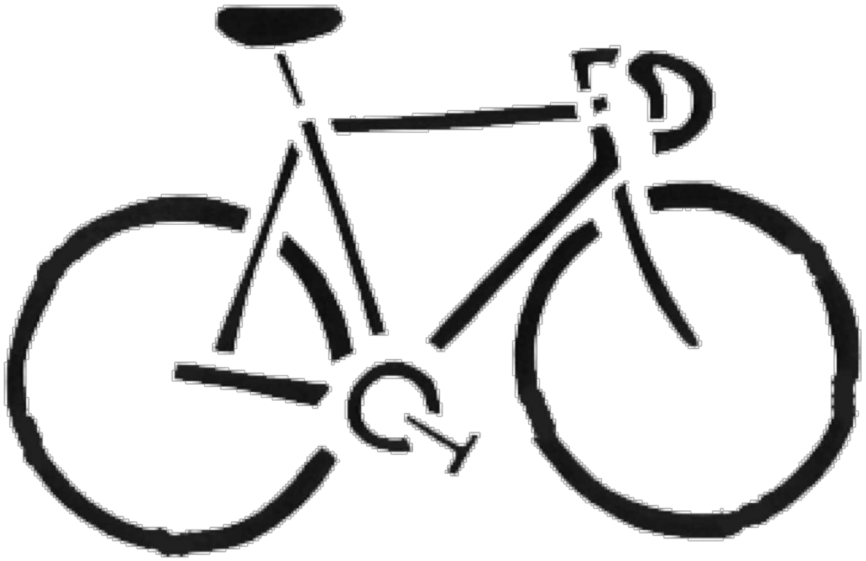
A fine-grained perspective onto object interactions

University of BRISTOL

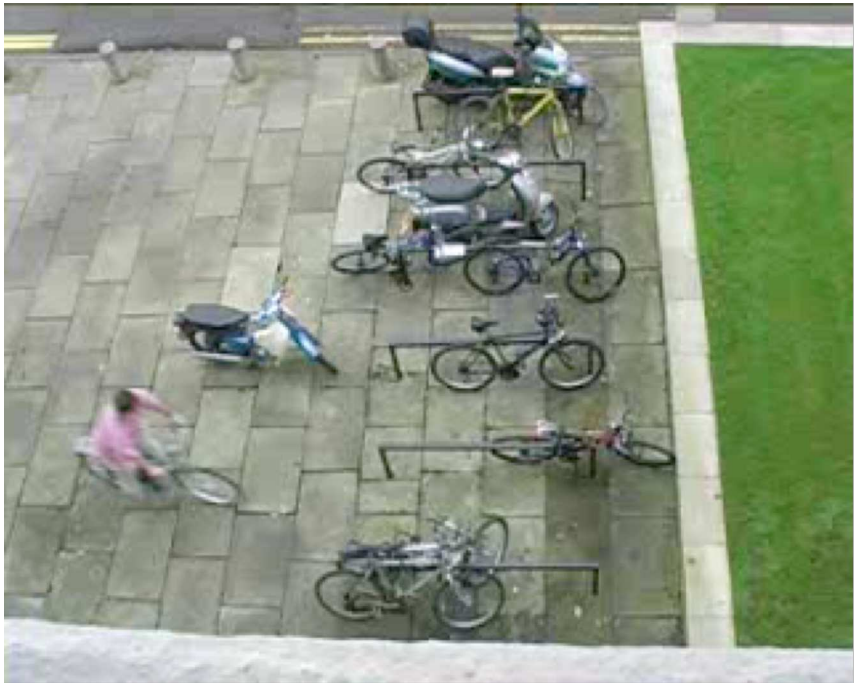
Dima Damen

15 March 2019

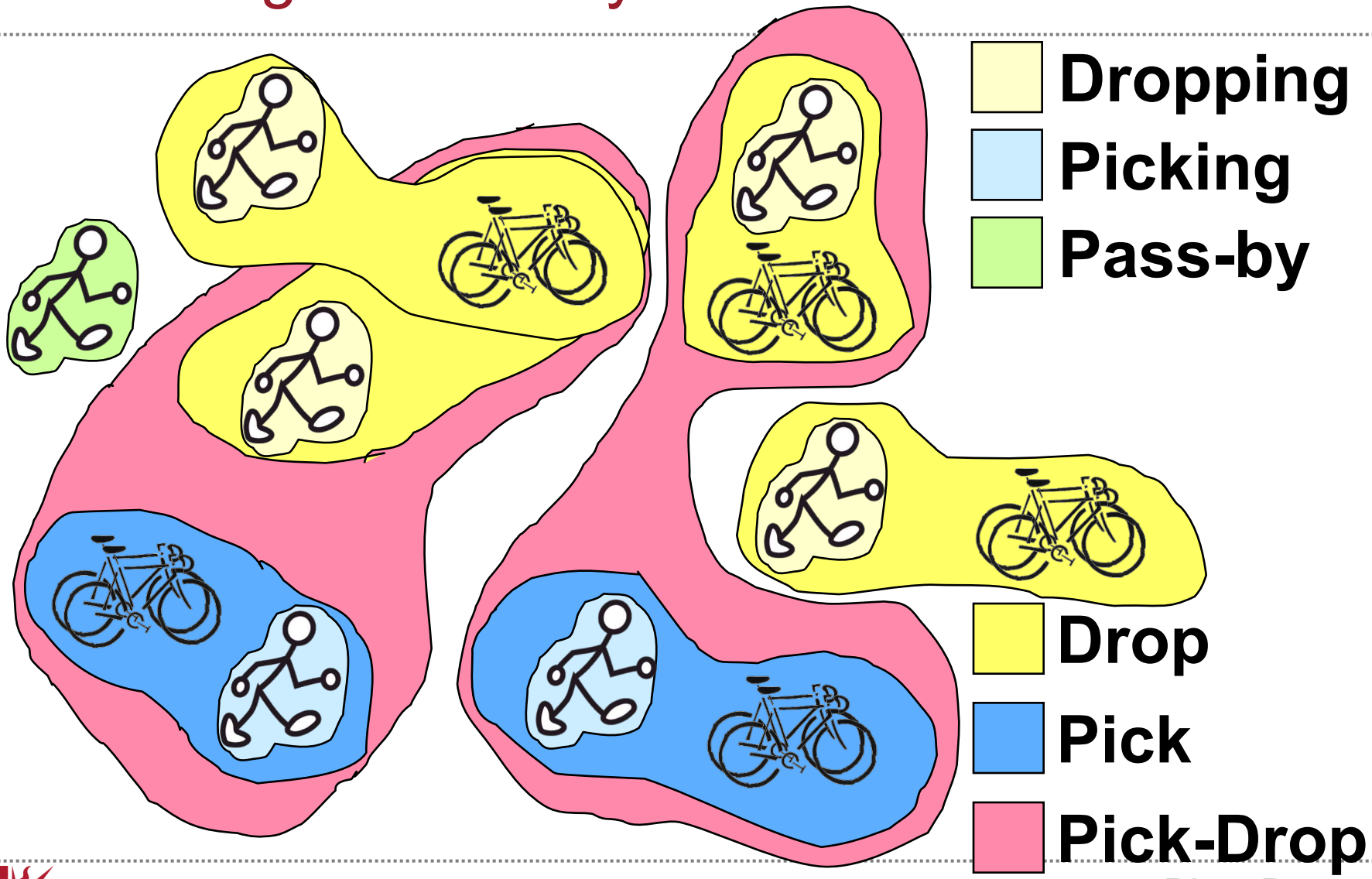
The Background Story…



The Background Story…

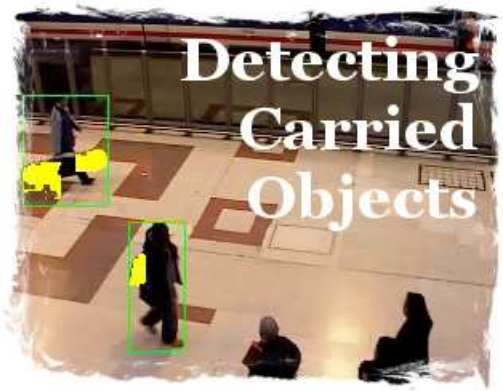


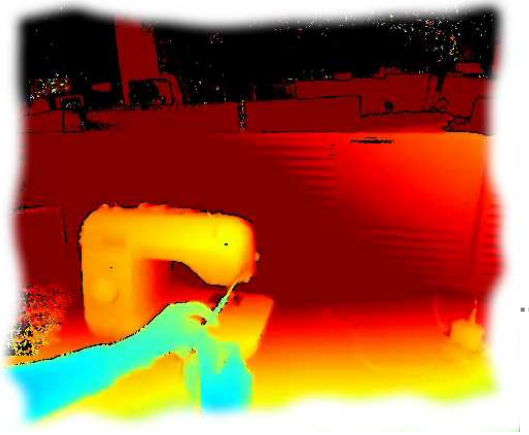
The Background Story…



The Background Story…

- A few more objects over the years…





Into First-Person Vision

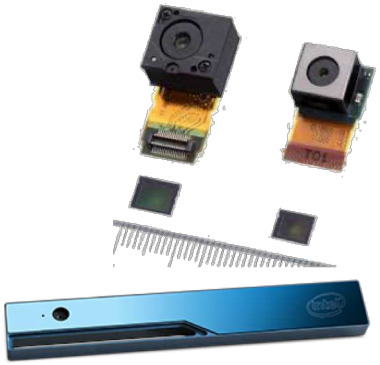
Visual Sensing – the landscape

Visual Sensing – the landscape

Visual Sensing – the landscape

Visual Sensing – the landscape





Most Wearable!

Hand-Held Wireless

Hand-Held Wired

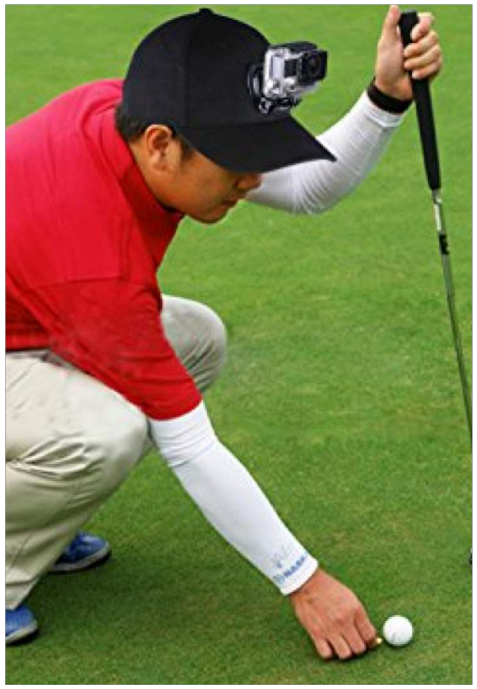
Least Static

Wearable

Wearable?

Wearable?

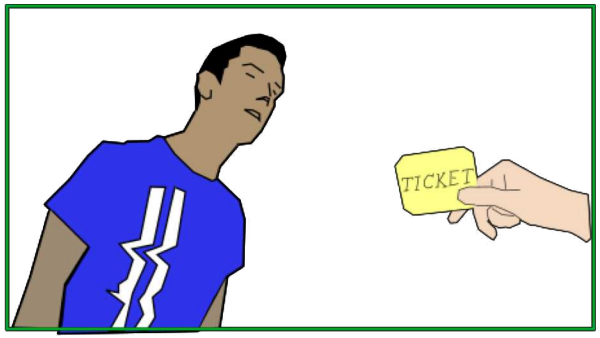
Wearable?



Wearable?

Wearable Cameras vs First Person View

- OPV (Ordinal-Person Views)
  - FPV (First-Person View)
  - SPV (Second-Person View)
  - TPV (Third-Person View)

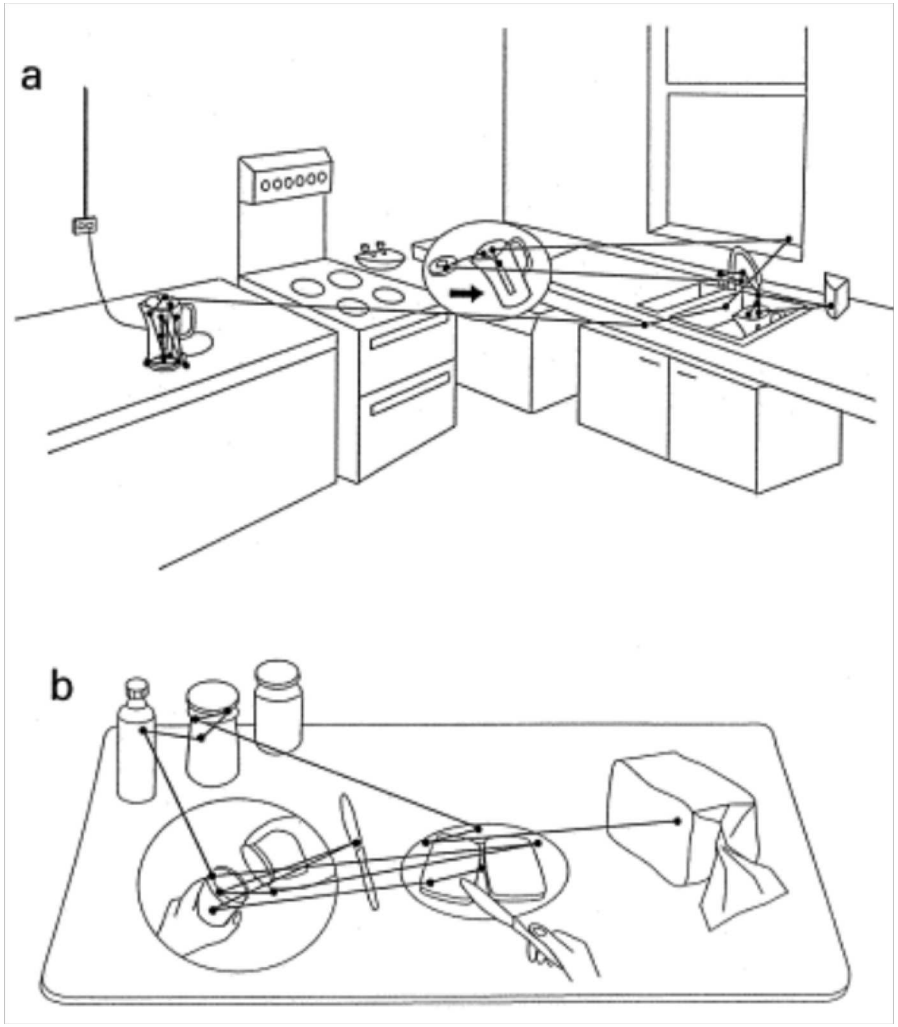


See for yourself!

- Videos...

