LATVIJAS UNIVERSITĀTE
DATORIKAS
FAKULTĀTE

ELEKTRONIKAS UN
DATORZINĀTŅU
INSTITŪTS

INSTITUTE OF
ELECTRONICS AND
COMPUTER SCIENCE

Development of Robot Cognition Through Applications of Machine Learning

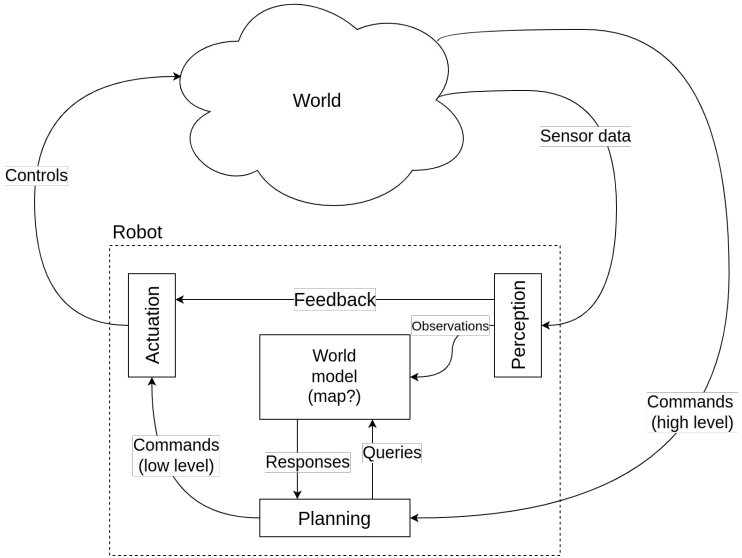
Thesis progress report

Author: Peteris Racinskis
Supervisor: Dr.sc.comp. Modris Greitans

Contents

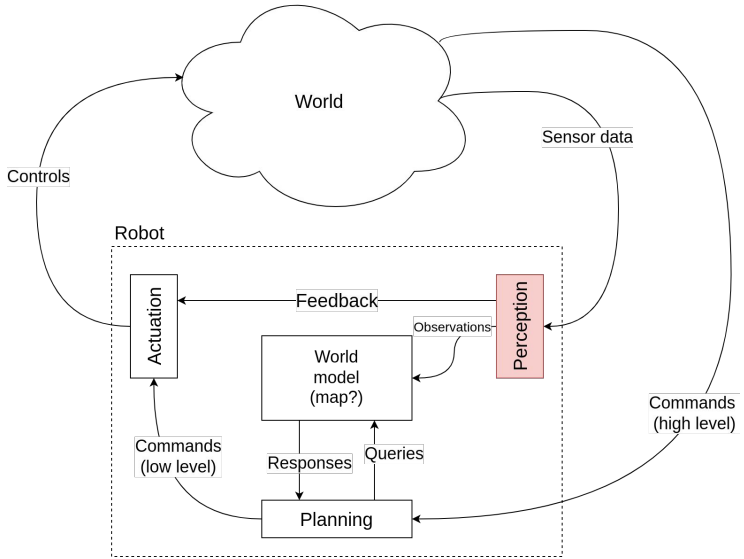
Topic

- Development of Robot Cognition through the Application of Machine Learning
  - “*Mašīnmācīšānās metožu pielietojums robotu kognitīvo spēju attīstīšanā*”
- Three main directions of study
  - Perception
  - Environment mapping / modelling
  - Planning
  - All exist on a continuum!



Topic – perception

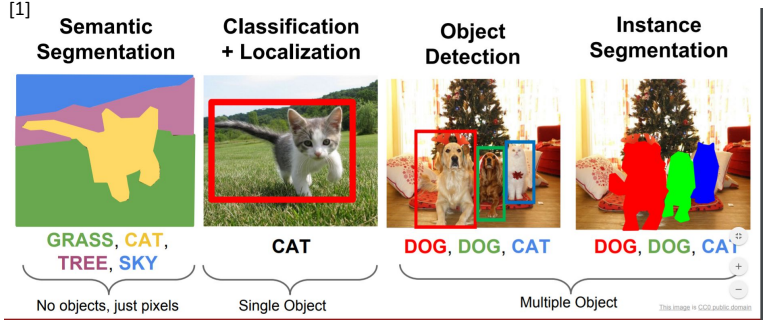
- Given raw sensor measurements, produce meaningful observations
- Visual inputs
  - Classification
  - Object detection
  - Segmentation
  - ...
- Other modalities
  - Radar
  - LiDAR
  - Audio
  - ...



