioned objects are commonly found in the home.
- The datasets were custom made, and consisted of about 30,000 images. Due to the large amount of time labelling the images takes, 11,000 images from the dataset were chosen for training. The images were labeled using Sloth, a program that allows users to draw bounding boxes around objects in images. These boxes tell the neural network which pixels are considered part of the object. It converts Sloth's JSON output to the Darknet format, which is then used for training.
- For training the detector, we utilized the YOLO (You Only Look Once) Object Detection Algorithm on the Darknet Neural Network deep learning framework, developed at the University of Washington by Joseph Redmon. YOLO works by predicting a bounding box, and the confidence of there being an object of a given class there. If the confidence is above a certain threshold, a box is drawn on the image, "bounding" the pixels that Darknet thinks are an object of a certain class.
- We trained the detector on the Kamiak High Performance Computing Cluster. We used 4 Tesla K80 GPUs, which have 24GB of DDR5 memory and 4,992 NVIDIA CUDA dual-GPU cores each. It trained for 4 days, and went through approximately 100,000 iterations on a training set of 10,260 images, with a validation set of 1,140 images.

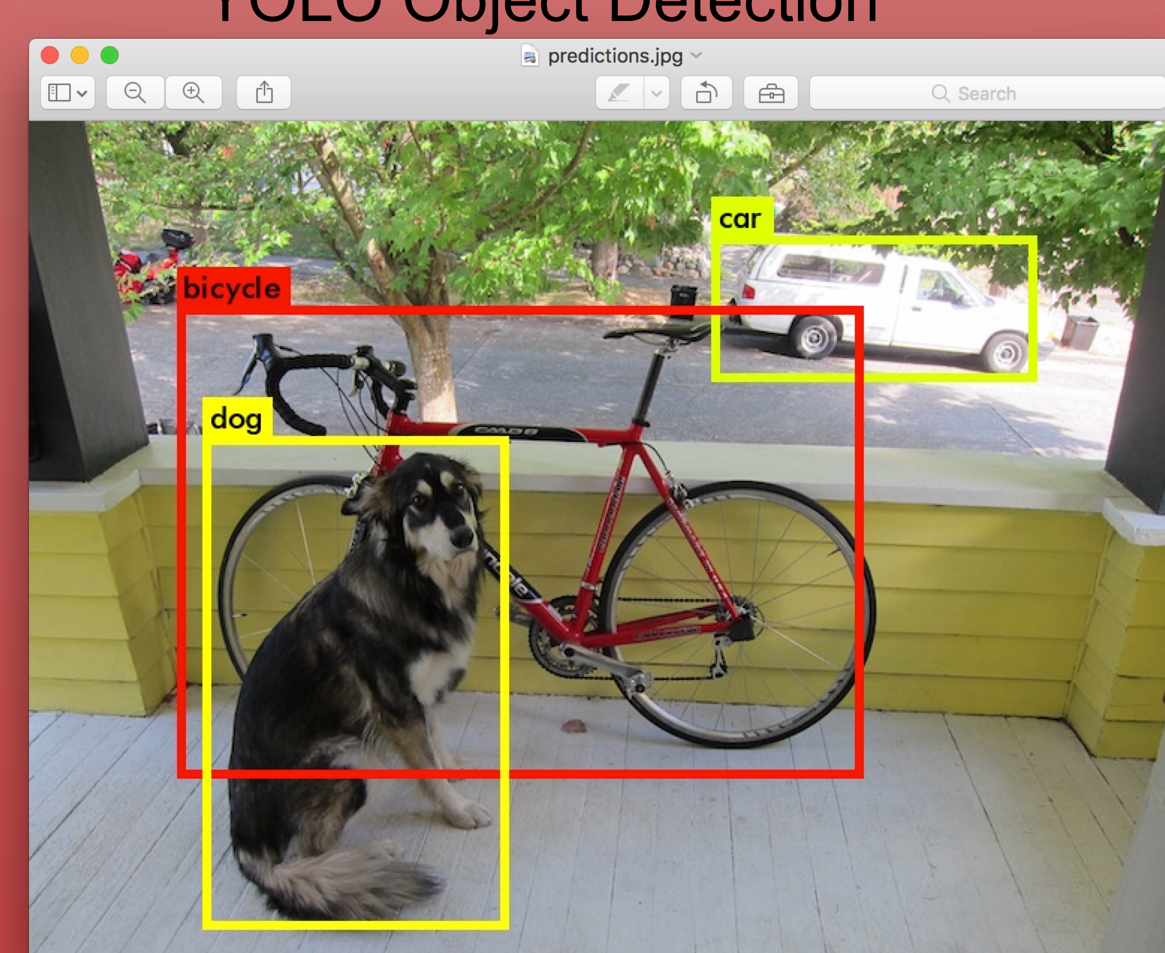
Overview of RAS



Object Detection For Dummies (Conceptual Example)