Top-Left: Fathi – Disney dataset for social interactions 2012
Top-Right Damen EPIC-Kitchen 2017
Bottom-Left Damen BEOID 2014
Bottom-Right Kitani (Youtube) Sports 2011

Quick introduction to human gaze



Land and Hayhoe (2001) In what ways do eye movements contribute to everyday activities? Vision Research

Quick introduction to human gaze

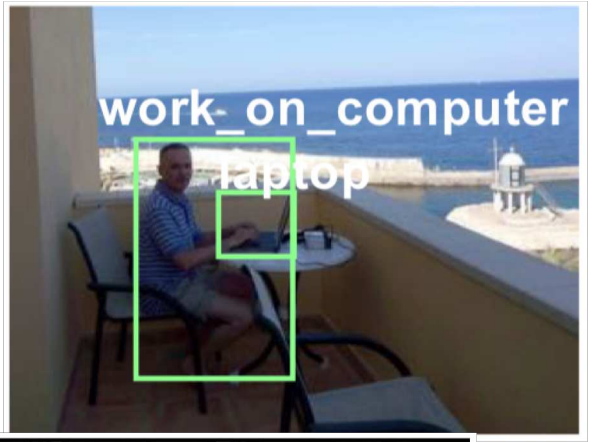
Hence…

a fine-grained perspective

to object interactions

from first-person video

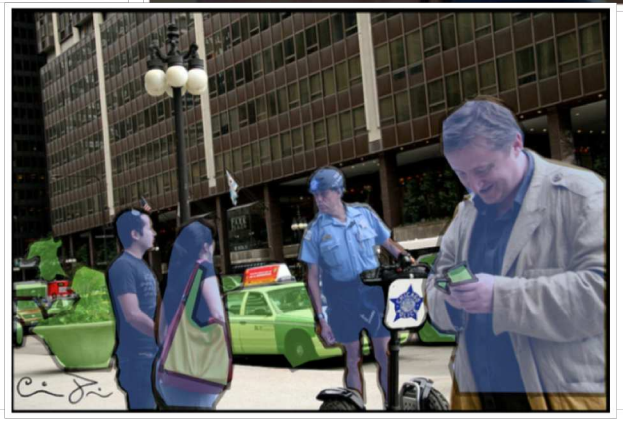
Action Recognition/Understanding



ATTACKING	
AGENT	MAN
VICTIM	PLAYER
TOOL	CHAIR
PLACE	STADIUM

SIGNING	
AGENT	MAN
SIGNEDITEM	BOOK
TOOL	PEN
PLACE	SHOP

THROWING	
AGENT	MAN
ITEM	BASEBALL
DESTINATION	CATCHER
PLACE	BA

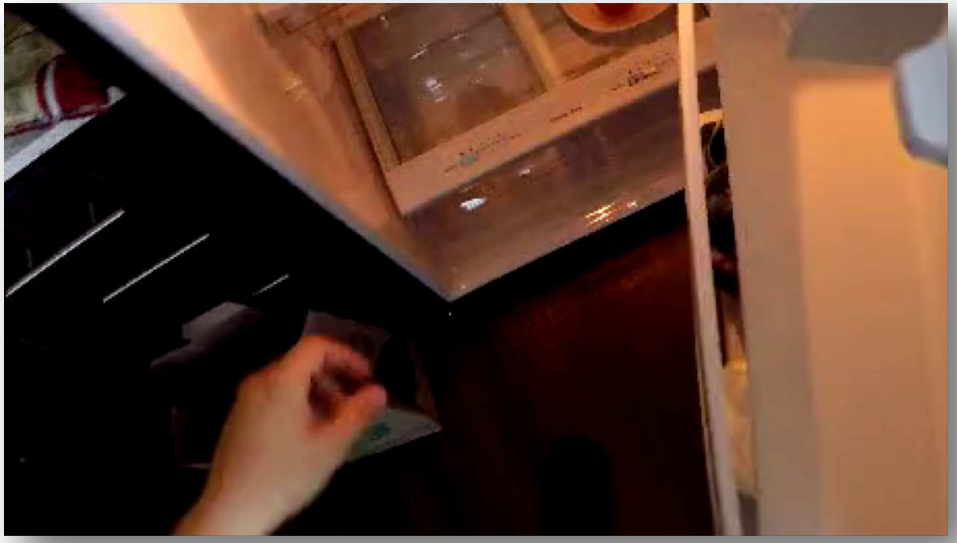


A man reading a paper and two people talking to a officer.

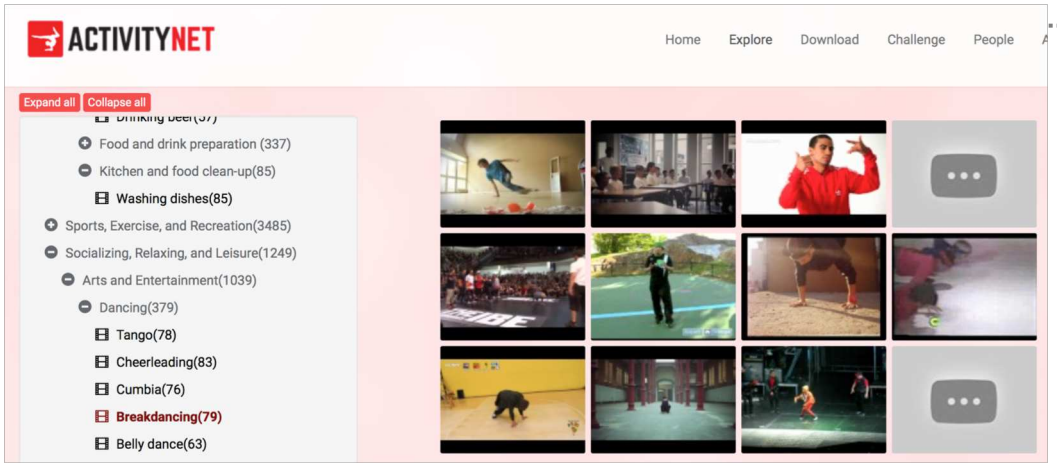
A man in a yellow jacket is looking at his phone with three others are in the background.

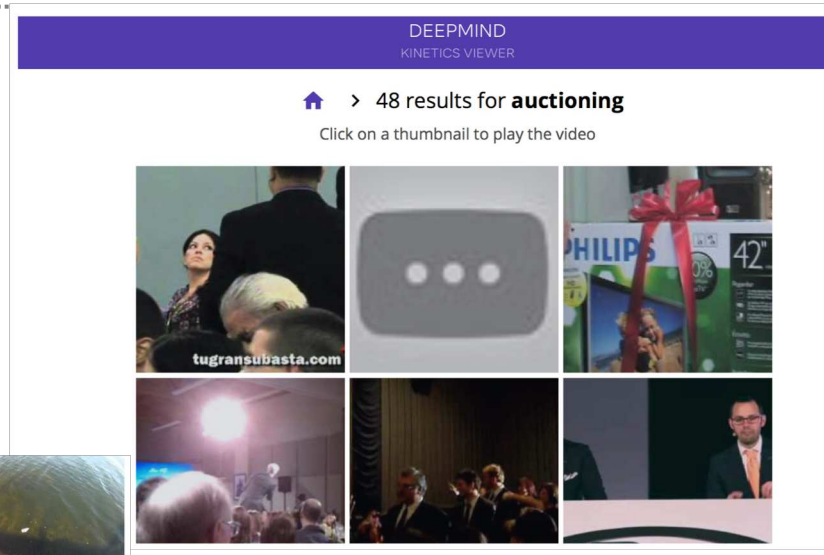
[1] Gkioxari et al (2017) Detecting and Recognizing Human-Object Interactions. CVPR
[2] Yatskar et al (2016) Situation Recognition. CVPR
[3] Ronchi and Perona (2015) Describing Common Human Visual Actions in Images. BMVC

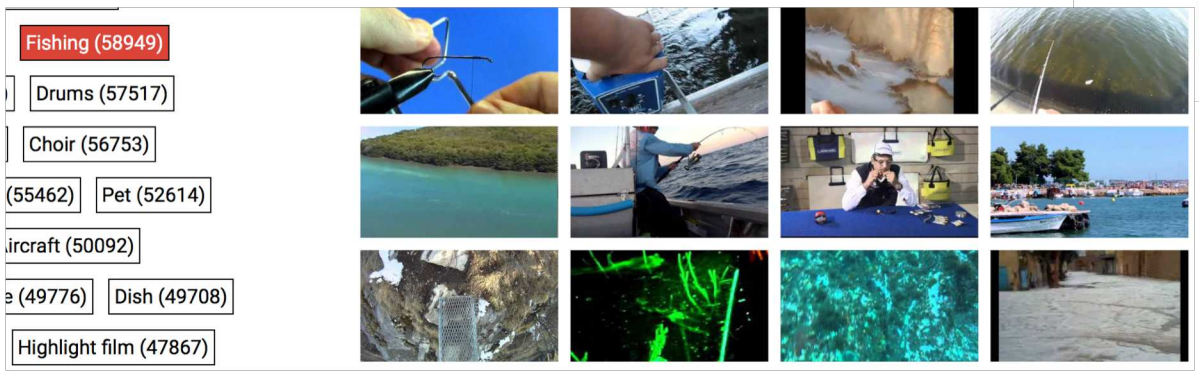
Action Recognition/Understanding

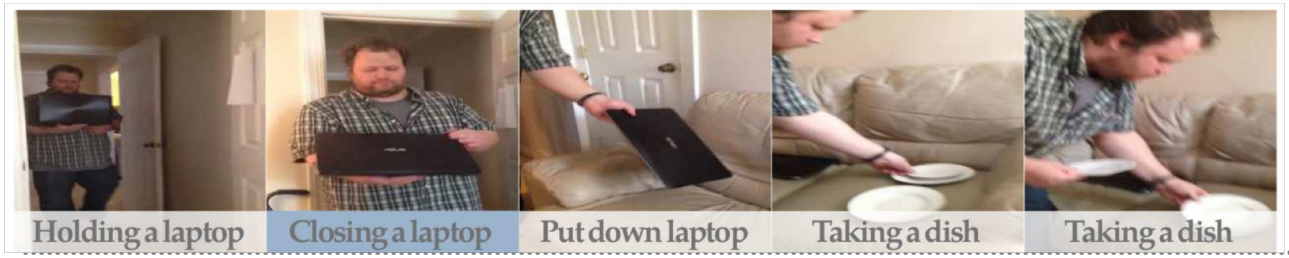


Action Recognition/Understanding









[1] Caba Heilbron et al (2015) Activity-Net. CVPR
[2] Kay et al (2017) Kinetics. Arxiv
[3] Abu-El-Haija et al (2016). YouTube-8M. Arxiv
[4] Gunner et al (2016) Hollywood in Homes. ECCV

In this talk

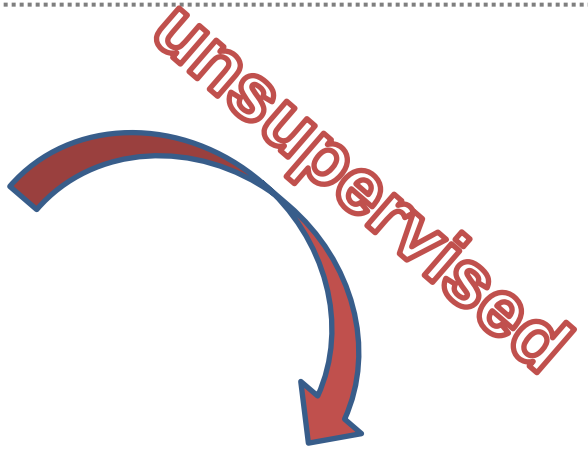
- Why fine-grained?
  - You-Do, I-Learn
- Fine-Grained Problems
  - how ‘well’: Skill Determination in Video
  - when: Action Completion
  - when: Trespassing the Boundaries
  - which: Unequivocal Representation of Actions
- EPIC-KITCHENS 2018
  - Dataset collection and Challenges

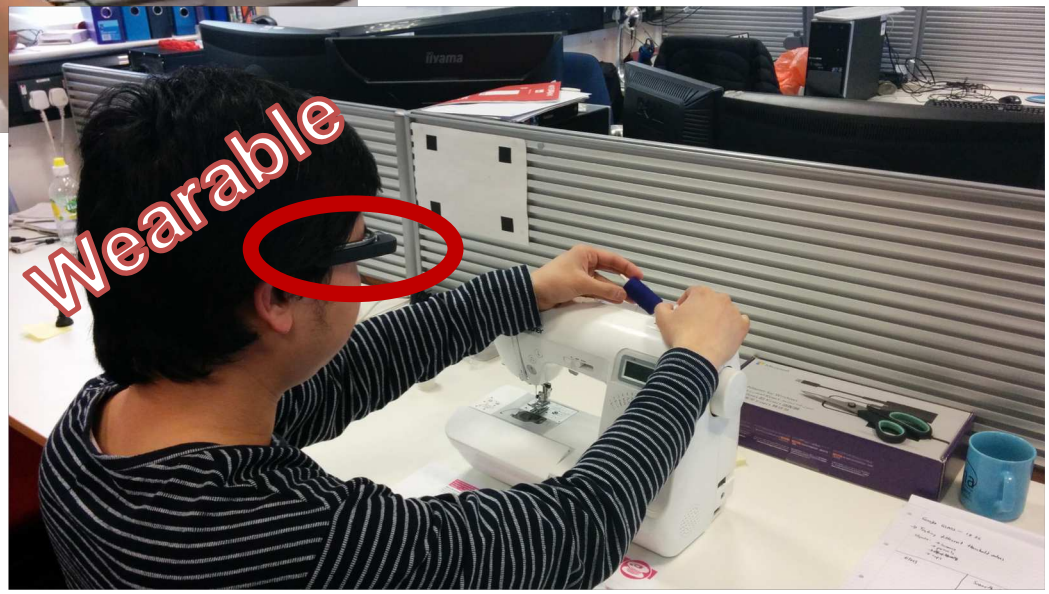
In this talk

- Why fine-grained?
  - You-Do, I-Learn
- Fine-Grained Problems
  - how ‘well’: Skill Determination in Video
  - when: Action Completion
  - when: Trespassing the Boundaries
  - which: Unequivocal Representation of Actions
- EPIC-KITCHENS 2018
  - Dataset collection and Challenges

You-Do, I-Learn

You-Do, I-Learn





with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You Do, I Learn - Demonstration

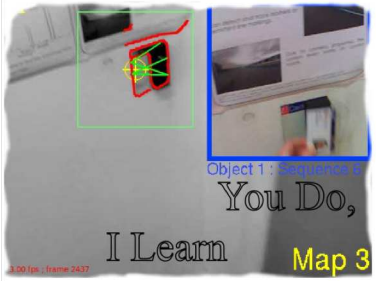
D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

You Do, I Learn – Google Glass Prototype

D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

More info…

Project You-Do, I-Learn



Video1 (2014), Video2 (2017)

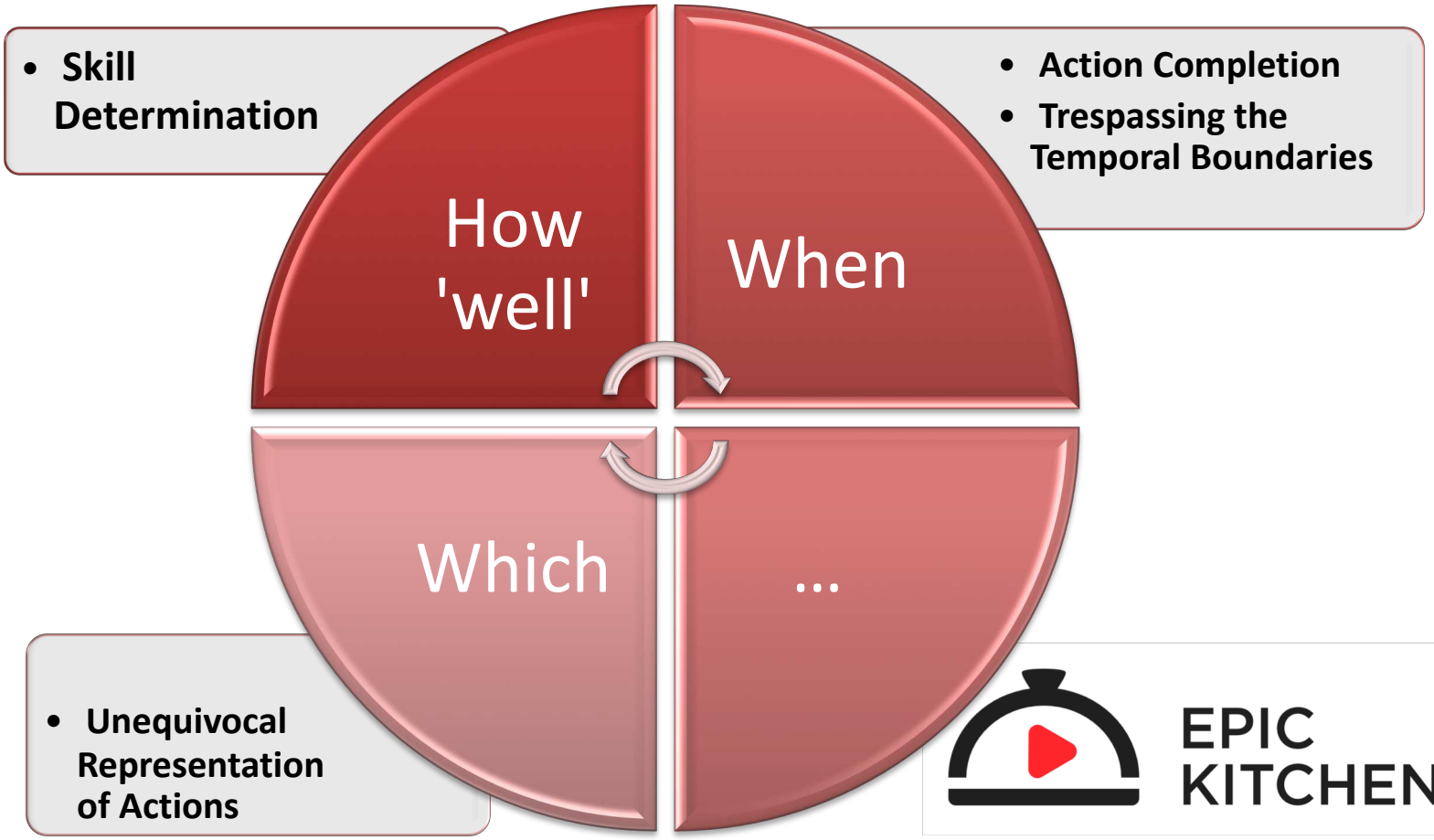
Automated capture and delivery of assistive task guidance with an eyewear computer: The GlaciAR system. T Leelasawassuk, D Damen, W Mayol-Cuevas. Augmented Human, Mar 2017 pdf

