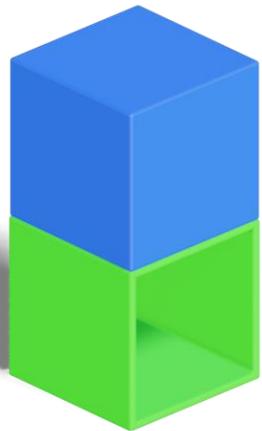AVHRC 2020

Combining Reinforcement Learning with Supervised Deep Learning for Neural Active Scene Understanding

Dano Roost, Ralph Meier, Giovanni Toffetti Carughi,
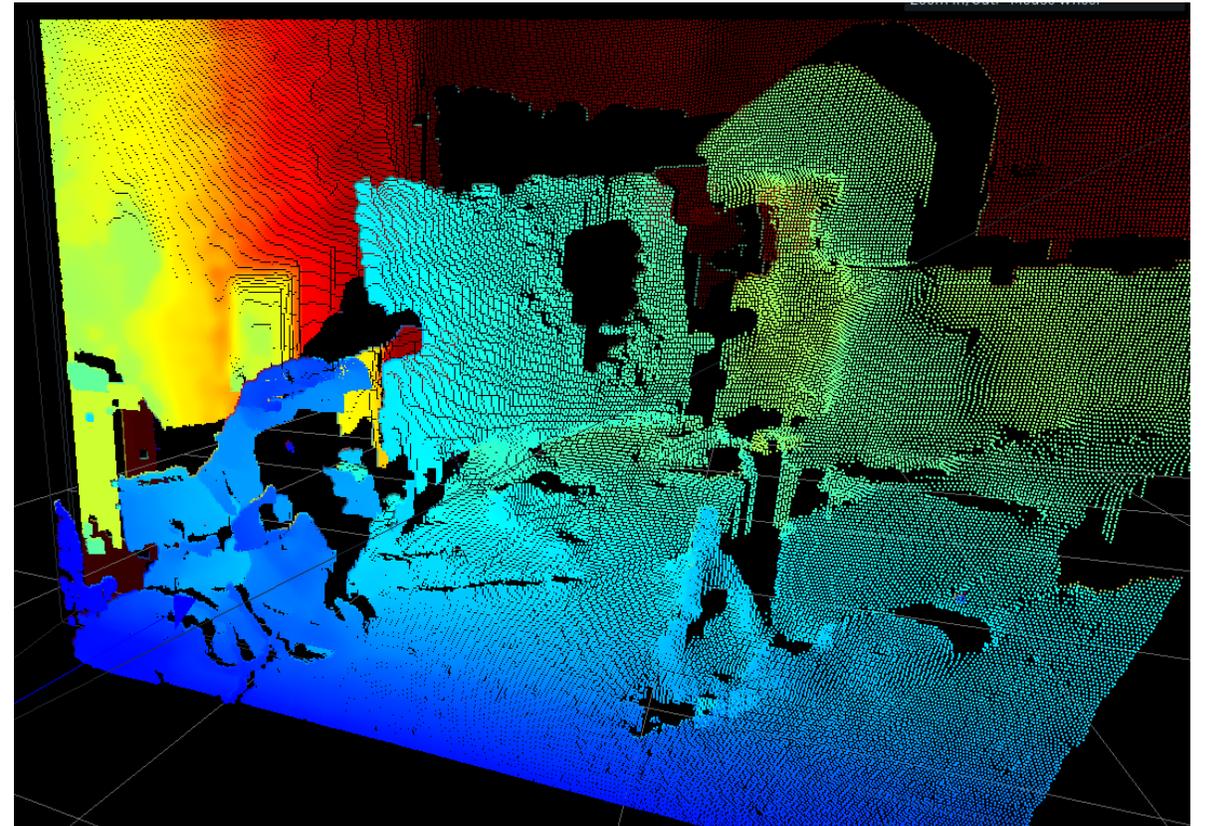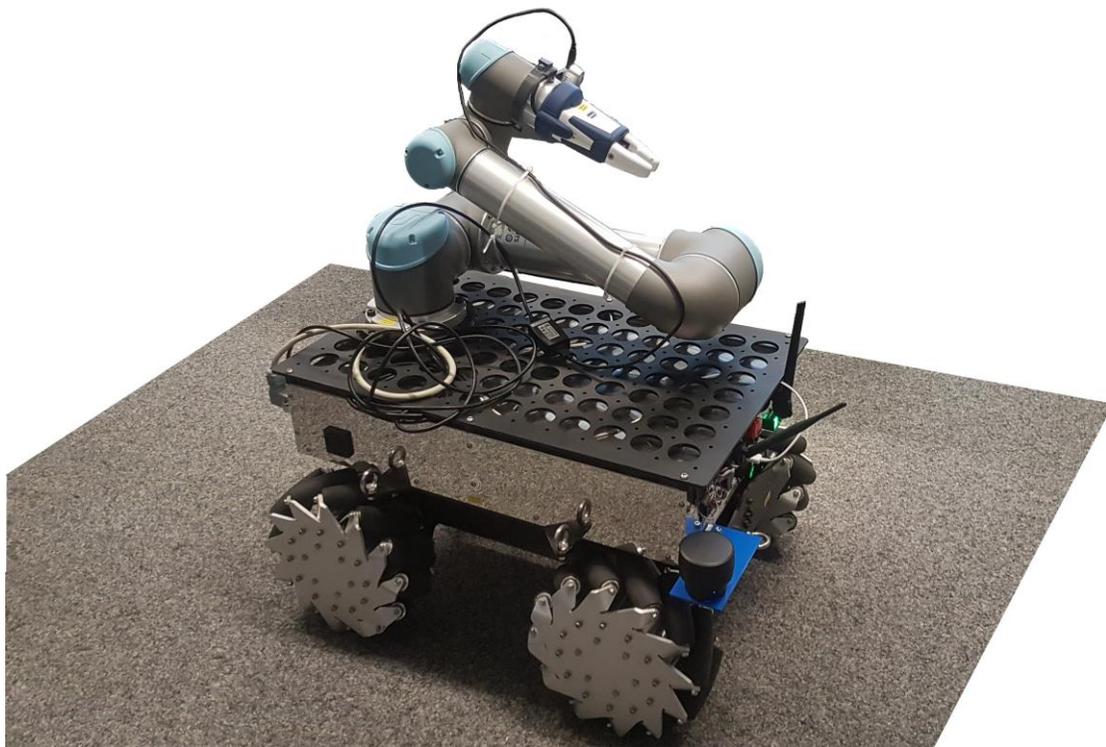and Thilo Stadelmann