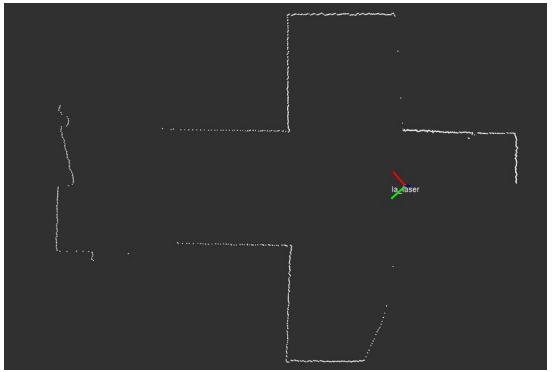
Topic – perception

- Given raw sensor measurements, produce meaningful observations
- Visual inputs
  - Classification
  - Object detection
  - Semantic segmentation
  - Instance segmentation
  - ...
- Other modalities
  - Radar
  - LiDAR
  - Audio
  - ...

[1]



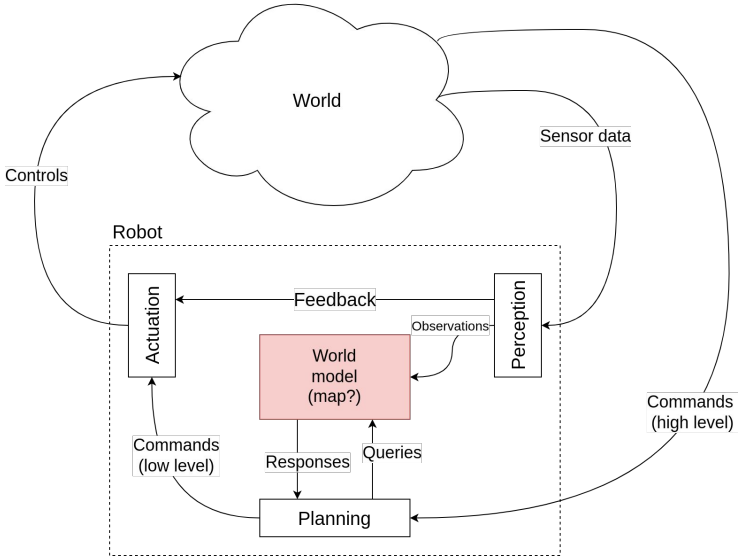
[2]



[1] https://data-flair.training/blogs/wp-content/uploads/sites/2/2020/09/segmentation-types.png
[2] https://www.researchgate.net/figure/2D-laser-scan-left-taken-with-a-Sick-LMS-100-Photo-credit-IST-TU-Graz_fig15_304987927

Topic – environment modelling

- Given meaningful observations, construct an *actionable* model of the environment
- What is the state of the world at a given place, time? Needed for planning
- Most common form - maps
  - Metric
  - Topological
  - Semantic
  - Implicit
- Many AI-based planning approaches still skip this step!



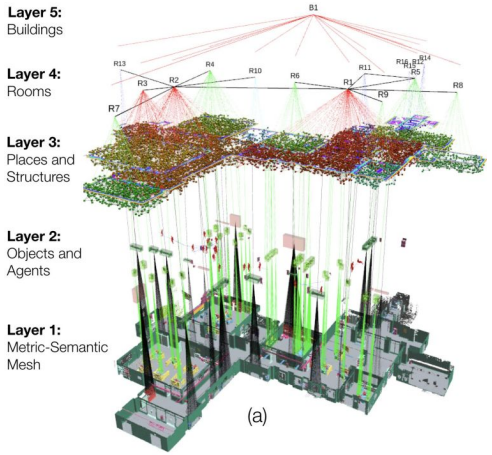
Topic – environment modelling

PGO SLAM demo

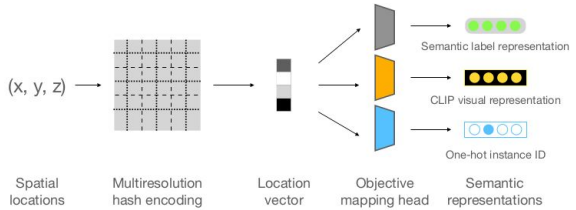
Topic – environment modelling

- Given meaningful observations, construct an *actionable* model of the environment
- What is the state of the world at a given place, time? Needed for planning
- Most common form - maps
  - Metric
  - Topological
  - Semantic
  - Implicit
- Many AI-based planning approaches still skip this step!