You-Do, I-Learn: Discovering Task Relevant Objects and their Modes of Interaction from Multi-User Egocentric Video. D Damen, T Leelasawassuk, O Haines, A Calway, W Mayol-Cuevas. British Machine Vision Conference (BMVC), Sep 2014. PDF | Abstract | Dataset

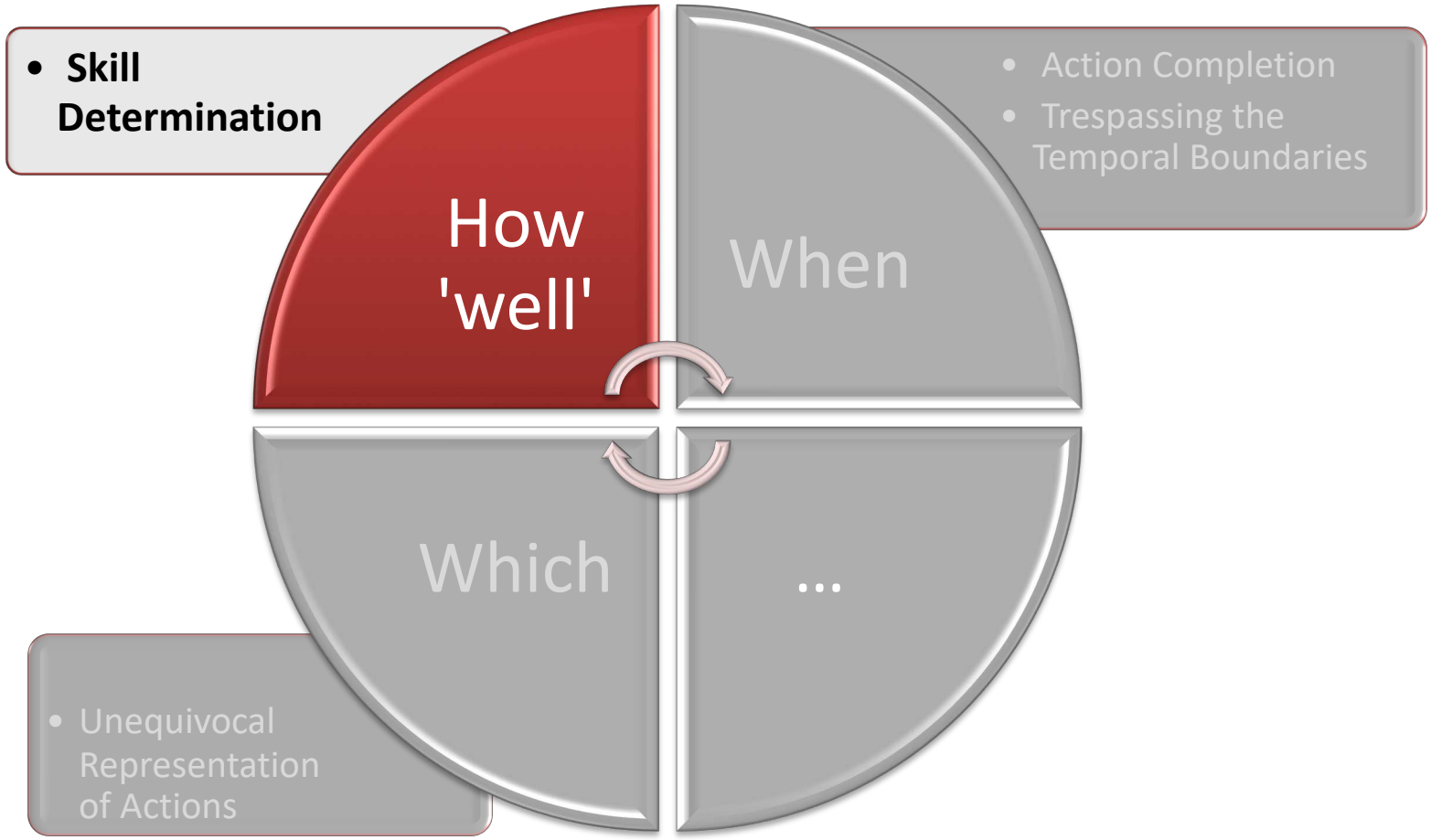
Multi-user egocentric Online System for Unsupervised Assistance on Object Usage. D Damen, O Haines, T Leelasawassuk, A Calway, W Mayol-Cuevas. ECCV Workshop on Assistive Computer Vision and Robotics (ACVR), Sep 2014. PDF Preprint

Estimating Visual Attention from a Head Mounted IMU. T Leelasawassuk, D Damen, W Mayol-Cuevas. International Symposium on Wearable Computers (ISWC), Sep 2015. PDF

Fine-Grained Object Interactions



Fine-Grained Object Interactions



Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

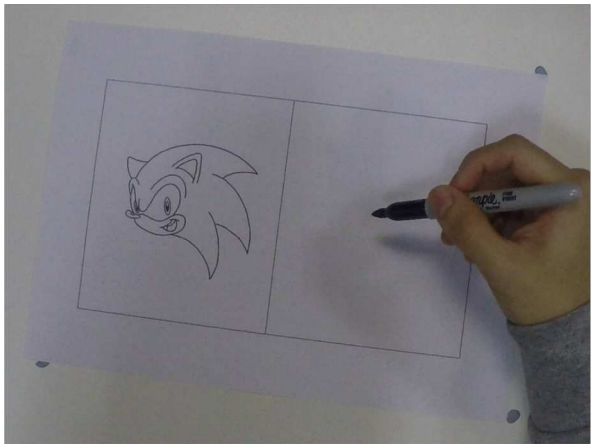
with: Hazel Doughty
Walterio Mayol-Cuevas

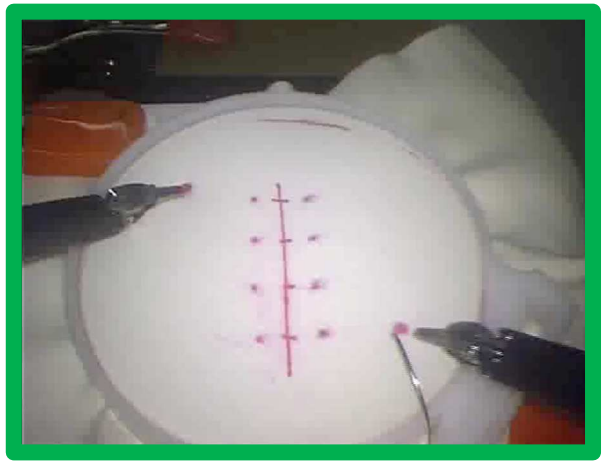
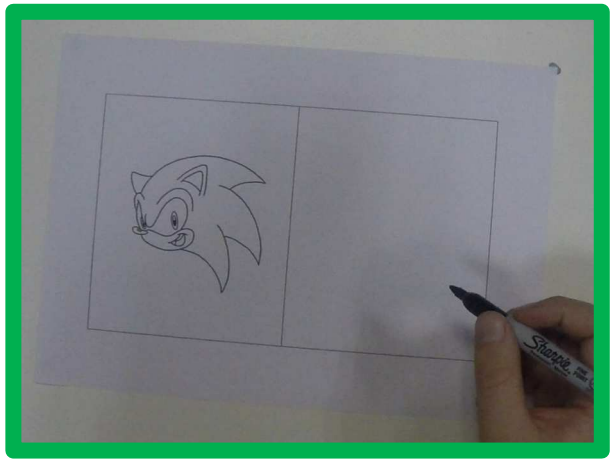
Assess relative skill for a collection of video sequences, applicable to a variety of tasks.

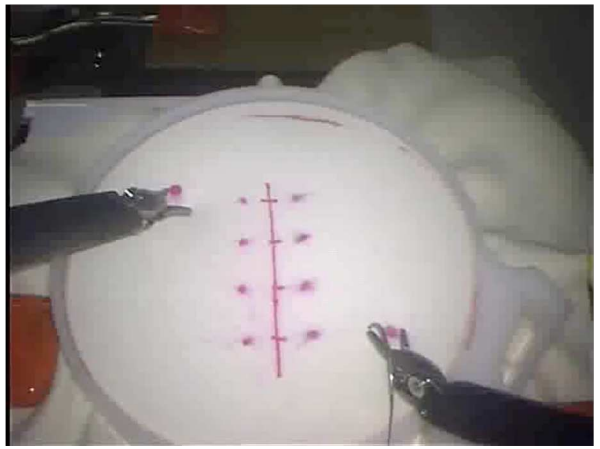
Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

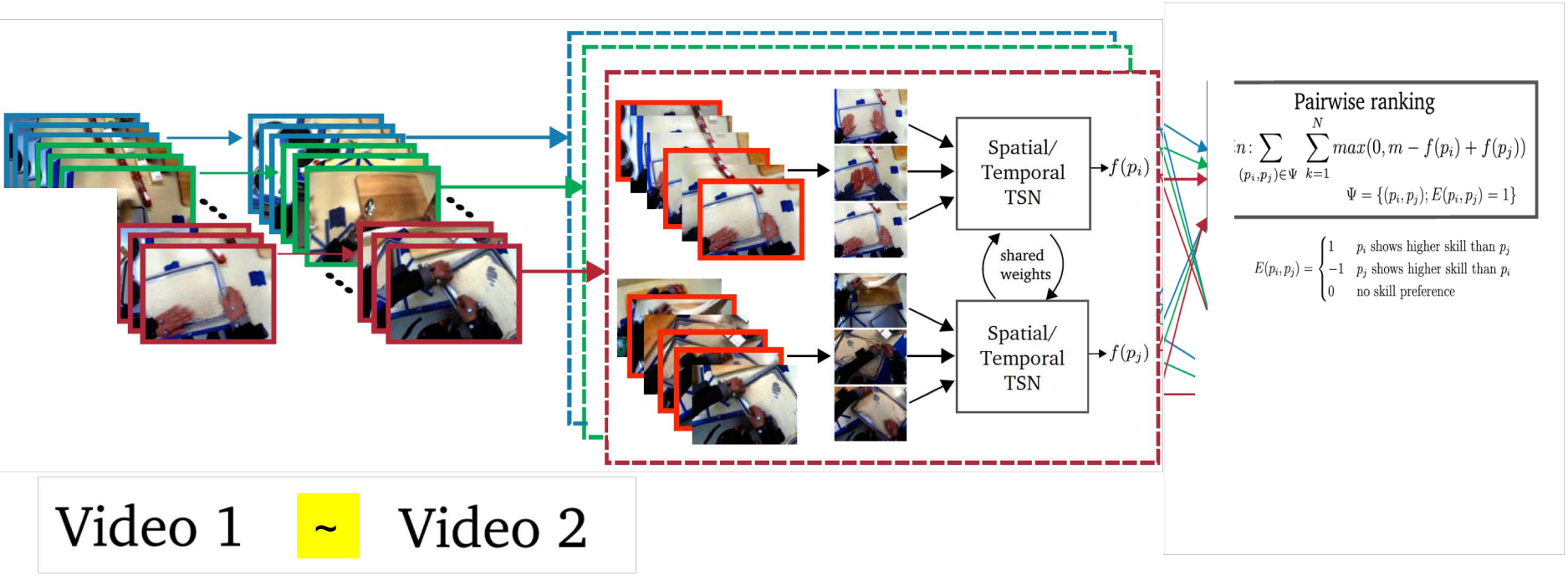
Input: Pairwise annotations of videos, indicating higher skill or no skill preference



