visuAAL

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

ESR12. AI for Dementia Care

3rd Doctoral Seminar

Vienna, Austria
01.12.2023

Irene Ballester Campos
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

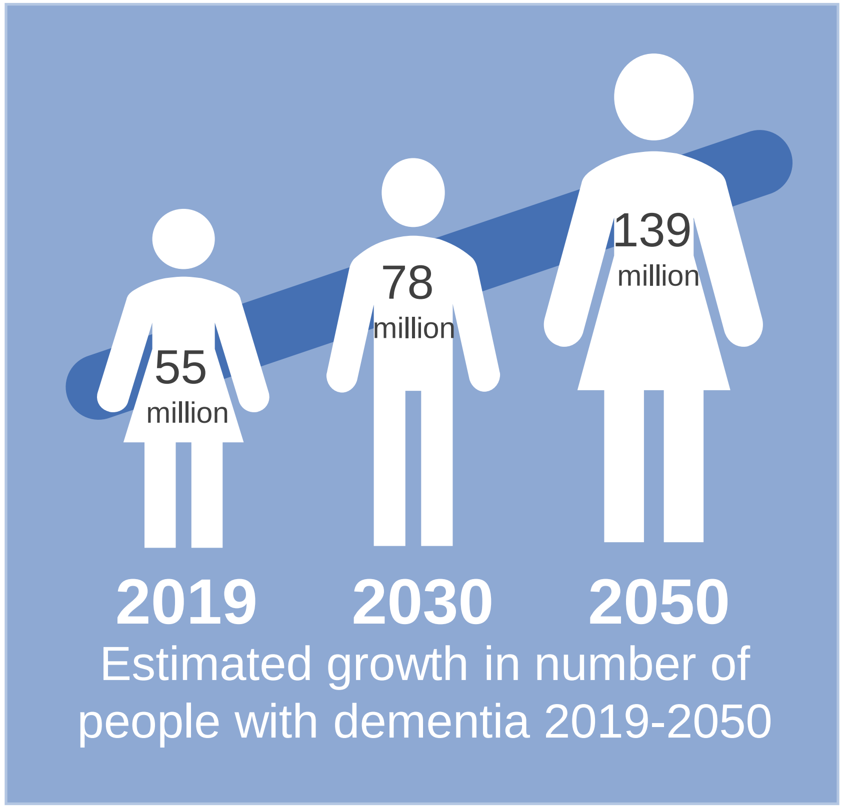
TU WIEN

What’s dementia?

- Syndrome in which there is a **deterioration in cognitive functioning beyond what might be expected from normal ageing** [1]

Why dementia?

- Behavioural changes **strongly correlated** with the degree of functional and cognitive impairment [2]. **Behavioral and Psychological Symptoms of Dementia (BPSD):** agitation, aberrant motor behaviour, anxiety, irritability, depression, apathy, delusions, changes in sleep or appetite [2].
- One of the **major causes of dependency** among older people [3]



Data source: WHO [1]

[1] World Health Organization https://www.who.int/news-room/fact-sheets/detail/dementia (accessed April 25, 2022)
[2] Joaquim Cerejeira, Luısa Lagarto, and Elizabeta Blagoja Mukaetova-Ladinska. “Behavioral and psychological symptoms of dementia”. In: *Frontiers in neurology 3 (2012*), p. 73.
[3] **Global status report on the public health response to dementia**. World Health Organization (2021)

AI for behaviour analysis from unobtrusive sensor data

Goal: Development of AI methods for **measuring** the **behaviours** of care home residents with **dementia** using **unobtrusive sensors (depth maps)**

In order to:

1. Detect and **measure behavioural changes** indicative of dementia (BPSD) in the mid and long term *(weeks, months, years)*

2. Provide **assistance with ADLs** for people with dementia in the short term *(seconds, minutes)*

Thesis Overview: Research Questions

RQ1. Robust performance for real-world HAR

RQ2. Different inputs for behaviour measurement

RQ3. Providing assistance with ADLs in the short term } **ToiletHelp**

RQ1. Robust performance for real-world HAR

RQ2. Different inputs for behaviour measurement

RQ3. Providing assistance with ADLs in the short term } **ToiletHelp**

Goal

Step-by-step guidance for people with mild dementia in the toilet using a depth camera

RQ3. Providing assistance with ADLs in the short term

1. How can we use depth-based methods to detect the **need for assistance** in people with dementia, for example, in a toilet setting?
2. How **interaction** with the user must be designed to be **effective for people with dementia**?

ToiletHelp demo

Interaction design

1. **Focus groups** with health professionals: *How the system should communicate with people with dementia?*

2. Prototype developed with hand-washing and acknowledgement module

3. Validation in the lab: functional testing

- 98.5% avg. accuracy in action recognition
- Interaction: 100% in fixed scenarios, 8/10 correct in open scenarios

4. New visualizations and audio

2 publications:

"**RITA: A privacy-aware toileting assistance designed for people with dementia**” by Irene Ballester, Tamar Mujirishvili and Martin Kampel, In *Proceedings of the 15th EAI International Conference on Pervasive Computing Technologies for Healthcare,* 2021

"**Automated vision-based toilet assistance for people with dementia**” by Irene Ballester and Martin Kampel, AHFE 2022 - *13th International Conference on Applied Human Factors and Ergonomics,* July 24-28, 2022, New York, USA

+ collaboration in the DIANA project

Evaluation of the action recognition module

Methodology

Action is detected **correctly** **if:**
H1 < fID_detected – fID_gt< H2

Results

Toilet Module evaluated against **human annotators** for action recognition with **20 people with dementia** and **50 sequences**, with 182 actions:

- Hospital group accuracy: **85%**
- Day center accuracy: **70%**
- **Overall accuracy: 81%**

Dataset	% of successfully recognized actions									
	Next to the toilet bowl		*Sat down on the toilet bowl*		*Stood up from the toilet bowl*		*Next to the basin*		**Total**	
	N actions	Acc.	N actions	Acc.	N actions	Acc	N actions	Acc.	N actions	Acc.
S1 (N=1)	6	33%	10	90%	10	90%	7	71%	33	**76%**
S2 (N=1)	10	70%	10	100%	10	90%	10	100%	40	**90%**
S3 (N=1)	10	100%	10	80%	10	100%	6	100%	36	**94%**
Hospital (N=3)	26	**73%**	30	**90%**	30	**93%**	23	**92%**	109	**87%**
Day center (N=17)	17	**59%**	19	**95%**	19	**68%**	18	**56%**	73	**70%**
All (N=20)	43	**67%**	49	**92%**	49	**84%**	41	**76%**	182	**80%**

Submitted: I. Ballester, M. Gall, T. Münzer, and M. Kampel, “Vision-Based Toilet Assistant for People with Dementia in Real-Life Situations” *Pervasive-Agetech Workshop at PERCOM2024*

Evaluation of the interaction module

Methodology

Install the prototype in a semi-public toilet and ask caregivers and older adults about their opinion via questionnaires

Conclusions

End-users (30 older adults,17 in DC1 and 13 in DC2)

- Increased **feeling of safety and independency** and not feeling feel afraid nor annoyed
- **Participants with dementia** affirm they **understand the instructions**
- Changes in rating as the user uses the system more times

Caregivers (14 professionals, 4 in DC1 and 10 in DC2)

- Positive rating in terms of being **useful for end-users and reducing caregivers' workload**
- Positive **validation of the interaction modalities**

Submitted: I. Ballester, M. Gall, T. Münzer, and M. Kampel, "ToiletHelp: an assistive technology to guide people with dementia in the toilet, *Journal of Ambient Intelligence and Humanized Computing*, 2023.

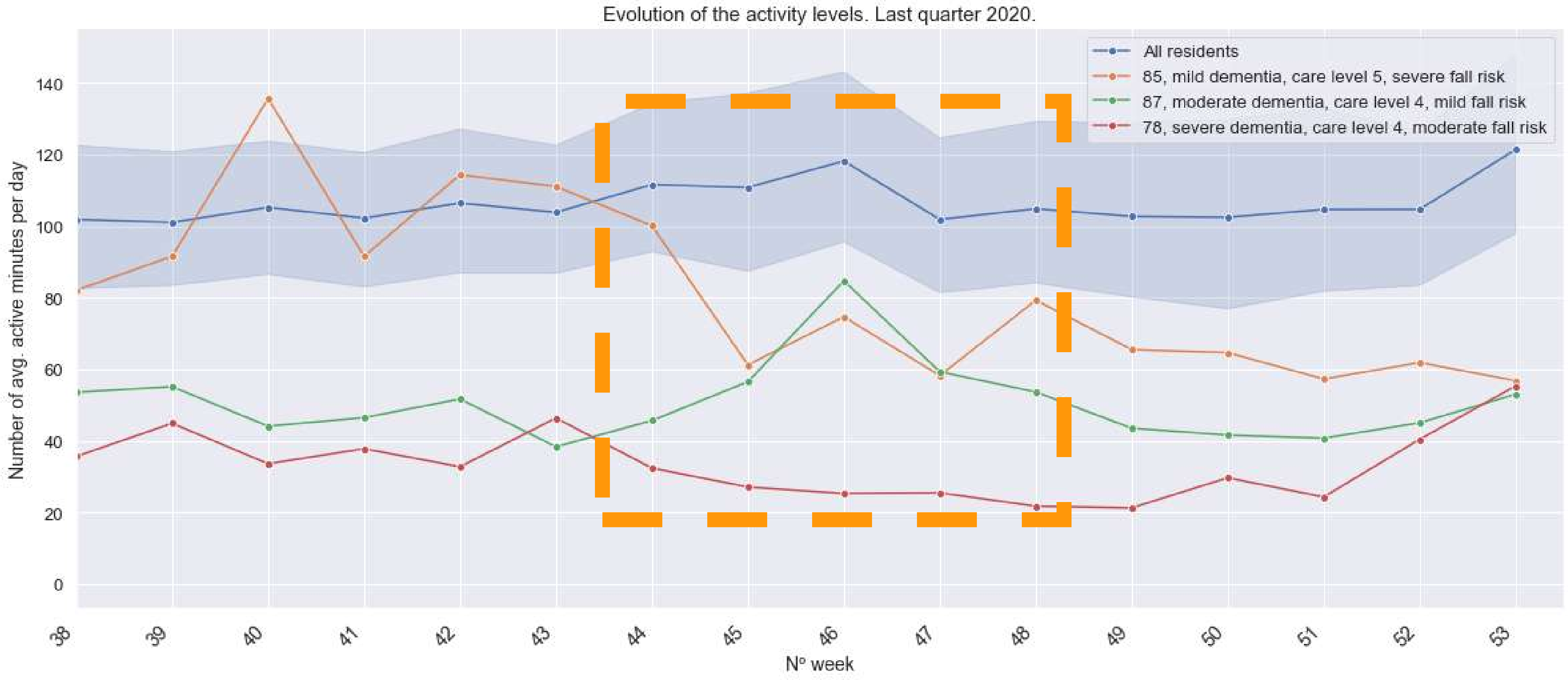
RQ1. Robust performance for real-world HAR