[1]



[2]

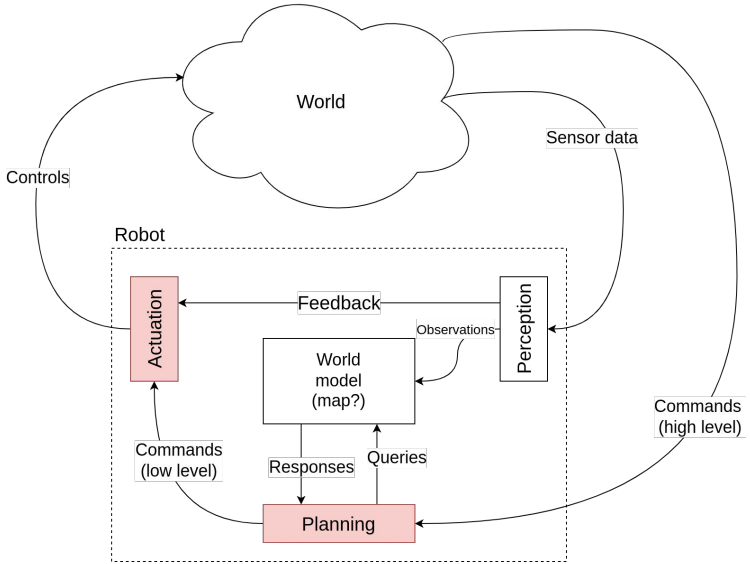


[1] Rosinol, Antoni, Arjun Gupta, Marcus Abate, J. Shi and Luca Carlone. "3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans." ArXiv abs/2002.06289 (2020): n. pag.

[2] Shafiullah, Nur Muhammad (Mahi), Chris Paxton, Lerrel Pinto, Soumith Chintala and Arthur D. Szlam. "CLIP-Fields: Weakly Supervised Semantic Fields for Robotic Memory." *ArXiv* abs/2210.05663 (2022): n. pag.

Topic - planning, actuation

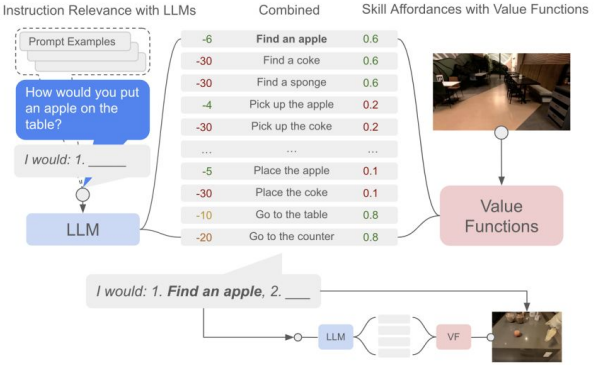
- Given commands and observations of the environment (possibly augmented by an internal world-model), produce a control output
- Classic approaches – find feasible / optimal trajectories in configuration space
- ML-based approaches
  - End-to-end learned policies
  - Motion primitives
- Planning vs control
  - Planner – construct sequence of commands or setpoints
  - Controller – physically drive the execution of commands, track setpoints through feedback



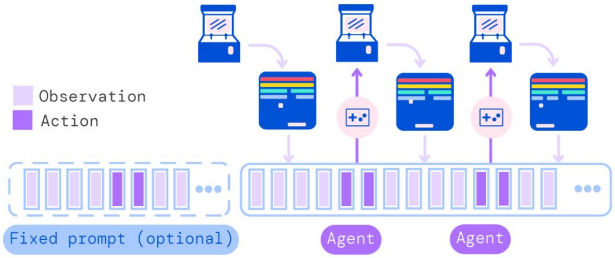
Topic - planning, actuation

- Given commands and observations of the environment (possibly augmented by an internal world-model), produce a control output
- Classic approaches – find feasible / optimal trajectories in configuration space
- ML-based approaches
  - End-to-end learned policies
  - Motion primitives
- Planning vs control
  - Planner – construct sequence of commands or setpoints
  - Controller – physically drive the execution of commands, track setpoints through feedback

[1]



[2]



[1] Ahn, Michael, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alexander Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil Jayant Joshi, Ryan C. Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego M Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, F. Xia, Ted Xiao, Peng Xu, Sichun Xu and Mengyuan Yan. "Do As I Can, Not As I Say: Grounding Language in Robotic Affordances." Conference on Robot Learning (2022).

