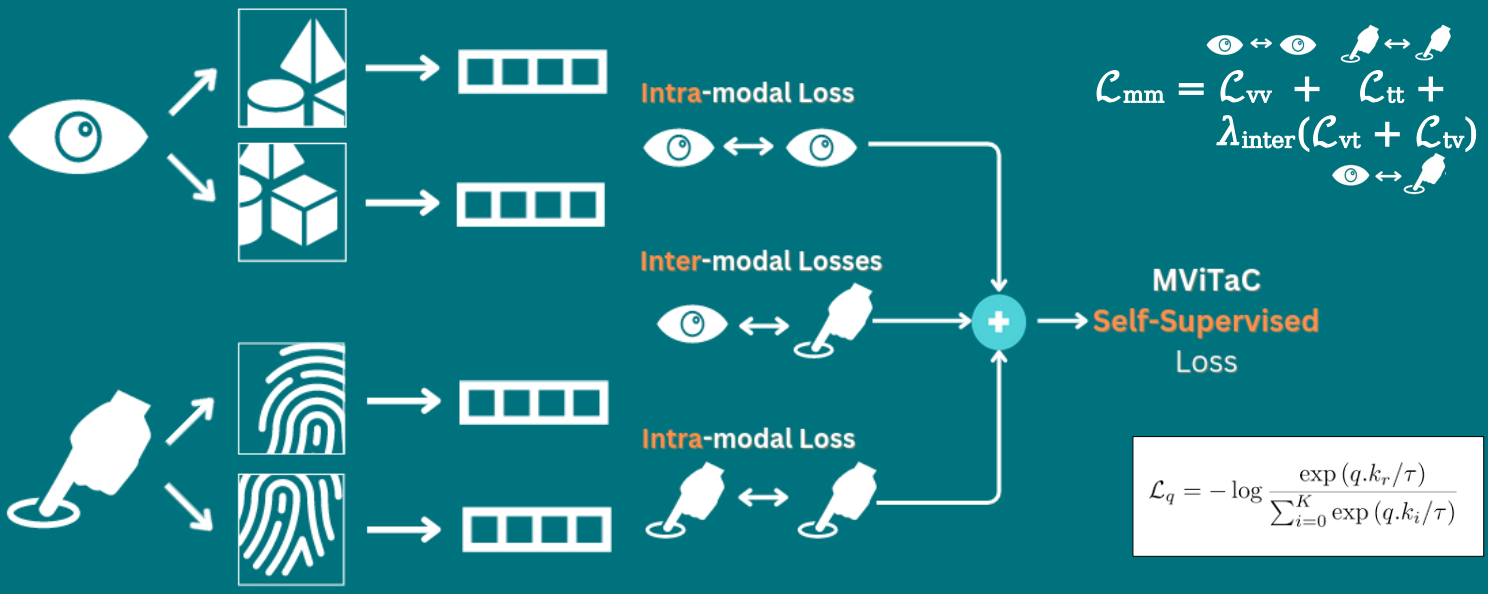
Table of contents

Multimodal Visual-Tactile Representation Learning through Self-Supervised Contrastive Pre-Training

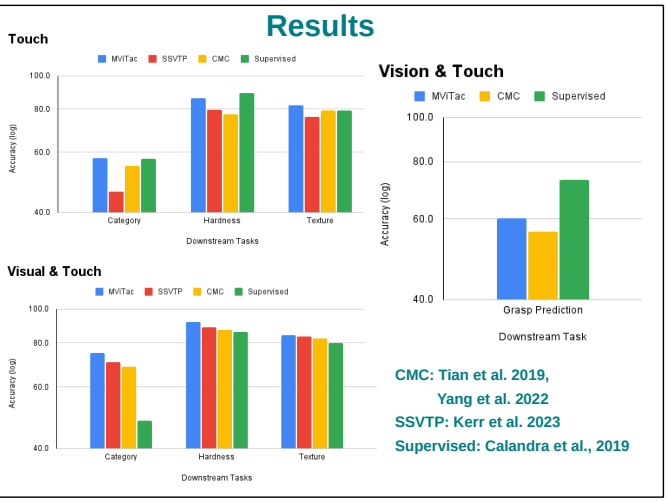
Vedant Dave* Fotios Lygerakis* Elmar Rückert *equal contribution

Chair of Cyber-Physical-System, Montanuniversität Leoben, Austria

Main Idea: Maximise the agreement of the representatons of similar samples from different modalities.

