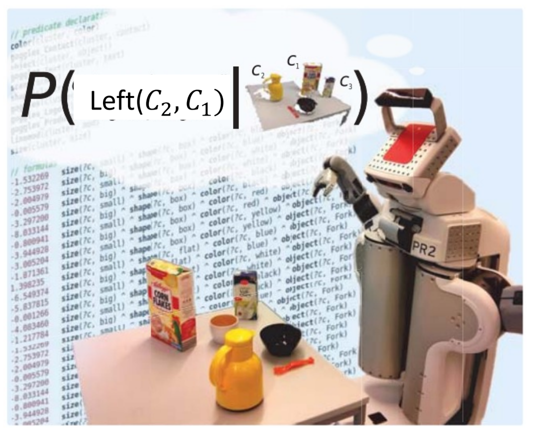
Understanding spatial relationships in tabletop scenarios for human-robot collaboration

Understanding Spatial Relationships in Tabletop Scenarios for Human-Robot Collaboration

Fabian Gracia[1,2], Kevin Nguyen[1,2], Mai Dao[1], Fujian Yan[2]
[1]Department of Mathematics and Statistics, [2]School of Computing, Wichita State University

Background

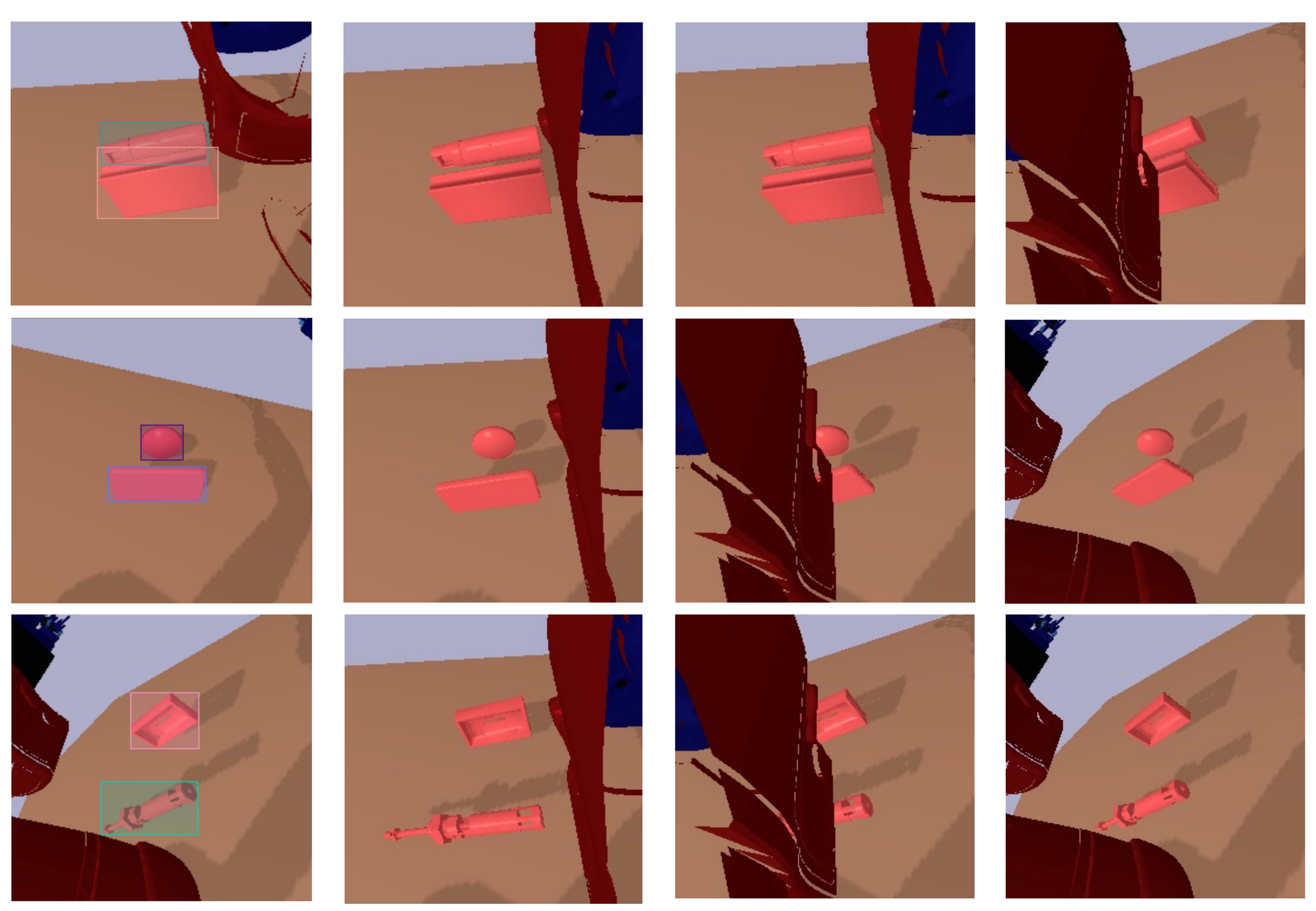
- Understanding spatial relations is *essential* for:
  - Task planning.
  - Fundamental descriptor of the environment.
  - Enhancing the adaptivity of robots.
- Hard-coding or emulating method *does not work* because of:
  - Limited generalization and adaptability.
  - Inherent ambiguity and uncertainty.