Zurich University
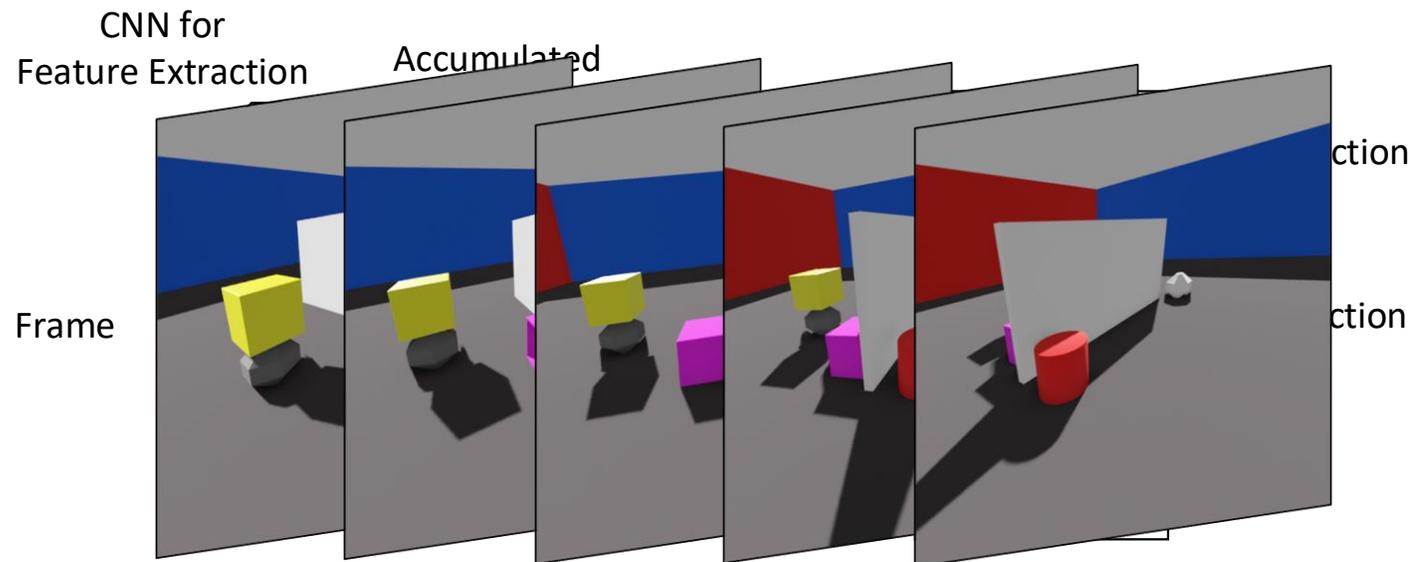of Applied Sciences

zh
aw

School of
Engineering

Traditional mobile robotics: 3D Mapping

→ Handcrafted pipeline to identify, locate & interact

Idea: From Handcrafted to Learned Capabilities

- Use sequences of RGB-D frames
- Accumulate information in **neural network** instead of **point cloud**
- Different outputs → Force network to capture a lot of information

Possible Questions for the System

- Color of object at given position?

- Relationship between two objects?

- List all present objects in the scene with another LSTM-layer
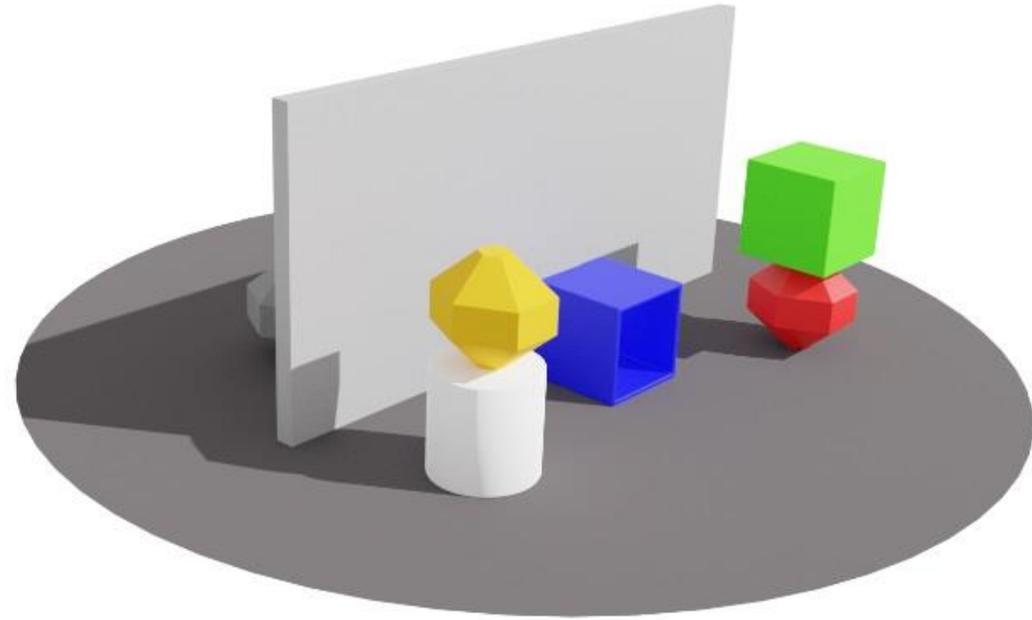
→ Each output has own loss