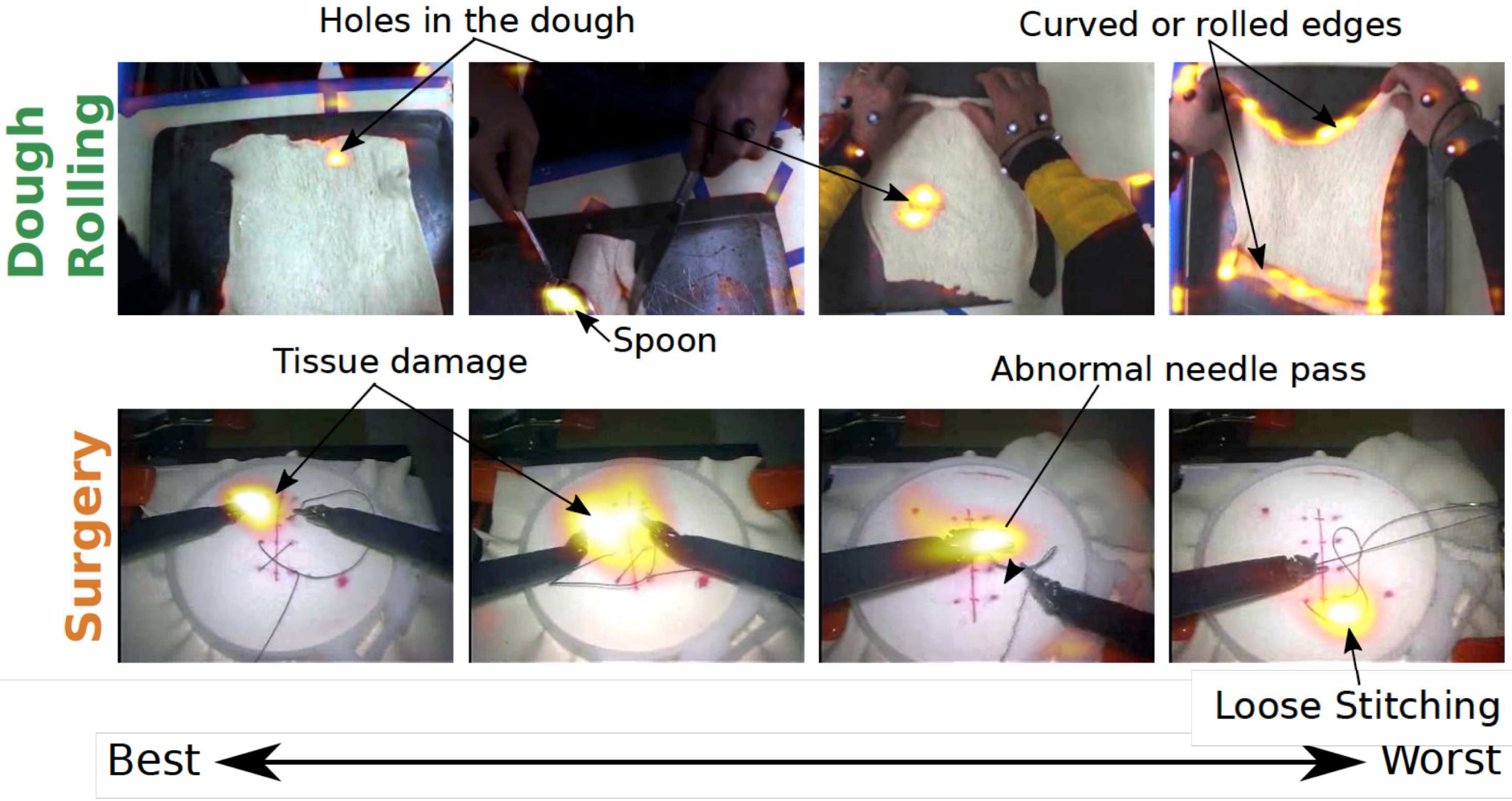


Who's Better? Who's Best? Skill Determination in Video using Deep Ranking



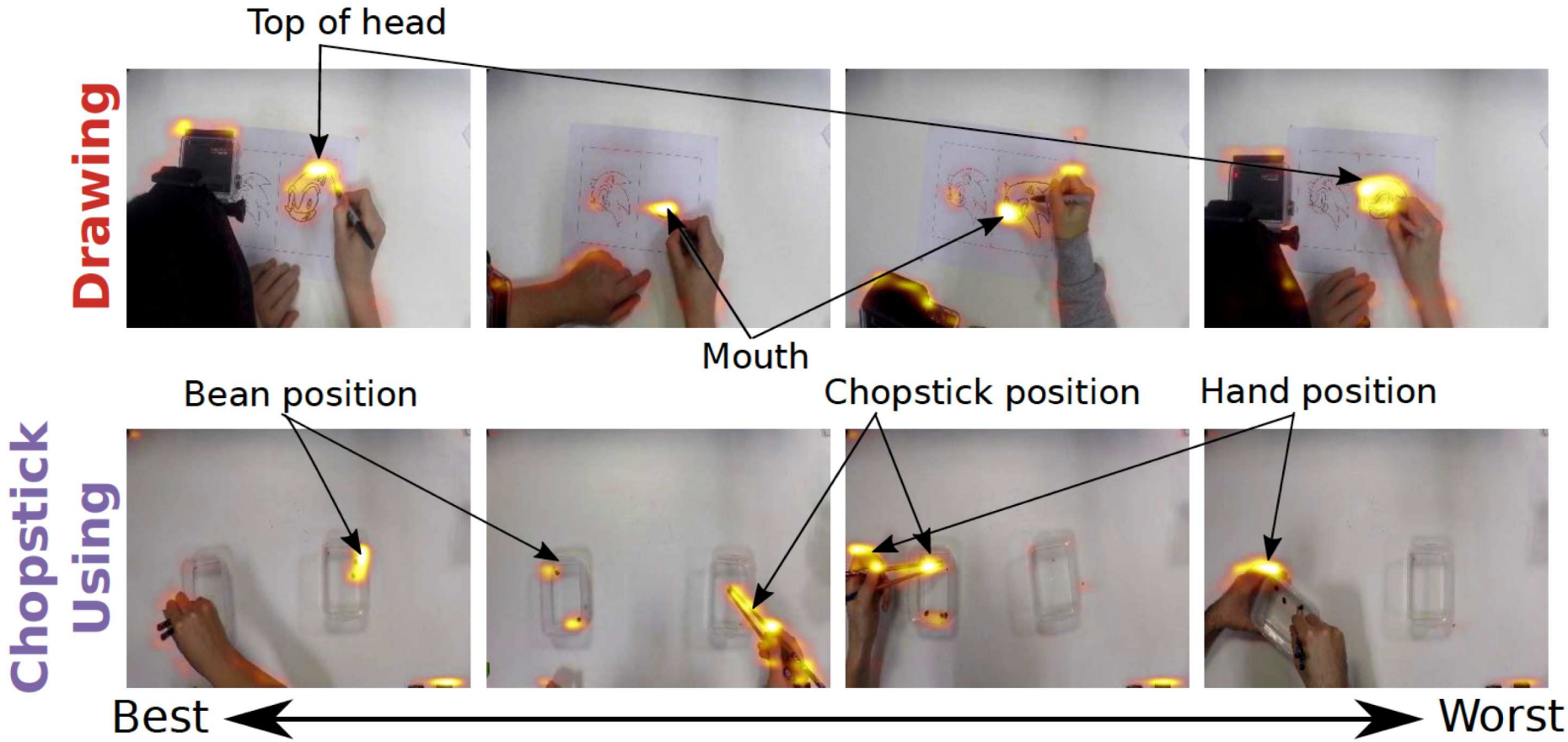
Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas



Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

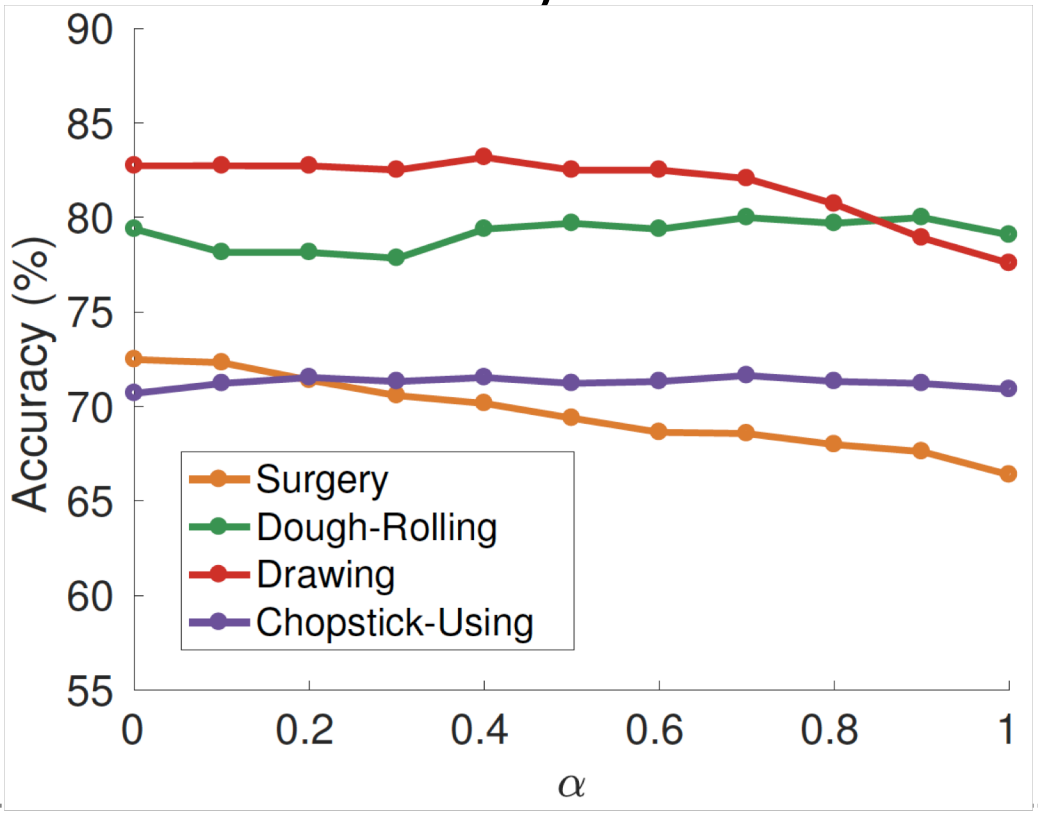
with: Hazel Doughty
Walterio Mayol-Cuevas



Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

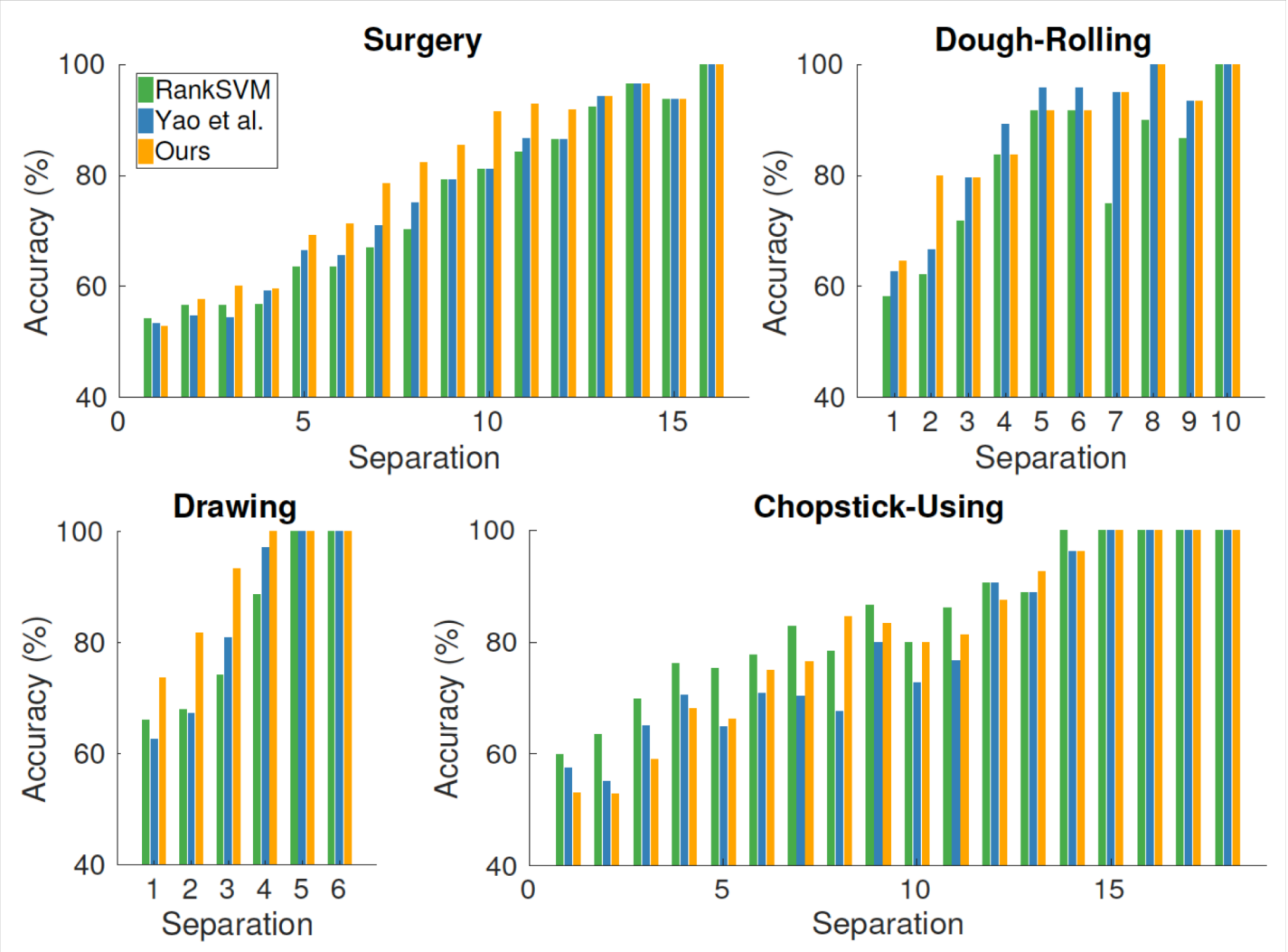
$$\frac{1}{\sigma}\sum_{j=1}^{\sigma} \boldsymbol{\alpha} f_s(p_{ij}) + (1-\boldsymbol{\alpha}) f_t(p_{ij})$$



