visuAAL

Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living

Lifelogging and Behaviour Modelling

ESR 10 – 2nd PhD seminar

Stockholm, Sweden
20 April, 2023

Wiktor Mucha
Computer Vision Lab
TU Wien

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091".

Universitat d'Alacant
Universidad de Alicante
Project Coordinator

RWTH AACHEN UNIVERSITY

Stockholm University

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

TU WIEN

Presentation agenda

1. Introduction
2. Motivation and applications
3. Research questions
4. Work until 1st PhD seminar
5. Work until 2nd PhD seminar
6. Present work
7. Plans and PhD schedule

Introduction

- Lifelogging → a **technology** that uses **wearable sensors** to gather and process **data** from the **daily lives** of an individual
- Using a wearable camera can **illustrate** in detail which **activities** the person wearing the camera has done during the **day**
- This **work** focuses on the possibilities of **egocentric** visual data processing for **health** and **lifestyle** improvement
- **Egocentric** → placing a camera on a **human body** giving a **view** from this **person's perspective**

User wearing a lifelogging device[1]

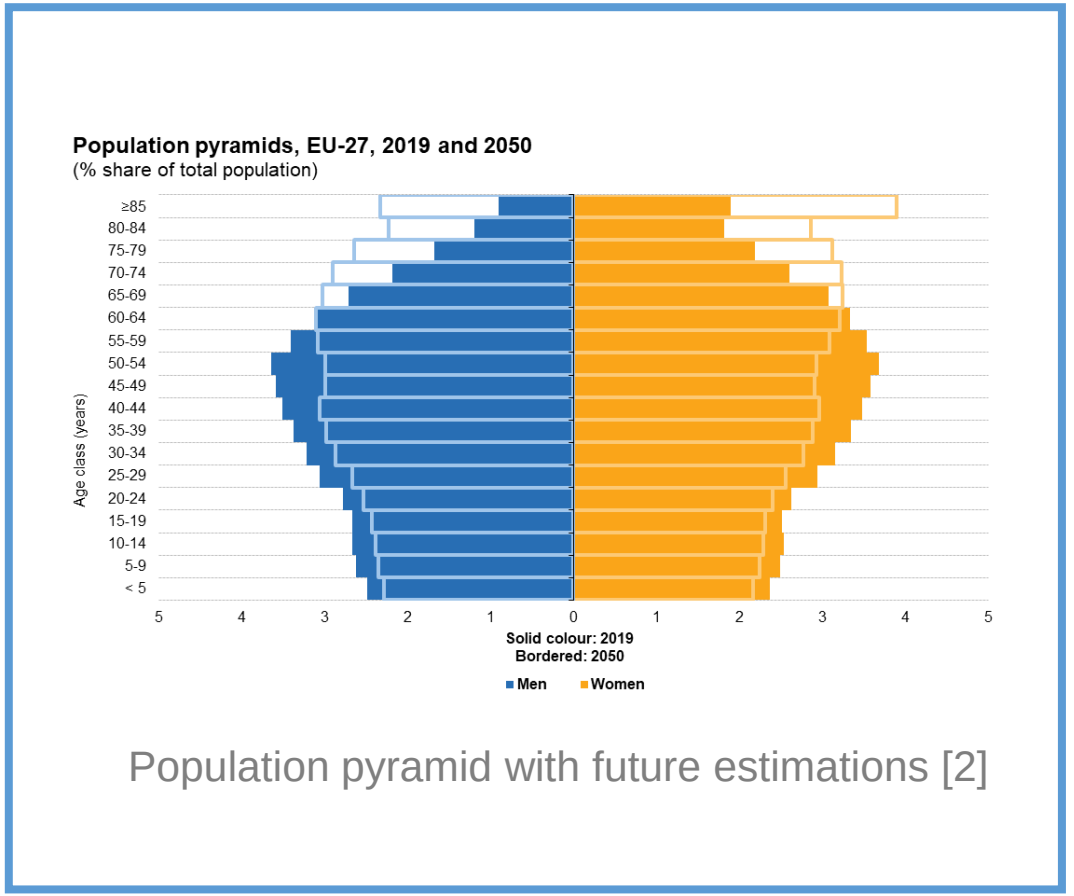
[1] https://newatlas.com/narrative-clip-2/35422/ visited on 20.01.2022

Motivation behind the thesis

- The **ageing population** demands new technological solutions to **reduce** the involved **medical personnel**
- The **technological progress** of electronic devices provides **new systems** for lifelogging and egocentric
- The **early stage** of **research** – no products on the market
- New devices on the market

RayBan Stories[3]

DJI Action 2 [4]



Population pyramid with future estimations [2]

[2] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Population_pyramids,_EU-27,_2019_and_2050_(%25_share_of_total_population)_AE2020.png visited on 24.09.2022
[3] https://www.ray-ban.com/canada/en/electronics/ray-ban%20stories%20%7C%20round-shiny%20black/8056597705035 visited on 22.09.2022
[4] https://www.gsmarena.com/the_dji_action_2_is_a_tiny_action_camera_made_big_by_its_multitude_of_accessories_and_mods-news-51608.php visited 14.04.2023

Application Examples

I. Action recognition

→ Task of discovering action in the image/clip

II. Activity recognition

→ **Activity** of Daily Living (ADL) **differs** from **action** detection in **length**