[2] Reed, Scott, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley D. Edwards, Nicolas Manfred Otto Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar and Nando de Freitas. "A Generalist Agent." *ArXiv* abs/2205.06175 (2022): n. pag.

Personal background

- BEng Mechatronics Engineering – RTU, 2020
- MSc Computer Science – LU, 2022
- Research Assistant at EDI – 2022-
  - Robotics and Machine Perception Laboratory, Robotics group
- Areas of relative strength
  - Electrical and mechanical engineering
  - Software development (primarily in Python, but not constrained by languages)
  - Practical skills in machine learning
- Weaknesses, things to work on
  - Lack of mathematical background

Plan - year 1 (initial)

- Mandatory courses – 8 credits
- Theoretical courses – 2 to 10 credits
  - Basic mathematics (analysis, algebra)
- Research activities – 30 to 38 credits
  - 6DOF pose estimation from RGBD data for industrial robotics – validation, performance improvements, porting to embedded hardware, algorithmic improvements as part of finalizing AI4DI
  - Start work on perception/mapping/control systems for autonomous mobile robots as part of RoLISe / EdgeAI
  - Publish at least one journal article and/or conference abstract
  - Start drafting theoretical part of PhD thesis

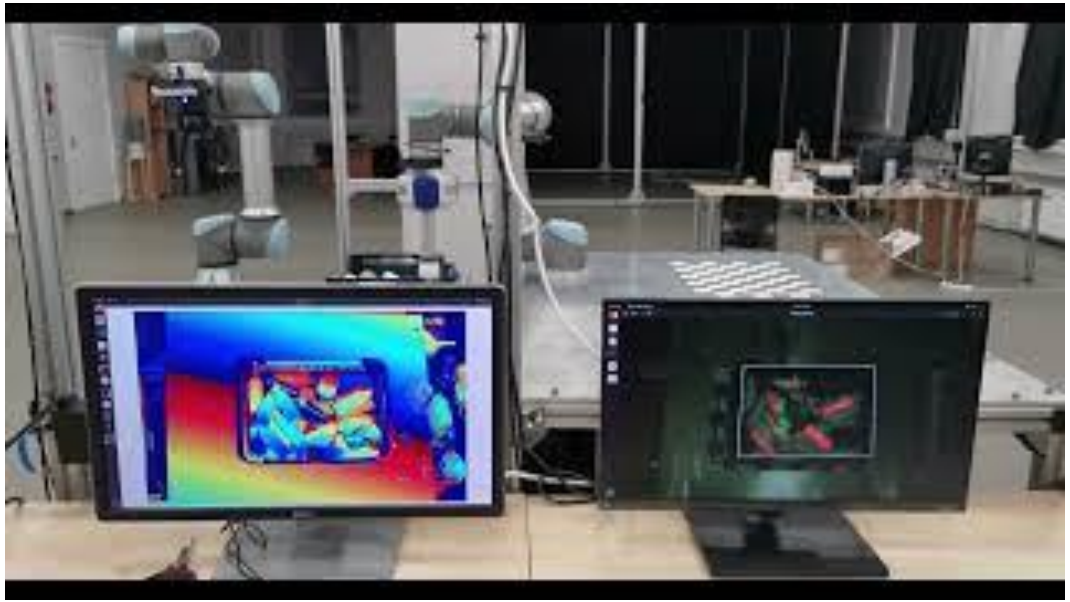
Plan - year 1 (revised)

- Mandatory courses – 8 credits
- Theoretical courses – pushed back to year 2 (took too long to negotiate)
- Research activities – 40+ credits
  - 6DOF pose estimation from RGBD data – validation, porting to FPGA, switch to direct inference of planar projection poses
    - paper at EDI conference
  - Semantic mapping overview article, technical specification – RoLISe + EdgeAI
  - Start working on technical implementation – RoLISe
  - Conference poster (drone detector)
  - Start drafting theoretical part of PhD thesis
- Educational activities – 2-4 credits (?)
  - Guest lectures – robotics seminar, image processing, deep learning(?)
  - Internship supervision
    - 1 x BSc mechanical engineering (done)
    - 1 x BSc CS (upcoming this summer)
  - Advising MSc theses in Intellectual Robotics (RTU)
    - NLP-conditioned control
    - Motion primitives for mobile manipulators