RQ2. Different inputs for behaviour measurement

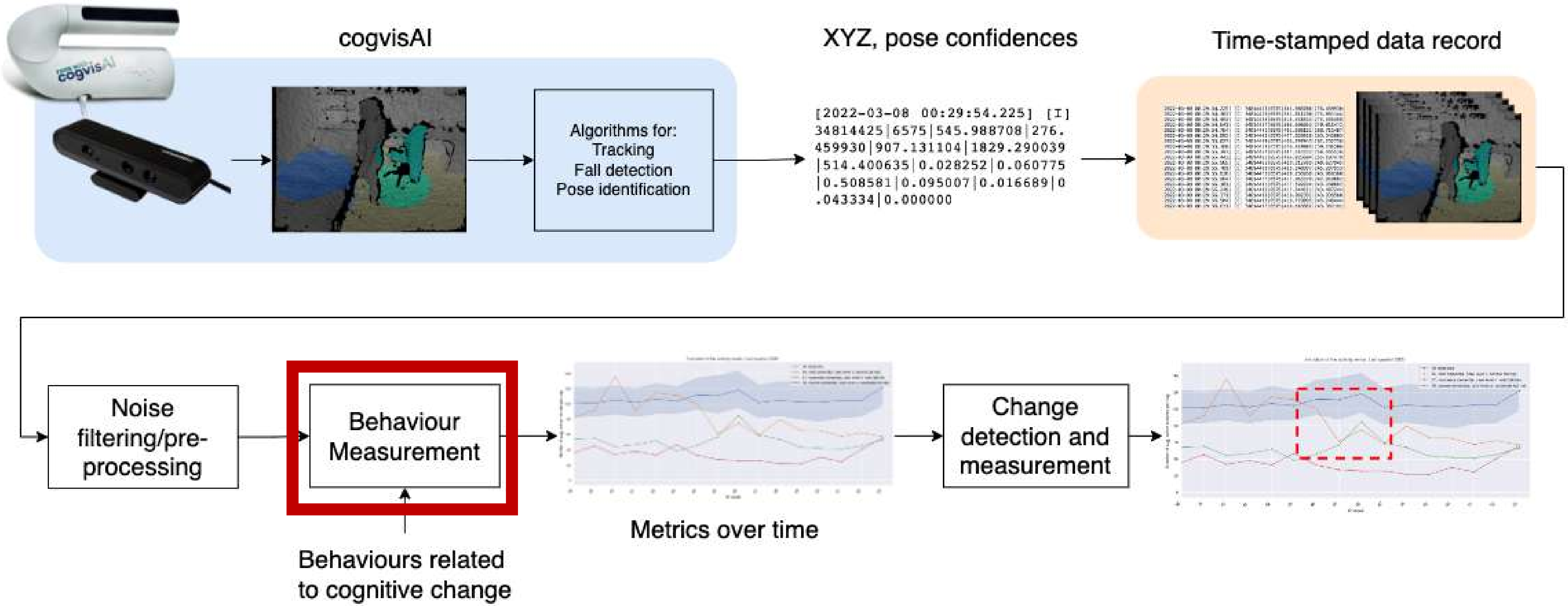
RQ3. Providing assistance with ADLs in the short term } **ToiletHelp**

Ultimate goal

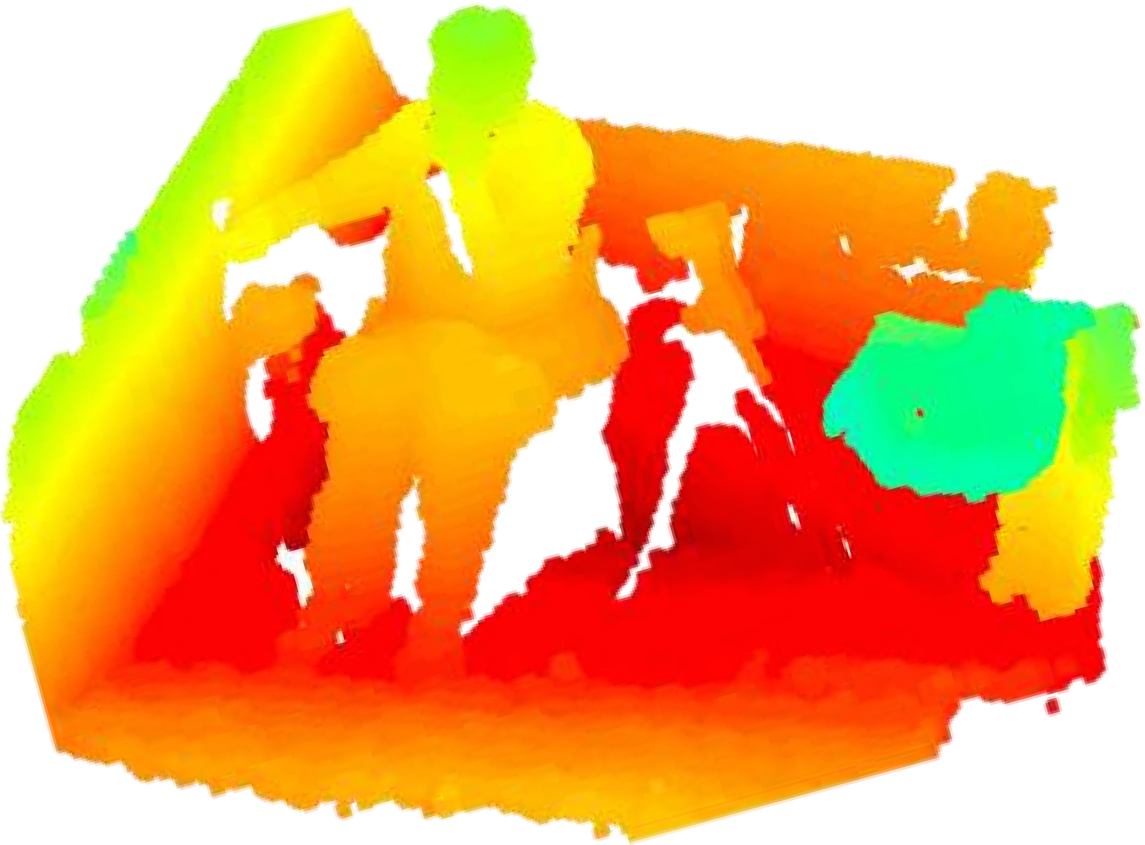
Detect and measure behavioural changes indicative of dementia in the mid and long term



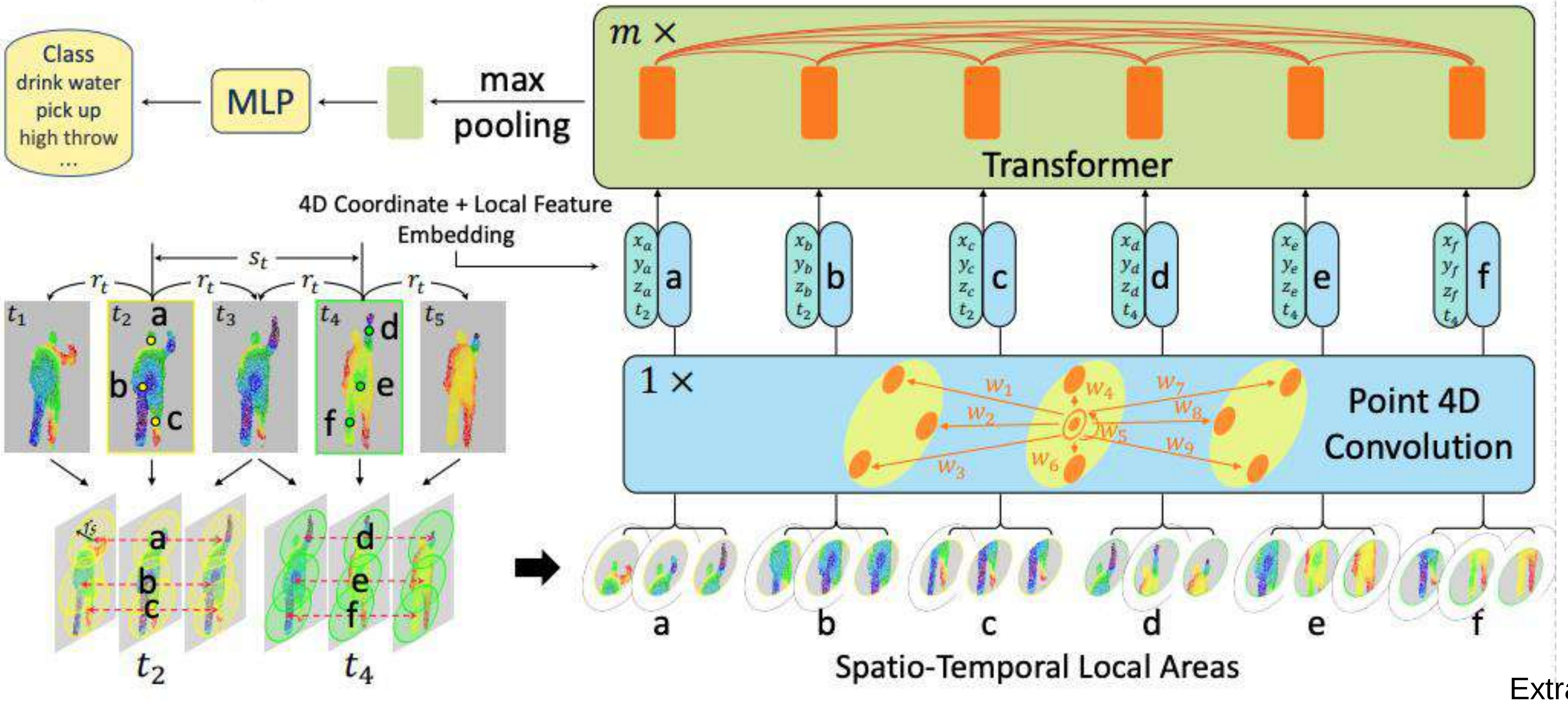
But…


cogvisAI
Algorithms for:
Tracking
Fall detection
Pose identification
XYZ, pose confidences
[2022-03-08 00:29:54.225] [I]
34814425|6575|545.988708|276.
459930|907.131104|1829.290039
|514.400635|0.028252|0.060775
|0.508581|0.095007|0.016689|0
.043334|0.000000
Time-stamped data record
Noise filtering/pre-processing
Behaviour Measurement
Behaviours related to cognitive change
Metrics over time
Change detection and measurement