III. Food scene monitoring

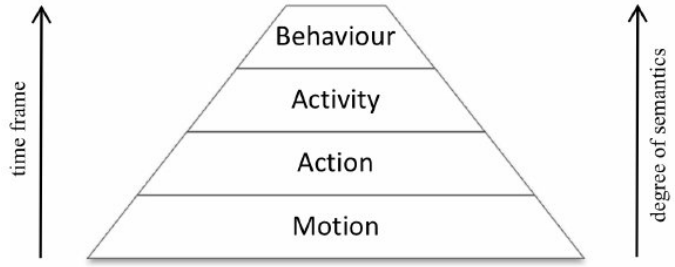
→ Food scene understanding, food detection, environment analysis

IV. Social interaction monitoring

→ Automated analysis for the social interaction pattern descriptions

→ Lack of social relations leads to a decrease in psychological well-being

Examples of actions in EPIC-KITCHEN dataset[5]



The difference between action and activity visualized [7]

[5] https://epic-kitchens.github.io/2021 visited on: 24.09.2022

[7] Nguyenet al., Recognition of Activities of Daily Living with Egocentric Vision: A Review. *Sensors (Basel, Switzerland)* vol. 16,1 72. 7 Jan. 2016

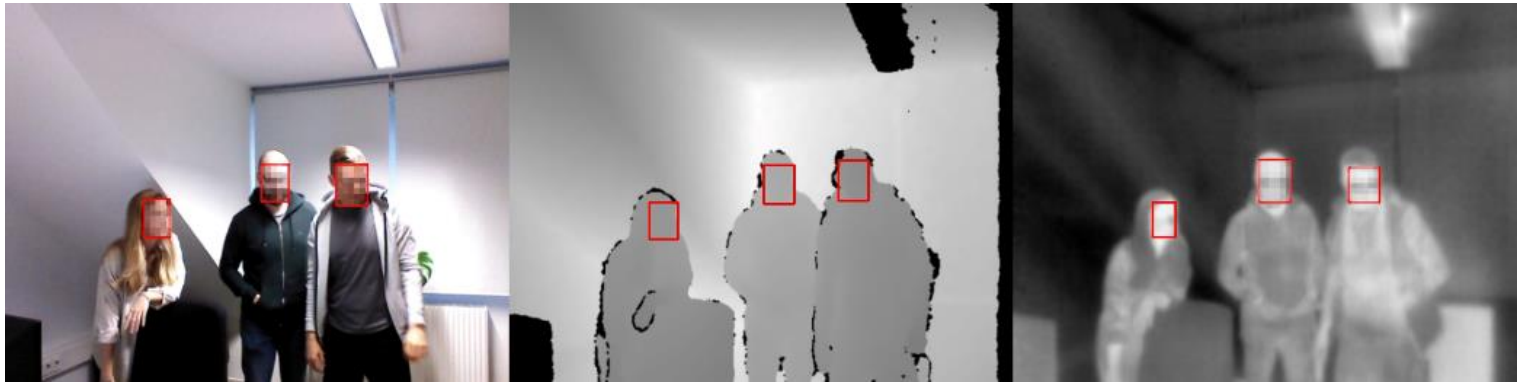
Research Questions

- **RQ1: What are actions and behaviours** impacting health and well-being **possible** to track for health improvement using **egocentric video-based lifelogging systems**?
- **RQ2:** Can egocentric images and lifelogging be employed to automatically **observe food intake**, to extract health factors and nutritional habits impacting health?
- **RQ3:** Is it feasible to process **egocentric image lifelogs** to **determine** the **fluid intake habits** of the subjects?
- **RQ4:** Is it possible to **automatically detect social signals** from the egocentric recordings of social interactions other than people's faces, e.g., posture, gaze, and emotions? Could these tools be used to **improve** the **mental well-being** of a camera wearer and people in the camera-wearer environment?

Before PhD seminar in Aachen

I. Research on various image modalities and their potential future applications for lifelogging

- Using additional modalities can lead to improvement in the results
- Comparison study of three image modalities (RGB, depth, thermal) in the example of Face Detection (FD) → FD is often a part of AAL systems, including egocentric lifelogging studies
- The RGB sensor is not always necessary, and it's superior in the hardest FD scenario



Example of a correct detection on each image modality, form left RGB, depth and thermal image

Mucha, W., & Kampel, M. (2022, February). **Depth and thermal images in face detection-a detailed comparison between image modalities**. In 2022 the 5th International Conference on Machine Vision and Applications (ICMVA) (pp. 16-21).

Before PhD seminar in Aachen

II. Privacy issues in lifelogging and AAL devices

- AAL vision-based life-logging systems raise privacy concerns due to monitoring indyviduals
- There are contradictory statements about privacy of depth cameras
- Contributions → Factors and scenarios affecting privacy in depth images
  → Face Recognition (FR) performance study to determine possibility of identification

Same scene visible in deptn and RGB image.

Before PhD seminar in Aachen

Conclusions

- Depth sensors preserve more privacy due to the lack of texture information. FR performs accurately only in laboratory environments with a small group of individuals and high-sensor resolution
- Lifelogging with depth sensors requires custom hardware, the market is evolving (e.g. mobile phones)
- Enhancing or replacing RGB images with depth modality is beneficial in certain scenarios (e.g., high privacy requirement, dietary monitoring), but at this moment it is restricted by hardware and data availability

Mucha, W., & Kampel, M. (2022, June). **Beyond Privacy of Depth Sensors in Active and Assisted Living Devices.** In Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments (pp. 425-429).