6DOF pose estimation from RGBD, trained on synthetic data

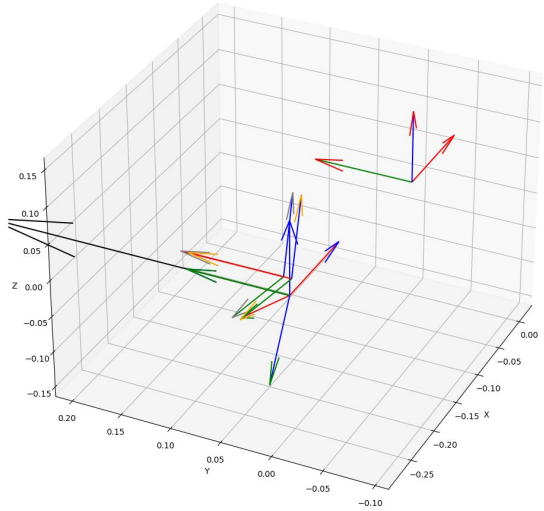
- Perception part of the robot cognition model
- 6DOF – six degrees of freedom (translation X rotation)
- RGDB – red-green-blue-depth imagery
- Project – AI4DI, “Artificial Intelligence for Digitising Industry”
- My involvement – starting in July 2022; project end – December 2022
  - “fire brigade” tasks
- Original approach
  - Point clouds from depth images
  - MaskRCNN – segmentation masks from images
  - Fragile, hand-crafted algorithm for extracting object direction from poses (only really worked for bottles)
  - 6DOF pose by backprojection

6DOF pose estimation from RGBD, trained on synthetic data



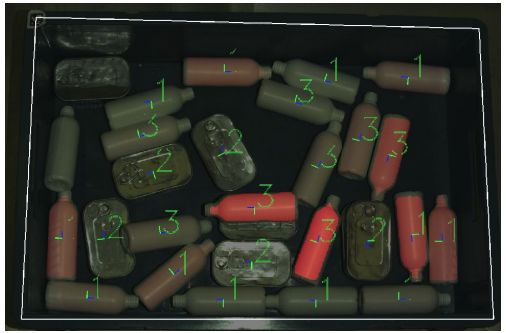
AI4DI – validation

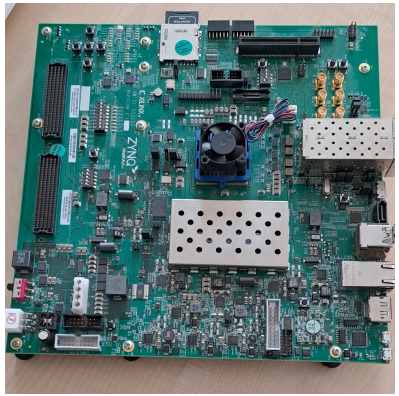
- Models trained on synthetic data
- Use synthetic ground truth to quantify model performance
- Validation pipeline
  - Generate evaluation data sets
  - Compute “optimal” grasps
    - not trivial when rotational symmetries involved!
  - Compare with system predictions
  - Devise and compute metrics of interest



AI4DI – refactoring, embedded execution

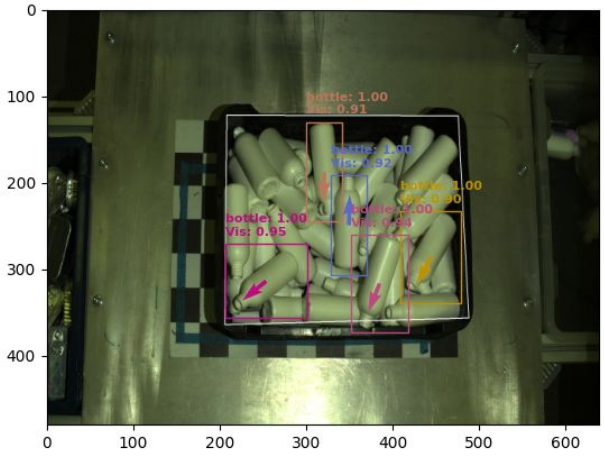
- Rewrite entire inference pipeline
  - Fragile filesystem-based “protocol” -> REST API
  - ~300x speedup of direction estimation algorithm through reimplementation
  - Allow pose estimation for multiple objects
  - Bring system to workable state by eliminating a long list of implementation flaws and overlooked edge cases
- Switch out inference back-end for embedded execution
  - MaskRCNN – infeasible on FPGA accelerator
  - Train SOLO model, implement inference server in C++





