chist-era

CORSMAL

Collaborative Object Recognition, Shared Manipulation And Learning

ORMR - Object Recognition and Manipulation by Robots: data sharing and experiment reproducibility (2019-2022)

CHIST-ERA Projects Seminar 2021

Queen Mary University of London

SORBONNE UNIVERSITÉ

EPFL

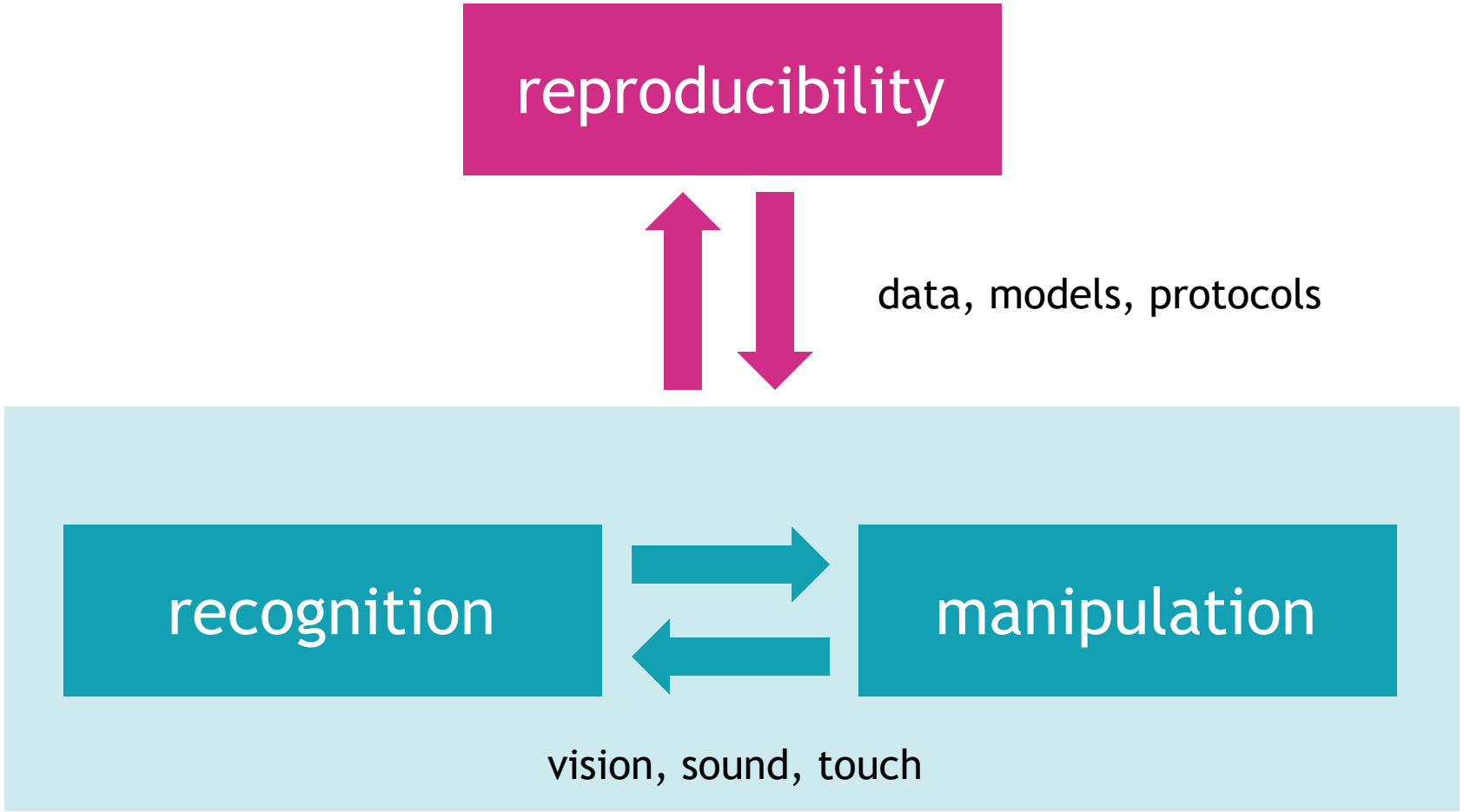
Aims

To create an open dataset and an evaluation protocol for the recognition and manipulation of previously unseen object instances

To explore multiple sensing modalities (touch + sound + vision) to accurately and robustly estimate the physical properties of objects