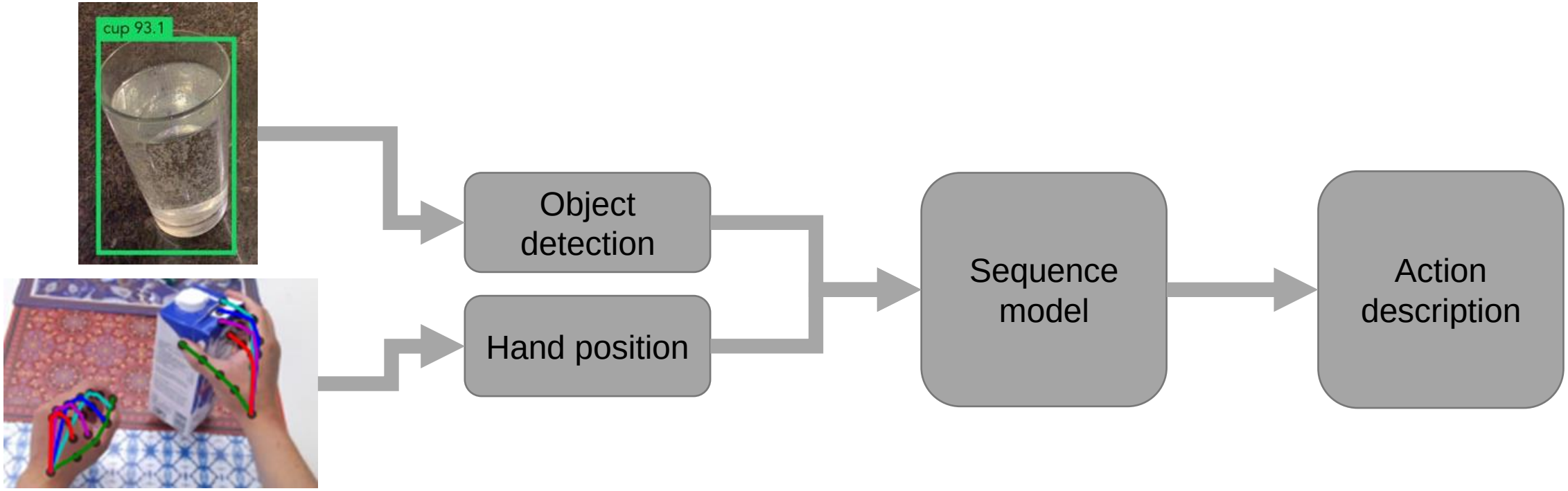
Mucha, W., & Kampel, M. (2022, July). **Addressing Privacy Concerns in Depth Sensors**. In Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part II (pp. 526-533). Cham: Springer International Publishing.

Work after Aachen seminar

- **Proficiency Evaluation**
  - First milestone in doctoral studies at TU Wien: Research proposal presented in front of a committee including national experts
  - Present work focuses on **activity** recognition from **longer time series** or **action** recognition from **short video clips** → Instead of tracking activities performed by subjects, there is a large space for solutions tracing specific actions regarding health:
    - Hand hygiene routine evaluation[16]
    - Recognition of cooking-related actions[17]
    - Medication intake
    - Alcohol consumption
- **Egocentric action recognition**
  - Hand pose modelling
  - Hand-based action recognition

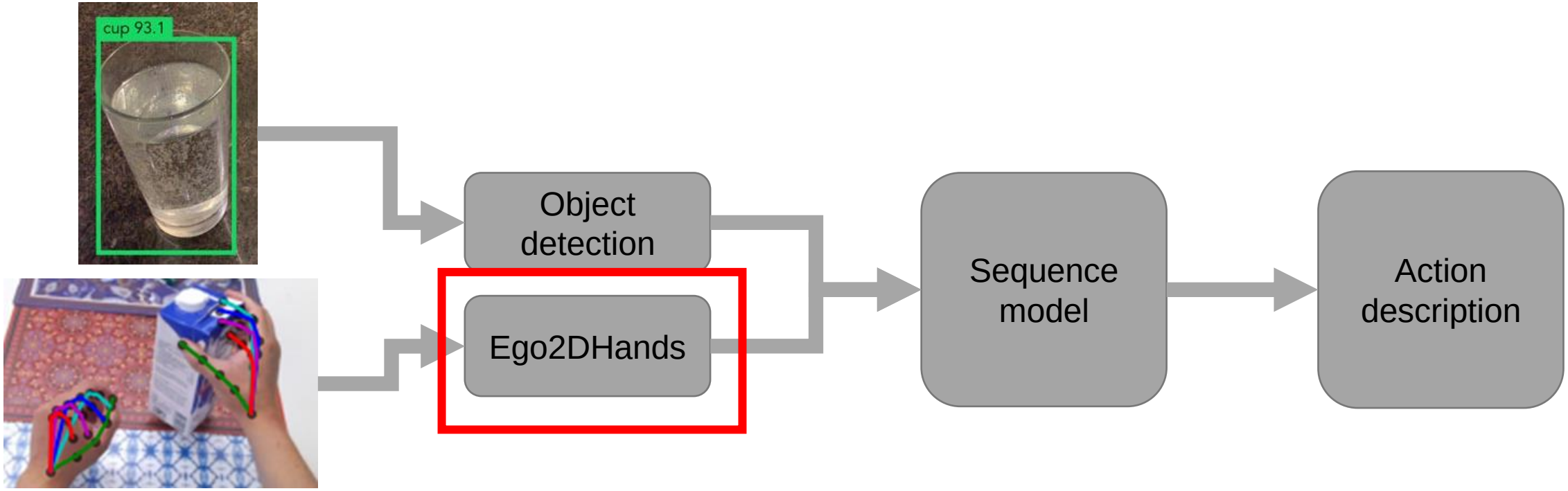
[16] C. Zhong et al., Hand-hygiene activity recognition in egocentric video. *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*. 2019
[17] S. Michibata et al., Cooking activity recognition in egocentric videos with a hand mask image branch in the multi-stream CNN. *Workshop on Multimedia for Cooking and Eating Activities*. 2020, pp. 1–6.