H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

Who's Better? Who's Best? Skill Determination in Video using Deep Ranking

with: Hazel Doughty
Walterio Mayol-Cuevas

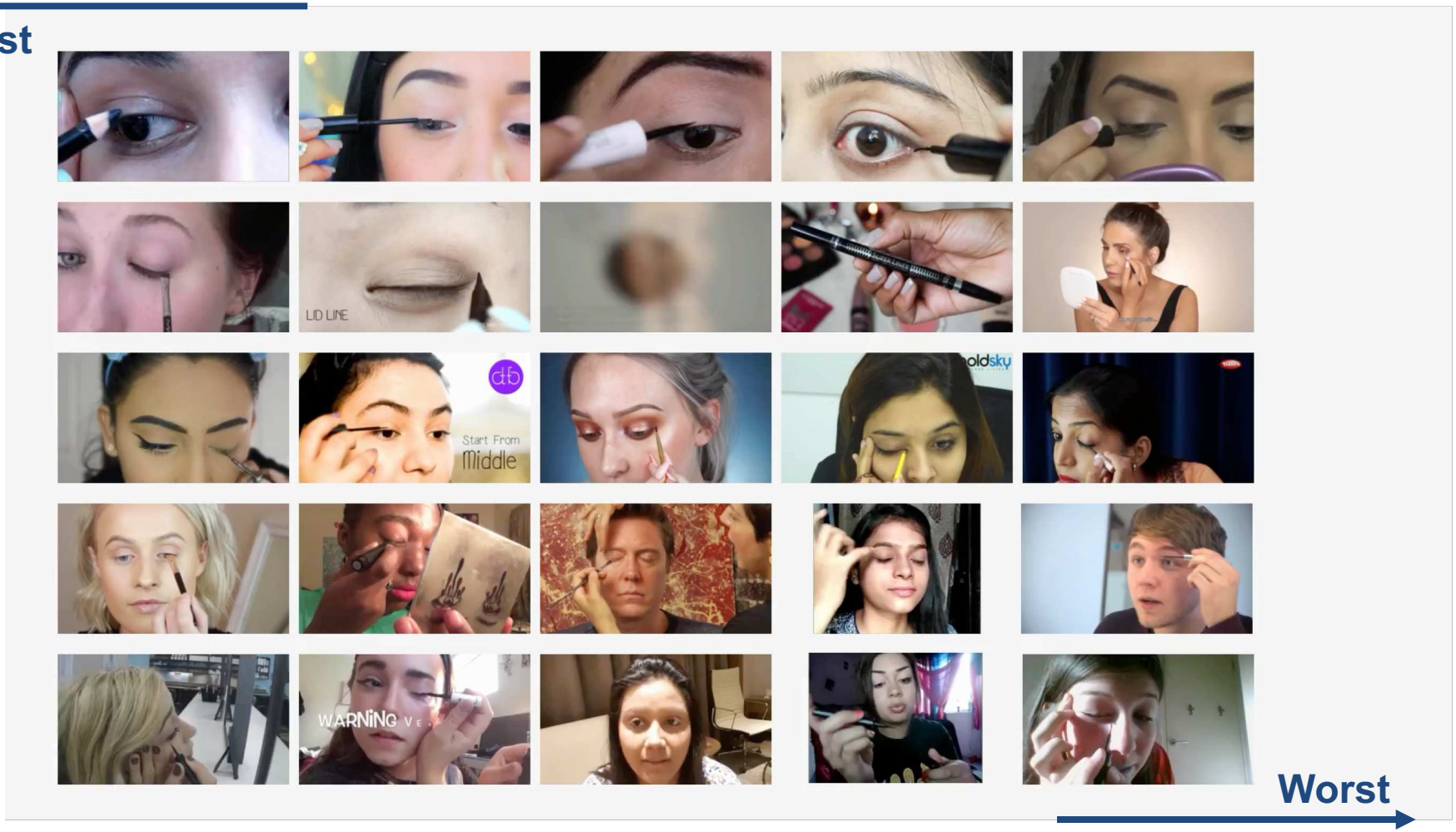


H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

The Pros and Cons: Rank-Aware Temporal Attention

with: Hazel Doughty
Walterio Mayol-Cuevas

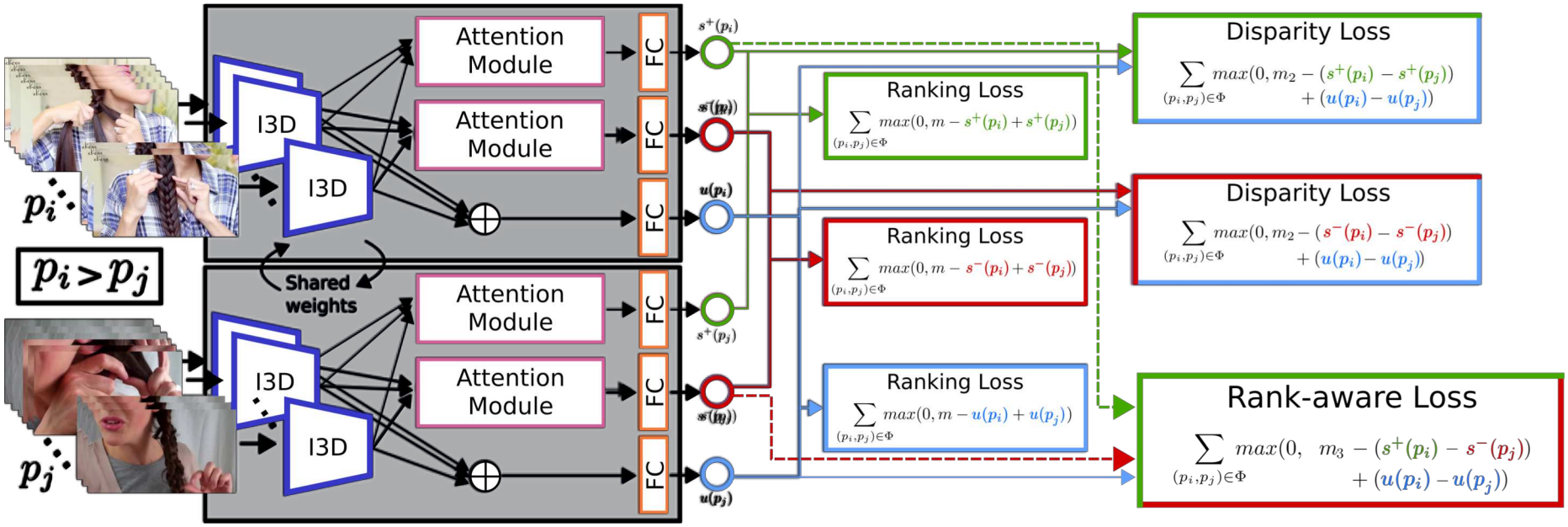
Best



Worst

The Pros and Cons: Rank-Aware Temporal Attention

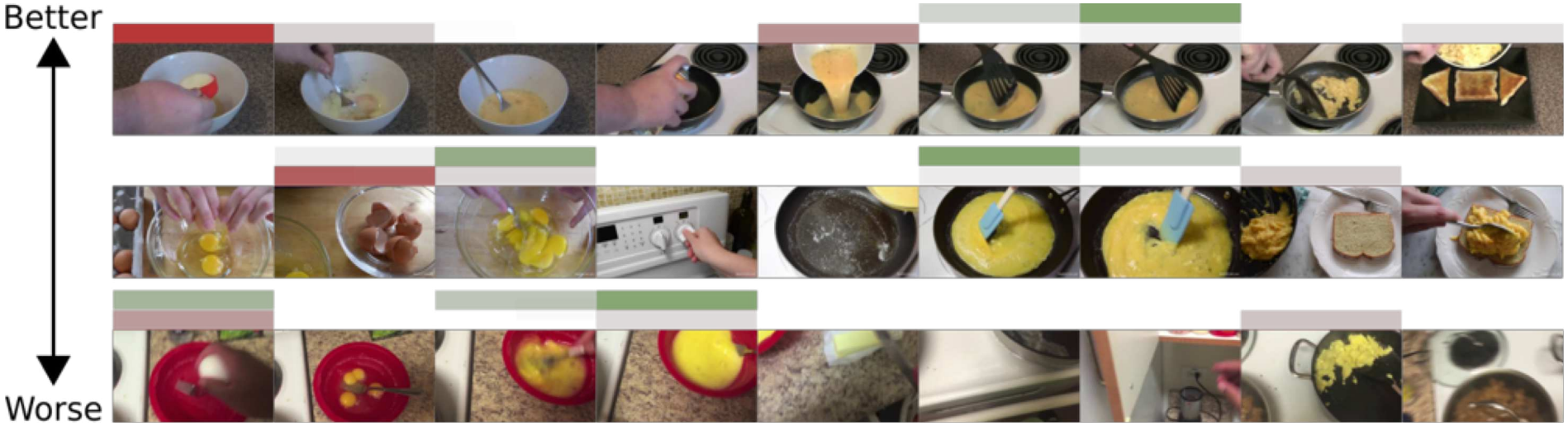
with: Hazel Doughty
Walterio Mayol-Cuevas



H Doughty, W Mayol-Cuevas, D Damen (2019). The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos. *Computer Vision and Pattern Recognition (CVPR)*

The Pros and Cons: Rank-Aware Temporal Attention

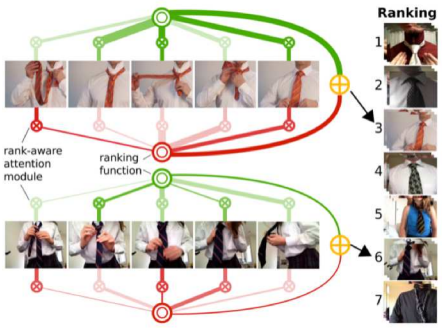
with: Hazel Doughty
Walterio Mayol-Cuevas



H Doughty, W Mayol-Cuevas, D Damen (2019). The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos. *Computer Vision and Pattern Recognition (CVPR)*

More info…

Project Skill Determination in Video

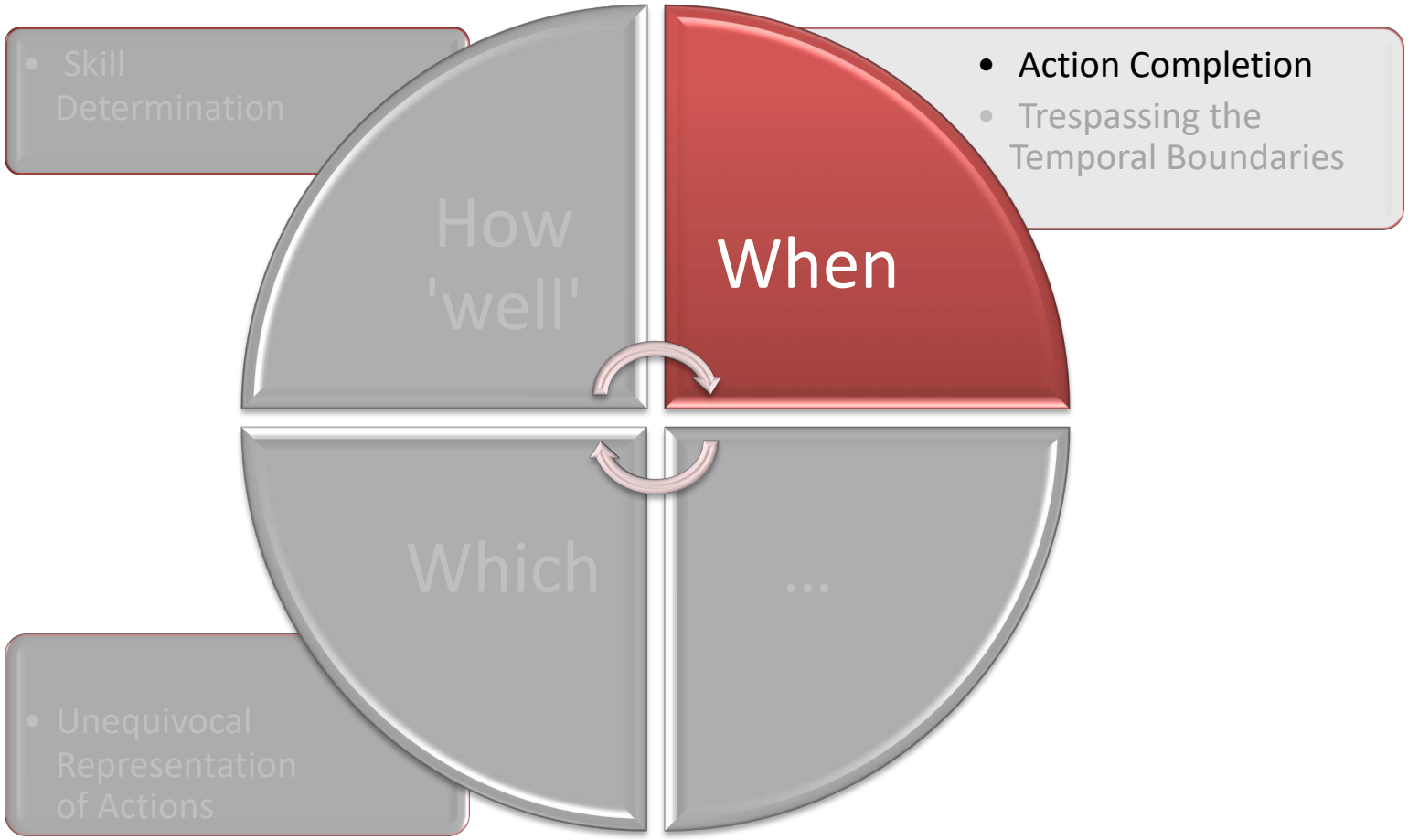


Video

The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos. H Doughty, W Mayol-Cuevas, D Damen. Arxiv (2018). arxiv

Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. H Doughty, D Damen, W Mayol-Cuevas. CVPR (2018). PDF | arxiv | Dataset

Fine-Grained Object Interactions



with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

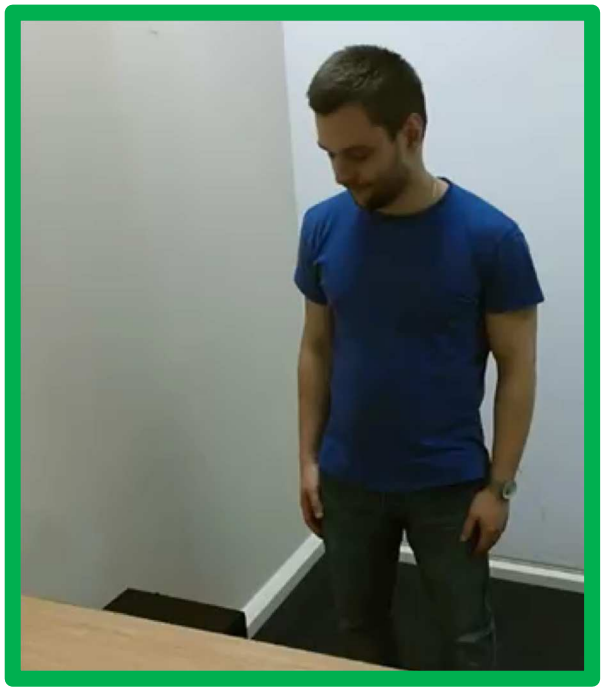
F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

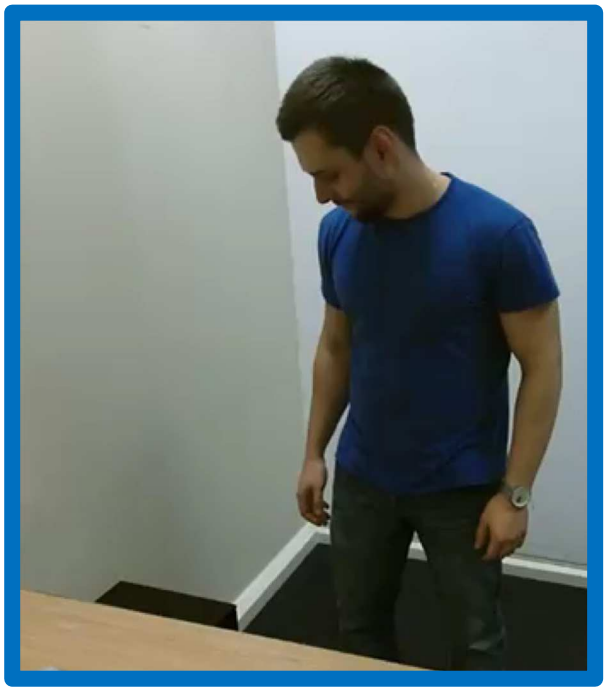
Understanding Completion

What if the observed action is not fully completed!?