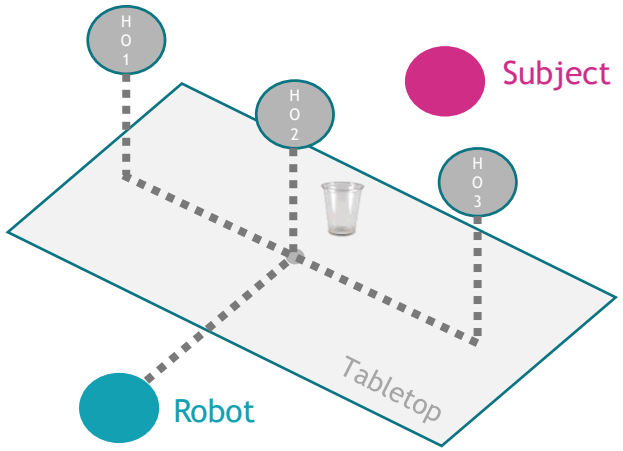
The task

Human grasping → Human manipulation → Human-robot handover → Robot manipulation → Robot delivery


reproducibility
data, models, protocols
recognition
manipulation
vision, sound, touch

reproducibility

Benchmark and **baseline algorithms** for human-to-robot handovers distributed to (and used by) the community



CORSMAL *Protocol and Code*
github.com/CORSMAL/Benchmark
ieeexplore.ieee.org/document/8968407
IEEE RA-L
Open Access