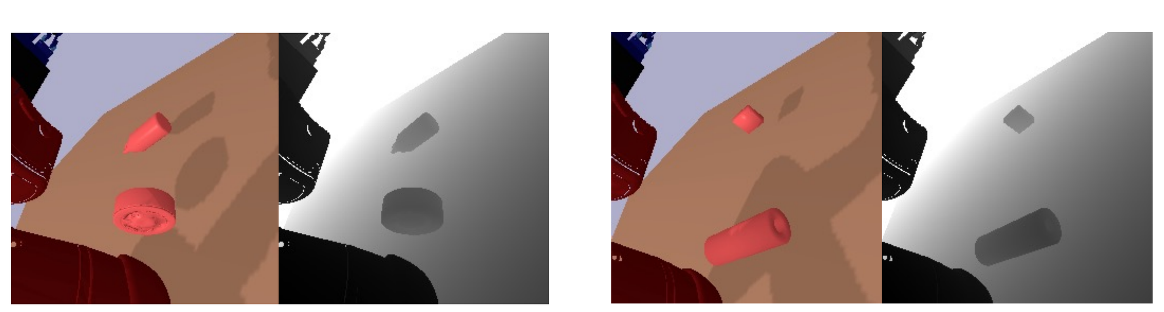
Data Preparation

The dataset was obtained using an automated script:

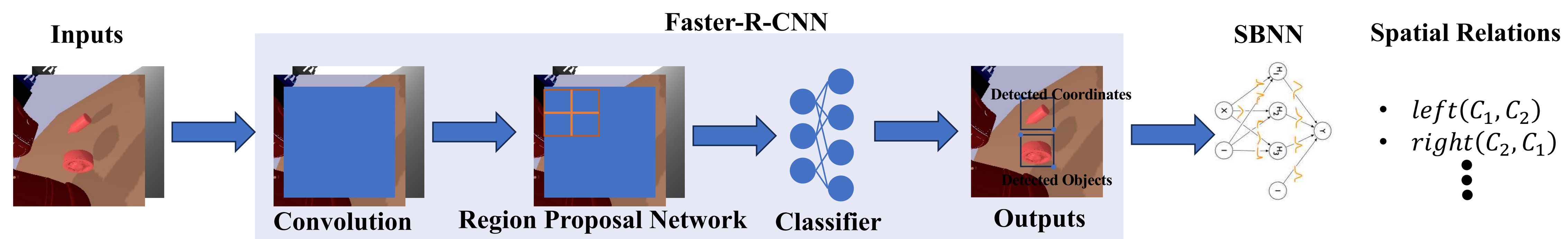
- **56** images of each pair of objects at **six** different distances.
- A total of **13,104** object pair images.
- The dataset comprises of *common household objects*, such as books, bottles, cardboard boxes, and other items.
- They are carefully selected to simulate *realistic scenarios and tasks* encountered in common daily household settings.
- Dataset insights:



- Both RGB and Depth images are captured in the dataset.