https://www.thatsnerdalicious.com/funny/not-hotdog-app/

YOLO Object Detection



https://pjreddie.com/media/image/Screen_Shot_2016-11-17_at_11.14.54_AM.png

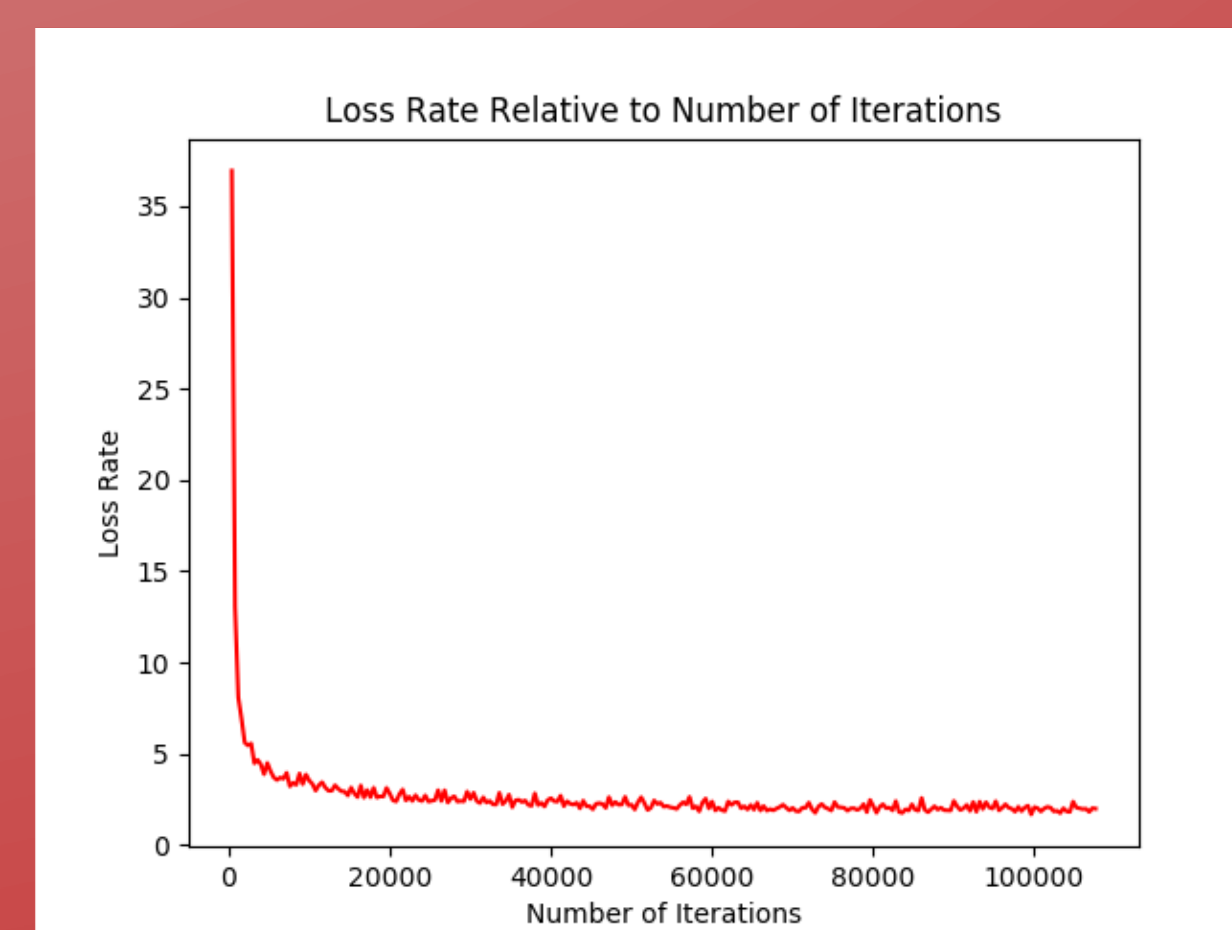
Results

- Successfully trained detectors to recognize a black Expo marker and a pill bottle, toothbrush, and 4 colors of Expo marker, respectively.
- Loss rate for training the deep network was ~2%
- Average confidences were 88-93% for objects in the validation set
- Graphed loss rate relative to number of iterations (loss being a function of the area of the predicted bounding box, and how much it intersects with the actual bounding box of a give object in a given image)
- Learned a lot about machine learning, object detection, Python programming, robotics, computer vision, gerontechnology, research methodology, and high performance computing!

(No image due to technical issues with connection to the high performance computing cluster.)

Future Work

- Continue to test and validate the proof of concept network that was trained
- Test the object detector on a live webcam, to verify real time detection
- Compare to other frameworks and algorithms (TensorFlow, RCNN) to verify which detection algorithm is the fastest and most accurate
- Improve autonomous mapping on the RAS robot
- Write ROS package for object detection (to interface with the robot)
- Purchase and equip the RAS onboard computer with an external GPU so images can be quickly detected
- Add code to tag the environment map with the locations of tagged objects
- Test various image qualities
- Build a larger, more comprehensive dataset for the already trained networks (tag and train on the remaining 19,000 images)
- Build a large, comprehensive dataset of useful objects; train a detector to detect said objects



This work was supported by the National Science Foundation's REU program, and NIH/NIA grant 1R25AG046114