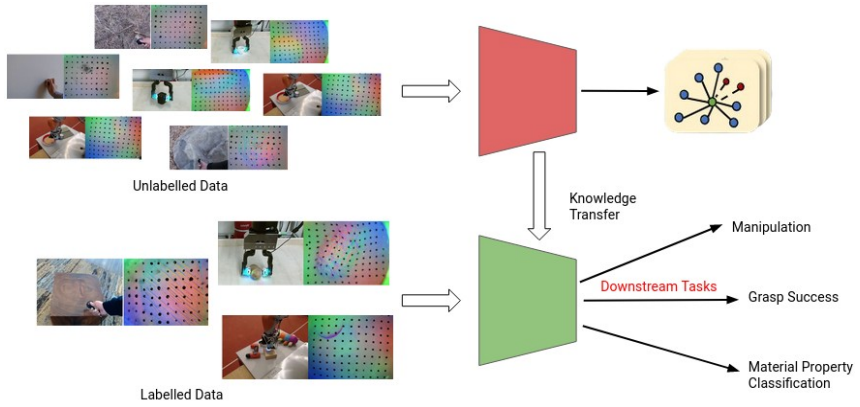


Results

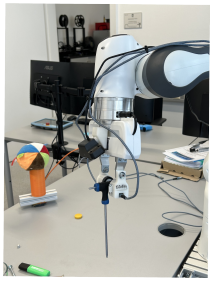


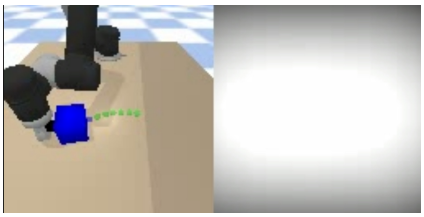
CMC: Tian et al. 2019, Yang et al. 2022
SSVTP: Kerr et al. 2023
Supervised: Calandra et al., 2019

Why is self-supervised learning important?



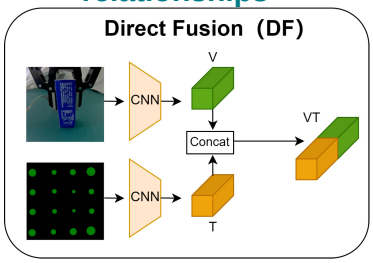
Future Work



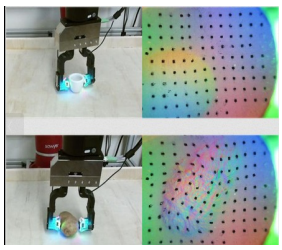


Lygerakis Fotios, Vedant Dave, and Elmar Rueckert. "M2CURL: Sample-Efficient Multimodal Reinforcement Learning via Self-Supervised Representation Learning for Robotic Manipulation." 2024 21st International Conference on Ubiquitous Robots (UR), New York, USA, 2024.

Unexplored modality relationships

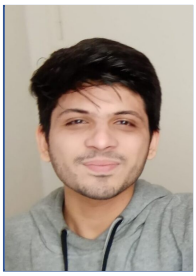


Data collection is costly!



Calandra et al.

Yang et al.



PhD Candidate
Vedant Dave, M.Sc.
Chair of Cyber-Physical-Systems
Vedant.dave@unileoben.ac.at

Vedant Dave*, Fotios Lygerakis*, Elmar Rueckert., "Multimodal Visual-Tactile Representation Learning through Self-Supervised Contrastive Pre-Training." 2024 IEEE International Conference on Robotics and Automation (ICRA), Tokyo, Japan, 2024.

CPS UR 2024 NEW YORK

ICRA2024 YOKOHAMA | JAPAN