Current Work: Hand-based action recognition



- Usage of hands and objects as input for supervised sequence model
- Allows to use of pre-trained models reducing the learning costs
- Allows adoption for various health-related tasks which involve hands manipulation
- Privacy enhancement with employment of only skeleton and object data

Current Work: Hand-based action recognition



- Usage of hands and objects as input for supervised sequence model
- Allows to use of pre-trained models reducing the learning costs
- Allows adoption for various health-related tasks which involve hands manipulation
- Privacy enhancement with employment of only skeleton and object data

Ego2DHands: Introduction

Task:

→ Hand keypoint descriptor for **egocentric** vision estimating the position of **21 keypoints** in **2D** space using **RGB** image

Motivation:

→ **Applicable to tasks:**

- Simplifies hand gesture recognition
- Grasp analysis
- Object-hand interaction
- Rehabilitation

→ **Device restrictions:**

- Most existing solutions use RGB-D
- Wearable devices work with RGB images
- New RGB devices on the market

→ new data

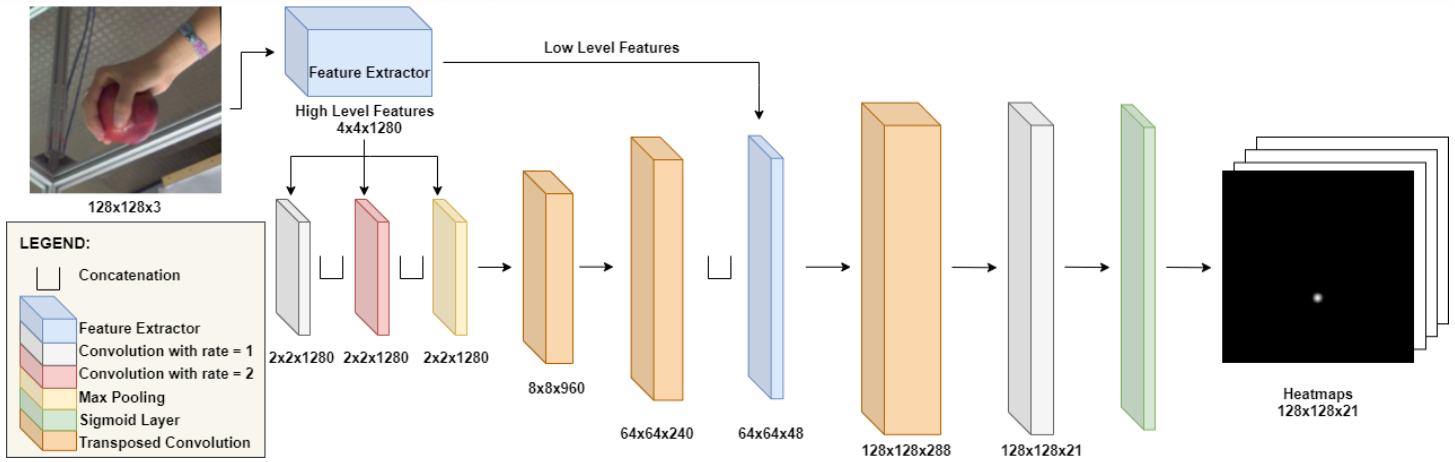
Contribution:

- A **novel** 2D hand keypoints prediction model
  → ***EffHandWfall***
  ✓ Consists of a previously **unused** features extractor and a **new** form of **features decompression** to **heatmaps**
- Accurately adapted to the egocentric perspective
- Performance is evaluated in scenes with movement, objects, and actions
- Results outperform state-of-the-art methods

