Development of Computer Vision in Quadruped Robots for Construction Applications

Ryan Bender, Andrew Schalk, Jared Jess, Dr. Chris Gordon

URCA Undergraduate Research and Creative Activities

Southern Illinois University Edwardsville Established 1957

Abstract

Quadruped robots like Boston Dynamics SPOT are gaining traction in construction for site monitoring and automation, but they can be made even better with accurate object recognition. This project develops a computer vision pipeline using TensorFlow to improve SPOT's ability to identify construction-related objects. We collected data, trained models, and tested in real-world conditions. Key challenges included real-time performance and environmental variability. Future work will focus on refining detection and integrating it with autonomous functionality.

Methods

Object Recognition

- To start, a simple dog toy is used as a proof of concept. The robot is used to capture 223 training images. A human labels each image with a bounding box around object.
- The labeled images are used with TensorFlow to create a neural network for object identification.

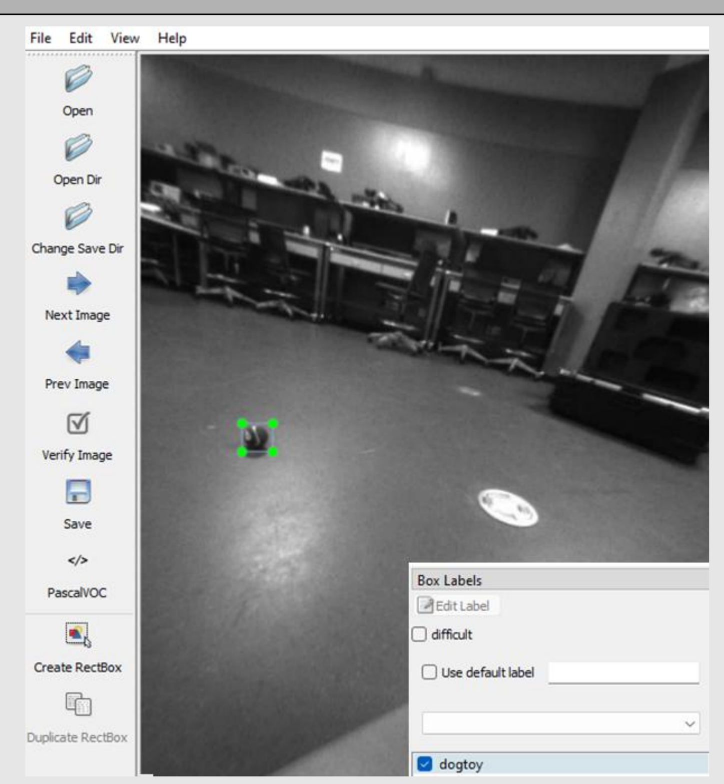

Figure 1: Image Labeling Software

- Upon completing training on the dog toy, the process was easier to streamline. A construction helmet was put through the same process as the dog toy, training with 252 images.

High Performance Computing Cluster

- Training neural networks for object recognition requires significant computing resources. The SIUE high-performance computing cluster provides access to many CPU cores and NVIDIA A100 GPUs.
- With access to these resources, the neural network can be trained quickly – in around a half hour – allowing for on-the-fly modification of parameters and quick validation of model success.

TensorFlow Pipeline

- The trained network is loaded onto a development computer where images will be processed in real-time.