Datasets

CORSMAL *Containers*
doi.org/10.17636/101CORSMAL1

CORSMAL *Containers Manipulation*
doi.org/10.17636/corsmal2

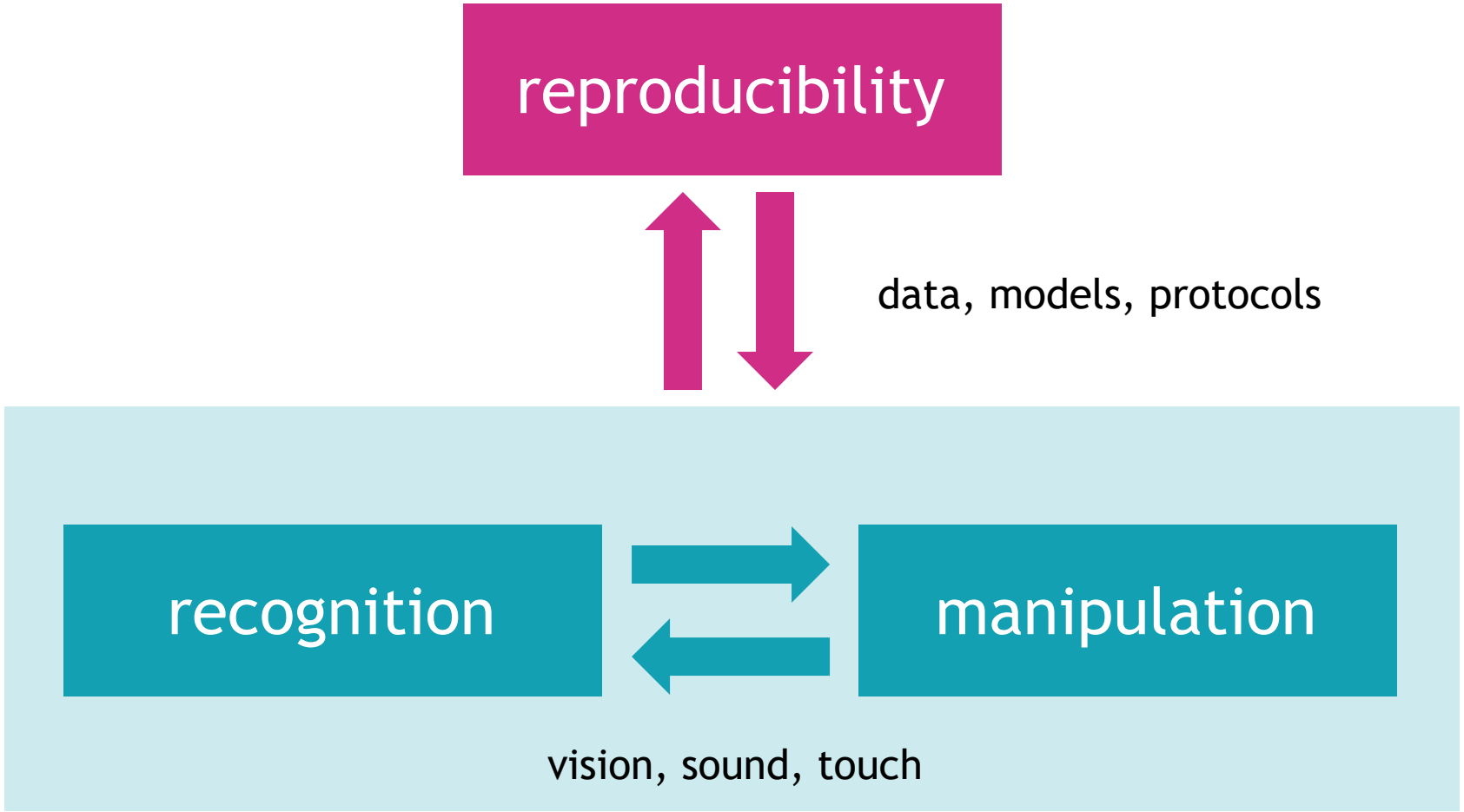
Models

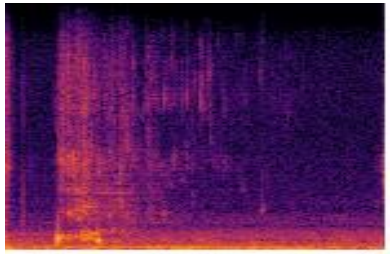
CORSMAL *Pre-trained Models*
zenodo.org/record/4518951#.YC9-z-qnw5k

Platform for participants to submit the results of the **CORSMAL Challenge** and progressive update of the **Leaderboard**

Team	Description	Task 1	Task 2	Task 3	Public	Private	Overall
Because It's Tacti...	GRU+ Random Forest for filling properties estimation. LoDE with RGB-D-IR data from selected frames in a video for volume estimation.	✓	✓	✓	64.98	65.15	65.06
HVRL	Log-Mel spectrogram-based audio features as input to VGG-based CNN and LSTM for filling properties estimation. Container volume from the shape approximation as cuboid of the 3D point cloud obtained with RGB-D data and object detection with Mask R-CNN.	✓	✓	✓	63.32	61.01	62.16
Concatenation	Multi-modal learning with audio features and prior of container categories through object detection for inferring container capacity and fluid properties.	✓	✓	✓	52.80	54.14	53.47
NTNU-ERC	MFCC features in a 20s-window + neural network to classify filling type. Object detection and selection of the closest contours (up to 700 mm) in the depth data + regression with a CNN for container capacity.		✓	✓	38.56	39.80	39.18
Random	Baseline with random estimations for each task.	✓	✓	✓	38.47	31.65	35.06
Challengers	Sound-based classification of filling type and level with STFT and 5-layers fully connected neural network.	✓	✓		29.25	23.21	26.23
SCC-Net	Sound-based hierarchical ensemble of DNNs to jointly classify filling type and level.	✓	✓		28.02	22.92	25.47
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		19.46	9.59	14.53
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		15.15	9.96	12.56
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		17.28	6.99	12.14
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		12.95	10.25	11.60

CORSMAL *Challenge and Evaluation Toolkit*
corsmal.eecs.qmul.ac.uk/challenge.html
github.com/CORSMAL/CORSMALChallengeEvalToolkit

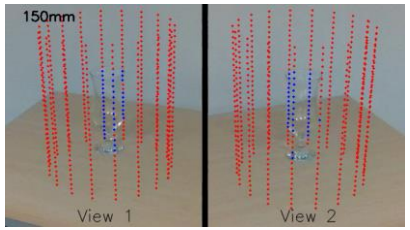

reproducibility
data, models, protocols
recognition
manipulation
vision, sound, touch