Ego2DHands: Methodology

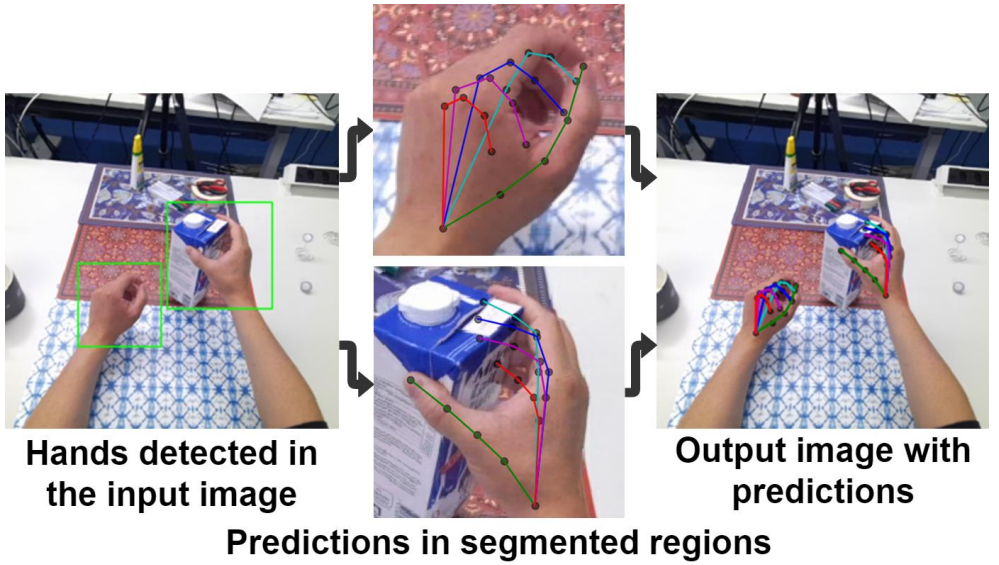
1st step → Single Hand network:

- Feature extractor:
  → *EfficientNetV2-S* [18]
- Prediction head



2nd step → Egocentric Adoption:

- Pre-trained hand detector in the egocentric input image
- Hand pose prediction in segmented regions
- Broadcasting prediction to input frame

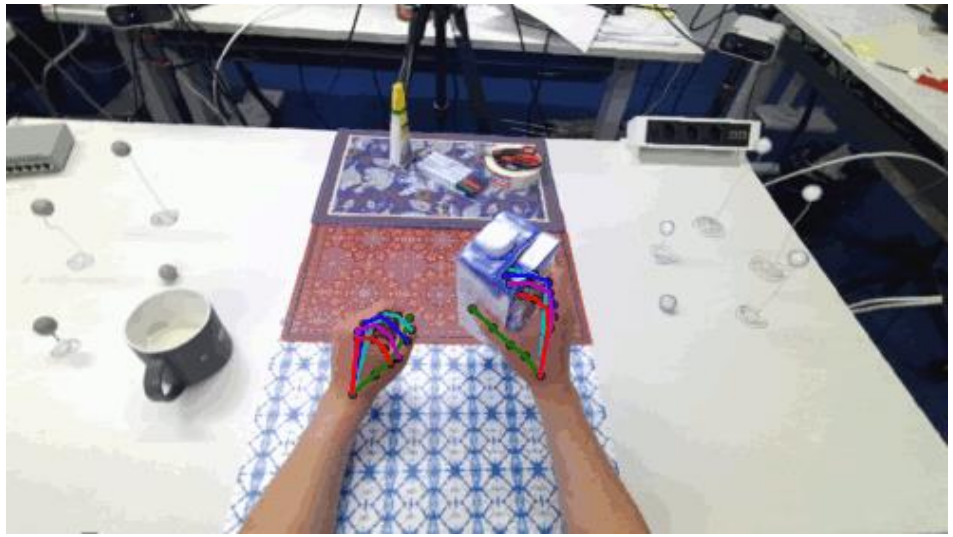


[18] Mingxing Tan and Quoc Le, "Efficientnetv2: Smaller models and faster training," in International Conference on Machine Learning. PMLR, 2021, pp. 10096–10106.

Ego2DHands: Results

Table 1. Different results on *FreiHAND dataset* and on egocentric *H2O dataset* including ours *EffHandWfall* .

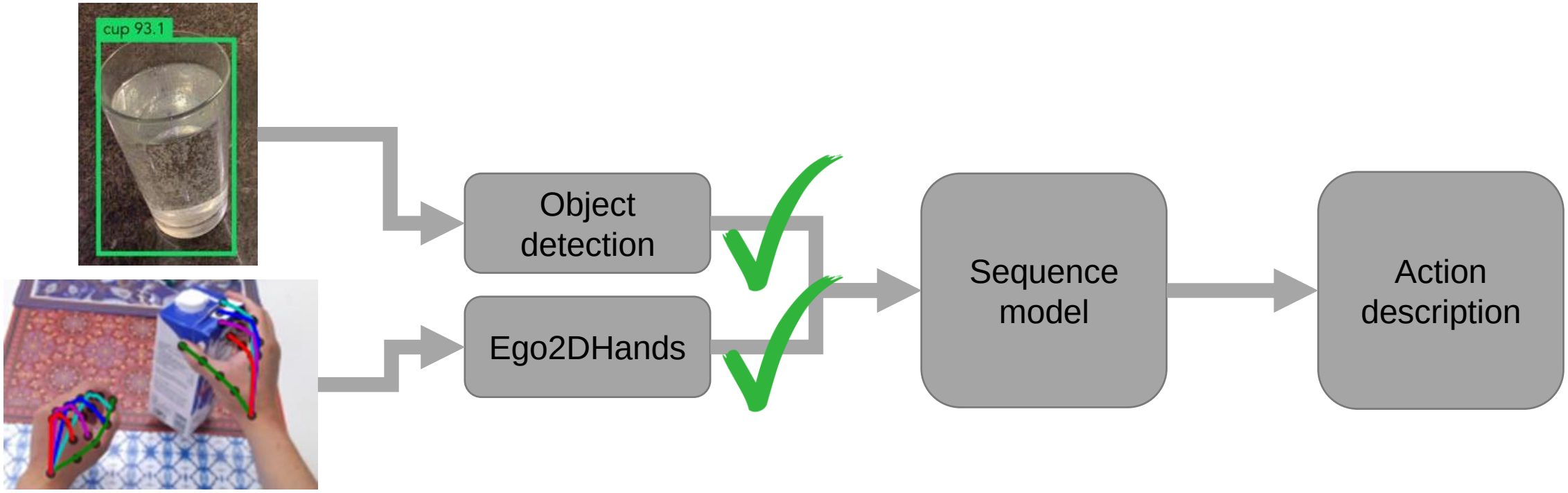
	PCK@0.2	EPE (Mean)	AUC
FreiHAND test subset:			
PoseResNet50 [20]	99.20%	3.27	0.868
Santavas et al. [13]	-	4.00	0.870
EffHandWafall	98.70%	2.24	0.921
EffHandWfall+P	**99.32%**	**1.72**	**0.929**
FreiHAND final evaluation subset:			
EffHandWfall	86.40%	4.61	0.853
EffHandWfall+P	**90.32%**	**3.85**	**0.874**
H2O Dataset - egocentric adoption:			
EffHandWfall	**93.58%**	**7.53**	**0.881**