→ Idea: Use **total loss reduction** as **reward** signal for reinforcement learning based active vision
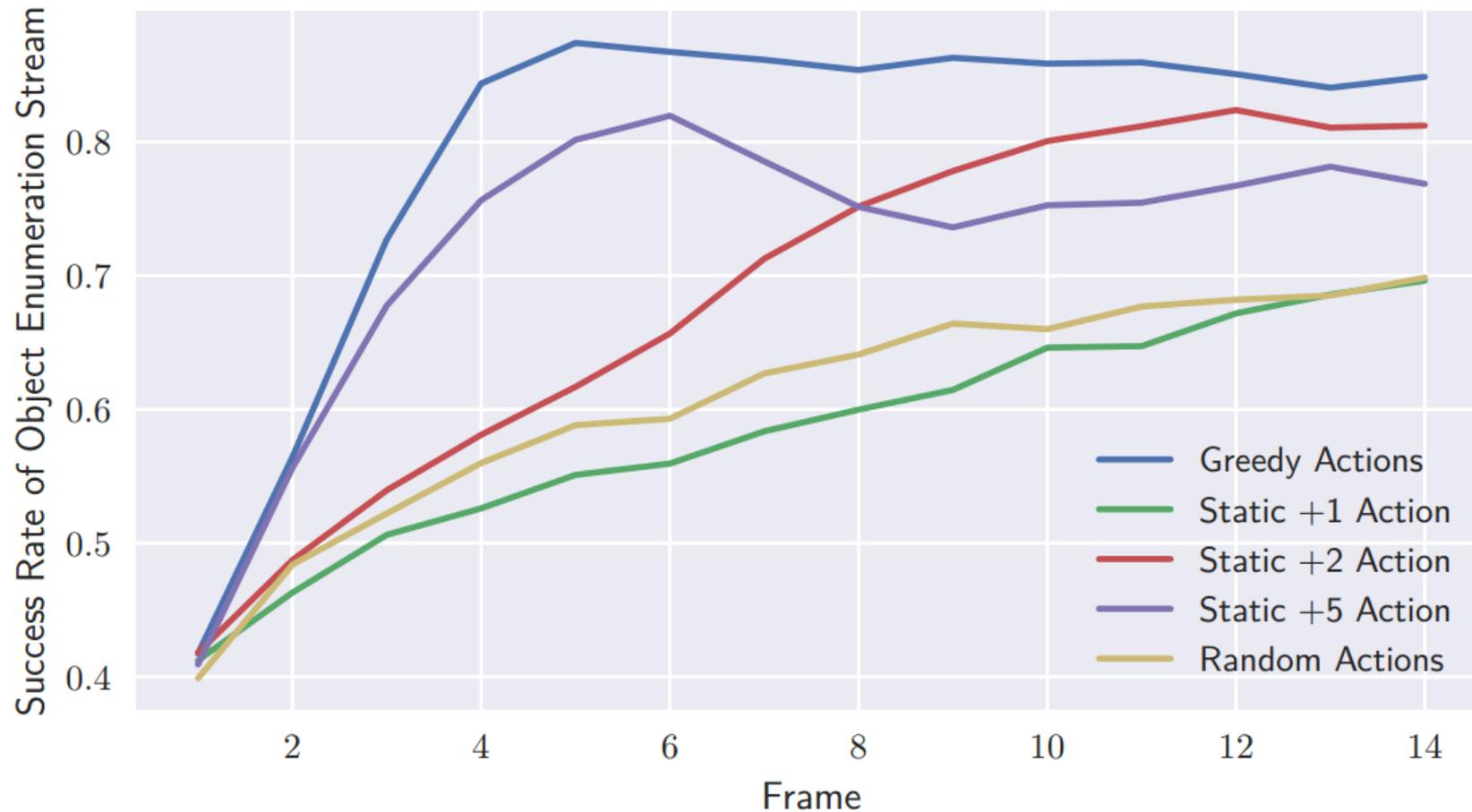
Control Camera with RL

