CORSMAL
Collaborative object recognition, shared manipulation and learning

Audio-visual object classification for human-robot collaboration

Poster number: 9401

Alessio Xompero, Yik Lung Pang, Timothy Patten, Ahalya Prabhakar, Berk Calli, Andrea Cavallaro
{a.xompero,y.l.pang,a.cavallaro}@qmul.ac.uk, timothy.patten@uts.ed.au, ahalya.prabhakar@epfl.ch, bcalli@wpi.edu

1. Introduction

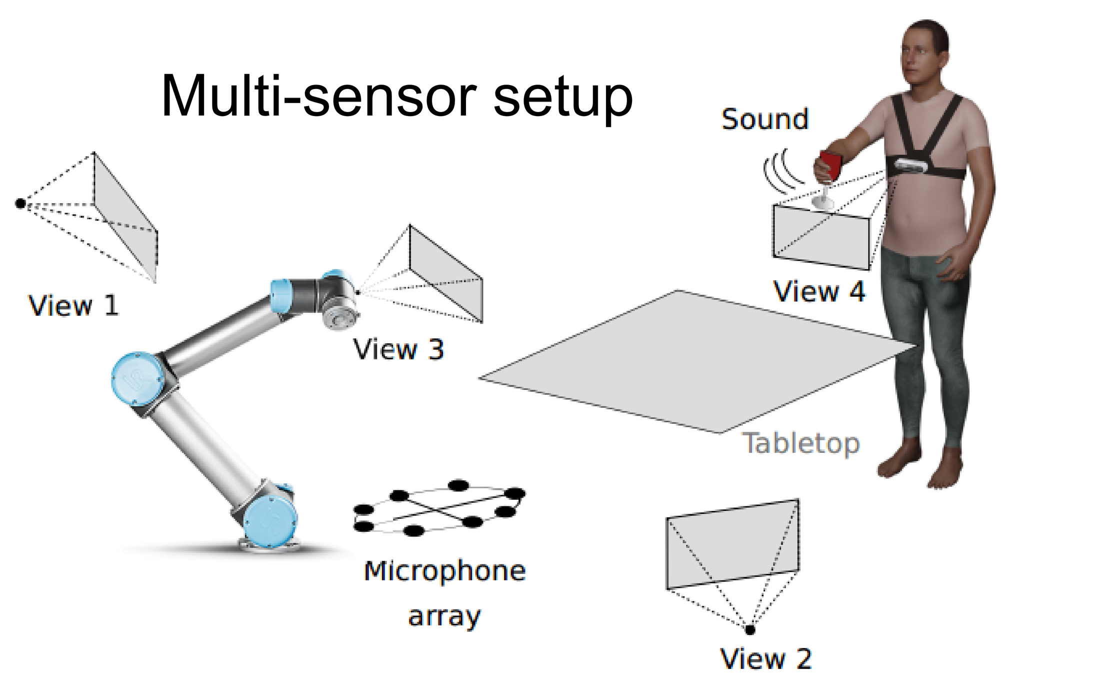
Assistive scenarios at home or workplace (human-to-robot handovers)

Challenges:
- varying physical properties of household containers
- different types and amount of contents
- hand occlusions

CORSMAL simulator (R2S) [1]:
- assessing the accuracy of the estimations
- visualizing the safeness of human-to-robot handovers