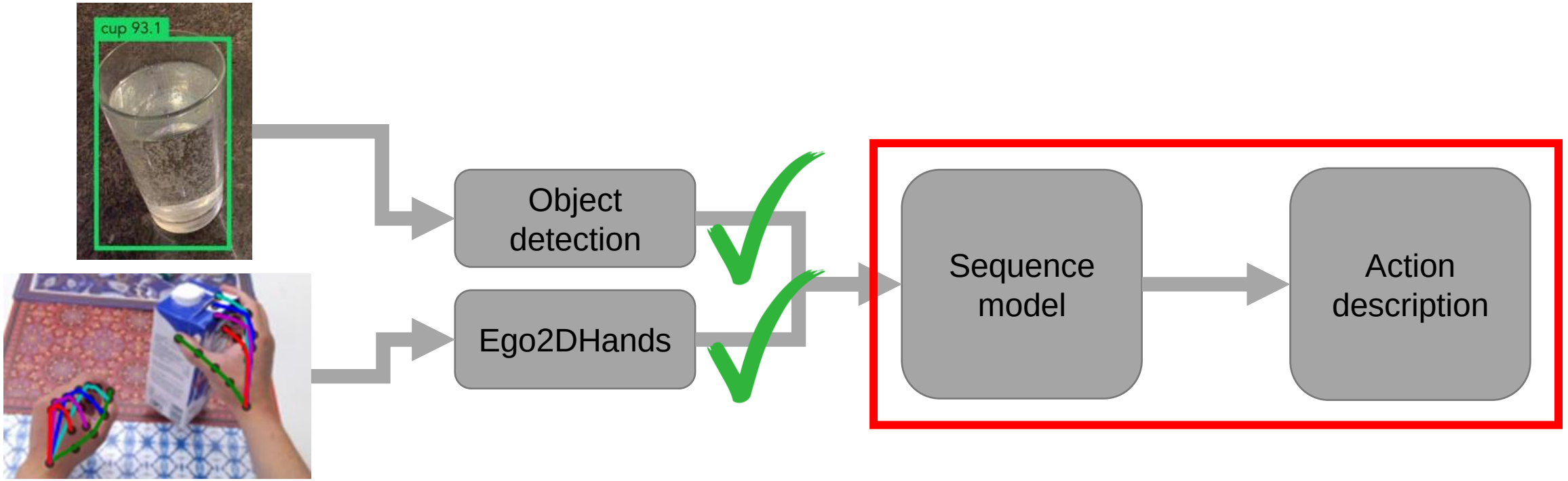
- ✓ **Egocentric hand pose detection is accurate and suitable for higher inference tasks**

Submitted: Mucha W., Kampel M. (2023) "**Ego2DHands: Generating 2D Hand Skeleton in Egocentric Vision**", International Conference on Image Processing - ICIP 2023, October 8-11, 2023, Kuala Lumpur, Malaysia

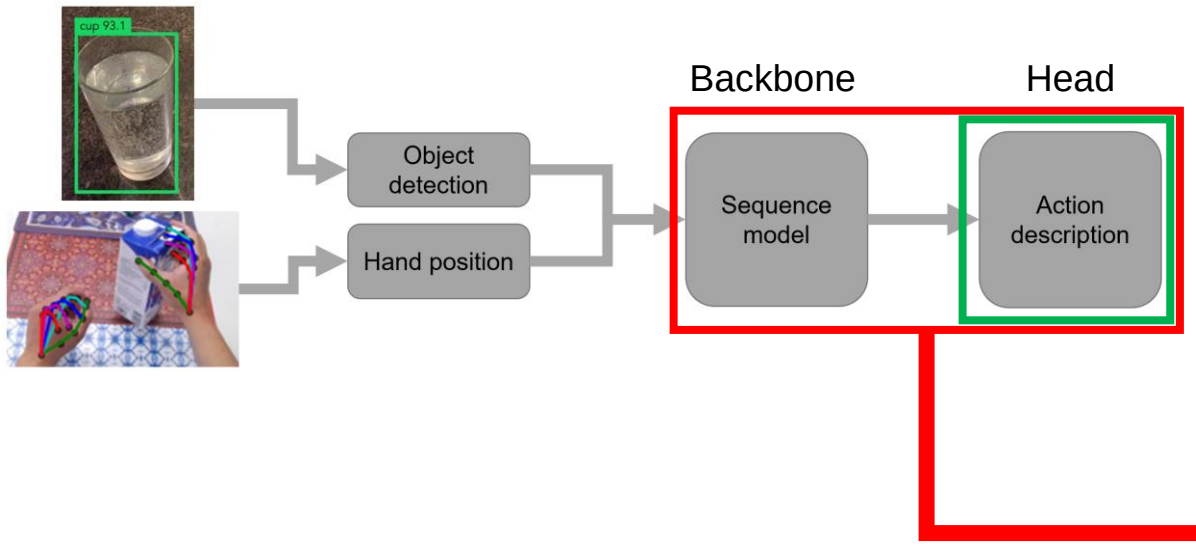
Current Work: Hand-based action recognition