Complete *pull*



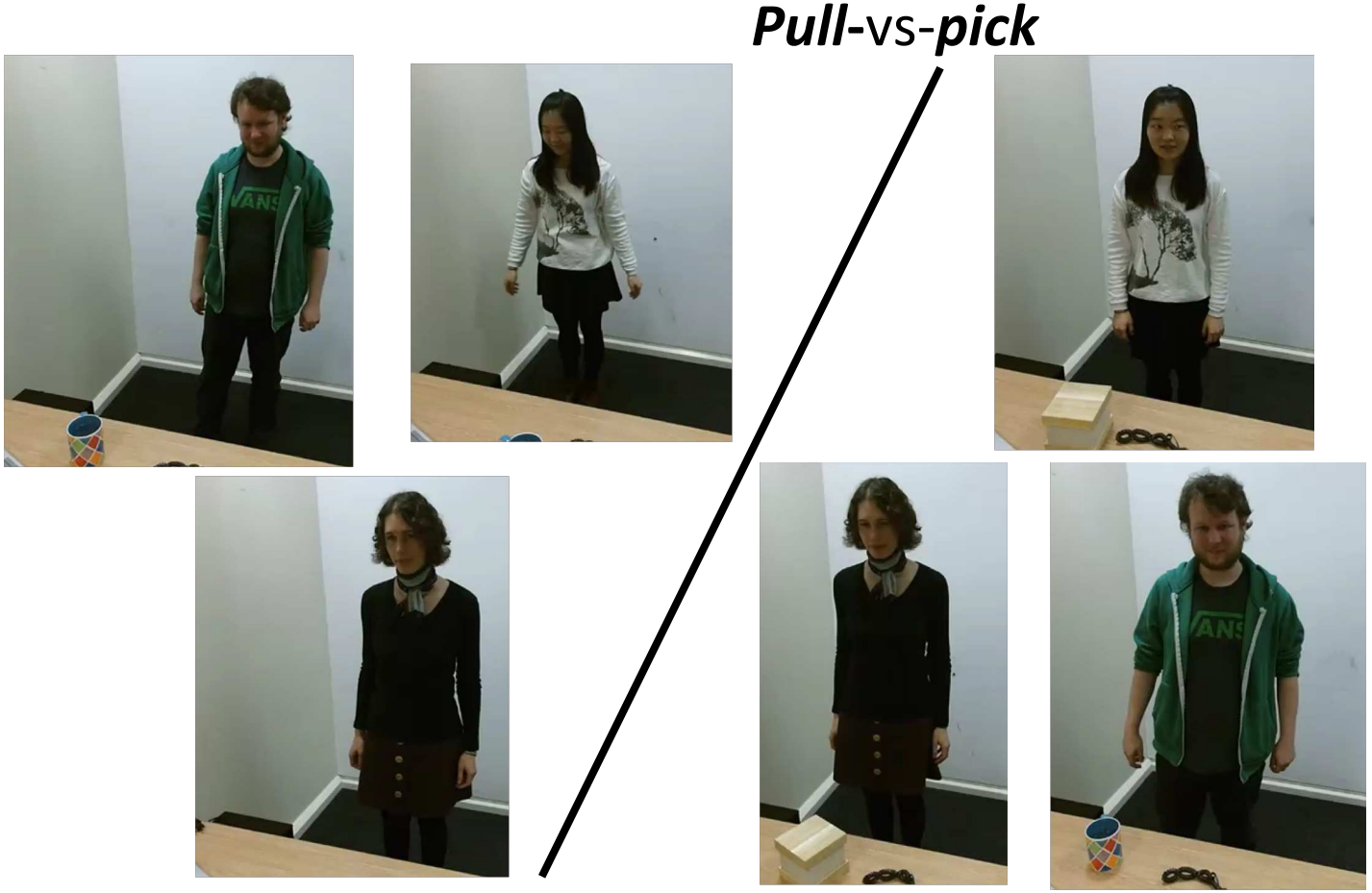
Incomplete *pull*



F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Understanding Completion

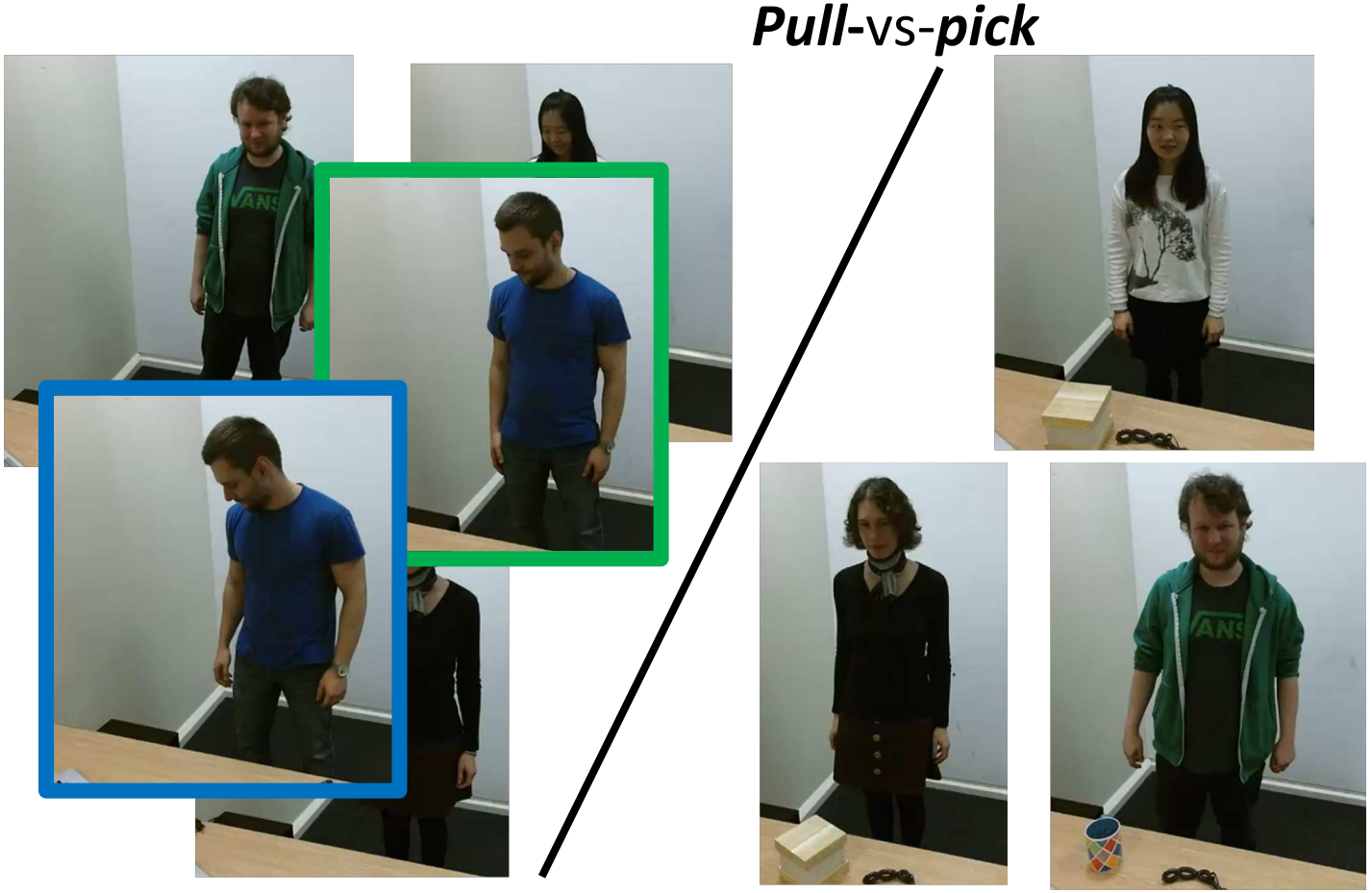


Complete *pull* and incomplete *pull* are introduced to *pull*-vs-*pick* classifier.

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Understanding Completion



Both complete *pull* and incomplete *pull* are classified as *pull*.

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

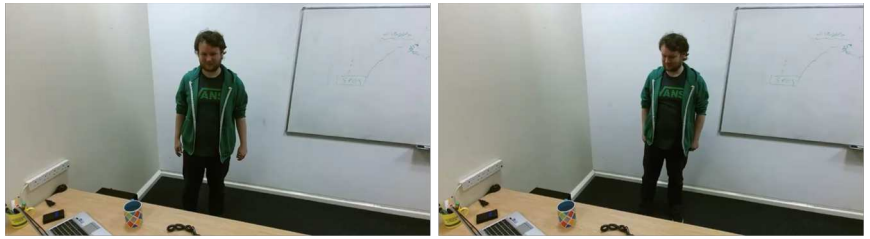
with: Farnoosh Heidarivincheh
Majid Mirmehdi

Bristol Action Completion Dataset 2016

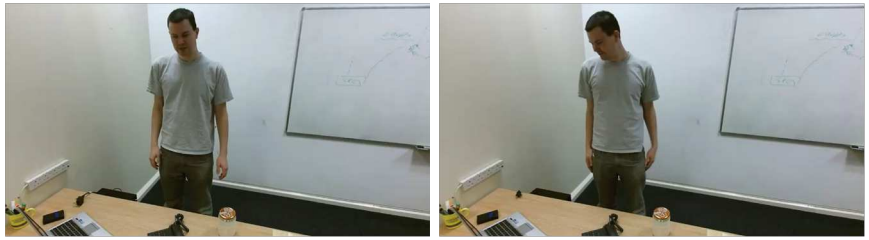
complete incomplete

switch

complete incomplete



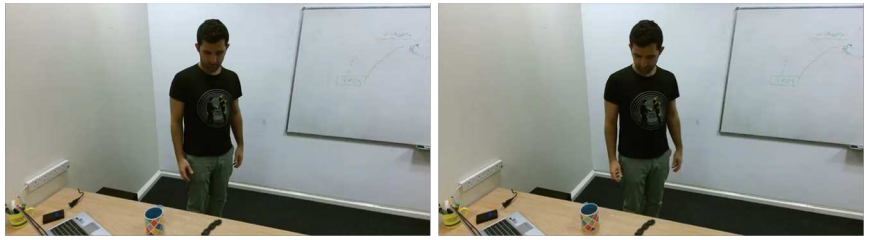
pull



plug

pick

open

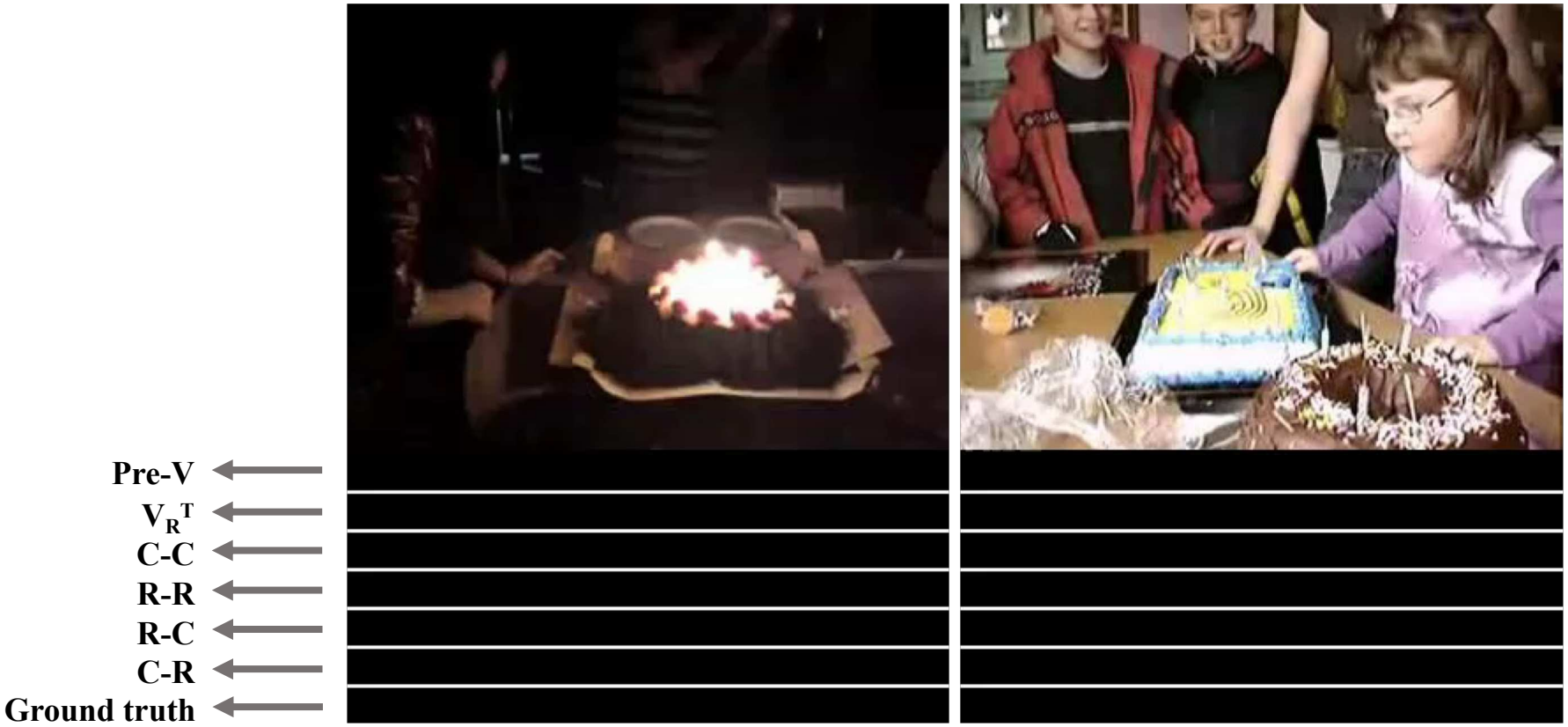


drink

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

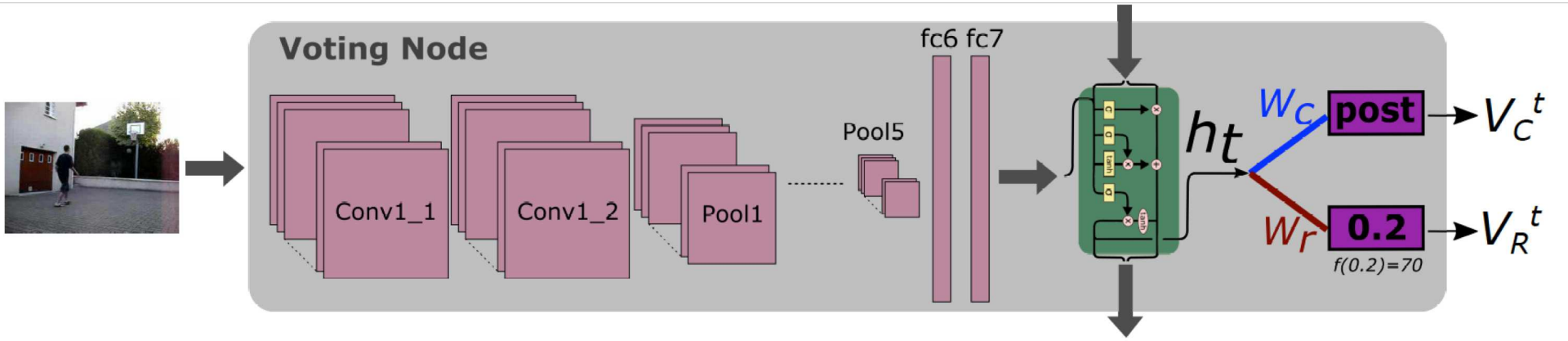


F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

- Each frame in the sequence, contributes to the completion moment detection via ‘voting’

