

Masterarbeit

Thema: Skelettbasierte Aktionserkennung mithilfe von Word Embeddings

Beschreibung:

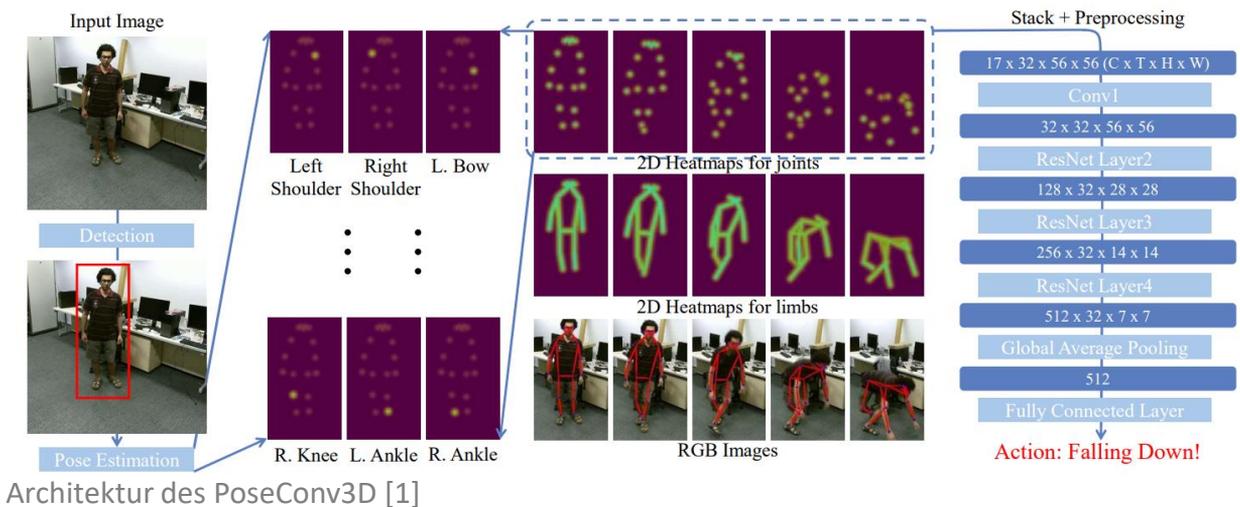
In der derzeitigen Forschung ist die Kollaboration von Menschen und Roboter ein immer wichtiger werdendes Aufgabenfeld. Der variantenreiche Montageprozess ist ein konkretes Anwendungsszenario in dem Projekt „Engineering for Smart Manufacturing“ (E4SM). Dabei soll ein Roboter den Menschen unterstützen, indem er Situationen erkennt und Folgeschritte prädiziert. Hierfür ist eine Aktionsklassifizierung [1] hilfreich, bei welcher für getrimmte Skelettsequenzen eine Klasse zugeordnet werden soll.

In typischen skelettbasierten Aktionserkennungsverfahren treten oft verschiedene Probleme auf. Zum einen skaliert das Verfahren schlecht mit unterschiedlichen Skelettarten wie Ganzkörperskeletten oder Handskeletten. Diese Verfahren verbrauchen mehr Speicher, werden größer und langsamer. Auch die Hinzufügung von Objekten wie einem Gelenkpunkt als Schwerpunkt verursacht dasselbe Problem wie bei mehreren Skeletten.