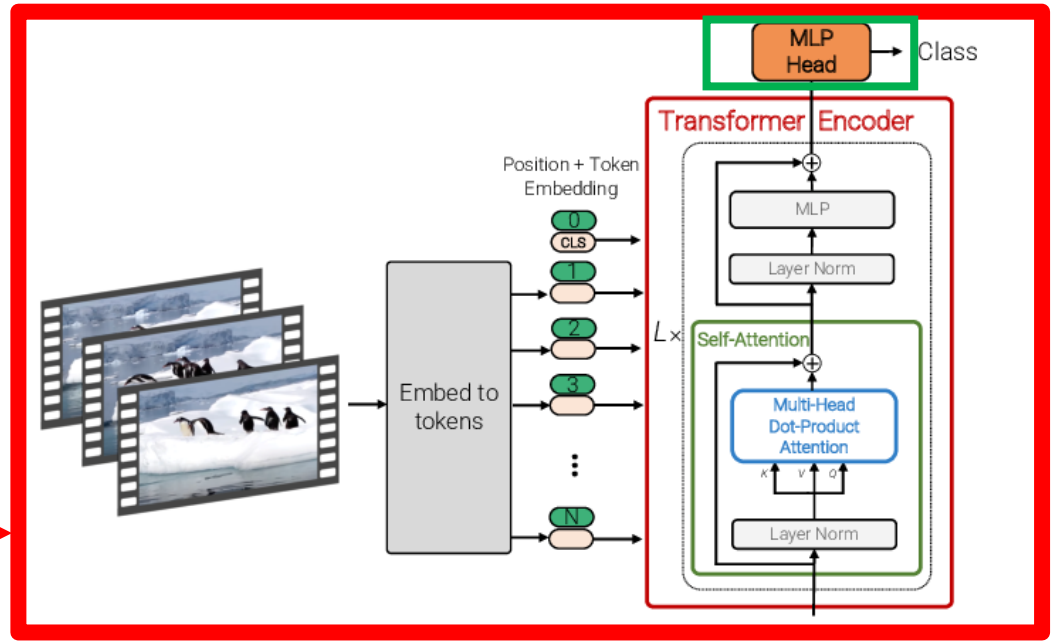


Current Work: Hand-based action recognition



Current Work: Hand-based action recognition





Transformer architecture example from[19]

Sequence classification

- Backbone: Transformer architecture performs better than the LSTM model
  → self-attention mechanism between the input sequence
- Each registered hands pose, and object is passed as an input token
- Head: Classification using Multilayer Perceptron (MLP)
- Data augmentation: horizontal flipping, random cropping, random frames sampling

[19] Arnab et al., ViViT: A Video Vision Transformer. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*

Current Work: Hand-based action recognition

Preliminary results:

- Action recognition on validation dataset of H2O Dataset
  → Accuracy: **94.26%** (36 action classes)

Next steps:

- Results on test data to place among state-of-the-art results
- Action recognition for medicament intake monitoring
- Adoption to medicament intake task:
  - Recordings of medicament intake-related actions (in progress)
  - Transfer learning for object detection and sequence classification

Method	Val accuracy	Test accuracy
C2D	76.10	70.66
I3D [1]	85.15	75.21
Slowfast [2]	86.00	77.69
H+O [3]	80.49	68.88
ST-GCN [4]	83.47	73.86
TA-GCN [5]	86.78	79.25

Available results on H20 Dataset [20]

[20] https://codalab.lisn.upsaclay.fr/competitions/4820 accesed 14.04.2023

Summary of last year activities

- 1st visuAAL secondment at the University of Alicante
- Proficiency evaluation at TU Wien
- Hand-based action recognition
  - ✓ **Hand pose paper → Mucha W., Kampel M. (2023) "Ego2DHands: Generating 2D Hand Skeleton in Egocentric Vision", International Conference on Image Processing - ICIP 2023, October 8-11, 2023, Kuala Lumpur, Malaysia (under review)**
  - ✓ Hand-based action recognition
- Subjects:
  - ✓ 199.096, VU, 2022W. From surviving to thriving: crafting your good professional life
  - ✓ 194.077, VU, 2022W. Applied Deep Learning
- Teaching:
  - ✓ 183.663 Deep Learning for Visual Computing. S22 – TU Wien
- Attended events:
  - ✓ **26TH Computer Vision Winter Workshop (CVWW), 15.-17.02.2023 Krems, Austria → presentation**
  - ✓ 17th Visual Computing Winter Workshop Austria, 6-8.03.2023, St. Johann im Pongau, Austria
- **Accepted workshop proposal**:
  - ✓ ViBRANT: Workshop on privacy-aware video-based AAL technologies, The 15th ACM SIGCHI Symposium on Engineering Interactive Computing Systems - EICS 2023

Next steps and planned work

Why egocentric medication monitoring?

