

MA/SA: Einsatz von Vision-Language-Action Models in der Soft-Robotik (m/w/d)

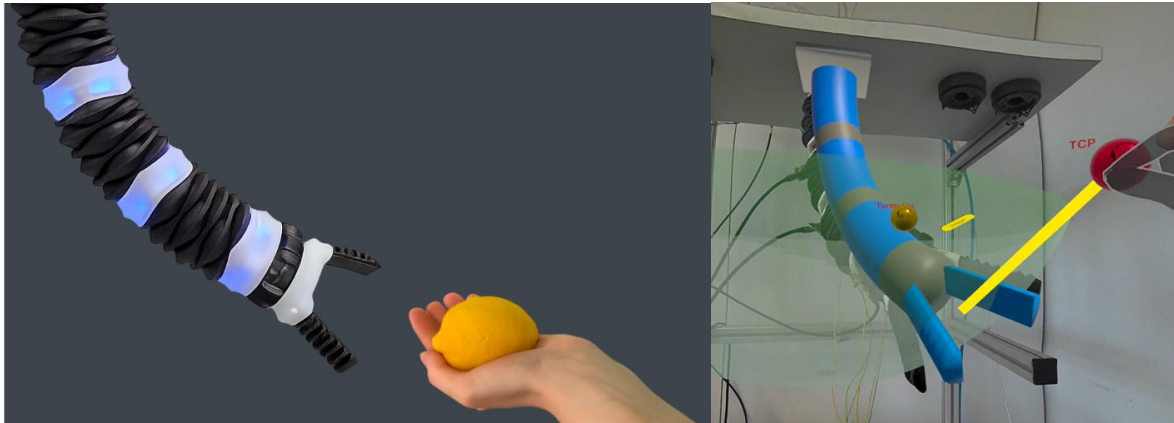


Abb. 1: Soft-robotischer Manipulator der 3. Gen. (links), POV einer menschlichen Interaktion in AR (rechts).

Vision-Language-Action (VLA) Modelle vereinen Wahrnehmung, Sprachverständnis und Handlung in einem gemeinsamen Framework und ermöglichen eine flexible Steuerung von Robotern anhand natürlicher Anweisungen. Bisher wurden diese Ansätze hauptsächlich auf starren Manipulatoren untersucht, während ihr Einsatz auf nachgiebigen (soft) Kontinuumsrobotern weitgehend unerforscht ist. Gerade diese bieten jedoch Vorteile wie intrinsische Sicherheit und Anpassungsfähigkeit in unstrukturierten Umgebungen [1], [2].

Ziel dieser Arbeit ist die Entwicklung und Untersuchung von Methoden zur Steuerung eines pneumatisch aktuierten, nachgiebigen Kontinuumsmanipulators mithilfe von VLA-Modellen. Ein Fokus liegt auf der Übertragung solcher Modelle auf Soft-Roboter („Embodiment Gap“) durch demonstrationsbasiertes Lernen.

Die Augmented-Reality (AR)-Umgebung dient dabei zur intuitiven Aufnahme von Demonstrationsdaten (Imitation Learning) mittels Meta Quest 3, welche zur Anpassung und zum Finetuning der Modelle verwendet werden.

Die Arbeit erfolgt in Zusammenarbeit mit dem Visualisierungsinstitut der Universität Stuttgart (VISUS).

Aufgaben

- Aufnahme von Demonstrationsdaten mit AR (Meta Quest 3).
- Anpassung und Finetuning von VLA-Modellen für Soft-Roboter.
- Validierung am realen System.

Voraussetzungen

In erster Linie sollst du Interesse für das Thema mitbringen. Des Weiteren legen wir Wert auf eine selbständige und strukturierte Arbeitsweise sowie frequente Kommunikation untereinander,

damit die Ziele erfolgreich erreicht werden können. Gute Kenntnisse in der Technischen Mechanik, Regelungstechnik und AR als auch in Python oder MATLAB sind vorteilhaft, können aber auch selbstständig erarbeitet werden. Bei Rückfragen kannst du dich gerne bei uns melden.