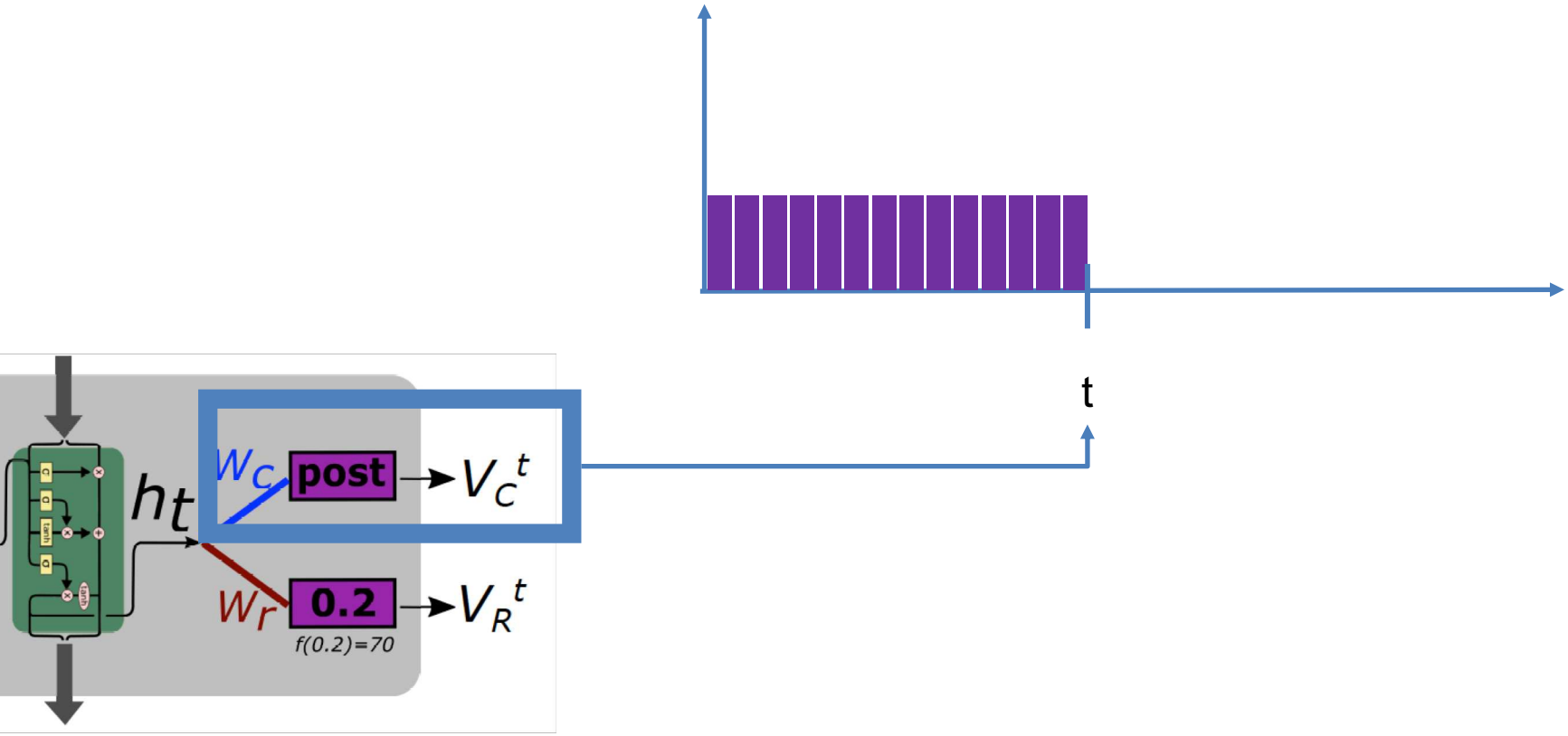


F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

1. Classification-Based Voting

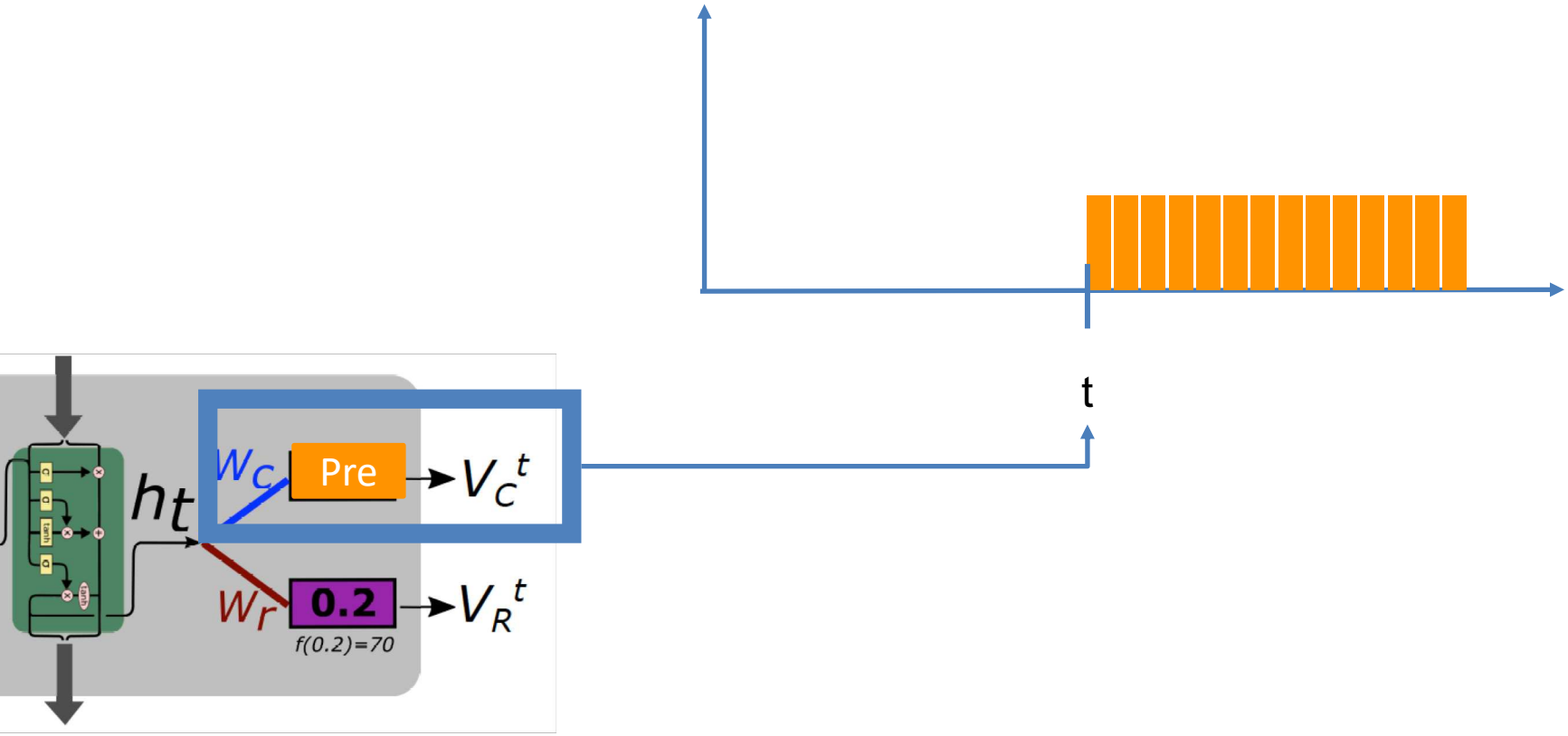


F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

1. Classification-Based Voting

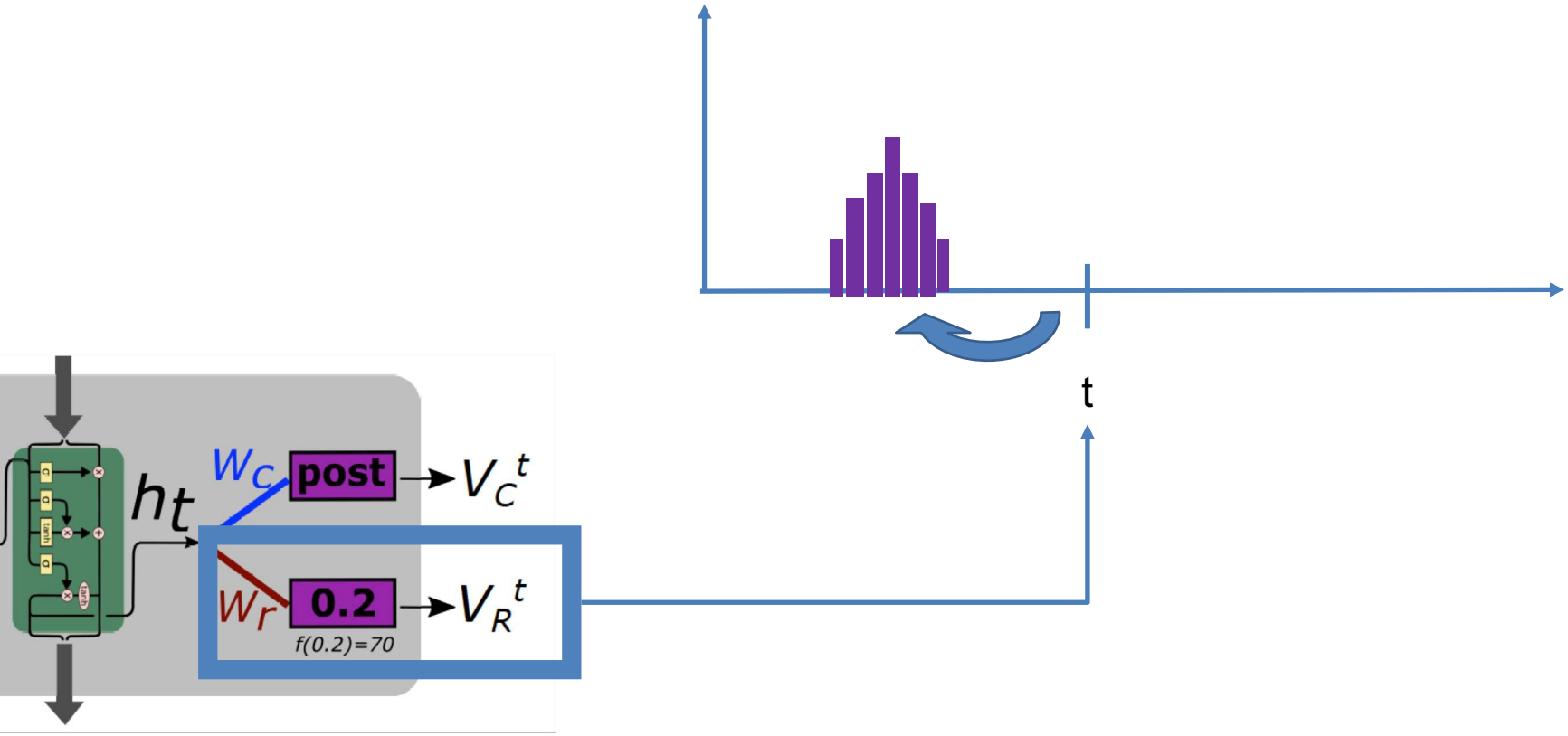


F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

2. Regression-Based Voting

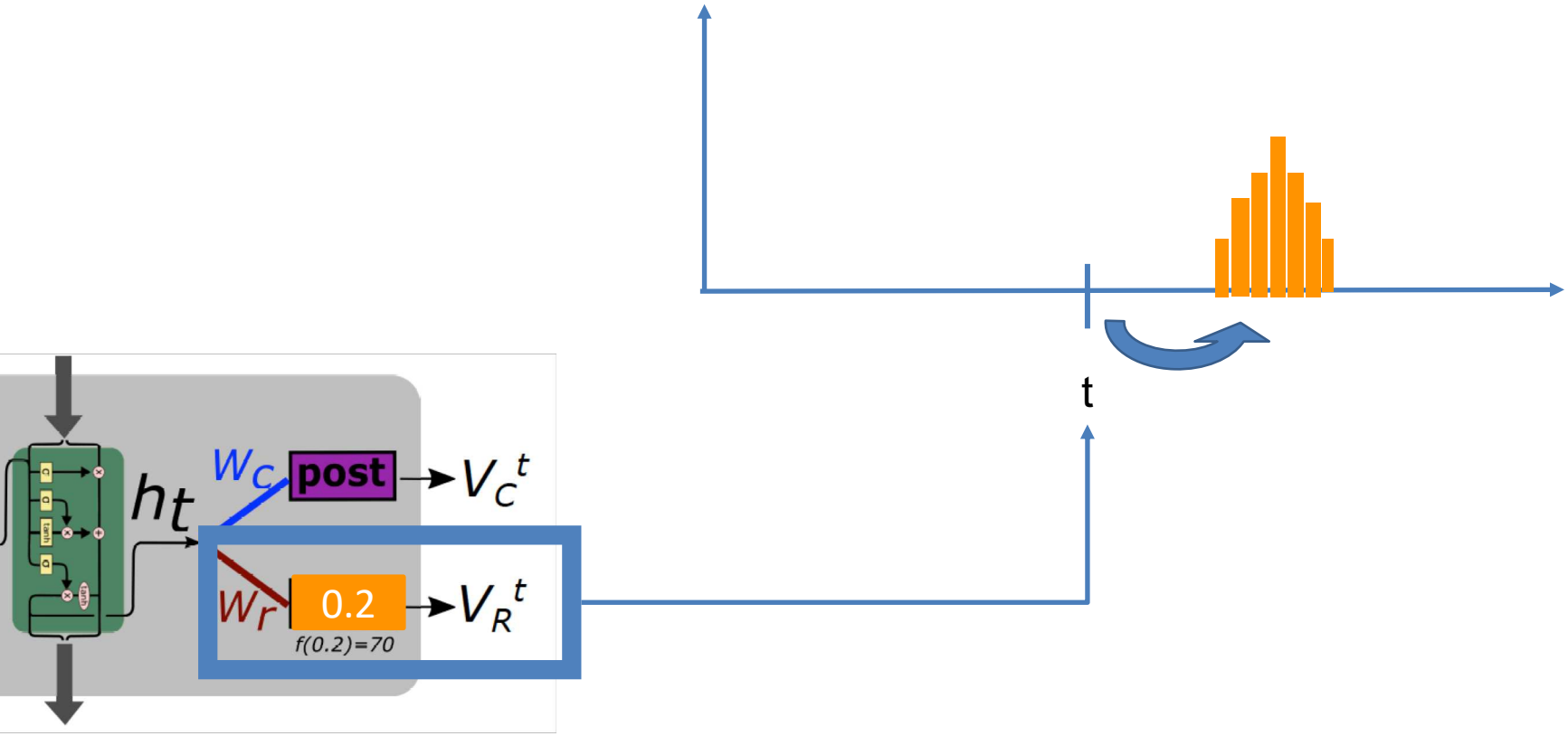


F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

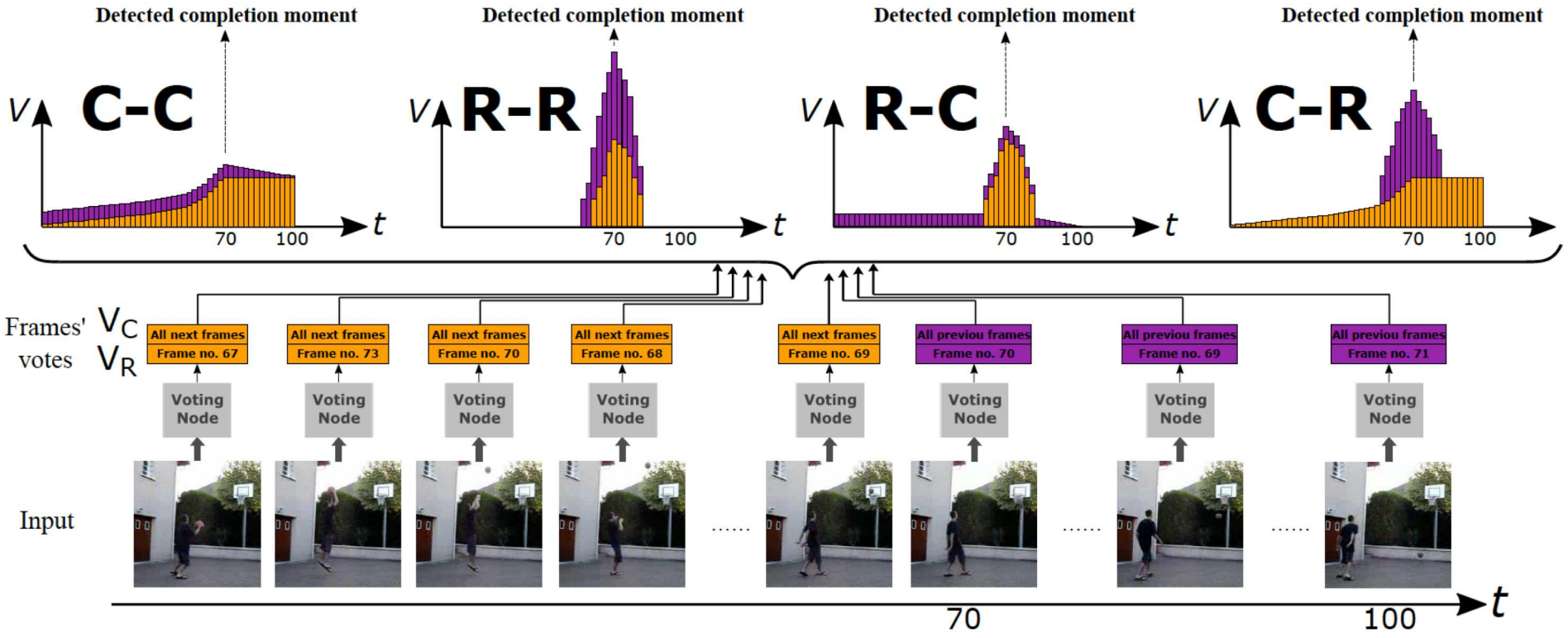
2. Regression-Based Voting



F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection



F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*

F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Action Completion Detection

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

with: Farnoosh Heidarivincheh
Majid Mirmehdi

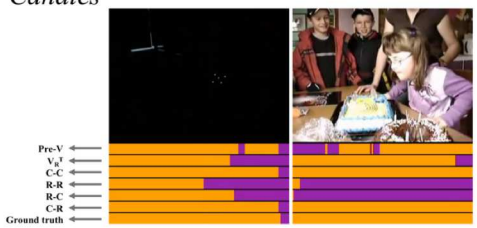
Action Completion Detection

F Heidarivincheh et al (2018). Action Completion: A Temporal Model for Moment Detection. *British Machine Vision Conference*
F Heidarivincheh et al (2016). Beyond Action Recognition: Action Completion in RGB-D Data. *British Machine Vision Conference*