- The system can store a visible image of any pill for additional manual verification
- Comfort of use - not dependent on place and position (e.g. wearable glasses)
- User cannot be identified - is not recording him/herself (privacy)
- Wearables allow further expansion to other activities (one tool for all)

Questions and open problems:

- Is the egocentric perspective feasible for such an application?
- What camera placement (head/glasses/chest) is best for this scenario?
- Pill recognition accuracy

Deadline:

- The 27th International Symposium on Wearable Computers (ISWC): 26.05.2023

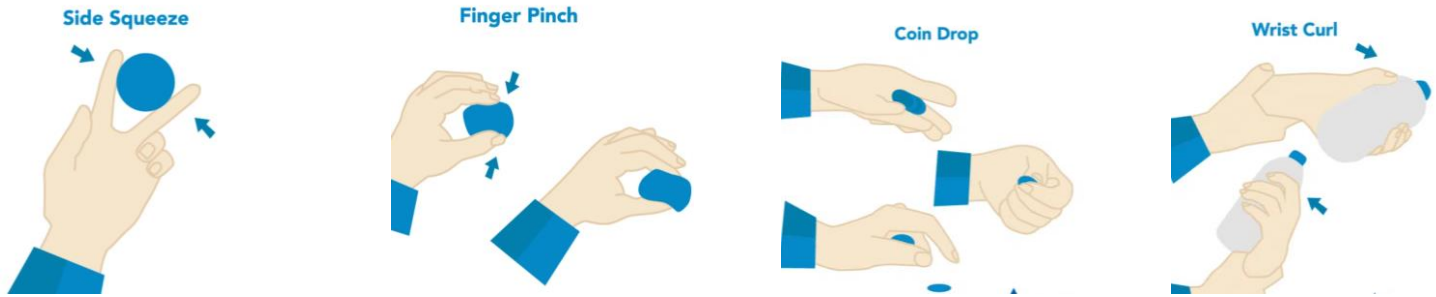
UBICOMP ISWC 2023 CANCUN, MEXICO

Next steps and planned work

Hand-based action model allows adoption for different tasks:

I. Hand rehabilitation monitoring

- Rehabilitation for individuals after a stroke or a spinal cord injury
- Distances between patients and hospitals obstruct rehabilitation outcomes
- Automatically monitoring rehabilitation exercises through hand-based action recognition systems can potentially lead to improvements in health.
- **Transfer Learning for sequential data**



Examples of hand excercices for recovery after stroke [21]

[21] https://www.saebo.com/blog/reclaim-your-dexterity-with-hand-exercises-for-stroke-recovery accesed 14.04.2023

Next steps and planned work

Hand-based acton model allows adoption for different tasks:

II. Water and fluid intake

- Tracking of fluid intake using egocentric action recognition

III. Social interaction tracking with egocentric vision

- Computer vision methods to read social signals between individuals

Potential conferences:

- International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP) (deadline 9th of October)
- PerCom: Pervasive Computing and Communications (Potential deadline - October)
- International Conference on Pattern Recognition (ICPR) (deadline 1st of May 2024)
- IEEE/CVF Winter Conference on Applications of Computer Vision (Potential deadline - July)

Estimated time schedule

Task	2021			2022				2023					2024				2025
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4	Q1
Initial research on life logging and AAL devices												Last visuAAL PhD seminar in Vienna, December 2023					
Modality comparison study			P														
Privacy with depth				P P													
Proficiency evaluation					R.P.	Pres.											
Egocentric hand pose paper								P									
Medication intake action recognition									P								
Hand rehabilitation with egocentric view											P						
Implementation for drinking module													P				
Secondments																	
Studying social interaction processing and implementation															P		
Thesis Writing																	Thesis

ViBRANT

ViBRANT - Workshop on Video-Based Respectful Assistive Novel Technologies

Part of EICS 2023, the 15th ACM SIGCHI Symposium on Engineering Interactive Computing Systems at Computational Foundry, Swansea, Wales, UK - June 27–30, 2023

Important Dates

Submission deadline: **29.05.2023**

Notification of acceptance: 09.06.2023

Planned workshop day: **27.06.2023** (half-day)

Call for Papers

1. **(Out-of-proceedings) research highlights and work-in-progress** (2-page extended abstract)
2. **Archival original work (**published by Springer LNCS)

- video-based technology for health monitoring
- fair and privacy-aware systems
- integration of engineering issues in the design process of video-based AAL
- legal requirements and privacy issues for data collection and processing in care
- user acceptance criteria for video-based AAL
- GDPR requirements for video-based AAL solutions
- best practices of interdisciplinary collaborations between legal, technical, social, health sciences

visuAAL **+ info:** wiktor.mucha@tuwien.at.ac and https://www.visuaal-itn.eu/events (coming soon**)**

Thank you!

Wiktor Mucha

TU Wien

wiktor.mucha@tuwien.ac.at

Conclusion

Planned contribution of this thesis

- Automatisation of lifestyle assessment tools:
- Novel research in frameworks and applications:
  - → Action recognition and monitoring
  - → Dietary monitoring
  - → Drinking activities monitoring
  - → Social interaction understanding
- Accuracy enhancement for existing methods/applications