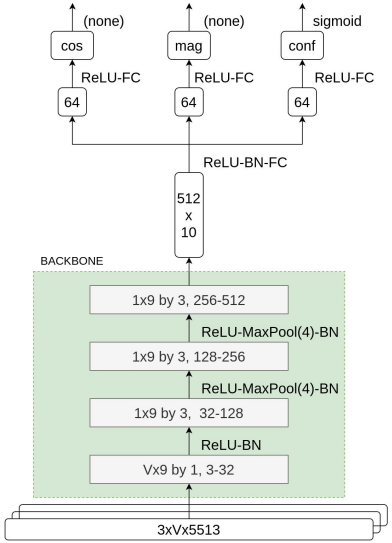
AI4DI – switch to direct prediction of planar poses

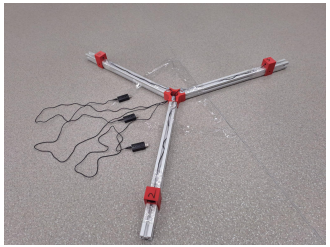
- Handcrafted direction finding algorithm – unreliable, reliant on features found in bottles
- After project end, system slated for use in Digital Innovation Hub
- Retrofit inference pipeline to directly predict object directions using a neural network
  - Base model – DETR
  - Add inference heads – direction, visibility
  - Sort outputs by visibility, backproject directly inferred directions and positions derived from bounding boxes
- Upcoming conference publication – June, EDI conference
- As far as I'm aware, this is a novel application of a neural network
- Idea, technical leadership - me; implementation – Andris Lapiņš

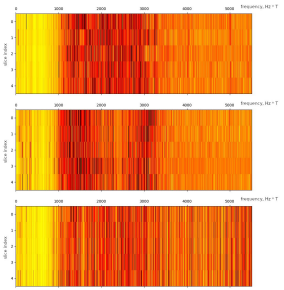


Audio drone detector

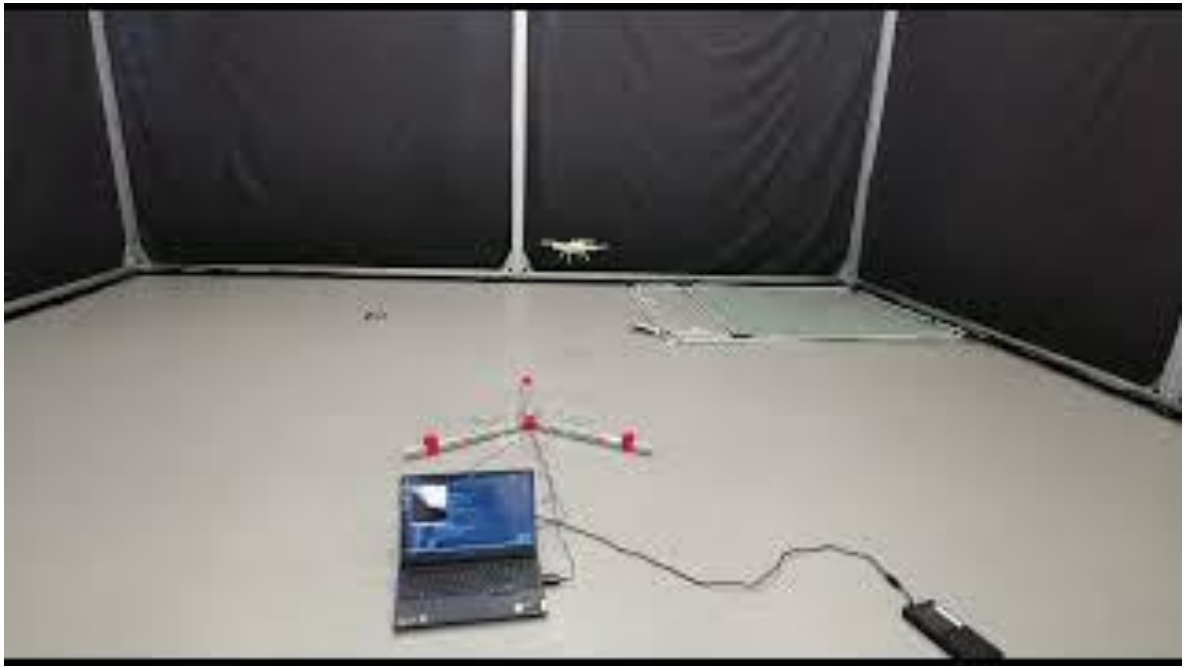
- Quick side project (~4 weeks of work)
- Original idea – at RTU defense makeathon 2022
- Sensors – array of 3 microphones
- Model – 2d-to-1d CNN
- Position ground truth data – motion tracking equipment
- Accepted as poster at DCOSS-IoT 2023 in Pafos, Cyprus