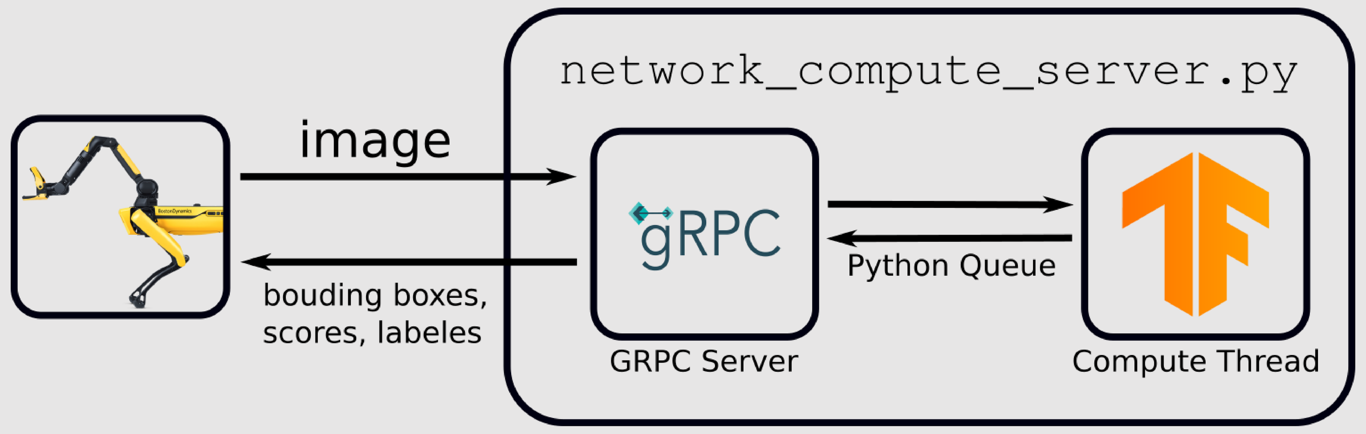

Figure 2: TensorFlow Pipeline

Results

Identifiable Objects

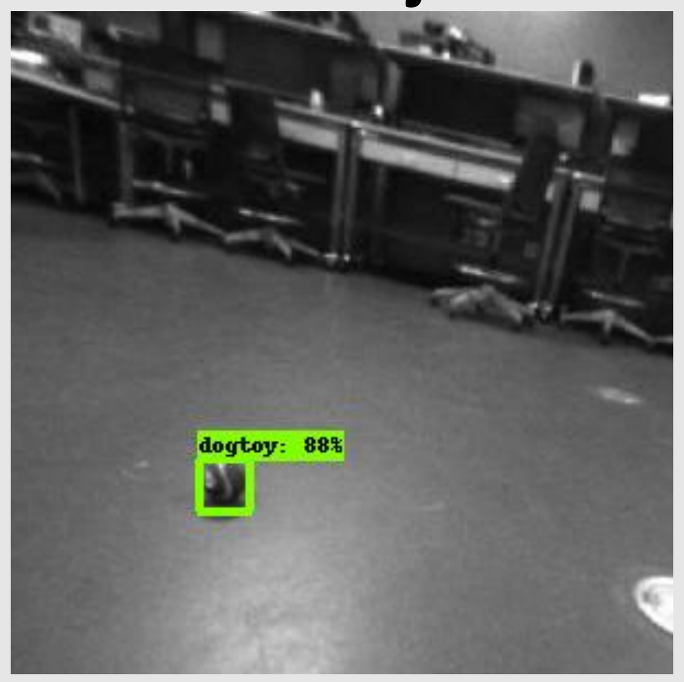

Figure 3: The model identifies a dog toy

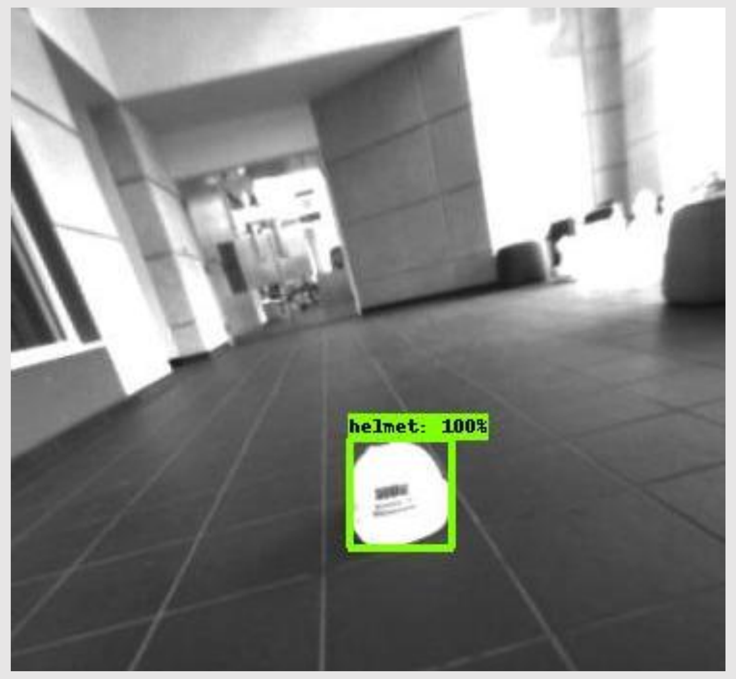

Figure 4: The model identifies a helmet

Helmet Recognition Accuracy

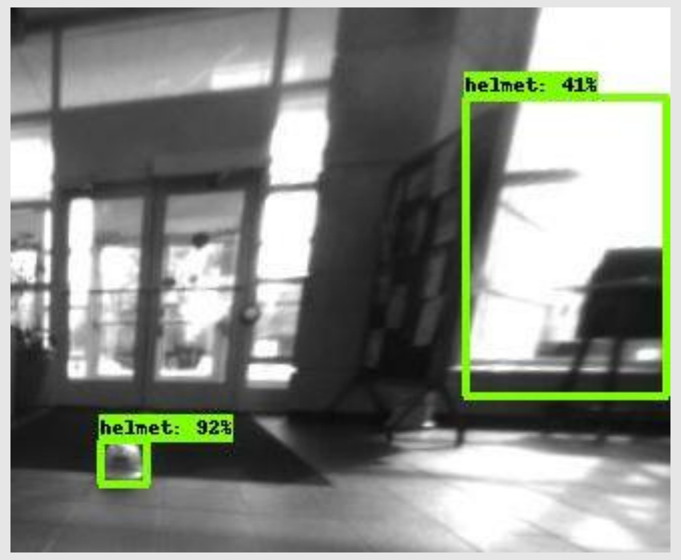

Figure 5: The model returns a false positive

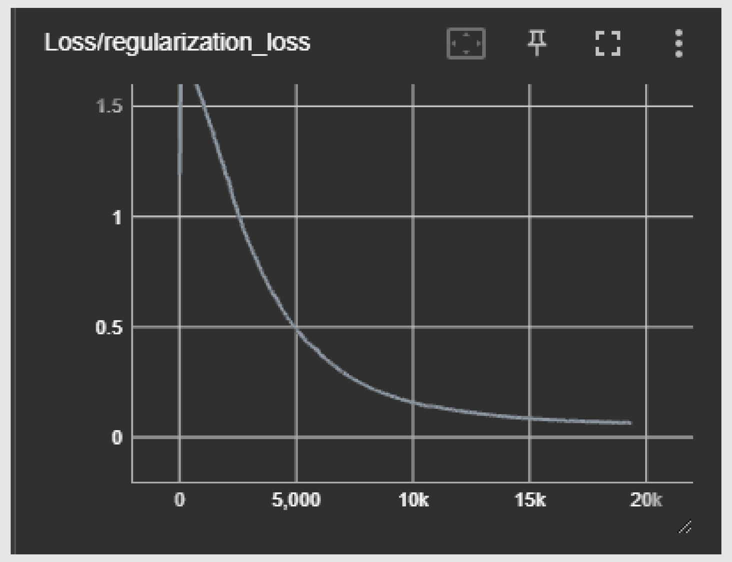

Figure 6: A graph seen within TensorBoard which shows the model's loss over training steps

- The model returned one false positive when it ran on the dataset. With a confidence of 41%, this can be ignored by setting a high confidence threshold.
- Twenty-seven images in the data set were never shown during training. Positive identification on these images ensures that the model was not overfit to the test data.
- 20,000 training steps were found to be an effective training length. The loss reaches an asymptote around this amount.
- The model had an average confidence of 97.4% in identifying the helmet and could identify it in 98.3% of the images in the set.
- The standard deviation of the dataset was 10.7%, so setting the threshold at 85% confidence would avoid false recognition.

Figure 7: BODE, SIUE's robot dog

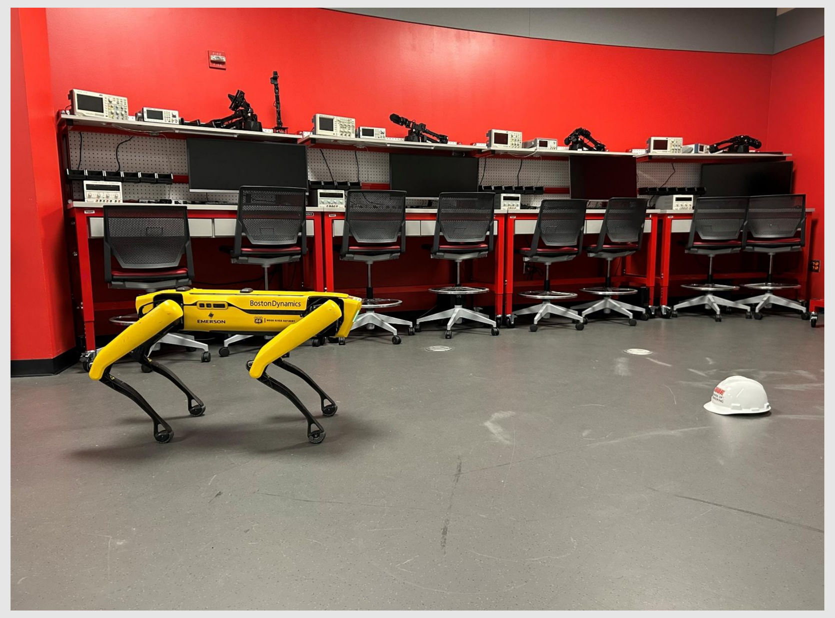

Figure 8: BODE looking at a helmet

Challenges

Deprecated Code

- Boston Dynamics provides documentation for robot operation and programming that provided guidance.
- Some of the programming examples and tutorials included poorly maintained code and were only discovered by contacting Boston Dynamics support when errors occurred.

Multi-system Integration

- The SPOT Core I/O was implemented, providing a companion computer, to be used for data acquisition and processing, hardware integration with the robot, running models/neural networks, etc.
- Implementing the model to the Core I/O proved challenging due to a strenuous debugging process and extensive upload times.

Learning Outcomes

Object Recognition on Robots

- Understand how to train object recognition on a robot and what application it can be used for.

Interactive Real-time Code

- Write Python scripts that are interactive and adaptive to the real-world environment.

Multi-System Integration

- Synchronize code, robot, and interfaces to work cooperatively to complete a task.

Future Directions

Following People & Fiducials

- Adding the ability to follow people and fiducials would drastically improve the robot's capabilities. This could make him a beneficial aide on construction site monitoring.

Inspection Application

- Jobsite monitoring could be utilized in many applications, such as: crew PPE/safety monitoring, site inspection/data capture, dangerous situations in lieu of humans, etc.

Hardware Integration

- Integrating an additional camera can improve the computer vision capabilities developed in this project.
- A speaker/microphone application could significantly improve social interaction when the robot is conducting a mission.