Literatur:

- [1] Kim et al., "OpenVLA: An Open-Source Vision-Language-Action Model", arxiv, 2024.
- [2] Su et al., "Bridging Embodiment Gaps: Deploying Vision-Language-Action Models on Soft Robots", arxiv, 2025.

MT/ST: Deploying Vision-Language-Action Models on Soft Robots (m/f/d)

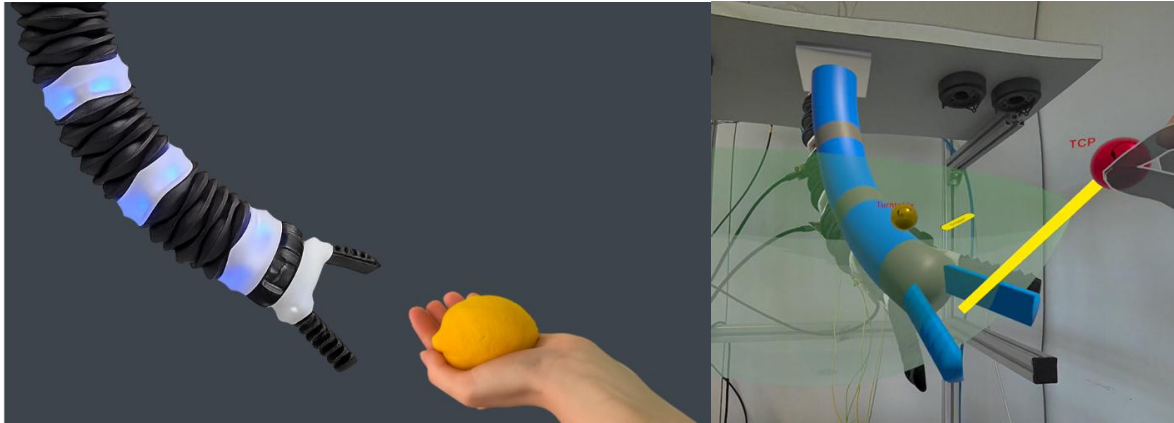


Figure 1: Soft manipulator of the 3rd generation (left) and a POV of human interaction in AR (right).

Vision-Language-Action (VLA) models integrate perception, language understanding, and action into a unified framework, enabling robots to execute tasks based on natural language instructions. While these approaches have mainly been studied on rigid manipulators, their application to compliant (soft) continuum robots remains largely unexplored. Soft robots offer advantages such as intrinsic safety and adaptability in unstructured environments [1], [2].

The goal of this thesis is to develop and investigate methods for controlling a pneumatically actuated soft continuum manipulator using VLA models. A key focus is on addressing the embodiment gap between rigid and soft robots through demonstration-based learning.

Augmented Reality (AR) is used as an intuitive interface for collecting demonstration data (imitation learning) via the Meta Quest 3, which is then used to adapt and fine-tune the models.

The thesis will be conducted in collaboration with the Visualization Institute of the University of Stuttgart (VISUS).

Tasks

- Collect demonstration data using AR (Meta Quest 3).
- Adapt and fine-tune VLA models for soft robots.
- Validation on the real system.

Prerequisites

First and foremost, you should be interested in the topic. Furthermore, we attach importance to an independent and structured way of working as well as frequent communication among each other so that the goals can be successfully achieved. Good knowledge of technical mechanics, control engineering and AR as well as MATLAB or python is advantageous, but can also be acquired independently. If you have any questions, please feel free to contact us.

[Samuel Pilch](mailto:samuel.pilch@isys.uni-stuttgart.de) | samuel.pilch@isys.uni-stuttgart.de

[Michael Pabst](mailto:michael.pabst@visus.uni-stuttgart.de) | michael.pabst@visus.uni-stuttgart.de

Literature:

- [1] Kim et al., "OpenVLA: An Open-Source Vision-Language-Action Model", arxiv, 2024.
- [2] Su et al., "Bridging Embodiment Gaps: Deploying Vision-Language-Action Models on Soft Robots", arxiv, 2025.