Sound-based Classification of the Content Level and Type in a Container

Improving Filling Level Classification with Adversarial Training

Redundant Features Can Hurt Robustness to Distribution Shift



Multi-view Shape Estimation of Transparent Containers

On Force Synergies in Human Grasping Behaviour

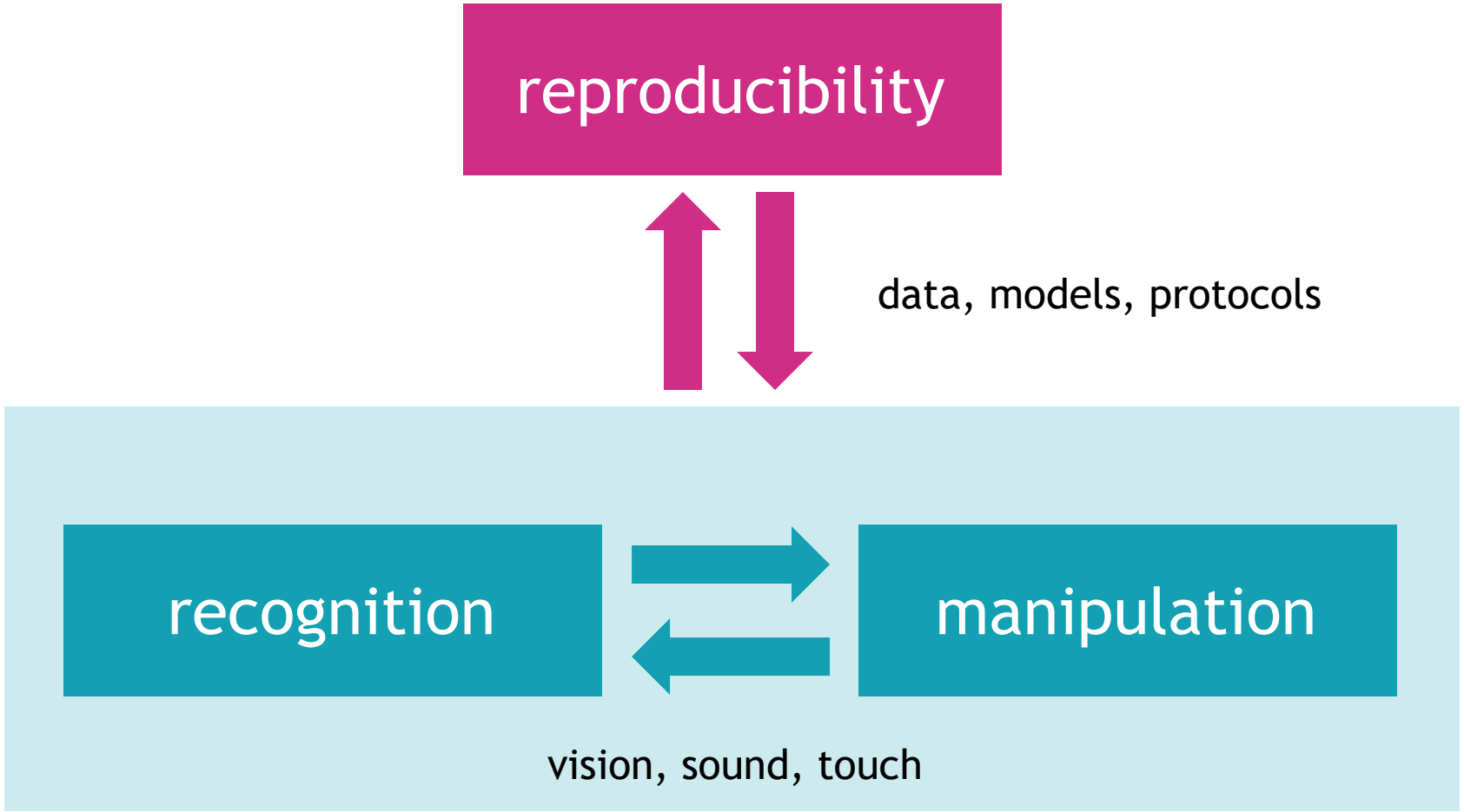
From Human Action Understanding to Robot Action Execution: How the Physical Properties of Handled Objects Modulate Non-verbal Cues