Point clouds from depth images



P4Transformer [4]:



Extracted from [4]

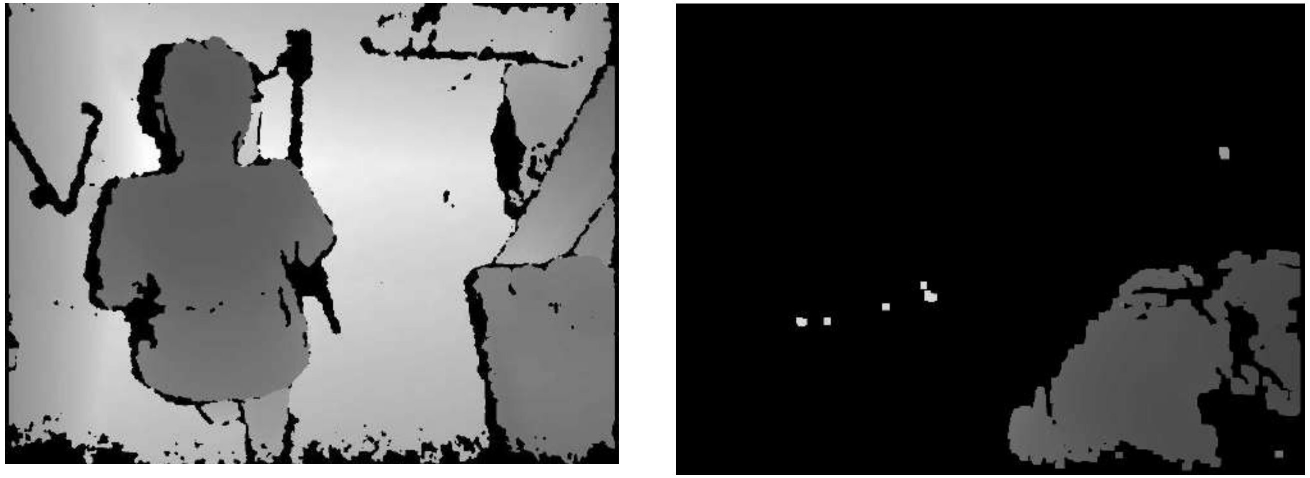
[4] Fan, H., Yang, Y., & Kankanhalli, M. (2021). Point 4d transformer networks for spatio-temporal modeling in point cloud videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14204-14213).

Bathroom Activities Dataset (BAD)

Description

- 50 full sequences
- **19 subjects with dementia** using the toilet, 60 to 97 years, 6 men and 13 women
- 7 classes: walking around, undressing, sitting down, sitting on the toilet, standing up, dressing, washing hands
- 2 different locations:
  - **BAD1**: 3 subjects (36k frames)
  - **BAD2:** 16 subjects (21k frames)
- **Unbalanced** dataset:
  E.g.: in BAD2: sitting: 8k frames vs. sitting down: 679 frames

BAD1 – 3 subjects



BAD2 – 16 subjects

P4Transformer

	F1-score	Acc.
MSR Action3D	0.882	90.94*
NTU RGB+D 60 – cross-subject	-	90.2*
NTU RGB+D 60 – cross-view	-	96.4*
NTU RGB+D 120 – cross-subject	-	86.4*
NTU RGB+D 120 – cross-view	-	93.5*
BAD1 (Train: S1-S3 Test: S1-S3)	0.8599	-
BAD1 – cross-subject	0.5315	-
BAD2 (Train: S7-S19, Test: S4-S6)	0.6363	-
Train in BAD1, tested in BAD2	0.0723	-
Train in BAD2, tested in BAD1	0.0790	-

Domain generalization problem!

*Results extracted from [4]

[4] Fan, H., Yang, Y., & Kankanhalli, M. (2021). Point 4d transformer networks for spatio-temporal modeling in point cloud videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14204-14213).

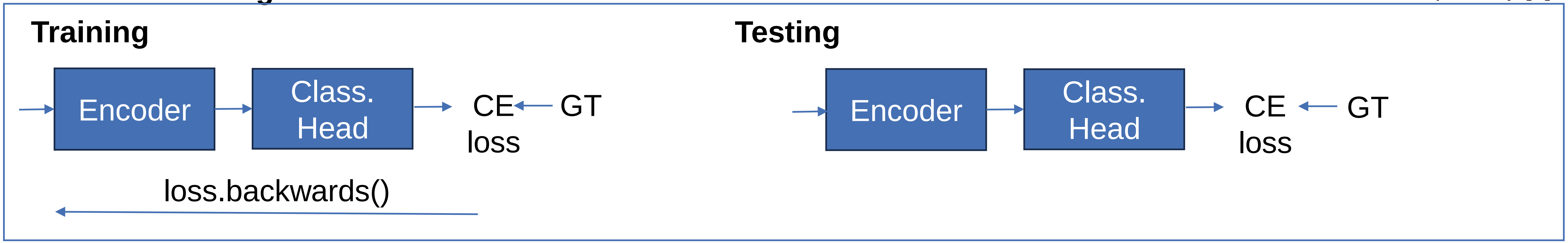
RQ1. Robust performance and domain generalization

1. What are the differences between data collected in **real-world scenarios with people with dementia** and data collected in controlled laboratory settings? How can models be developed to ensure **robust** performance **in real-world conditions**?
2. What strategies can facilitate **domain generalization** so that models can operate effectively in different domains? How can models be developed to account for **variation among individuals** and demonstrate the ability to generalize across subjects?
3. How does **self-supervised learning** contribute to this goal?

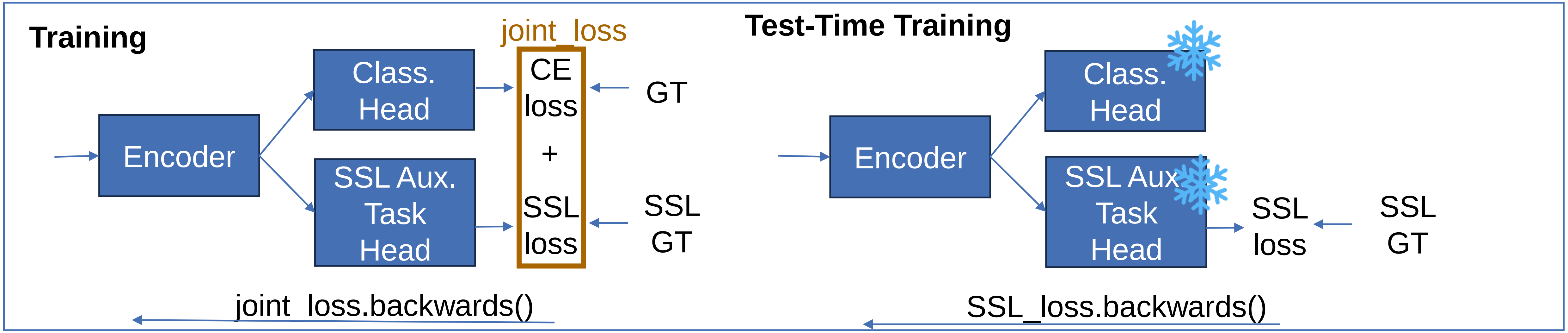
Test-Time Training (TTT)

Inspired by [5]

Standard Training

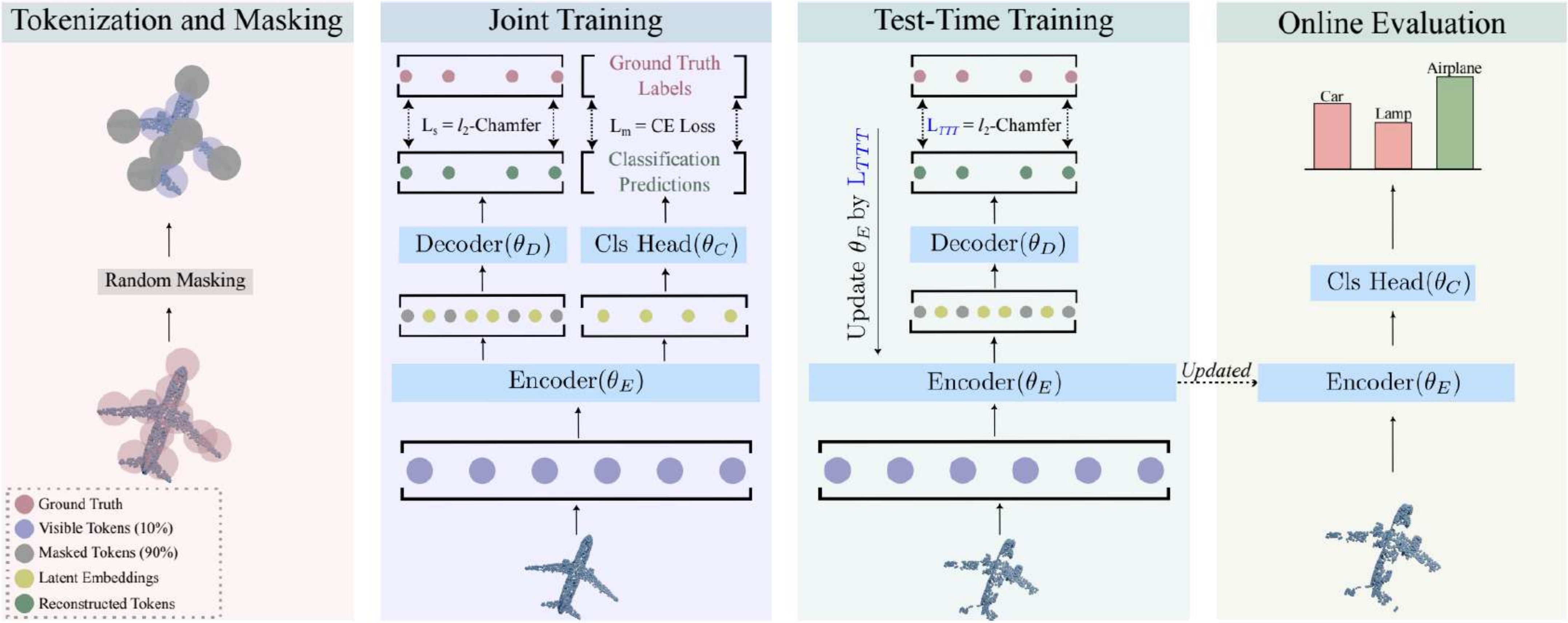


Test-Time Training



[5] Mirza, M. J., Shin, I., Lin, W., Schriebl, A., Sun, K., Choe, J., ... & Bischof, H. (2023). Mate: Masked autoencoders are online 3d test-time learners. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 16709-16718).