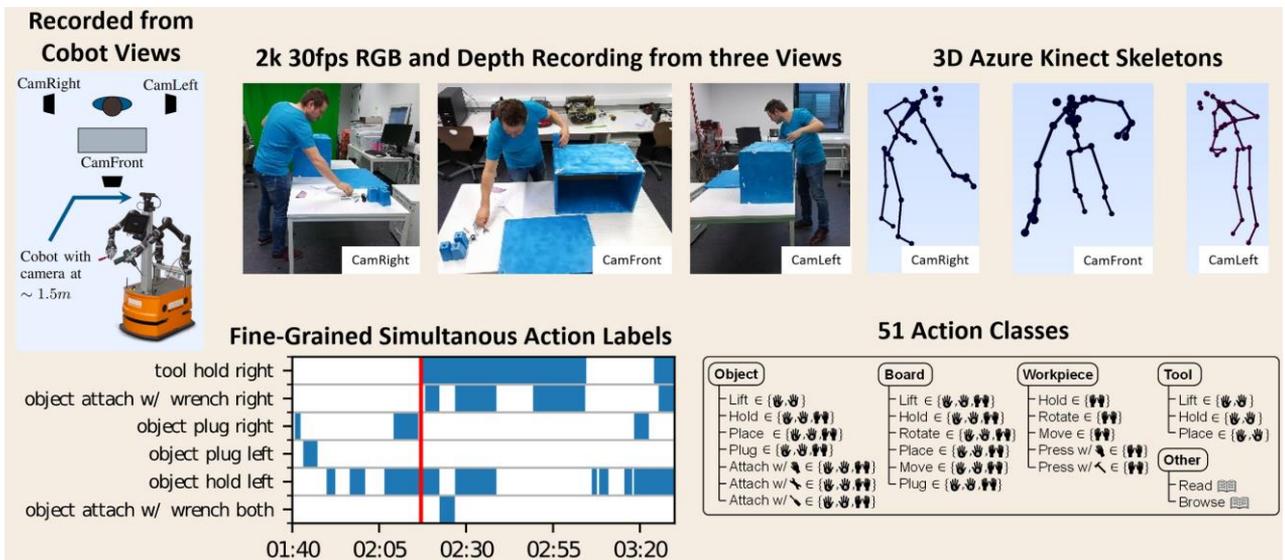
Ein weiteres Problem besteht darin, dass dem Netzwerk typischerweise nicht mitgeteilt wird, wie die verschiedenen Skelette in Relation zueinander stehen. Das Netzwerk muss diese Informationen selbst erarbeiten. Darüber hinaus sind diese Verfahren oft auf eine bestimmte Skelettprädiktion eingelernt, wie beispielsweise das Kinect-Skelett oder das COCO-Skelett.

In dieser Masterarbeit wird daher eine Lösung für diese Probleme erarbeitet. Die Namen der Skelettknoten oder die Namen der verwendeten Objekte sollen im Vorfeld mittels eines Word Embeddings eingebettet werden. Auf diese Weise erhält das Aktionsklassifizierungsnetzwerk zusätzliche hilfreiche Informationen, indem ihm die Word Embeddings als Eingabe gegeben werden.

Hierfür sollen Verfahren ähnlich des PoseConv3D [1] oder des Video Swin Transformer [2] erweitert werden.



Architektur des PoseConv3D [1]



ATTACH Datensatz [3]

Aufgabenstellung:

- Aufarbeitung des State of the Art zu folgenden Themenfeldern:
 - Word Embeddings
 - Dimensionsreduktion von Word Embeddings
 - Alternative Skelettrepräsentationen für die Klassifizierung
- Anpassen geeigneter skelettbasierter Aktionsklassifizierer (z.B. [1])
- Systematische Experimente und Evaluation auf dem ATTACH Datensatz [3] und IKEA ASM Datensatz [4]
- Untersuchung zu der Kombination von Skeletten mit Objekten mithilfe von Word Embeddings
- Untersuchung zu der Kombination von unterschiedlichen Skeletten mithilfe von Word Embeddings
- Ausarbeitung von Präsentationen für den Eröffnungs-, und Abschlussvortrag
- Anfertigen der Masterarbeit entsprechend der Vorgaben des FG NI&KR

Erforderliche Vorkenntnisse:

- Vorlesung „Deep Learning for Computer Vision“ besucht oder Erfahrungen im Bereich Deep Learning

Für weitere Recherchen zu verwendende Quellen:

- [1] Duan et al.: Revisiting Skeleton-Based Action Recognition, CVPR 2022
- [2] Liu et al.: Video Swin Transformer, CVPR 2022
- [3] Aganian et al.: ATTACH Dataset: Annotated Two-Handed Assembly Actions for Human Action Understanding, ICRA 2023
- [4] Ben-Shabat et al.: The IKEA ASM Dataset: Understanding People Assembling Furniture through Actions, Objects and Pose, WACV 2021
- Google Scholar scholar.google.com
- Proceedings der relevanten Konferenzen (CVPR, ICCV, ECCV, BMVC, AVSS, ICPR, ICIP, IROS, ICRA, ...)

Betreuer: Dustin Aganian, M.Sc. (dustin.aganian@tu-ilmenau.de)
Betr. Hochschullehrer: Prof. Dr. H.M. Groß
Bearbeiter: Erik Franze