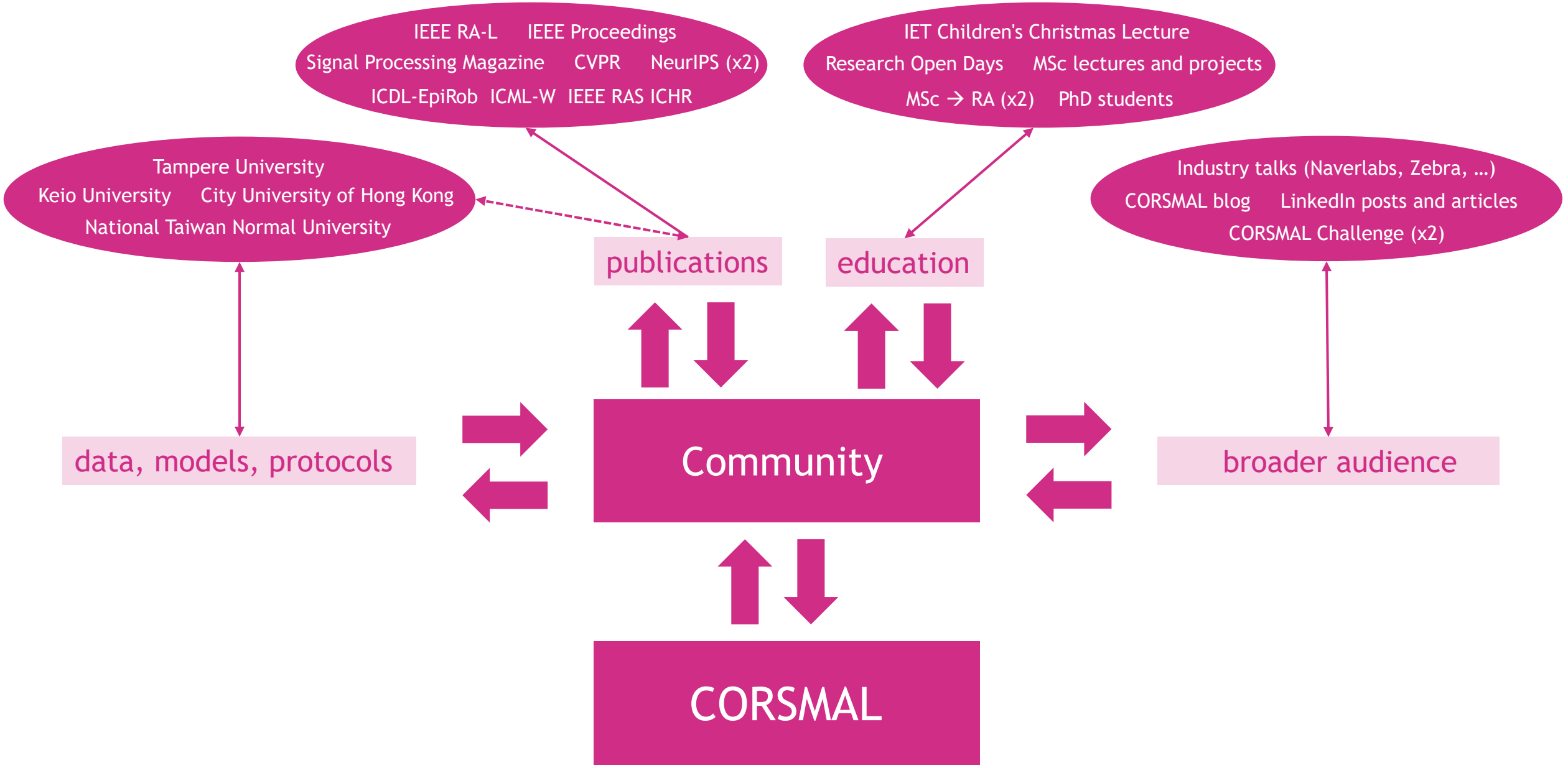
recognition ⇄ manipulation

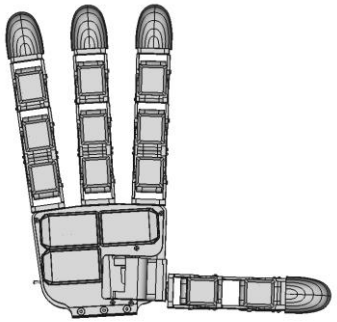
vision, sound, touch

Full list of publications at: corsmal.eecs.qmul.ac.uk/publications.html

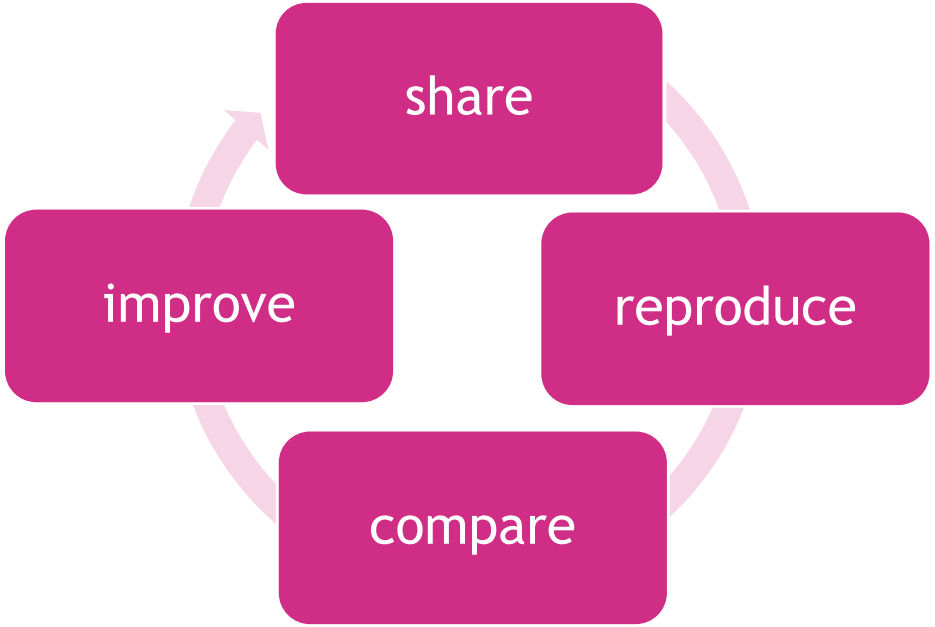
Includes publications by external teams that used CORSMAL datasets, protocols & models and, in turn, shared their annotations & models!

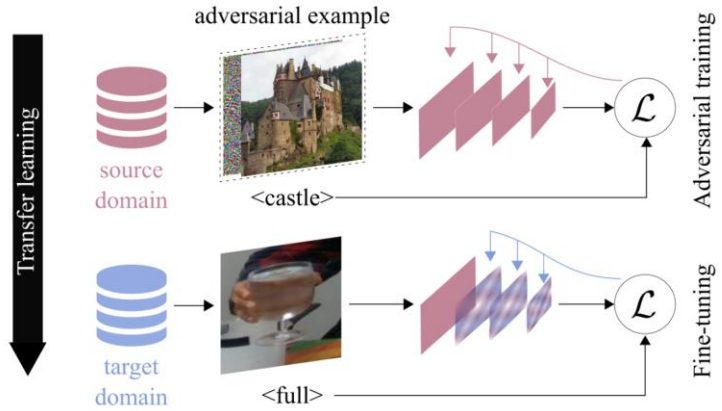

reproducibility
data, models, protocols
recognition
manipulation
vision, sound, touch


IEEE RA-L
IEEE Proceedings
Signal Processing Magazine
CVPR
NeurIPS (x2)
ICDL-EpiRob
ICML-W
IEEE RAS ICHR
IET Children's Christmas Lecture
Research Open Days
MSc lectures and projects
MSc → RA (x2)
PhD students
Tampere University
Keio University
City University of Hong Kong
National Taiwan Normal University
Industry talks (Naverlabs, Zebra, …)
CORSMAL blog
LinkedIn posts and articles
CORSMAL Challenge (x2)
publications
education
data, models, protocols
Community
broader audience
CORSMAL



Gripper → Hand





Shared Models → Transfer Learning

Team	Description	Task 1	Task 2	Task 3	Public	Private	Overall
Because It's Tacti...	GRU+ Random Forest for filling properties estimation. LoDE with RGB-D-IR data from selected frames in a video for volume estimation.	✓	✓	✓	64.98	65.15	65.06