Methodology

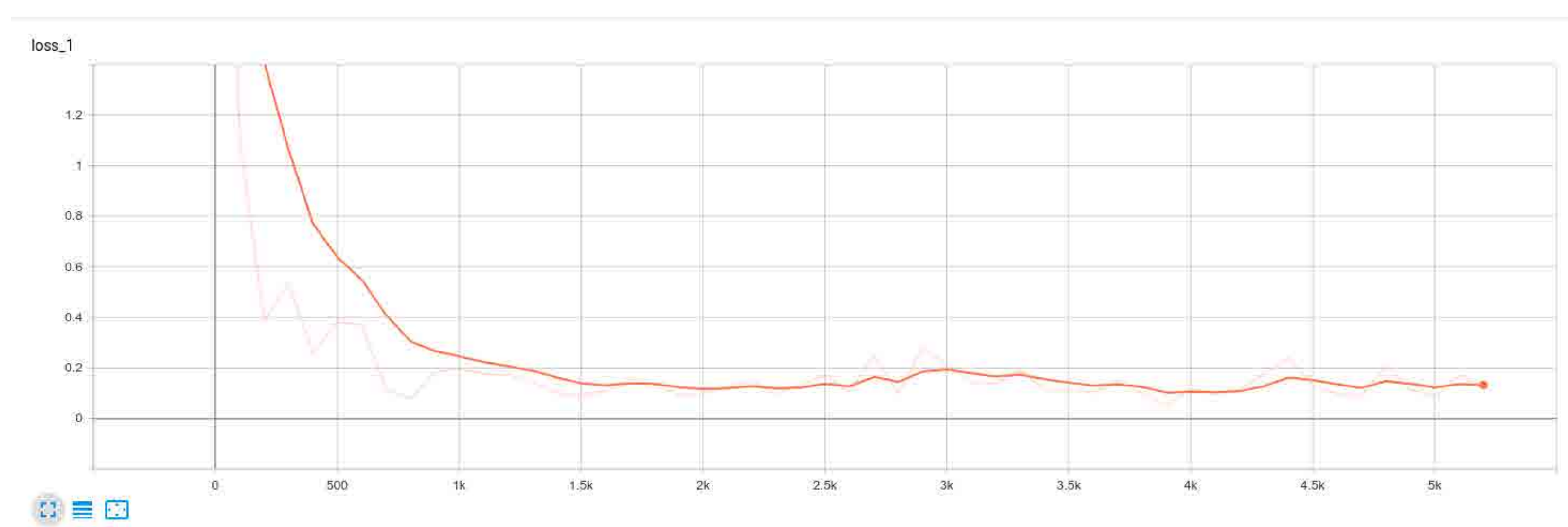


- **Fast R-CNN:** Recent advances in object detection are driven by *region proposal* methods and region-based convolutional neural networks (R-CNN). Fast R-CNN achieves near real-time rates using very deep networks, while ignoring the time spent on region proposals, the main computational *bottlenecks*.
- **Region Proposal Network:** A region proposal network (RPN) takes an image as input and outputs a set of rectangular object proposals, each with an *objectness score*. It is implemented as a fully-convolutional network and trained end-to-end by back-propagation and stochastic gradient descent.
- **Faster R-CNN:** This algorithm learns the *shared convolution layers* between the RPN and Fast R-CNN.
  - Train the RPN with an ImageNet pre-trained model and fine-tune it end-to-end.
  - Train a separate detection network by Fast R-CNN using the proposals generated by step 1 RPN.
  - Use the detector network to initialize RPN training. Fix the shared convolution layers and only fine-tune the layers unique to RPN.
  - Keep the shared convolution layers fixed, fine-tune the layers of the Fast R-CNN.

