Out-of-Distribution Detection for Safe and Efficient Human-Robot Collaboration

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik

Background

We are on the verge of a technological revolution, where robots are becoming increasingly integrated into our everyday lives. New robots must interact in highly dynamic environments and guarantee safety for themselves and others. For this, measuring the environment as accurately as possible is essential. Cameras are cheap and flexible sensors for most robotic applications with the potential to measure the entire environment. Deep neural networks have been demonstrated to be the most sufficient solution for extracting information from images. Unfortunately, we cannot give strong bounds on the disturbances that act on the output of a deep neural network. For safety-critical perception like human detection and prediction, we often cannot rely on deep neural networks, and we restrain ourselves from using them altogether.

To alleviate this, we propose a framework for a safe robotic stack that integrates deep neural network predictions into formal methods to give strong safety guarantees. We aim to deploy this framework in human-robot collaboration settings where a manipulation robot interacts with humans.

Deep learning underlies two sources of uncertainties: aleatoric uncertainties are due to inherent randomness and imprecision in the learning process, whereas epistemic uncertainties mainly stem from states unseen during training time. Predicting aleatoric uncertainties became a common standard in most deep learning domains. Recent advances in out-of-distribution detection [2, 4, 1] predict epistemic uncertainties, allowing us to predict when the model output is reliable. As shown in previous works [5], these methods can then be used to guarantee the safety of autonomous robots by falling back to a failsafe trajectory if an out-of-distribution state is detected.

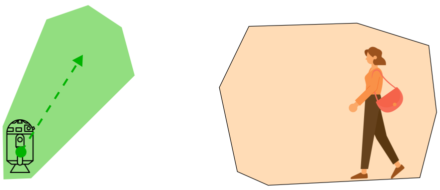

figure 1: The human is correctly detected and the set of all future states is safe.

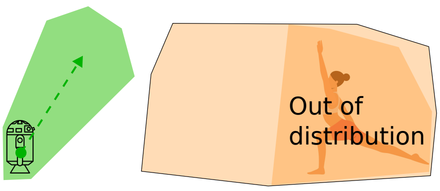

figure 2: The human is not correctly detected due to an out-of-distribution state. However, since the human is so far away and we still have a valid measurement from the last time step, we can still verify the robot action as safe.

Our proposed safety framework utilizes set-based reachability analysis [3, 6]. Its core idea is to predict all possible future states of the robot and the environment. We then validate the safety of this set of possible future states. If a subset of future states is recognized as potentially unsafe, we fall back to a failsafe trajectory, bringing the robot to a safe state. We propose using neural network prediction to measure parts of the robotic stack. The predicted aleatoric uncertainty will be used to determine the size of our reachable sets. We can discard false measurements by detecting out-of-distribution inputs of the neural network. Our set-based reachability analysis propagates the uncertainties in the system forward in time, regardless of the last measurement time. We can, therefore, guarantee continuous safety regardless of out-of-distribution states.

We want to show that safe human-robot collaboration is possible based on visual inputs

Supervisor:
Prof. Dr.-Ing. Matthias Althoff
Prof. Dr. Marco Pavone

Advisor:
Jakob Thumm
Rohan Sinha

Research project:
CONCERT

Type:
MA

Research area:
Out-of-distribution detection, human pose estimation, safe human-robot collaboration

Programming language:
PyTorch / TensorFlow, Python, C++

Required skills:
Strong experience in deep learning, PyTorch or TensorFlow, Linux, Python or C++.

Language:
English, German

Date of submission:
14. März 2024

For more information please contact us:

Phone:

E-Mail: jakob.thumm@tum.de

Internet: www.ce.cit.tum.de/cps

alone. For this, we propose two prediction networks, one for measuring the current human pose and one for predicting future human motion. Both networks will be checked for aleatoric and epistemic uncertainties to ensure a prediction within known bounds. By integrating this prediction into our reachability analysis, we can shrink the human reachable sets, making the robot control less restrictive while maintaining high robustness.

Technische Universität München

Fakultät für Informatik

Lehrstuhl für Echtzeitsysteme und Robotik

Tasks

- Perform a literature review on
  - human pose estimation
  - uncertainty prediction in human pose estimation
  - out-of-distribution detection
- Design a deep neural network for human pose estimation that outputs aleatoric uncertainties, mainly from existing literature.
- Apply a state-of-the-art out-of-distribution detection algorithm to the pre-trained network.
- Incorporate the outputs of the previously trained networks into our existing safety shield for safe human-robot collaboration.
- Extend the framework to predict the human motion into the future.
- Real-world experimental evaluation with a manipulator.

Key learnings

- Human pose estimation from images.
- Estimating aleatoric uncertainties of neural network predictions.
- Out-of-distribution detection in deep learning.
- Formal methods and set-based reachability analysis.
- Safety for real-world robotic applications.

References

[1] Jogendra Nath Kundu, Siddharth Seth, Pradyumna YM, Varun Jampani, Anirban Chakraborty, and R. Venkatesh Babu. Uncertainty-aware adaptation for self-supervised 3D human pose estimation. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 20448–20459, 2022.

[2] Liu Qihao, Adam Kortylewski, and Alan Yuille. Poseexaminer: Automated testing of out-of-distribution robustness in human pose and shape estimation. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 672–681, 2023.

[3] Sven Schepp, Jakob Thumm, Stefan B. Liu, and Matthias Althoff. SaRA: A tool for safe human–robot coexistence and collaboration through reachability analysis. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 4312–4317, 2022.

[4] Megh Shukla, Roshan Roy, Pankaj Singh, Shuaib Ahmed, and Alexandre Alahi. Vl4pose: Active learning through out-of-distribution detection for pose estimation, 2022.

[5] Rohan Sinha, Edward Schmerling, and Marco Pavone. Closing the loop on runtime monitors with fallback-safe MPC. In *Proc. of the IEEE Conf. on Decision and Control (CDC)*, pages 6533–6540, 2023.

[6] Jakob Thumm and Matthias Althoff. Provably safe deep reinforcement learning for robotic manipulation in human environments. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 6344–6350, 2022.