HVRL	Log-Mel spectrogram-based audio features as input to VGG-based CNN and LSTM for filling properties estimation. Container volume from the shape approximation as cuboid of the 3D point cloud obtained with RGB-D data and object detection with Mask R-CNN.	✓	✓	✓	63.32	61.01	62.16
Concatenation	Multi-modal learning with audio features and prior of container categories through object detection for inferring container capacity and fluid properties.	✓	✓	✓	52.80	54.14	53.47
NTNU-ERC	MFCC features in a 20s-window + neural network to classify filling type. Object detection and selection of the closest contours (up to 700 mm) in the depth data + regression with a CNN for container capacity.		✓	✓	38.56	39.80	39.18
Random	Baseline with random estimations for each task.	✓	✓	✓	38.47	31.65	35.06
Challengers	Sound-based classification of filling type and level with STFT and 5-layers fully connected neural network.	✓	✓		29.25	23.21	26.23
SCC-Net	Sound-based hierarchical ensemble of DNNs to jointly classify filling type and level.	✓	✓		28.02	22.92	25.47
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		19.46	9.59	14.53
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		15.15	9.96	12.56
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		17.28	6.99	12.14
Mask R-CNN + R...	Vision baseline for filling properties estimation.	✓	✓		12.95	10.25	11.60