TTT for static point clouds

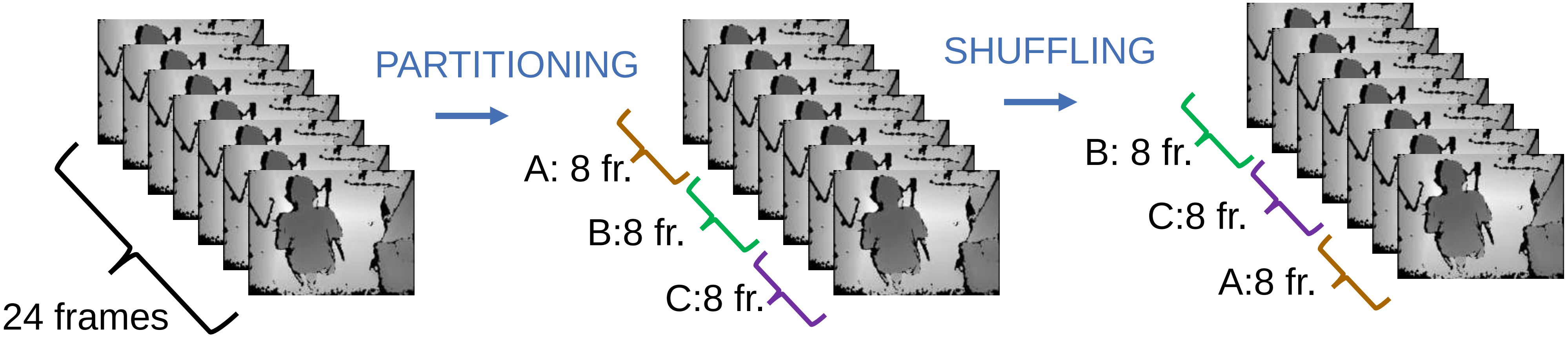


Extracted from [5]

[5] Mirza, M. J., Shin, I., Lin, W., Schriebl, A., Sun, K., Choe, J., ... & Bischof, H. (2023). Mate: Masked autoencoders are online 3d test-time learners. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 16709-16718).

Inspired by [6]

- Auxiliary task: frame order classification (LSTM)



Preliminary results

F1-score

	Standard Training	TTT
Trained in BAD1, tested in BAD2	0,0723	**0,1073**
Trained in BAD2, tested in BAD1	0,079	0,0632

[6] Wang, H., Yang, L., Rong, X., Feng, J., & Tian, Y. (2021). Self-supervised 4d spatio-temporal feature learning via order prediction of sequential point cloud clips. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3762-3771).

Next steps

Development of **self-supervised learning** methods to improve domain and cross-subject generalization for **dynamic point clouds**

1. Literature review on self-supervised methods for:
   - Video understanding [7]
   - Point clouds [8, 9]
2. Implementation and evaluation of different models for point cloud-based activity recognition [10,11]
3. Pre-processing and annotation of the bedroom dataset

[7] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P. (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*.
[8] Saltori, C., Galasso, F., Fiameni, G., Sebe, N., Ricci, E., & Poiesi, F. (2022, October). Cosmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation. In *European Conference on Computer Vision* (pp. 586-602). Cham: Springer Nature Switzerland.
[9] Yi, L., Gong, B., & Funkhouser, T. (2021). Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15363-15373).
[10] Yi, L., Gong, B., & Funkhouser, T. (2021). Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15363-15373).
[11] Fan, H., Yang, Y., & Kankanhalli, M. (2022). Point spatio-temporal transformer networks for point cloud video modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 2181-2192.

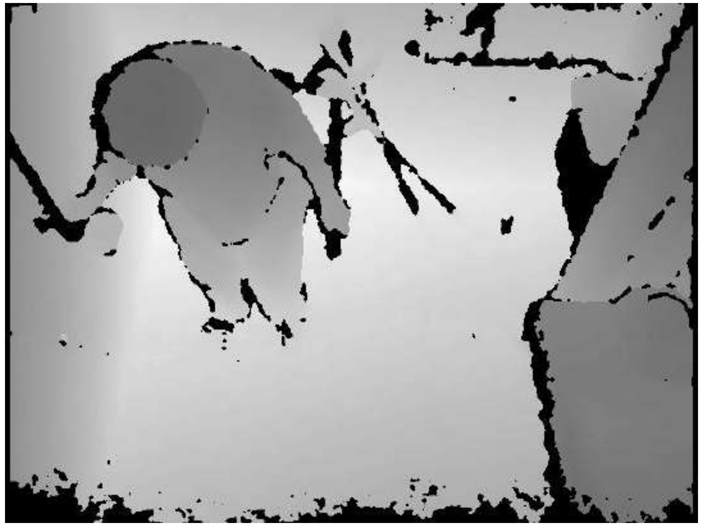
RQ1. Robust performance for real-world HAR

RQ2. Different inputs for behaviour measurement

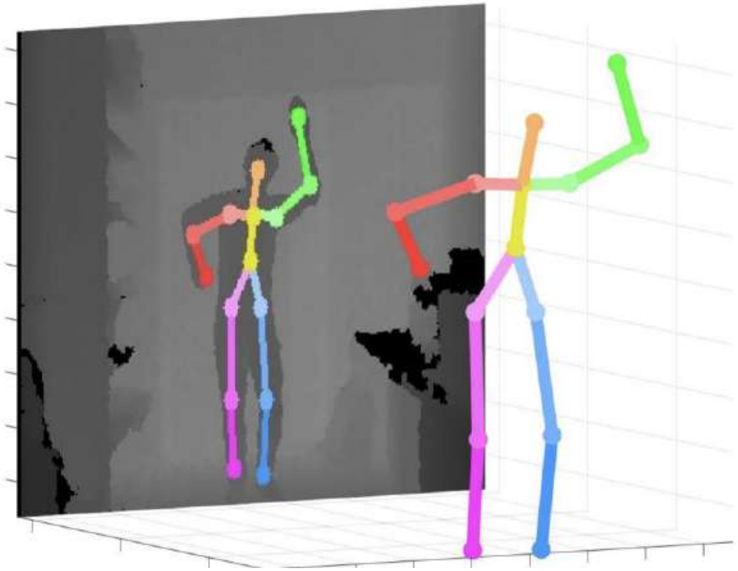
RQ3. Providing assistance with ADLs in the short term } **ToiletHelp**

Different inputs

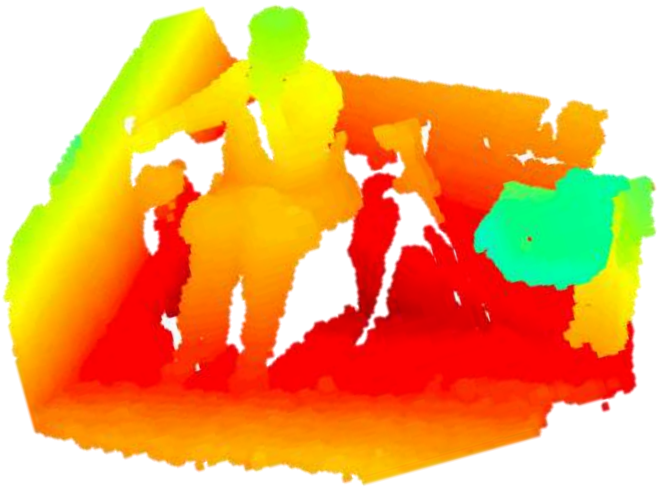
(Raw) depth images



Skeletons/mesh from depth



Point clouds from depth



Extracted from [12]

[12] Martínez-González, A., Villamizar, M., Canévet, O., & Odobez, J. M. (2020, October). Residual pose: A decoupled approach for depth-based 3D human pose estimation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 10313-10318). IEEE.

RQ2. Different inputs for behaviour measurement

What **combination of computer vision deep learning models and input data types** (raw depth images, skeletons, or point clouds) yields the best performance for processing depth image sequences?

There is no conclusion on which combination works better

→ **Exploration of different models + inputs**

Table 2. Action recognition accuracy (%) on NTU RGB+D 60 [45] and NTU RGB+D 120 [30].

Method	Input	NTU RGB+D 60		NTU RGB+D 120	
		Subject	View	Subject	Setup
SkeleMotion [2]	skeleton	69.6	80.1	67.7	66.9
GCA-LSTM [33]	skeleton	74.4	82.8	58.3	59.3
FSNet [31]	skeleton	-	-	59.9	62.4
Two Stream Attention LSTM [32]	skeleton	77.1	85.1	61.2	63.3
Body Pose Evolution Map [35]	skeleton	-	-	64.6	66.9
AGC-LSTM [48]	skeleton	89.2	95.0	-	-
AS-GCN [26]	skeleton	86.8	94.2	-	-
VA-fusion [64]	skeleton	89.4	95.0	-	-
2s-AGCN [47]	skeleton	88.5	95.1	-	-
DGNN [46]	skeleton	89.9	96.1	-	-
HON4D [40]	depth	30.6	7.3	-	-
SNV [62]	depth	31.8	13.6	-	-
HOG^2 [39]	depth	32.2	22.3	-	-
Li *et al.* [25]	depth	68.1	83.4	-	-
Wang *et al.* [57]	depth	87.1	84.2	-	-
MVDI [61]	depth	84.6	87.3	-	-
NTU RGB+D 120 Baseline [30]	depth	-	-	48.7	40.1
PointNet++ (appearance) [43]	point	80.1	85.1	72.1	79.4
3DV (motion) [59]	voxel	84.5	95.4	76.9	92.5
3DV-PointNet++ [59]	voxel + point	88.8	96.3	82.4	**93.5**
P4Transformer (ours)	point	**90.2**	**96.4**	**86.4**	**93.5**

Extracted from [4]

[4] Fan, H., Yang, Y., & Kankanhalli, M. (2021). Point 4d transformer networks for spatio-temporal modeling in point cloud videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14204-14213).

Comparison of different inputs

Depth has been forgotten….

