ASPIRE: Explosive Hazard Detection by Leveraging Multiple Sensing Modalities

M. Garcia, N. Isfar, C. Jayasekerage, S. Mollah, H. Kayman, J. Kusyk*, M. U. Uyar

Grove School of Engineering, the City College of New York, CUNY

(*) City College of Technology, CUNY

ABSTRACT

The Puppy-Pi is an Autonomous Quadruped Robot that utilizes a camera to detect objects. We designed and implemented a Fast Regional Convolutional Neural Network (FRCNN) in Python for object detection using the images provided by the US Army. FRCNN is trained to recognize images as one of the four categories, namely bushes, metal cans, plastic bottles and background. This FRCNN is then imported to Puppy-Pi platform to analyze images captured by a camera in real-time. We also implemented a CNN for audio speech recognition to identify a set of voice commands. Once imported into Puppy-Pi, which is equipped with a mini microphone, the CNN detects voice commands and controls Puppy-Pi motion based accordingly in real time.

BACKGROUND

Current explosive hazard detection systems struggle with accuracy, particularly in detecting buried or obscured threats in complex environments. Manual methods are slow, expose personnel to significant risks, and cannot scale effectively for large areas. Factors such as low light or cluttered terrains, further reduce the reliability of existing systems.

Figure 1: *COTS Puppy-Pi as hazardous material detection platform*

PURPOSE

We developed a system with a commercial off the shelf autonomous quadruped robot equipped with a camera for comprehensive multi-sensor data fusion. Using AI models, such as FR-CNN, we perform detection and classification of hazardous objects in real time.

METHODS

- The architecture of the FR-CNN model includes: 3 convolution layers with ReLU, Max Pooling, an RPN + ROI pooling.
- YOLO is used to implement object detection on the Puppy-Pi, with pre-trained weights stored into the memory of the Puppy-Pi.
- The YOLO model was implemented through the ONNX platform and integrated with the Raspberry-Pi and a camera on the Puppy-Pi to detect and follow red cones.

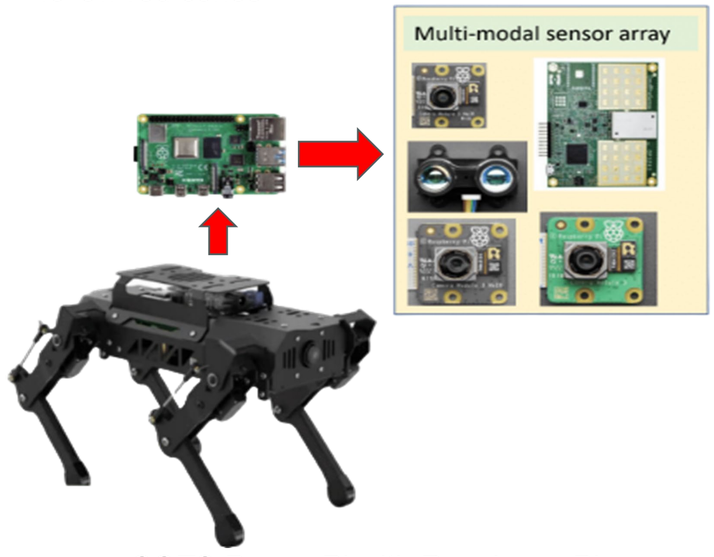
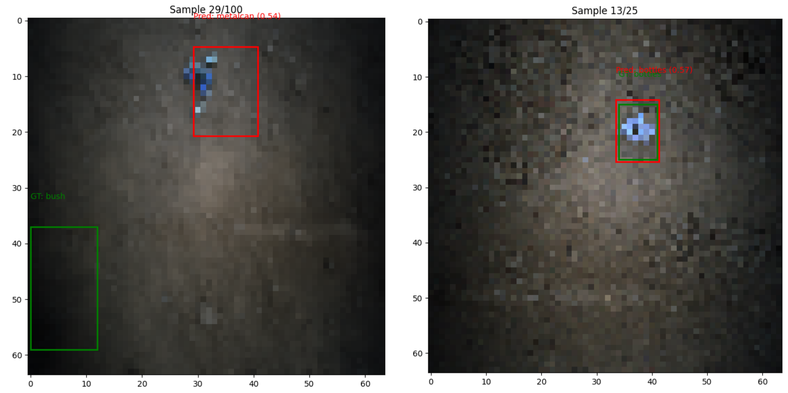


Figure 2: *COTS Puppy-Pi with Raspberry-Pi connected with multi-modality sensors*

- The architecture for the Audio CNN model includes: 3 convolutional layers with ReLU, BatchNormalization,Max pooling and a fully connected layer.
- The CNN for Audio Speech recognition undergoes a preprocessing data phase to extract raw audio of 4 classes: “Go”, “Left”, “Right”, and “Stop”.

RESULTS



Metric	Details/Results
Accuracy on Training Set: (cls)	84.75%
Accuracy on Testing Set: (cls)	80.40%
Average IoU Training:	0.5075
Average IoU Validation:	0.4785

Figure 3: *Images and Results of the FR-CNN*

CONCLUSION

Deliver a scalable, adaptable solution designed to minimize human risk while maximizing operational efficiency. The Puppy-Pi will demonstrate its ability to recognize an object through the YOLO model and detect voice commands from the CNN to control the motion of the Puppy-Pi . Looking ahead, we plan to enhance Puppy-Pi’s functionality by integrating advanced speaker recognition capabilities, and we aim to fuse the existing CNN with complementary modality sensors to create a comprehensive hazard detection system.

REFERENCES

1. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, **S.Ren, K. He, R. Girshick, J. Sun, 2015**
2. *Speech Command Recognition with Convolutional Neural Network,* **Stanford University**
3. *Audio Recognition using Mel Spectrograms and Convolution Neural Networks,* **University of California, San Diego, 2021**
4. *Ultralytics YOLO11*, **Glenn Jocher and Jing Qiu, 2024**