Compare → Leaderboad

CORSMAL
Collaborative object recognition, shared manipulation and learning

Summary

- Benchmark for Human-to-Robot Handovers
- Evaluation protocol
- Evaluation toolkit
- Baseline code + pre-trained models
- Datasets
  - CORSMAL Containers dataset
  - CORSMAL Containers Manipulation dataset
- Events at
  - IEEE Int. Conf. on Multimedia and Expo
  - Int. Conf. on Pattern Recognition
  - Intelligent Sensing Summer School
- **Join our community: participate in the CORSMAL Challenge!**

Partners

Queen Mary University of London

SORBONNE UNIVERSITÉ

EPFL

Sponsors

EPSRC Engineering and Physical Sciences Research Council

Agence Nationale de la Recherche ANR

FNSNF

chist-era

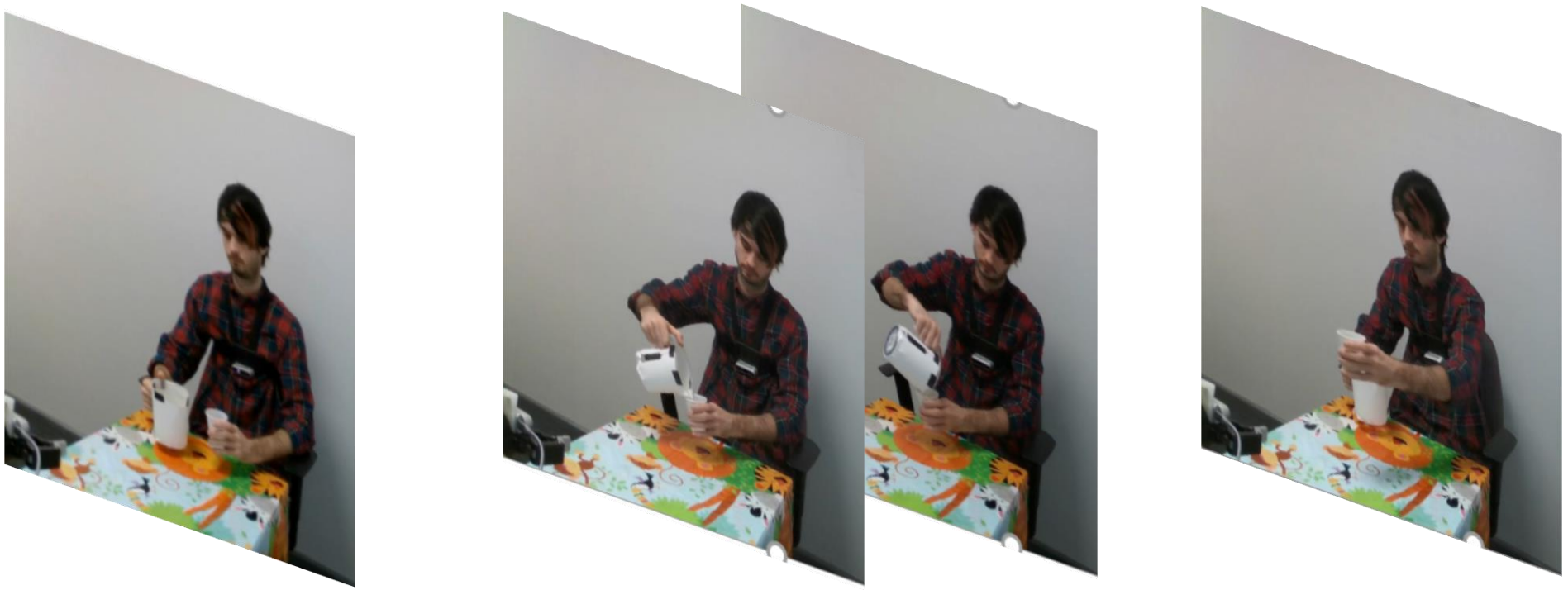
Contact:
a.cavallaro@qmul.ac.uk

CORSMAL
Collaborative object recognition, shared manipulation and learning

Human holds the container

Human pours a filling in the container

Human gives the container



Human holds the container

Human shakes the container

Human gives the container