Accuracy

	MSRDailyActivity3D	N–UCLA	UWA3D II	NTU RGB+D 60 (cs)	NTU RGB+D 60 (cv)
Depth	85 (Depthmaps CNN, 2015)	92 (Novel Viewpoints, 2016)	76.9 (Novel Viewpoints, 2016)	87.1 (Depth pooling, 2018)	87.3 (MVDI, 2018)
Other modalities	97.5 (DSSCA-SSLM, 2018 - RGB + D)	97.6 (LA-GCN, 2023 - Pose)	81.4 (VA-fusion, 2019 - Pose)	97.0 (PoseC3D, 2022 - RGB + Pose)	99.6 (PoseC3D, 2022 - RGB + Pose)

Next steps

1. Exploration and development of depth-based methods for activity recognition
2. Exploration and (development?) of depth-based methods for 3D pose estimation
3. Assessment of performance with real-world data
4. Comparison of models

Summary of (expected) main contributions

RQ1. Robust performance for real-world HAR

- General: **novel methods for domain generalization for point cloud sequences**
- Specific: robust methods dealing with real-world data for behaviour measurement for people with dementia

RQ2. Different inputs for behaviour measurement

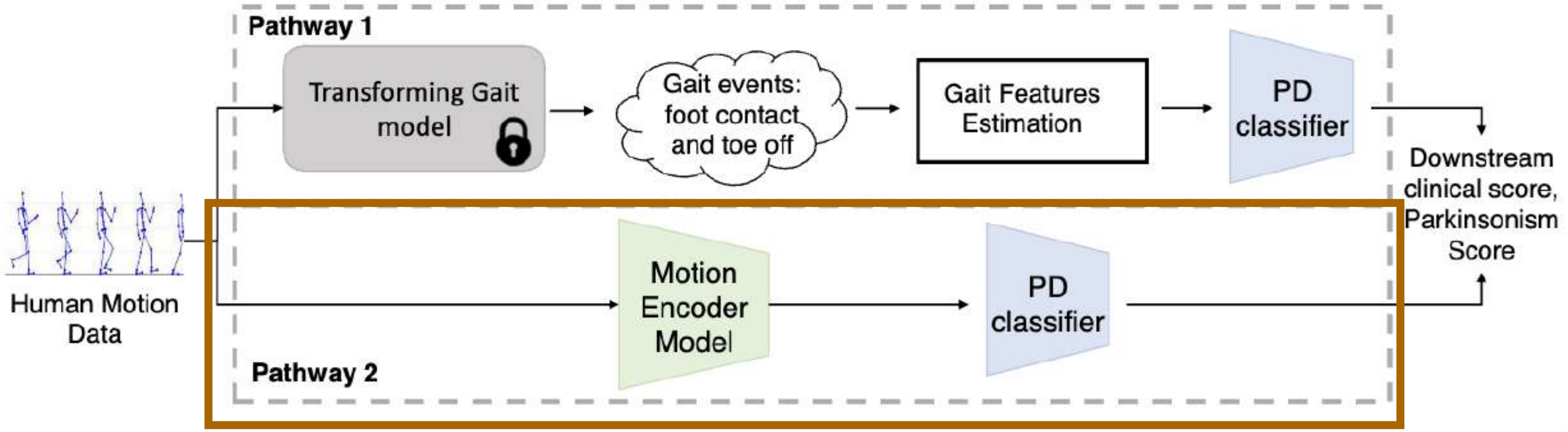
- Novel models for pose estimation and activity recognition from depth sequences
- Comparison of methods-inputs

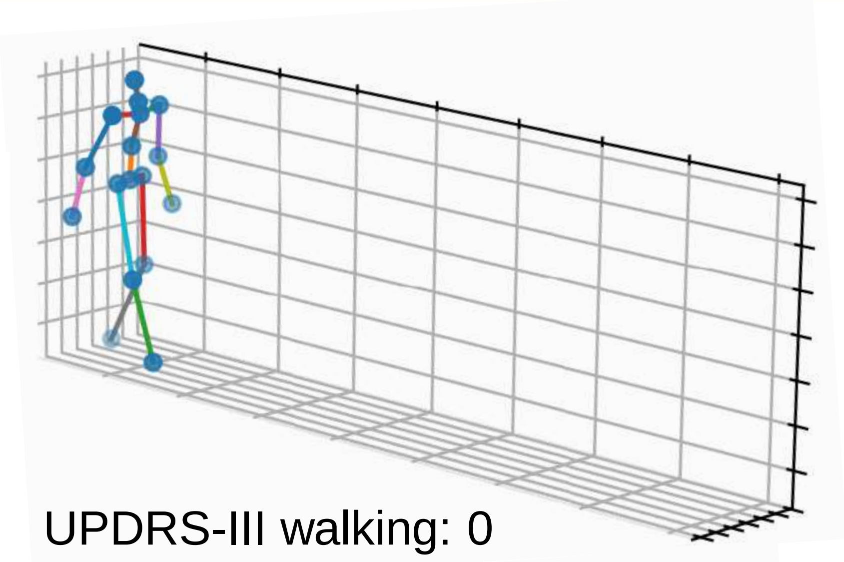
RQ3. Providing assistance with ADLs in the short term

- **ToiletHelp**: first-of-its-kind system to guide people with dementia in the toilet
- Guidelines for developing assistive systems for people with dementia

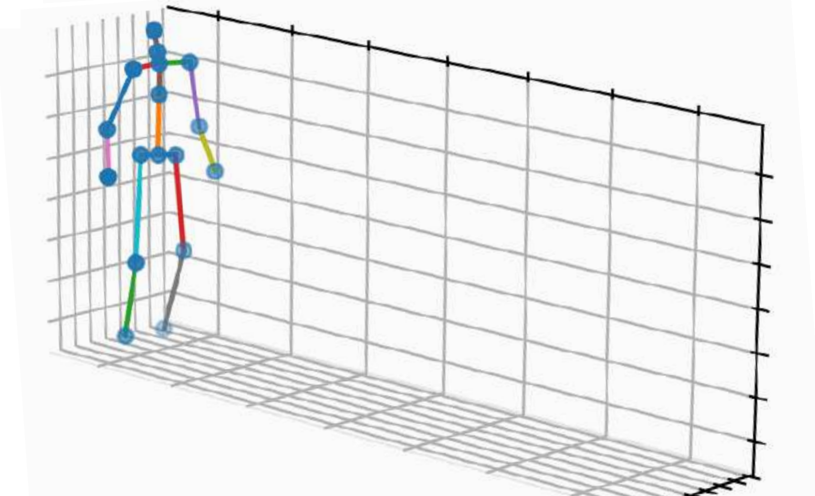
Secondment at UofT:
Gait Analysis for Parkinsonism Score Estimation from Motion-Capture data

Goal: compare performance **of hand-craft feature extraction** vs. **end-to-end deep learning** approach





UPDRS-III walking: 0



UPDRS-III walking: 2

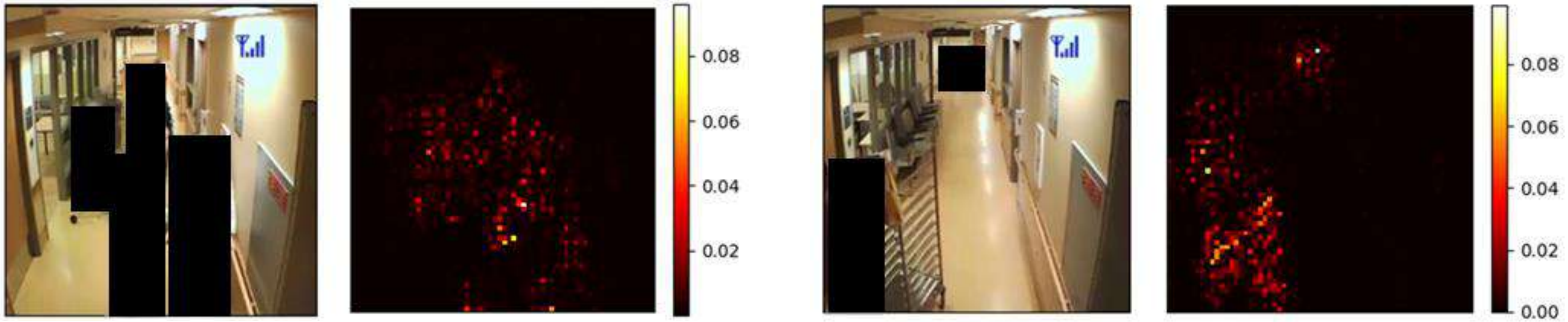
- Data set for gait analysis in Parkinson's disease[13]:
  26 subjects with UPDRS-III walking scores (0,1,2)
- 5 SoTA motion encoders (e.g. MotionBert[14], MotionAGFormer[15])

[13] Shida, T. K. F., Costa, T. M., de Oliveira,... & Coelho, D. B. (2023). A public data set of walking full-body kinematics and kinetics in individuals with Parkinson's disease. *Frontiers in Neuroscience*, *17*, 992585.
[14] Zhu, W., Ma, X., Liu, Z., Liu, L., Wu, W., & Wang, Y. (2023). MotionBERT: A unified perspective on learning human motion representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 15085-15099).
[15] Mehraban, S., Adeli, V., & Taati, B. (2023). MotionAGFormer: Enhancing 3D Human Pose Estimation with a Transformer-GCNFormer Network. *arXiv preprint arXiv:2310.16288*.

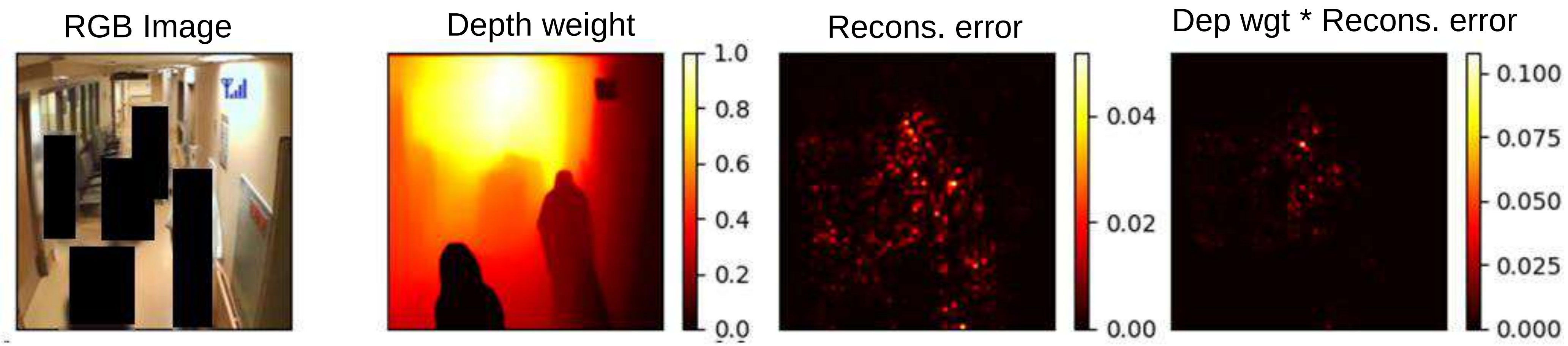
Secondment at UofT:
Agitation Detection in People with Dementia