http://corsmal.eecs.qmul.ac.uk/challenge.html

2. Dataset and tasks

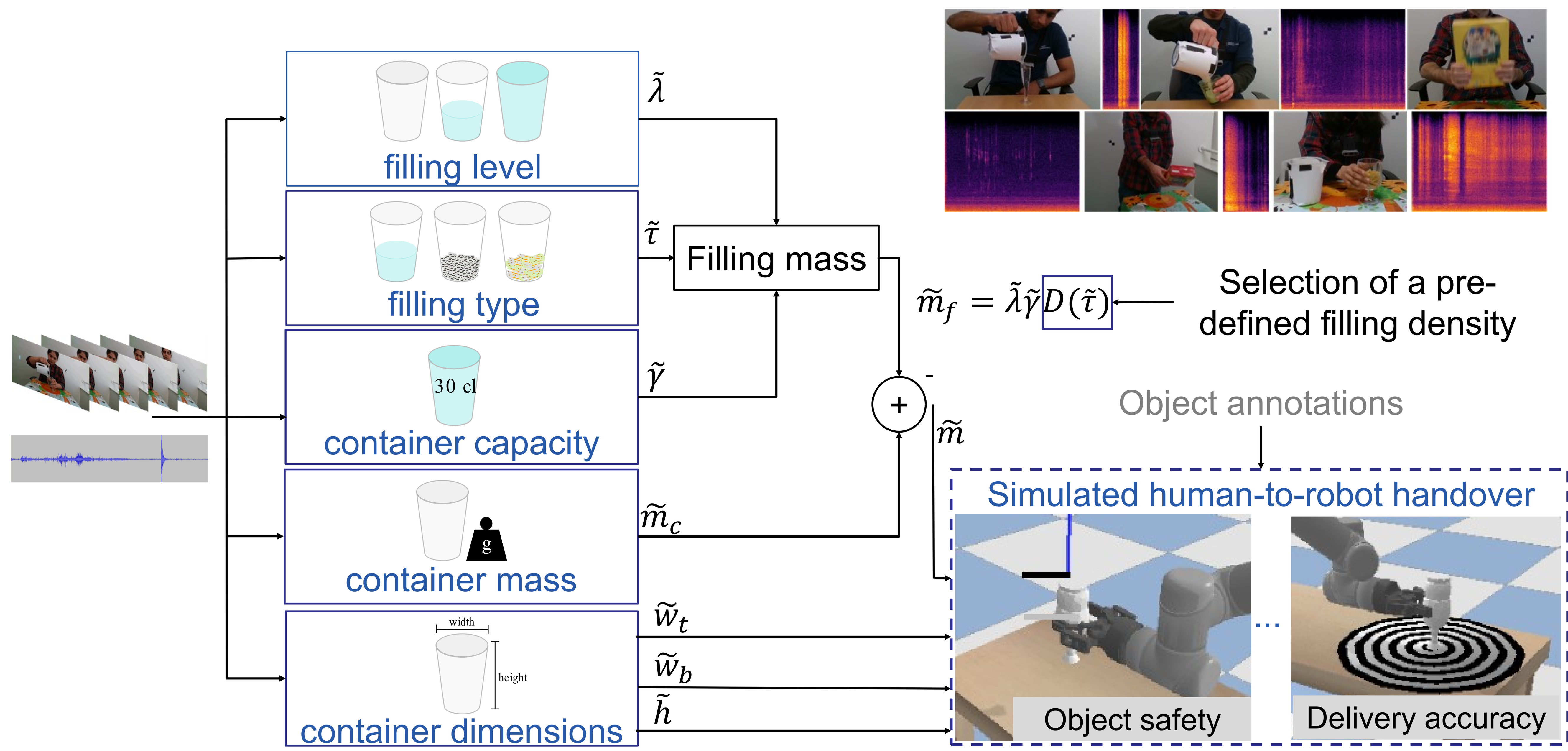


CORSMAL Containers Manipulation (CCM)

1,140 configurations
3 scenarios

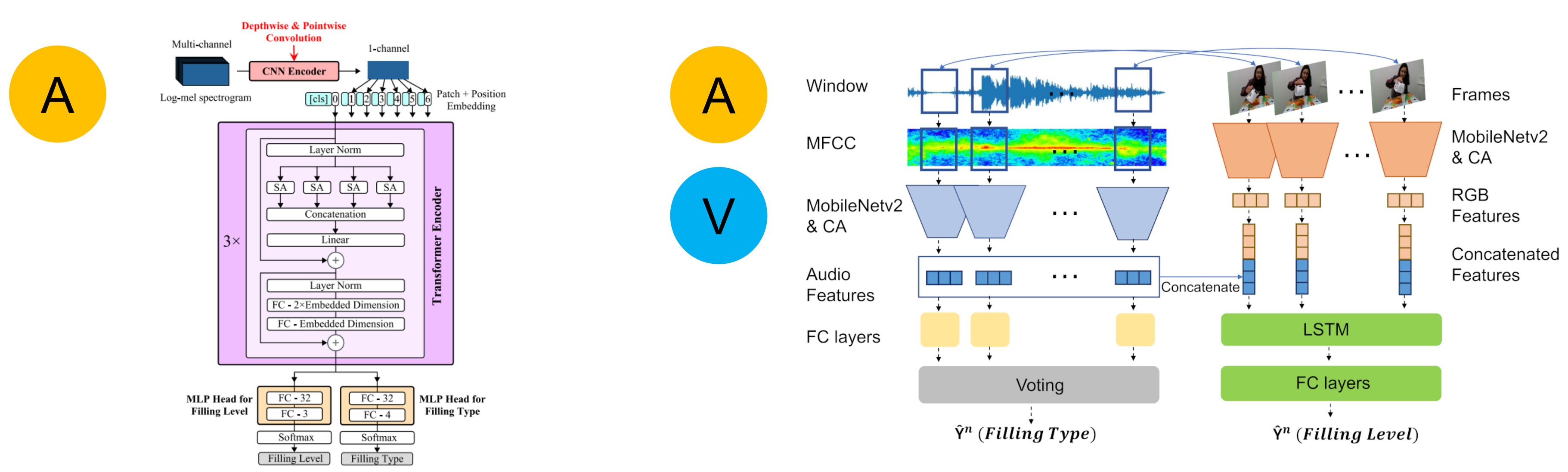
15 containers
3 filling types
3 filling levels