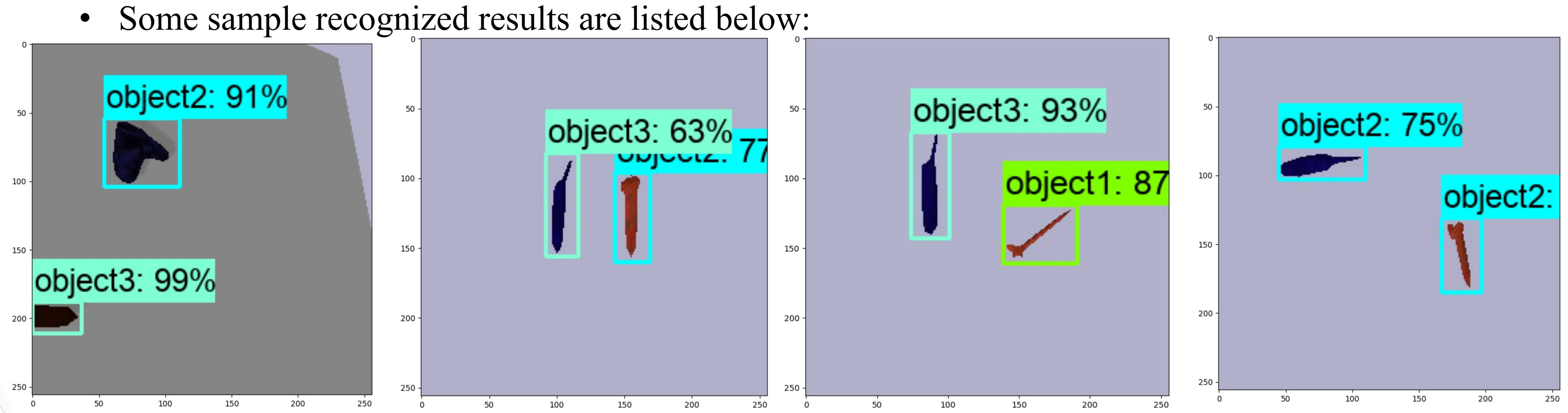
Results

Experiment Setup:

- Due to the long training time, we demonstrate the proposed model with the Random URDF dataset.

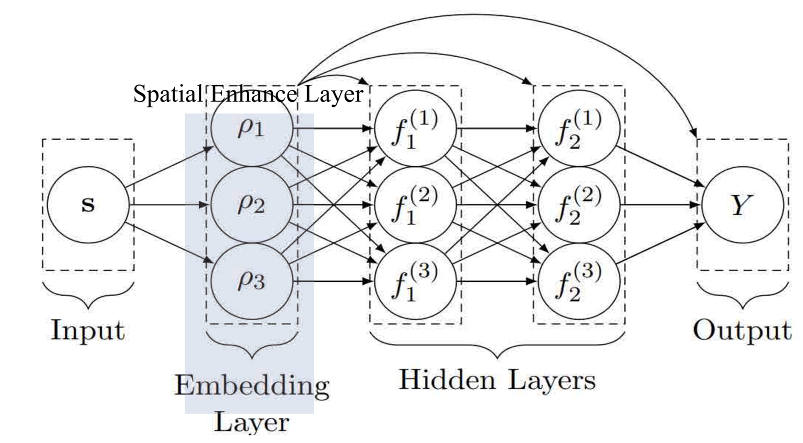


- Some sample recognized results are listed below:



Future Work

- **Execute the Three-Phase Algorithm:**
  - Detect and recognize objects in the scene.
  - Understand the spatial relations among objects in the scene.
  - Integrate the spatial relations comprehension into HRC.
- **Train the Spatial Bayesian Neural Network (SBNN):**
  - Inputs will be the coordinates of detected contours of objects and outputs will be the detected spatial relations.



- **Implement to Human-Robot Collaboration (HRC):**
  - More intuitive for human-robot interaction.
    - Understanding spatial relations helps robots interpret human gestures, body languages, and other non-verbal cues, allowing natural responses.
  - Collaboration in confined spaces.
    - In collaborative workspaces, particularly confined ones, like factories or labs, spatial relations are critical. Robots need to smoothly navigate these areas without disrupting human workflows.
  - Task coordination and efficiency.
    - Understanding spatial relations allows robots to recognize task-related spatial arrangements, such as identifying where to place items or when to assist with object retrieval.

Discussions

- Future research directions encompass a multifaceted approach aimed at advancing the capabilities of SBNNs and enhancing HRC systems for more complex, dynamic, and diverse collaborative tasks and environments.
- Additionally, investigating the integration of multi-model data sources, including RGB-D sensors and LiDAR, can improve the data input for SBNNs, leading to more robust and context-aware decision-making in collaborative scenarios.

References

1. *Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.*
2. *Zammit-Mangion, A., Kaminski, M. D., Tran, B. H., Filippone, M., & Cressie, N. (2024). Spatial Bayesian neural networks. Spatial Statistics, 100825.*
3. *Yan, F., Wang, D., & He, H. (2020, October). Robotic understanding of spatial relationships using neural-logic learning. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 8358-8365). IEEE..*
4. *Yan, F., Namboodiri, V., & He, H. (2021). Common reality: An interface of human-robot communication and mutual understanding. In Social Robotics: 13th International Conference, ICSR 2021, Singapore, Singapore, November 10–13, 2021, Proceedings 13 (pp. 319-328). Springer International Publishing.*