Goal: reduction of false positives in an autoencoder-based method for anomaly detection

Main: Objects/people close to the camera



Solution: **depth-weighted loss**



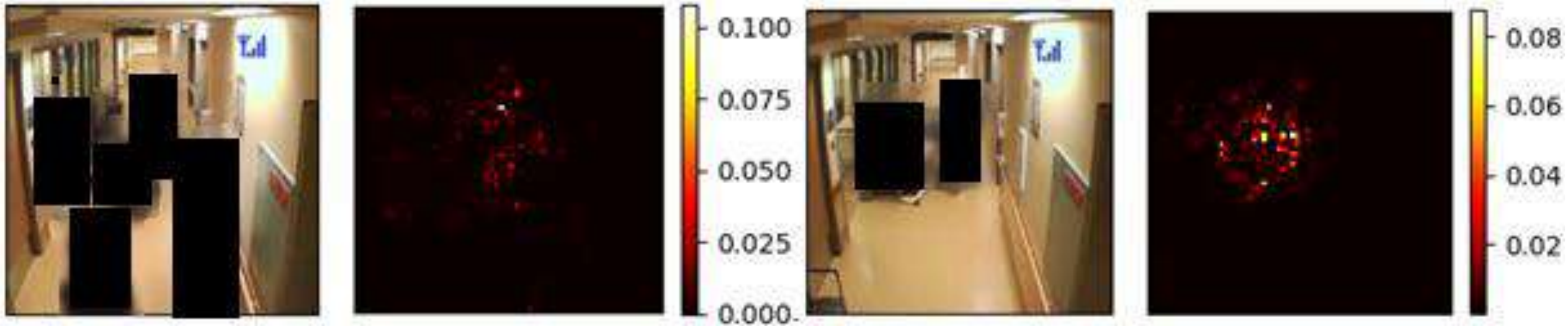
Secondment at UofT:
Agitation Detection in People with Dementia

Goal: reduction of false positives in an autoencoder-based method for anomaly detection

Solution: depth-weighted loss

✔ Common outlier windows:

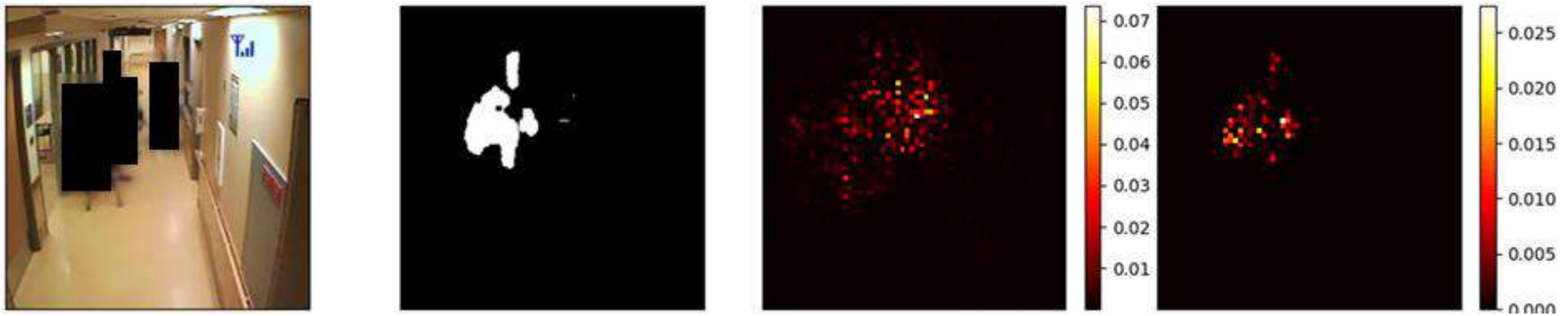
- Top 30: #0
- Top 40:#5
- Top 100: #20
- Top 1000: #704



New problem: crowded scenes far from camera

Solution II: Normalization using active pixels

RGB Image | Active pixels | Recons. error | Active pixels * Recons. error



Estimated timeline

	2021			2022				2023				2024				2025			
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
ToiletHelp - Interaction with people with dementia			Pervasive Health Paper																
Collaboration with DIANA project		MidTerm Deliv.						End Project Deliv.											
ToiletHelp - Functional Evaluation				AHFE Paper															
Proficiency Evaluation					Research Proposal Subm.	Present.													
ToiletHelp - Evaluation										JAIHC paper	PerCom Workshop Paper								
Domain Generalization for PC												ECCV paper			CVPR/ICCV paper?				
Activity recognition from depth sequences													CVPR-W paper						
Pose estimation from depth												ICPR paper							
Secondments					Alicante						UofT	FG paper	VAD paper						
Thesis writing																		Defense	

PhD Seminar in Aachen

PhD Seminar in Stockholm

Today

Thank you!

Irene Ballester Campos

TU Wien

irene.ballester@tuwien.ac.at

Summary of publications to date

Published:

Ballester, I., Kampel, M. (2022). Automated vision-based toilet assistance for people with dementia. In: Matteo Zallio (eds) Human Factors in Accessibility and Assistive Technology. AHFE (2022) International Conference. AHFE Open Access, vol 37. AHFE International, USA.

Ballester, I., Mujirishvili, T., Kampel, M. (2022). RITA: A Privacy-Aware Toileting Assistance Designed for People with Dementia. In: *Lewy, H., Barkan, R. (eds) Pervasive Computing Technologies for Healthcare. PH 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 431. Springer, Cham.

No proceedings

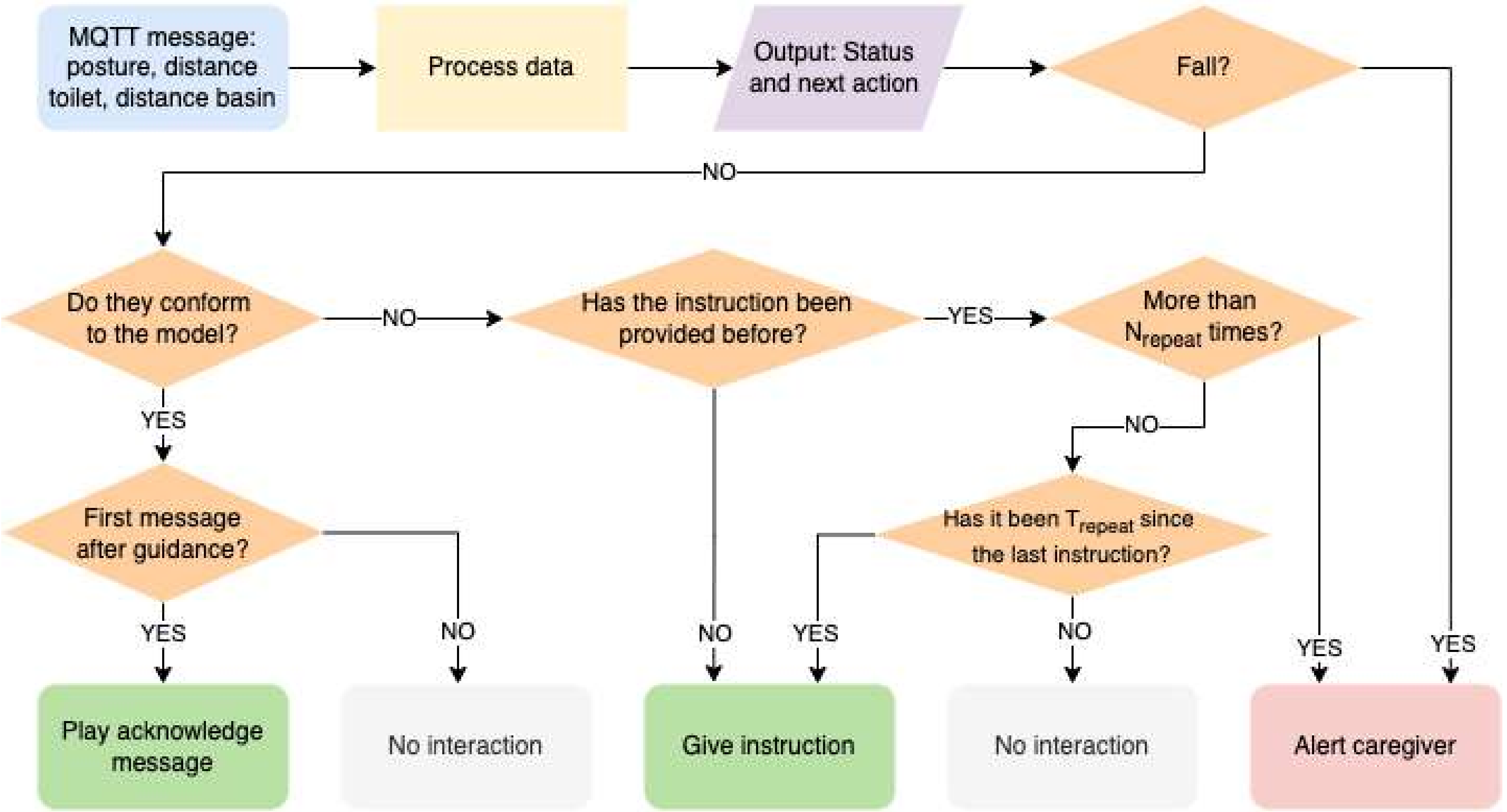
Ballester I., Gall M., Kampel M. (2023) “Interaction Design of a Toileting Assistive System for People with Dementia“, presented at WISH Symposium at the ACM CHI Conference on Human Factors in Computing Systems.

Submitted (under revision):

Ballester, I., Gall, M., Münzer, T., and Kampel, M., “ToiletHelp: an assistive technology to guide people with dementia in the toilet, *Journal of Ambient Intelligence and Humanized Computing*, 2023.

