Enabling Robots to Understand Natural-language Instructions for Object Manipulation

A Proof of Concept Developed at the NNL Hot Robotics Facility

Riza Batista-Navarro, PhD
Senior Lecturer
Department of Computer Science
University of Manchester

Viktor Schlegel, PhD (UoM)
Shayaan Sindhoo (UoM)
Cuebong Wong (NNL)
Brendan Perry (NNL)
Peter Sweetman (NNL)

UK Nuclear Academics Meeting 2022
7th September 2022

NNL Hot Robotics Facility

Testing & Demonstration Spaces

Decommissioning Cell ›

Laser Cutting Robot Facility ›

Sort and Segregation Facility ›

Water Dive Tank ›

NNL Hot Robotics Facility

Manipulators

Kinova Gen3 7DOF ›

KUKA KR 10 ›

KUKA KR 180 ›

KUKA KR120 ›

- XBOX One Controller
- Omnivision OV5640 Colour sensor
- 2-finger ROBOTIQ 2F-85 Gripper

NNL Hot Robotics Facility

Manipulators

Kinova Gen3 7DOF ›

KUKA KR 10 ›

KUKA KR 180 ›

KUKA KR120 ›

- Lightweight design, slender
- Suitable for glovebox environments

Using natural language to interface with a robot

Why?

- to enable robotic systems to respond to a human operator
- could allow robot to change its behaviour as it undertakes a certain task
- reduces the need to program a robot for every task

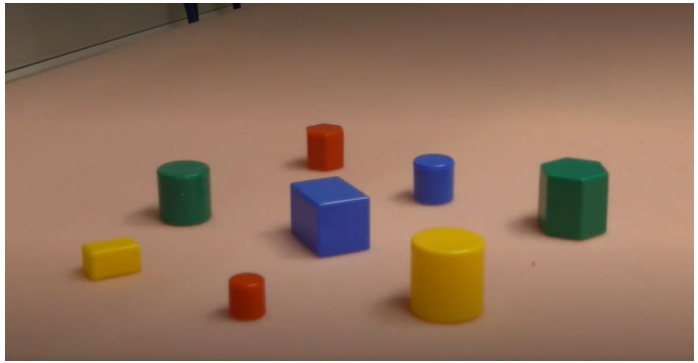
Using natural language to interface with a robot

Objectives:

1. To equip a robotic manipulator with a natural language interface that accepts spoken, **natural language instructions**
2. To integrate a semantic analysis component into the manipulator, in order to:
   - understand the meaning contained within the instructions
   - visually ground *referring expressions* to the current environment
   - execute appropriate actions

Natural Language Understanding (NLU)

Proof-of-concept-system: **Blocks world environment**



Speech-to-text Translation

"Place the blue cylinder on top of the green cylinder"

Intent Classification & Slot Filling

Type of action

Referring expressions, e.g., block shape, colour, position

intent: placeObject
colour: blue
shape: cylinder
target_colour: green
target_shape: cylinder
target_position: on top of

Natural Language Understanding

Speech-to-Text Translation

Implemented using **Rhasspy**[1]: on-device (offline) voice assistant

Uses the **Kaldi**[2] engine for automatic speech recognition

Intent Recognition & Slot Filling

Implemented using **Snips**[3]: on-device machine learning-based engine for NLU

Model was trained on a dataset with intents and slots manually labelled

[1] https://rhasspy.readthedocs.io/en/latest/
[2] https://kaldi-asr.org/
[3] https://snips-nlu.readthedocs.io/en/latest/

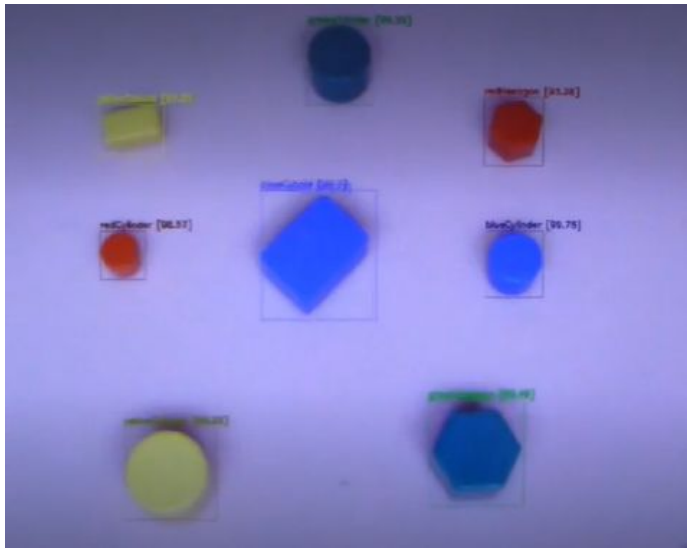
Visual Grounding

Object Recognition and Localisation

Implemented using the **YOLO**[4] (You Only Look Once) library

Model was trained on images with blocks annotated based on colour and shape:

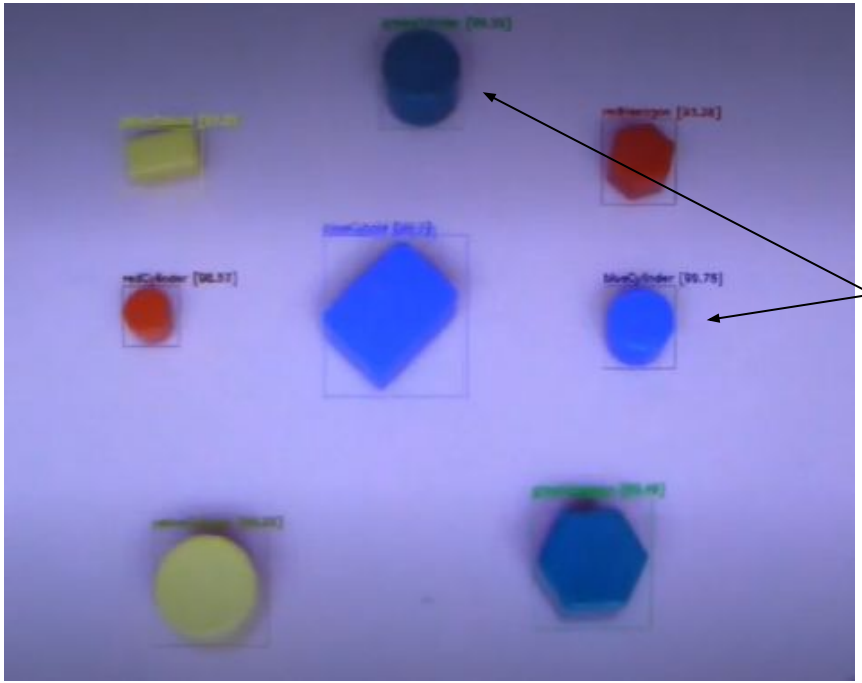
redCube	blueCube
redCuboid	blueCuboid
redHexagon	blueHexagon
redCylinder	blueCylinder
greenCube	yellowCube
greenCuboid	yellowCuboid
greenHexagon	yellowHexagon
greenCylinder	yellowCylinder



[4] https://pypi.org/project/yolo/

Visual Grounding

Object Recognition and Localisation



intent: placeObject
colour: blue
shape: cylinder
target_colour: green
target_shape: cylinder
target_position: on top of

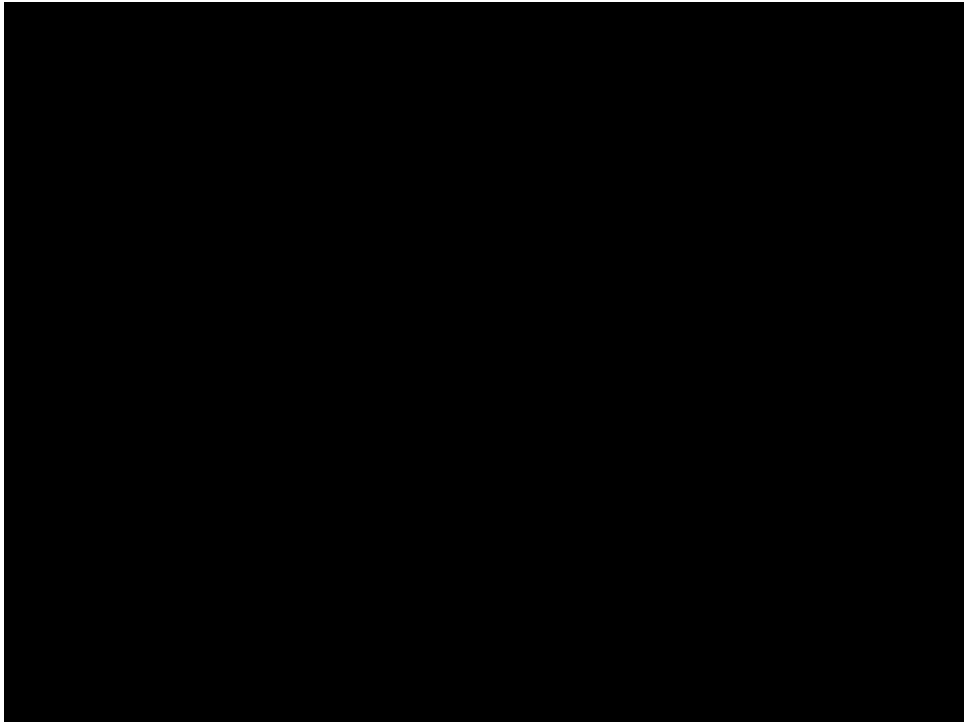
Manipulation

Planning

Real-world coordinates of the objects are calculated based on depth stream

Plan consisting of action sequence is created (based on intent and target position)

Plan executed by robot's control mechanism



Future Work and Acknowledgements

What's next?

To develop natural language interfaces to robotic manipulation systems specifically for nuclear environments

NNL Staff

Access to NNL Hot Robotics Facility at Workington for 20 days

Support from Cuebong Wong, Brendan Perry and Peter Sweetman

Any questions?

Please contact riza.batista@manchester.ac.uk