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

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.

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
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.

**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**