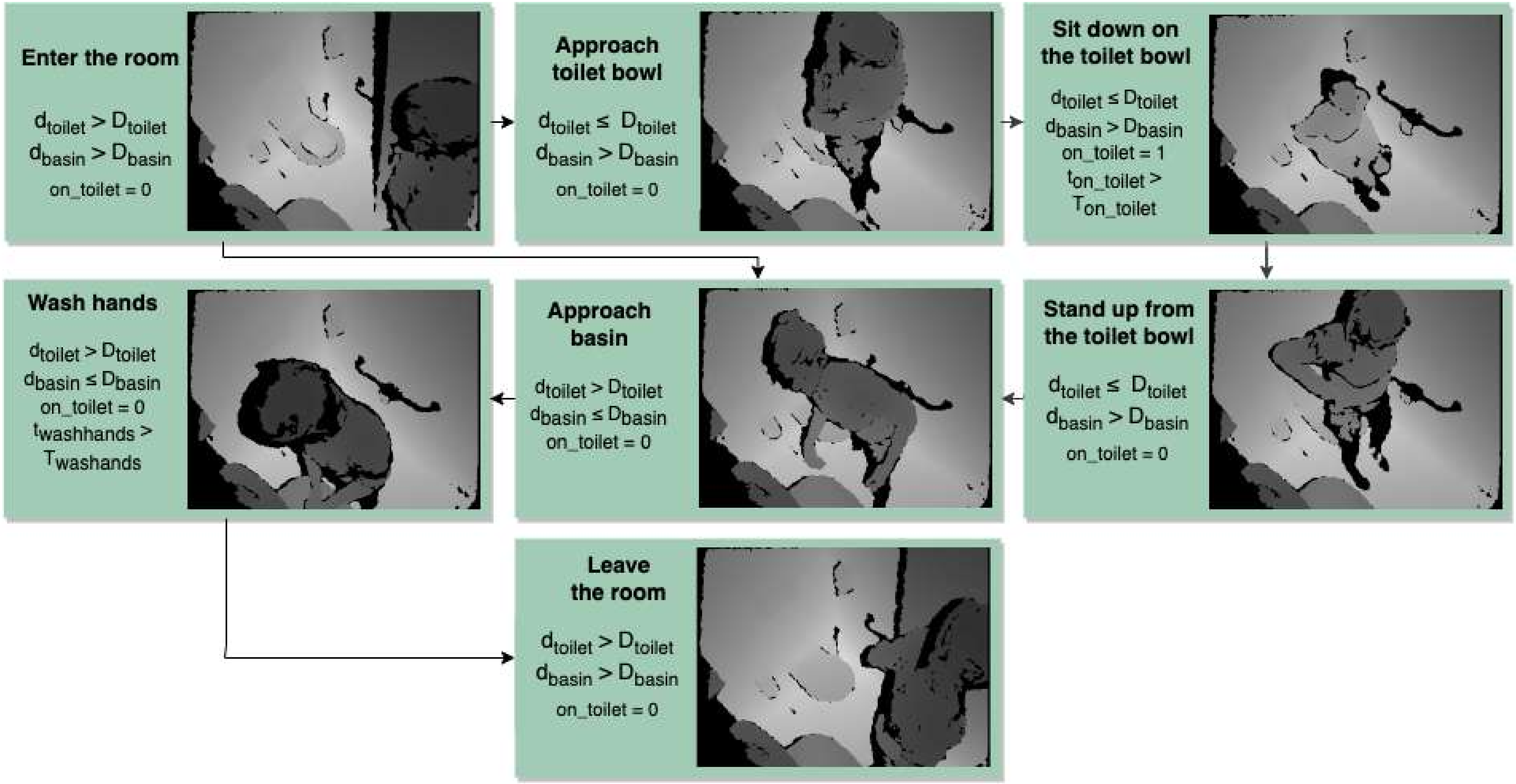
Ballester, I., Gall, M., Münzer, T., and Kampel, M.,, “Vision-Based Toilet Assistant for People with Dementia in Real-Life Situations” *Pervasive-Agetech Workshop at PERCOM2024*

ToiletHelp



visuAAL

ToiletHelp


Enter the room
$d_{toilet} > D_{toilet}$
$d_{basin} > D_{basin}$
on_toilet = 0
Approach toilet bowl
$d_{toilet} \leq D_{toilet}$
$d_{basin} > D_{basin}$
on_toilet = 0
Sit down on the toilet bowl
$d_{toilet} \leq D_{toilet}$
$d_{basin} > D_{basin}$
on_toilet = 1
$t_{on\_toilet} > T_{on\_toilet}$
Wash hands
$d_{toilet} > D_{toilet}$
$d_{basin} \leq D_{basin}$
on_toilet = 0
$t_{washhands} > T_{washands}$
Approach basin
$d_{toilet} > D_{toilet}$
$d_{basin} \leq D_{basin}$
on_toilet = 0
Stand up from the toilet bowl
$d_{toilet} \leq D_{toilet}$
$d_{basin} > D_{basin}$
on_toilet = 0
Leave the room
$d_{toilet} > D_{toilet}$
$d_{basin} > D_{basin}$
on_toilet = 0

Evaluation of the toilet action recognition module

Actions evaluated

1. The person is about to sit
2. The person sat down
3. The person stood up
4. The person is about to wash their hands

Inclusion criteria:

- Participant sits on the toilet
- Only one person in the room
- The room is empty at the beginning and the end of the sequence

Description of participants:

- 6 male, 14 female
- Age range: 60-97
- 19 participants with dementia, 1 without
- 6 participants use walking aids

Evaluation of the toilet action recognition module

Action is detected correctly if:

$$H1 < fID\_detected - fID\_groundtruth < H2$$

- **fID_detected**: frame in which the system detects the action
- **fID_groundtruth**: mean of the frames labelled by 3 annotators: mean(An1, An2, An3)

How to calculate the thresholds H1, H2?

- **H1= - uncertainty_action + min_system_delay   H2= uncertainty_action + max_system_delay**
  - Uncertainty_action = SD(An1, An2,An3) for each action-patient
  - System_delay = frames the system waits to ensure a robust recognition

Evaluation of the toilet action recognition module

Success assessment: Sensor *vs.* annotators

Example: S1 (Hosp), Action=Sitting down

H1 < fID_detected – fID_groundtruth < H2

	An1	An2	An3	Avrg = GT	Std Dev	Sensor	Sensor-GT	H2-H1	Success?
R 1	408	413	419	413.3	5.5	400	-13.3	35.0	
R 2	343	340	349	344.0	4.6	344	0.0	33.2	
R 3	455	453	458	455.3	2.5	400	-55.3	29.0	
R 4	436	433	437	435.3	2.1	440	4.7	28.2	
R 5	458	448	455	453.7	5.1	440	-13.7	34.3	
R 6	442	436	439	439.0	3.0	428	-11.0	30.0	
R 7	452	452	461	455.0	5.2	424	-31.0	34.4	
R 8	432	432	455	439.7	13.3	436	-3.7	50.6	
R 9	456	456	460	457.3	2.3	456	-1.3	28.6	
R 10	368	363	373	368.0	5.0	372	4.0	34.0	
Average	425.0	422.6	430.6	426.1	4.9	414.0	-12.1	33.7	
Std. Dev.	40	40.0	39.4	39.6		34.9	18.6	6.5	

Accuracy: 90%

EXTRA: Toilet Module: What has been done since then until now?

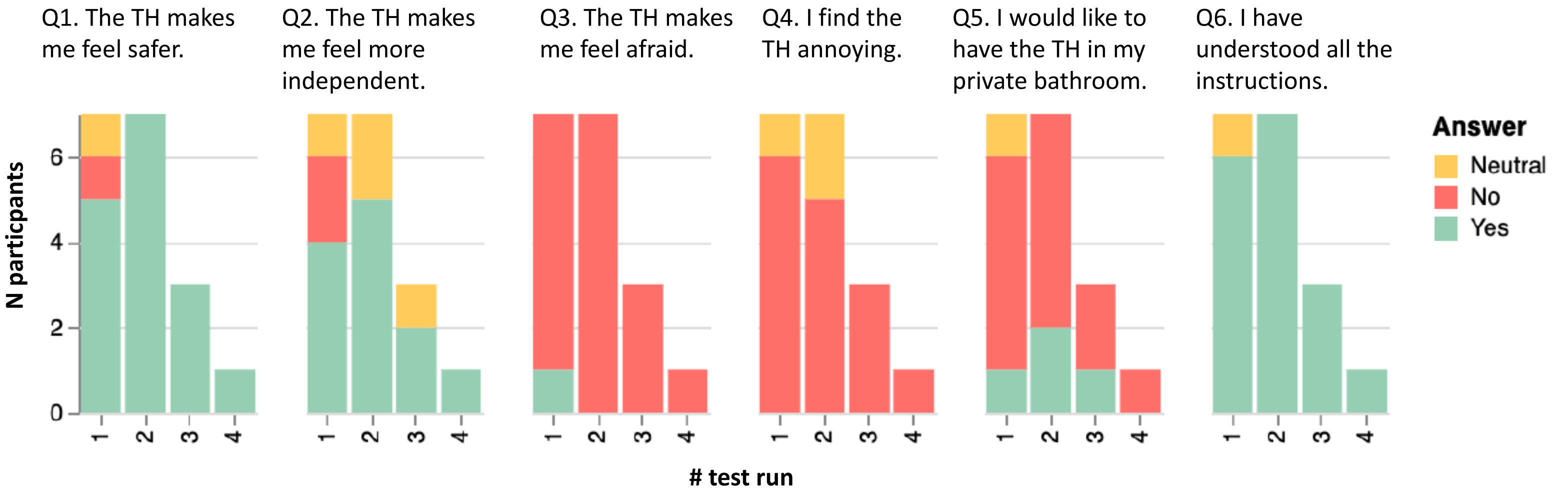
Evaluation of the interaction module

End-users in DC1 (St. Gallen, CH)

N = 17 older adults attending 1-2 days/week the day care centre

- 11 women, 6 men
- Dementia severity (6 measured by MMSE, 11 measured by MoCA) :
  - Mild: 11
  - Moderate: 5
  - Normal cognitive functioning: 1
- Mean age = 78,6. SD = 7.2, max. = 86, min. = 60
- Mean score of affinity for new technologies = 1.5 (0-2)

	1st time runs			All test runs		
	Yes	Neutral	No	Yes	Neutral	No
Q1. The TH makes me feel safer.	11	3	2	22	3	2
Q2. The TH makes me feel more independent.	11	0	5	19	3	5
Q3. The TH makes me feel afraid.	1	0	15	1	0	26
Q4. I find the TH annoying.	1	1	14	1	3	23
Q5. I would like to have the TH in my private bathroom.	2	1	13	5	1	21
Q6. I have understood all the instructions.	14	2	1	25	2	1



Times the participants tried the system: 1 time: 17, 2 times: 7, 3 times: 3, 4 times: 1

EXTRA: Toilet Module: What has been done since then until now?