- **Discrete camera positions**

- Actions: 5 left, 2 left, 1 left, stay here, 1 right, 2 right, 5 right

- Use **Q-Learning** to learn score for each action, given the accumulated information
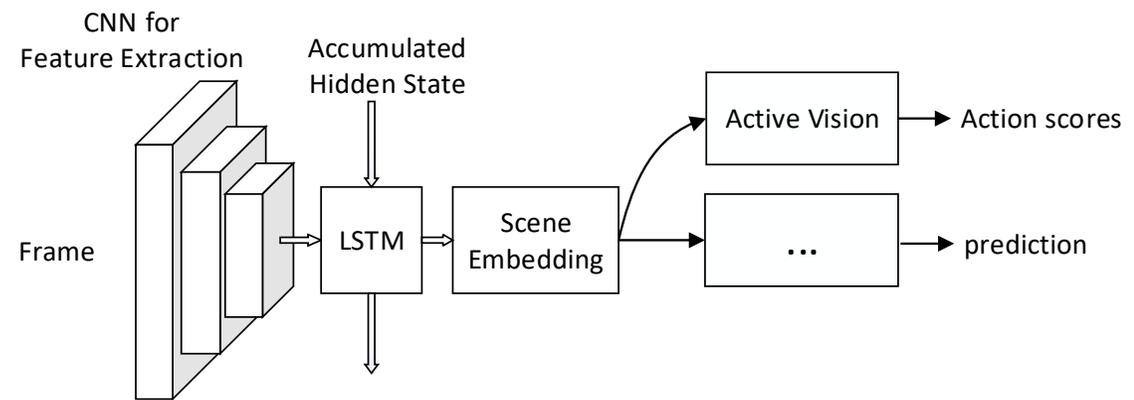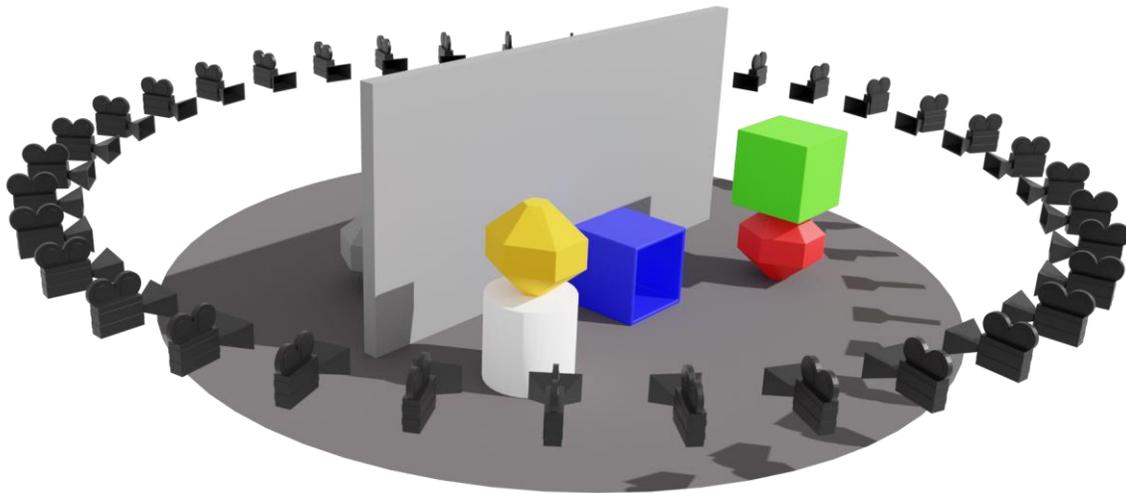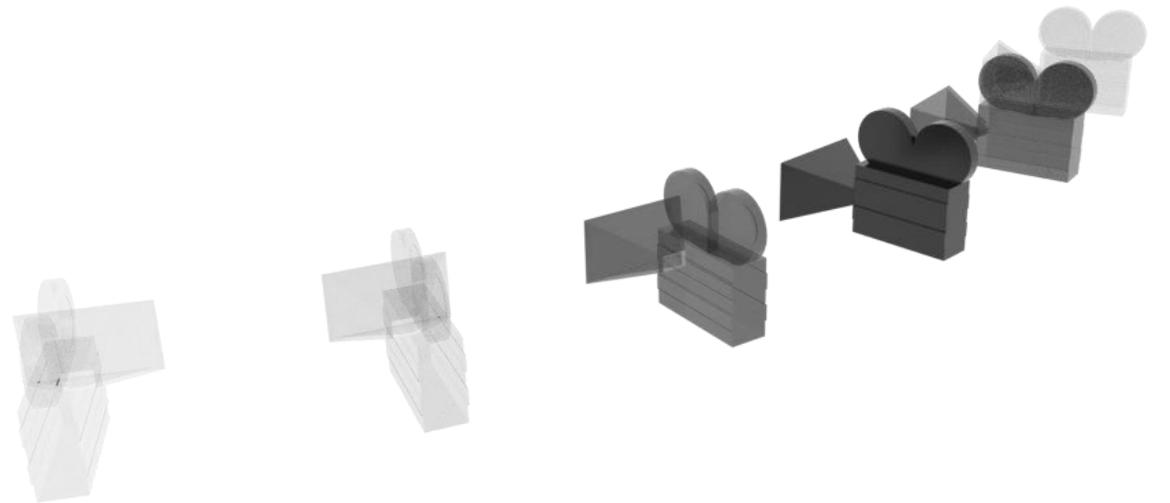
Impact of active vision

For benchmarking: We evaluate stream that should output all objects

Summary

- 40,000 simple synthetic scenes for training
- Easier to train additional outputs – only 200 scenes needed
- System is capable of remembering relative object positions, **even when camera moves and objects are occluded**

Next steps/further research

- Most important: Check, whether this works with real world data

- Maybe: use Transformer instead of LSTM

- Switch to Continous Camera Control

Thanks for joining!

Check out our work on Github:
https://github.com/Danoishere/ba-brain-net