https://doi.org/10.17636/101CORSMAL1

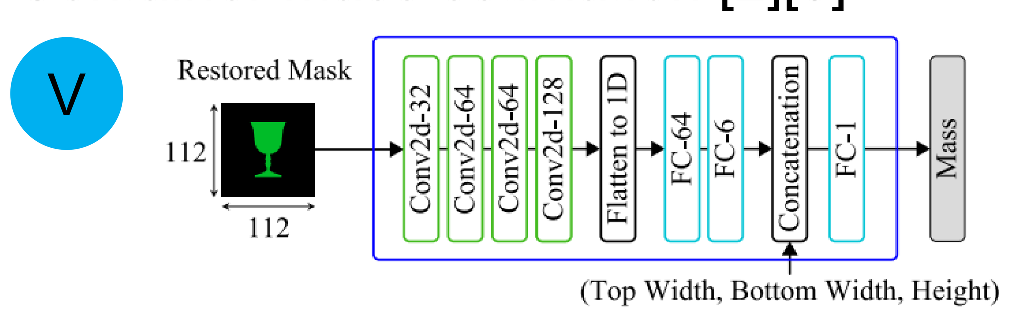


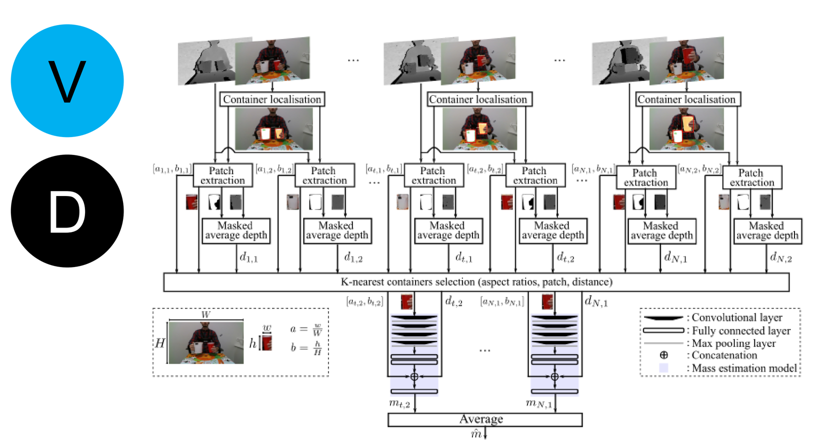
3. IEEE ICASSP 2022 Challenge entries

Filling type and level classification [3][4]

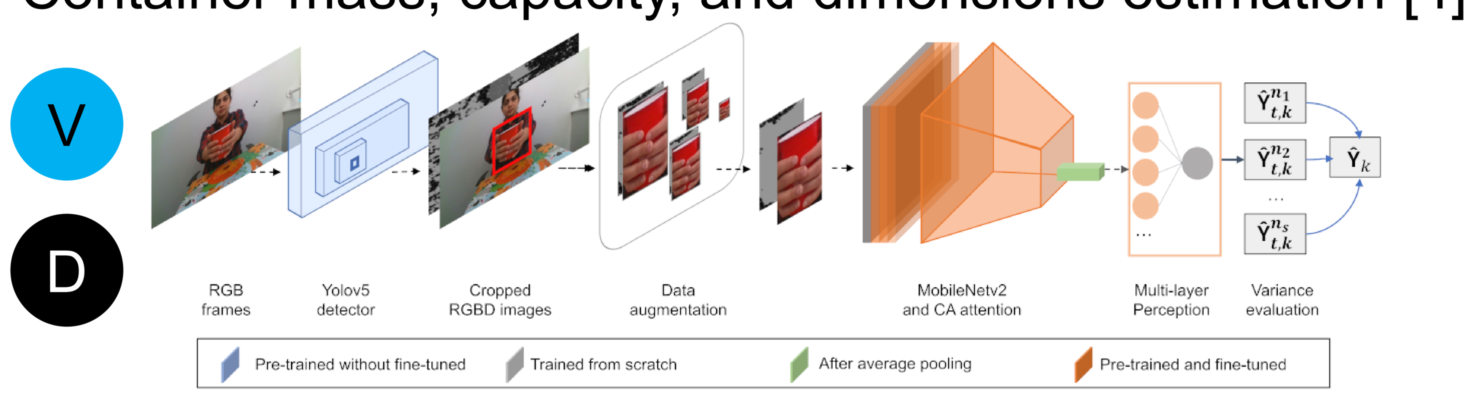


Container mass estimation [2][3]