Drone detector



RoLISe / Edge AI 4.1

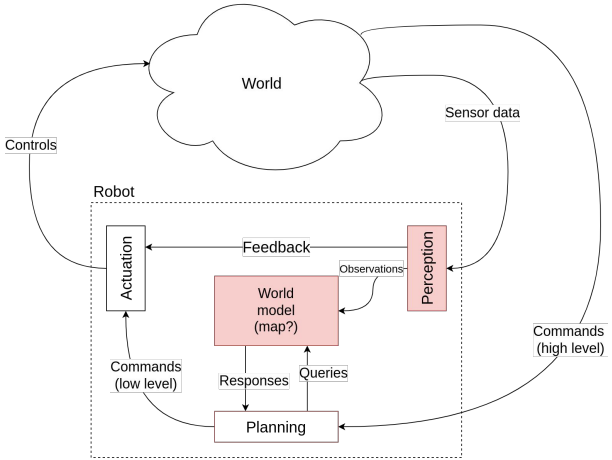
- Autonomous mobile robots, mobile manipulation
- RoLISe – part of VPP Fotonika / MOTE
  - Robotics, Internet of Things, Sensors
  - term – 2 years (end in Nov/Dec 2024)
  - focus of task 4.1 – perception, mapping framework
- Edge AI – EU-funded project
  - term – 3 years (end in Dec 2025)
  - focus of task 4.1 – NLP, planning, partner perception blocks, actuation
- Demonstrators
  - RoLISe – sensor suite and mapping framework, outdoor focus
  - Edge AI – autonomous agent using the above, potentially integrated with drone “scout”
- Role
  - RoLISe 4.1 - tech lead
  - Edge AI 4.1 - leadership of certain aspects

[1]

[1] https://robotnik.eu/wp-content/uploads/2021/02/home-gallery-mobile-manipulator-100121.jpg

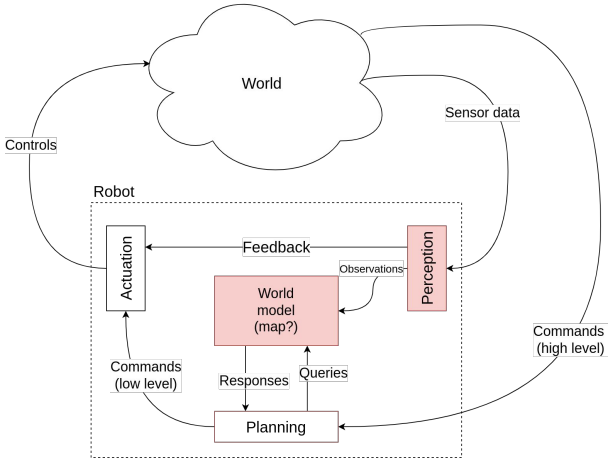
Perception and mapping framework (RoLISe)

- Input modalities
  - LiDAR
  - Panoptic open-set segmentation (LUMII partners)
  - Terrain segmentation (possibly radar-augmented?)
- Representations
  - Metric
  - Topological
  - Semantic
- Inspirations
  - Scene graph work from MIT
  - GNN-based approaches from TUM



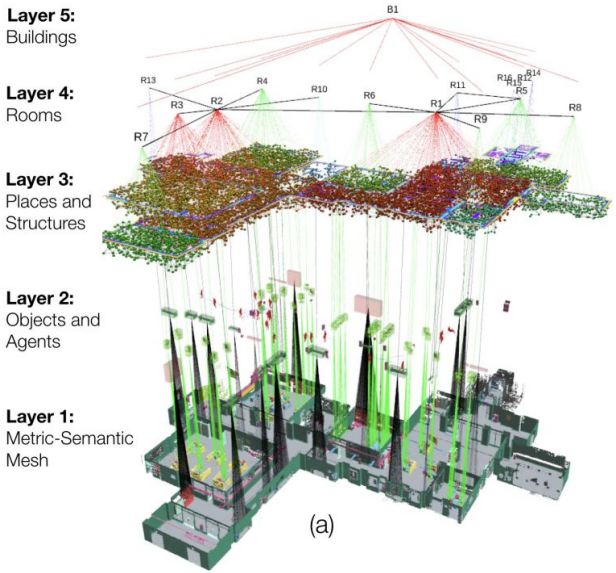
Perception and mapping framework (RoLISe)