Evaluation of the interaction module

End-users in DC2 (Coimbra, PT)

N = 13 older adults attending the day centre daily

- 9 women, 4 men
- No cognitive status reported
- Age distribution
  - 71-80 years: 3 participants
  - 81-90 years: 6 participants
  - +90 years: 4 participants
- Mean score of affinity for new technologies = 4.2 (SD=0.9, scale=1-5)

(1=I do not agree at all, 10= Totally agree)

	1st time runs (N=13)		2-4 time runs (N=8)		All test runs (N=21)	
	Mean	SD	Mean	SD	Mean	SD
Q1. The TH would make me feel safer.	**8,5**	2,7	**9,3**	0,7	**8,8**	2,2
Q2. The TH would make me feel more independent.	**8,5**	4,2	**8,8**	0,9	**8,6**	3,3
Q3. The TH would make me feel comfortable.	**8,3**	2,5	**8,9**	0,8	**8,5**	2,0
Q4. The TH would make me feel afraid.	**2,2**	2,5	**1,3**	0,7	**1,8**	2,1
Q5. I find the TH annoying.	**1,4**	0,9	**1,3**	0,5	**1,3**	0,7
Q6. I would prefer to be assisted by the TH rather than by a caregiver.	**5,7**	3,5	**6,0**	3,6	**5,8**	3,5
Q7. I would like to have the TH in my private bathroom.	**7,9**	2,9	**8,5**	1,9	**8,1**	2,6

Evaluation of the interaction module

Care staff in DC1 and DC2

N=14 care staff (N=4 from DC1 and N=10 from DC2)

- All women
- Affinity for technology (scale: 1-5)
  - Total: mean=4.1, SD=0.9
  - DC1: mean=3.3, SD=0.8
  - DC2: mean=4.4, SD=0.7

DC1				DC2			
Aged		Work experience		Aged		Work experience	
21-30 y.	0	<5 y.	0	21-30 y.	4	<5 y.	3
31-40 y.	0	6 to 10 y.	0	31-40 y.	1	6 to 10 y.	1
41-50 y.	1	10 to 16 y.	0	41-50 y.	3	10 to 16 y.	0
51-60 y.	3	16 to 20 y.	1	51-60 y.	1	16 to 20 y.	2
61-70 y.	0	>20 y.	3	61-70 y.	1	>20 y.	4

	Total (N=14)		DC1 (N=4)		DC2 (N=10)	
	Mean	SD	Mean	SD	Mean	SD
Q1. How useful would you rate the use of the TH for older adults with cognitive impairment?	**8,1**	1,7	**7,5**	0,6	**8,4**	2,0
Q2. How useful would you rate the use of the TH for reducing the workload for healthcare workers?	**8,1**	1,9	**6,3**	1,3	**8,7**	1,6
Q3. How adequate do you find the interaction modalities used in the TH to guide the user?	**8,2**	1,3	**7,5**	1,3	**8,5**	1,3
Q4. Is the vocabulary in the TH adequate to guide the user?	**8,8**	1,3	**8,5**	1,9	**8,9**	1,0
Q5. Is the tone used in the TH adequate to guide the user?	**8,6**	1,2	**8,3**	1,7	**8,7**	1,1
Q6. Are the videos used in the TH adequate to guide the user?	**8,4**	0,9	**8,8**	1,0	**8,3**	0,9
Q7. How intrusive do you think the TH is compared to a caregiver guiding the person with dementia?	**4,9**	3,1	**7,7**	1,5	**4,1**	3,0

Depth Weighted Loss

$$(AS)^{area} = \overline{(AS)}^{pixel} * A \quad (1)$$

In case of anomaly associated with a physical entity in camera view, its $(AS)^{area}$ should be the same regardless of its relative position to the camera.

We want,

$$(AS)_1^{area} = (AS)_2^{area} \quad (2)$$

what we have,

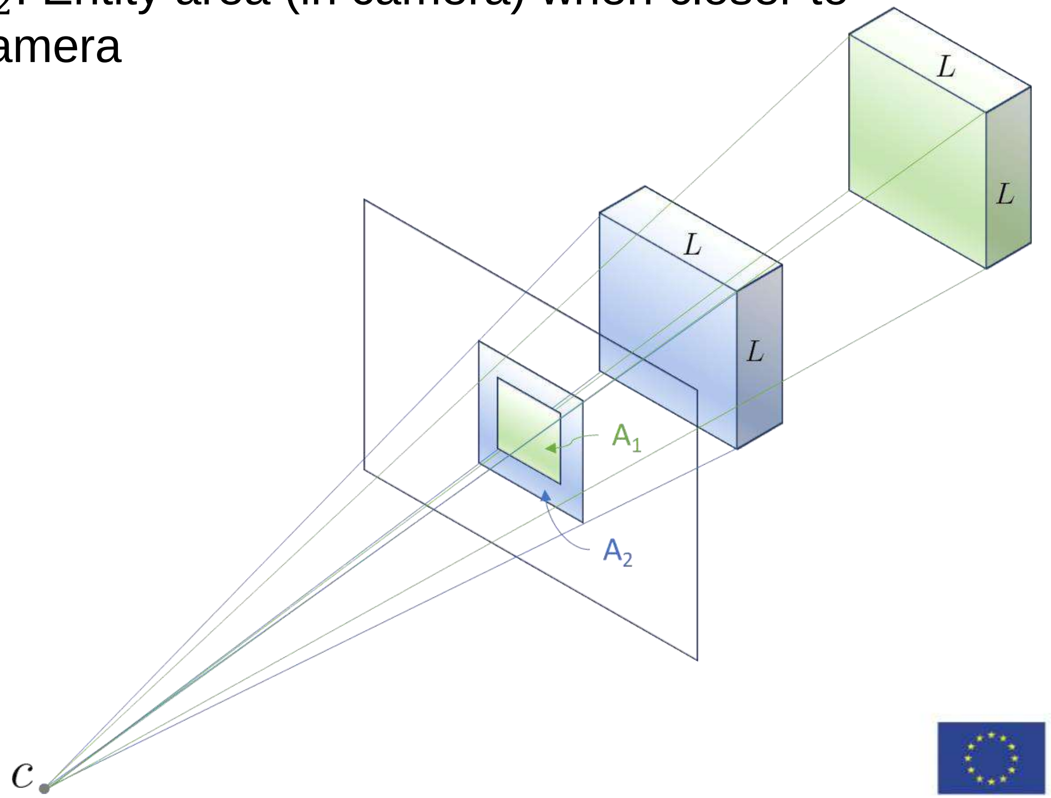
$$A_2 > A_1 \quad (3)$$

$$\overline{(AS)}_1^{pixel} = \overline{(AS)}_2^{pixel} = \overline{(AS)}^{pixel} \quad (4)$$

Thus, we need factors K1 and K2 to achieve (2):

$$A_1 * K_1 = A_2 * K_2 \quad (5)$$

AS: Anomaly score
$(AS)^{area}$: Total anomaly score of the area
$\overline{(AS)}^{pixel}$: Average reconstruction error per pixel
A : Area
A_1: Entity area (in camera) when away from camera
A_2: Entity area (in camera) when closer to camera



Depth Weighted Loss

How can A be calculated?

$$A = (u_2 - u_1)^2 \quad (6)$$

Image projection $u = f\frac{x}{z}$

where u : 2D image coordinates,
f : focal length (camera intrinsic),
x : 3D world coordinates,
z : depth

$$u_1 = f * \frac{x_1}{z_1}; u_2 = f * \frac{x_2}{z_2} \quad (7)$$

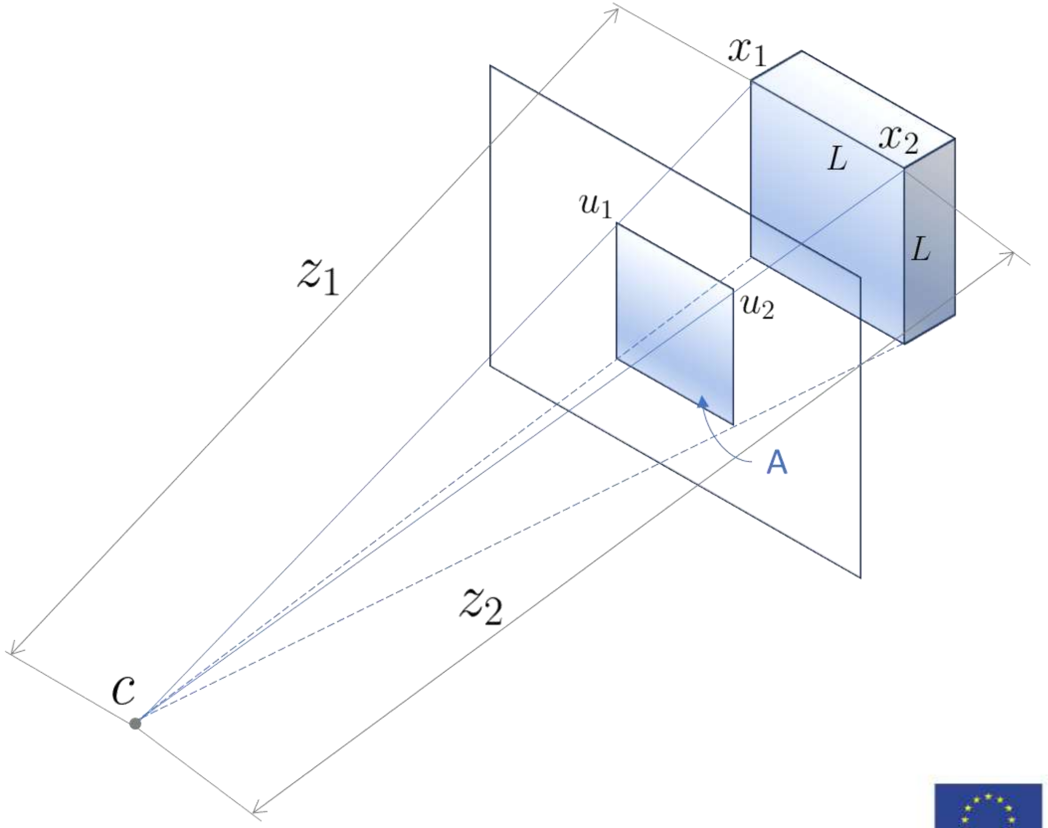
If we assume $z_1 \approx z_2 = z$. Thus:

$$A = (u_2 - u_1)^2 = (\frac{f}{z})^2(x_2 - x_1)^2 \quad (8)$$

$$A = (fL)^2(\frac{1}{z^2}) \quad (9)$$

L: $x_2 - x_1$

f and L are constants



Depth Weighted Loss

From (5):

$$K_1(fL)^2\frac{1}{{z_1}^2} = K_2(fL)^2\frac{1}{{z_2}^2} \qquad (10)$$

Thus,

$$K_1 = {z_1}^2; K_2 = {z_2}^2 \qquad (11)$$

→ In order to achieve depth invariance, we must multiply reconstruction error for each pixel by a factor equal to z^2