Container mass, capacity, and dimensions estimation [4]



A audio
V RGB
D depth

4. Performance scores and challenge results (combined CCM test sets)

T1	T2	T3	T4	T5	Description	Unit	Measure	Score	Weight	Type	R2S	Random	Average	[2]	[3]	[4]
●	○	○	○	○	Filling level		λ^j	$s_1 = \bar{F}_1(\lambda^1, \dots, \lambda^J, \hat{\lambda}^1, \dots, \hat{\lambda}^J)$	$\pi_1 = 1/8$	Direct	○	37.62	33.15	-	65.73	**77.40**
○	●	○	○	○	Filling type		τ^j	$s_2 = \bar{F}_1(\tau^1, \dots, \tau^J, \hat{\tau}^1, \dots, \hat{\tau}^J)$	$\pi_2 = 1/8$	Direct	○	24.38	23.01	-	80.72	**99.13**
○	○	●	○	○	Capacity	mL	γ^j	$s_3 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j e^{-\varepsilon^j(\gamma^j, \hat{\gamma}^j)}$	$\pi_3 = 1/8$	Direct	○	24.58	40.73	-	**72.26**	59.51
○	○	○	●	○	Container mass	g	m_c^j	$s_4 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j e^{-\varepsilon^j(m_c^j, \hat{m}_c^j)}$	$\pi_4 = 1/8$	Direct	○	29.42	22.06	49.64	40.19	**58.78**
○	○	○	○	●	Width at the top	mm	w_t^j	$s_5 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j \sigma_1(w_t^j, \hat{w}_t^j)$	$\pi_5 = 1/24$	Direct	○	32.33	76.89	-	69.09	**80.01**
○	○	○	○	●	Width at the bottom	mm	w_b^j	$s_6 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j \sigma_1(w_b^j, \hat{w}_b^j)$	$\pi_6 = 1/24$	Direct	○	25.36	58.19	-	59.74	**76.09**
○	○	○	○	●	Height	mm	h^j	$s_7 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j \sigma_1(h^j, h^j)$	$\pi_7 = 1/24$	Direct	○	42.48	64.32	-	70.07	**74.33**
●	●	●	○	○	Filling mass	g	m_f^j	$s_8 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j e^{-\epsilon^j(m_f^j, \hat{m}_f^j)}$	$\pi_8 = 1/8$	Indirect	○	35.06	42.31	-	**70.50**	65.25
●	●	●	●	●	Object mass	g	m^j	$s_9 = \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_j \psi^j(m^j, \hat{m}^j)$	$\pi_9 = 1/8$	Indirect	●	56.31	58.30	53.54	60.41	**71.19**
●	●	●	●	●	Pose at delivery	(mm,°)	(α^j, β^j)	$s_{10} = \frac{1}{J}\sum_{j=1}^{J} \Delta_j(\alpha^j, \beta^j, \eta, \phi)$	$\pi_{10} = 1/8$	Indirect	●	72.11	70.01	60.54	73.17	**79.32**
●	●	○	○	○	Joint filling type and level			$s_{11} = \bar{F}_1(\lambda^1, \tau^1, \dots, \hat{\lambda}^1, \hat{\tau}^1, \dots)$	-	Direct	○	10.49	8.88	-	59.32	**78.16**
○	○	●	○	●	Container capacity and dimensions			$s_{12} = s_3/2 + (s_5 + s_6 + s_7)/6$	-	Direct	○	28.99	53.60	-	**69.28**	68.16
●	●	●	●	●	Overall score			$S = \sum_{l=1}^{10} \pi_l s_l$	-	Indirect	-	39.11	44.51	-	66.16	**73.43**

$\mathbb{I}$: indicator function

T: task

5. Conclusion

- A framework to design audio-visual solutions for the estimation of physical properties of manipulated containers
- 5 tasks, 13 performance scores, 10 leaderboards (facilitate comparisons)
- Simulator: assessing the accuracy of the estimations & visualizing the safeness of human-to-robot handovers

References

[1] Y. Pang, A. Xompero, C. Oh, A. Cavallaro, "Towards safe human-to-robot handovers of unknown containers", *IEEE RO-MAN*, 2021
[2] T. Apicella, et al., "Container localization and mass estimation with an RGB-D camera", *IEEE ICASSP*, 2022
[3] T. Matsubara, et al., "Shared transformer encoder with mask-based 3D model estimation for container mass estimation", *IEEE ICASSP*, 2022
[4] H. Wang, et al., "Improving generalization of deep networks for estimating physical properties of containers and fillings ", *IEEE ICASSP*, 2022

Acknowledgment

This work is supported by the CHIST-ERA program through the project CORSMAL, under UK EPSRC grant EP/. S031715/1 and Swiss NSF grant 20CH21_180444.

chist-era