- Current work – overview article, semantic mapping
  - Low level concepts – SLAM, segmentation
  - Representations – dense (points), hierarchical (graphs), implicit (radiance fields)
  - Applications – indoor, outdoor, specialized (e.g., autonomous driving)
  - Done – this month
- Current work – technical specification
  - KPIs
  - Requirements
  - Preliminary architecture
  - Done – June

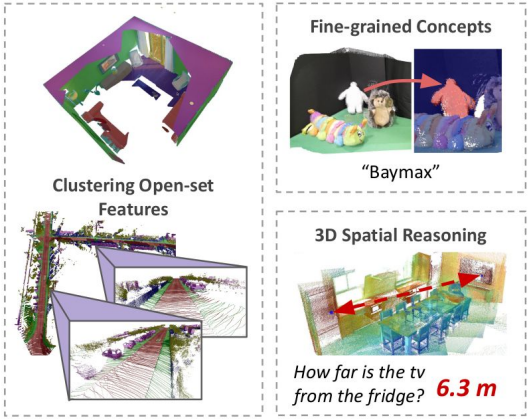


Perception and mapping framework (RoLISe)

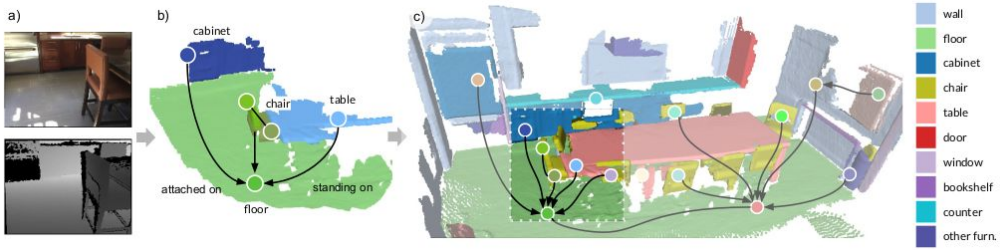
[1]



[2]



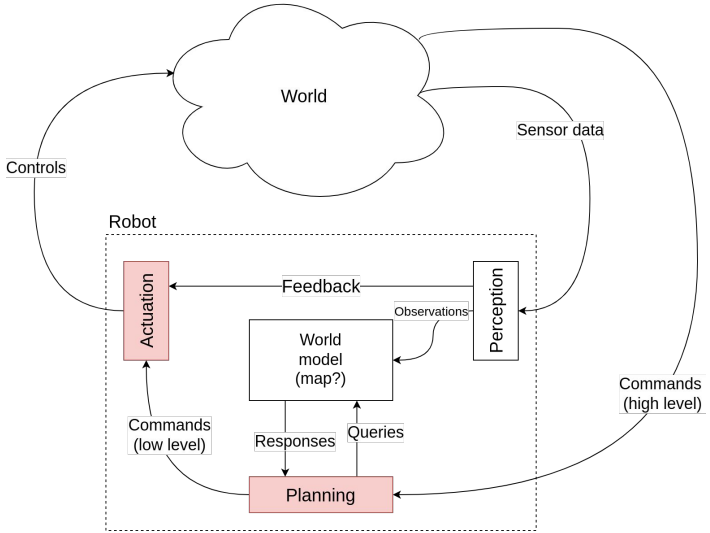
[3]



[1] Rosinol, Antoni, Arjun Gupta, Marcus Abate, J. Shi and Luca Carlone. "3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans." ArXiv abs/2002.06289 (2020): n. pag.
[2] Jatavallabhula, Krishna Murthy, Ali Kuwajerwala, Qiao Gu, Mohd Omama, Tao Chen, Shuang Li, Ganesh Iyer, Soroush Saryazdi, Nikhil Varma Keetha, Ayush Tewari, Joshua B. Tenenbaum, Celso M. de Melo, M. Krishna, Liam Paull, Florian Shkurti and Antonio Torralba. "ConceptFusion: Open-set Multimodal 3D Mapping." ArXiv abs/2302.07241 (2023): n. pag.
[3] Wu, Shun-cheng, Johanna Wald, Keisuke Tateno, Nassir Navab and Federico Tombari. "SceneGraphFusion: Incremental 3D Scene Graph Prediction from RGB-D Sequences." 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 7511-7521.

NLP-conditioned planning, mobile manipulator actuation, collaborative robotics (Edge AI)

- Advisory capacity – master's thesis on NLP-conditioned planning
- Advisory capacity – master's thesis on motion primitives for mobile robot control (tentative)
- Potential BSc thesis – integration with drone "scout" through active perception
- Technical leadership – oversee integration of planning, actuation solutions with mapping framework



Thank you for your attention!