More info…

Project Action Completion

Action Completion Detection – Blowing Candles

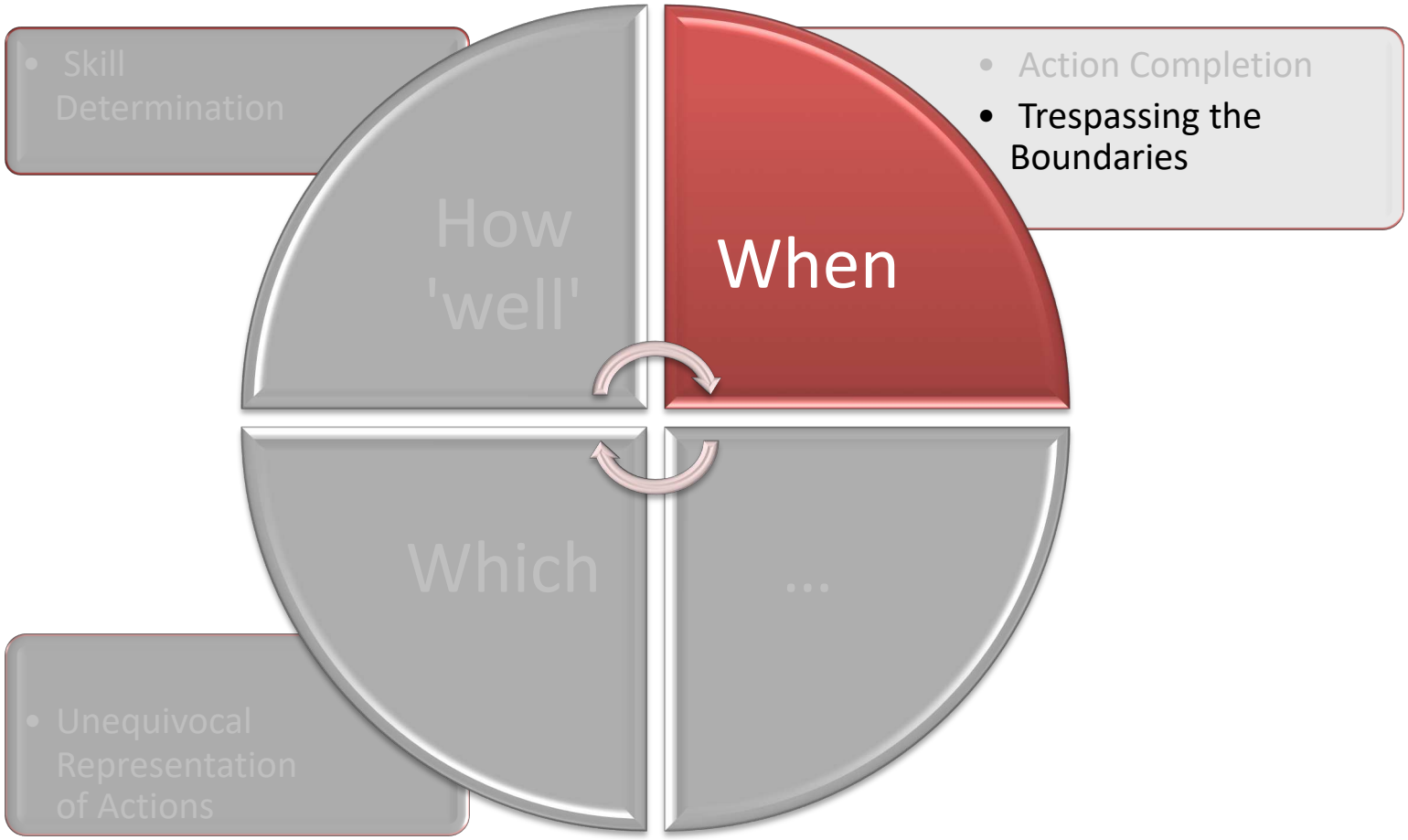


Video

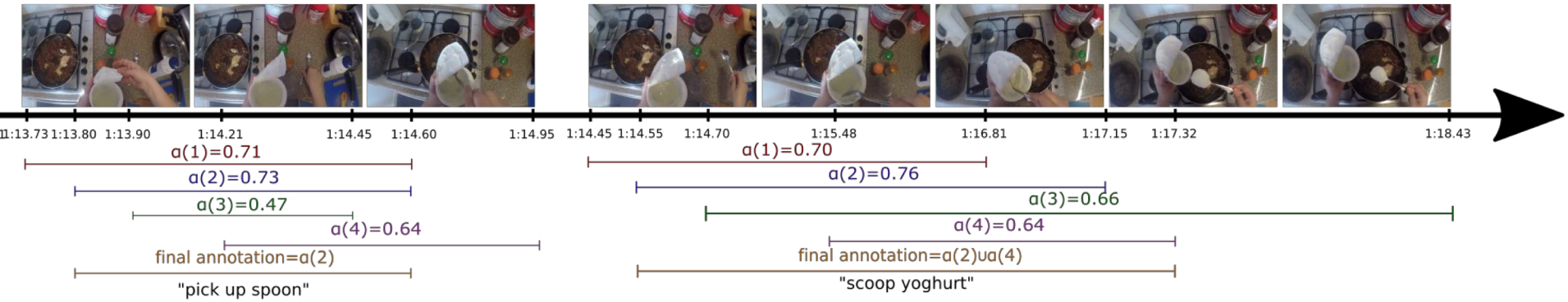
Action Completion: A Temporal Model for Moment Detection. F Heidarivincheh, M Mirmehdi, D Damen. Arxiv (2018) Arxiv | Dataset

Beyond Action Recognition: Action Completion in RGB-D Data. F Heidarivincheh, M Mirmehdi, D Damen. British Machine Vision Conference (BMVC), Sep 2016. pdf | abstract | Dataset

Fine-Grained Object Interactions



Temporal Boundaries for Object Interactions



- How robust are current state-of-the-art approaches to annotated boundaries in test segments?
- Modify test segment boundaries, maintaining significant overlap of segments IoU > 0.5
- **Correct in Green** – **Incorrect in Red**

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

with: Davide Moltisanti
Michael Wray

Trespassing the Boundaries

GTEA Gaze+

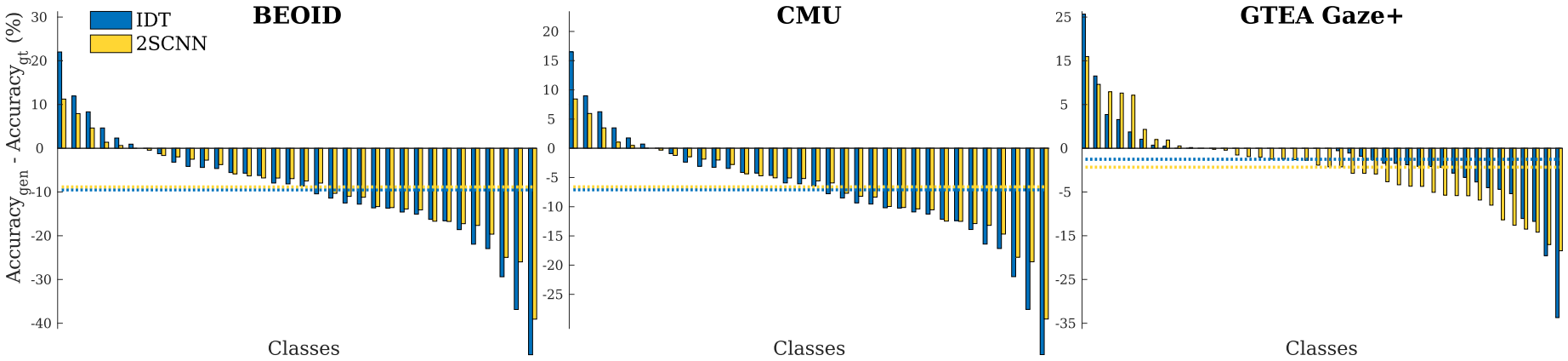
ground truth

predicted class: take knife

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

with: Davide Moltisanti
Michael Wray

Trespassing the Boundaries

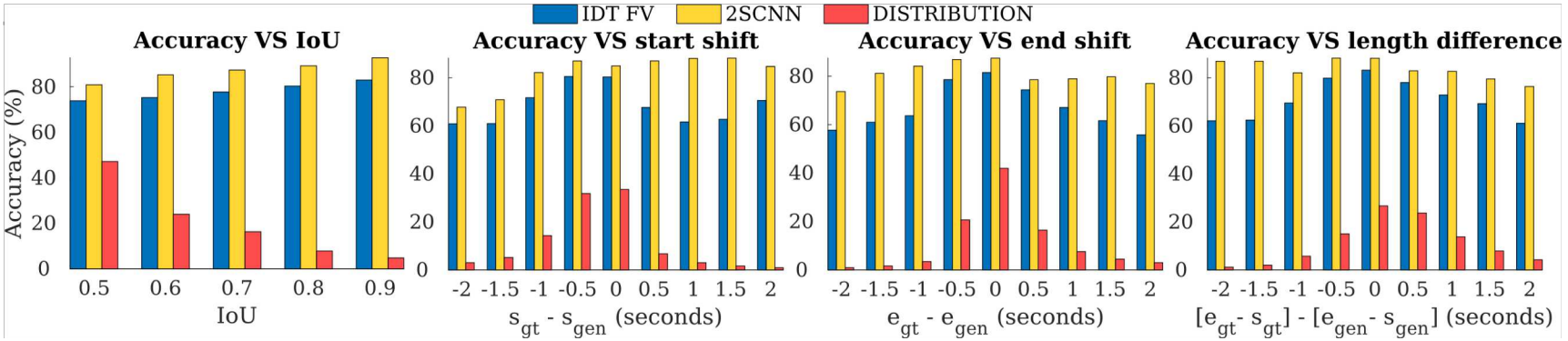


D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

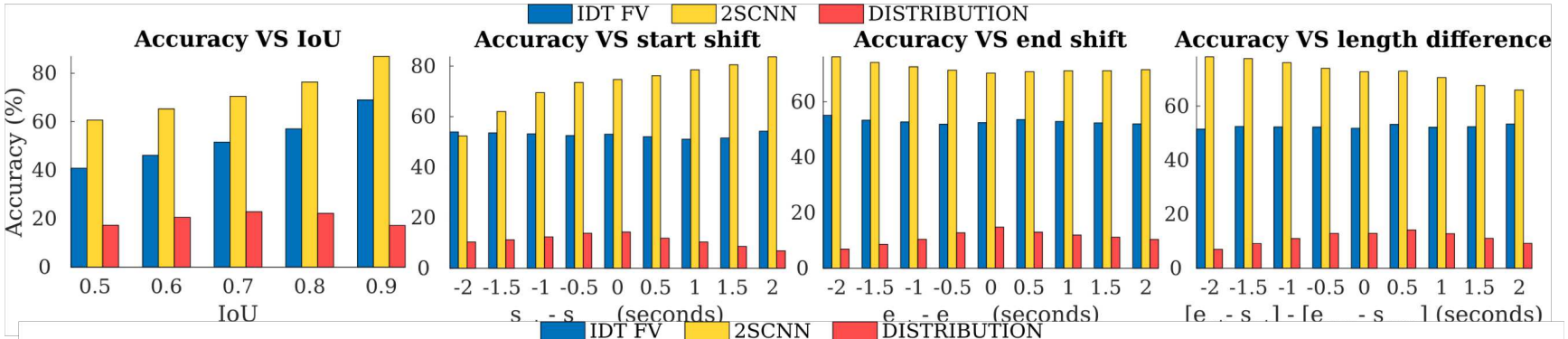
with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

Trespassing the Boundaries

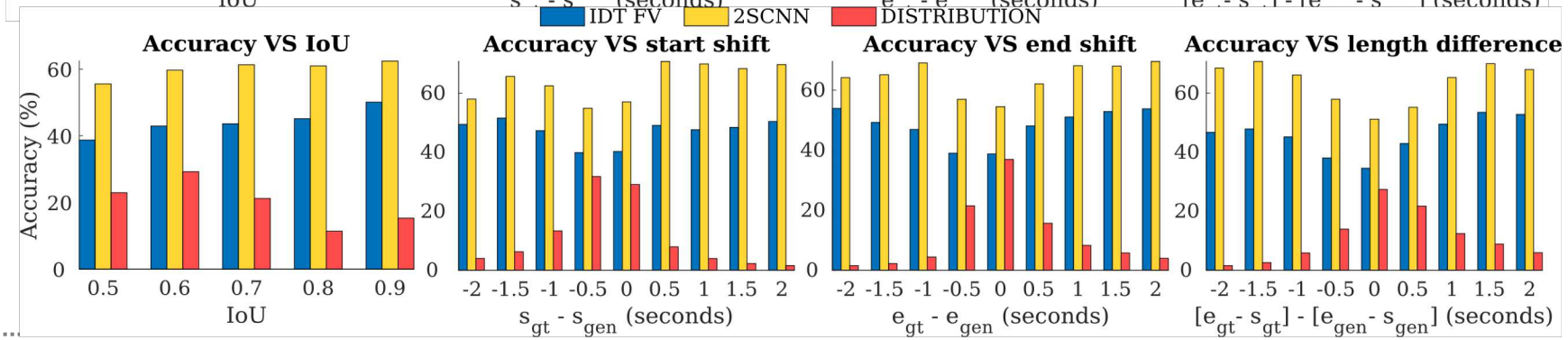
BEOID



CMU



GTEA+



University of BRISTOL

D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

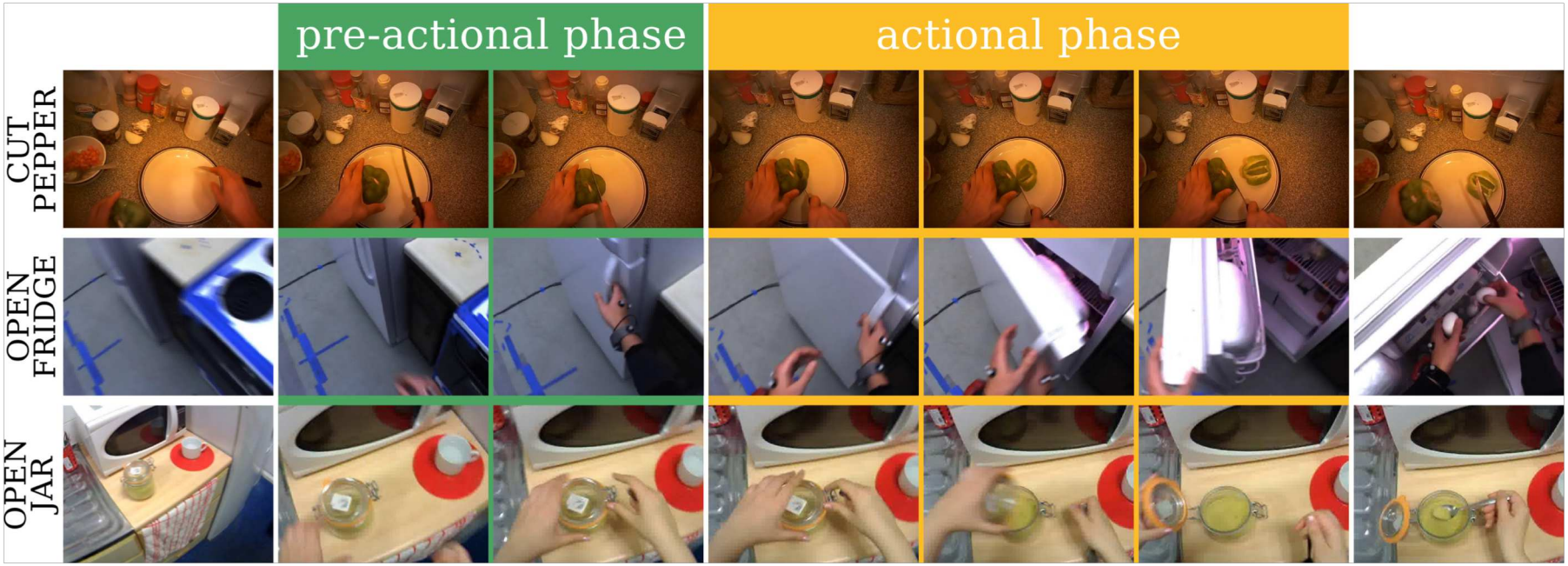
The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

- Labelling approach proposal for temporally consistent annotations
- Decomposes an object interaction into two phases:
  - *pre-actional* phase
  - *actional* phase

The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

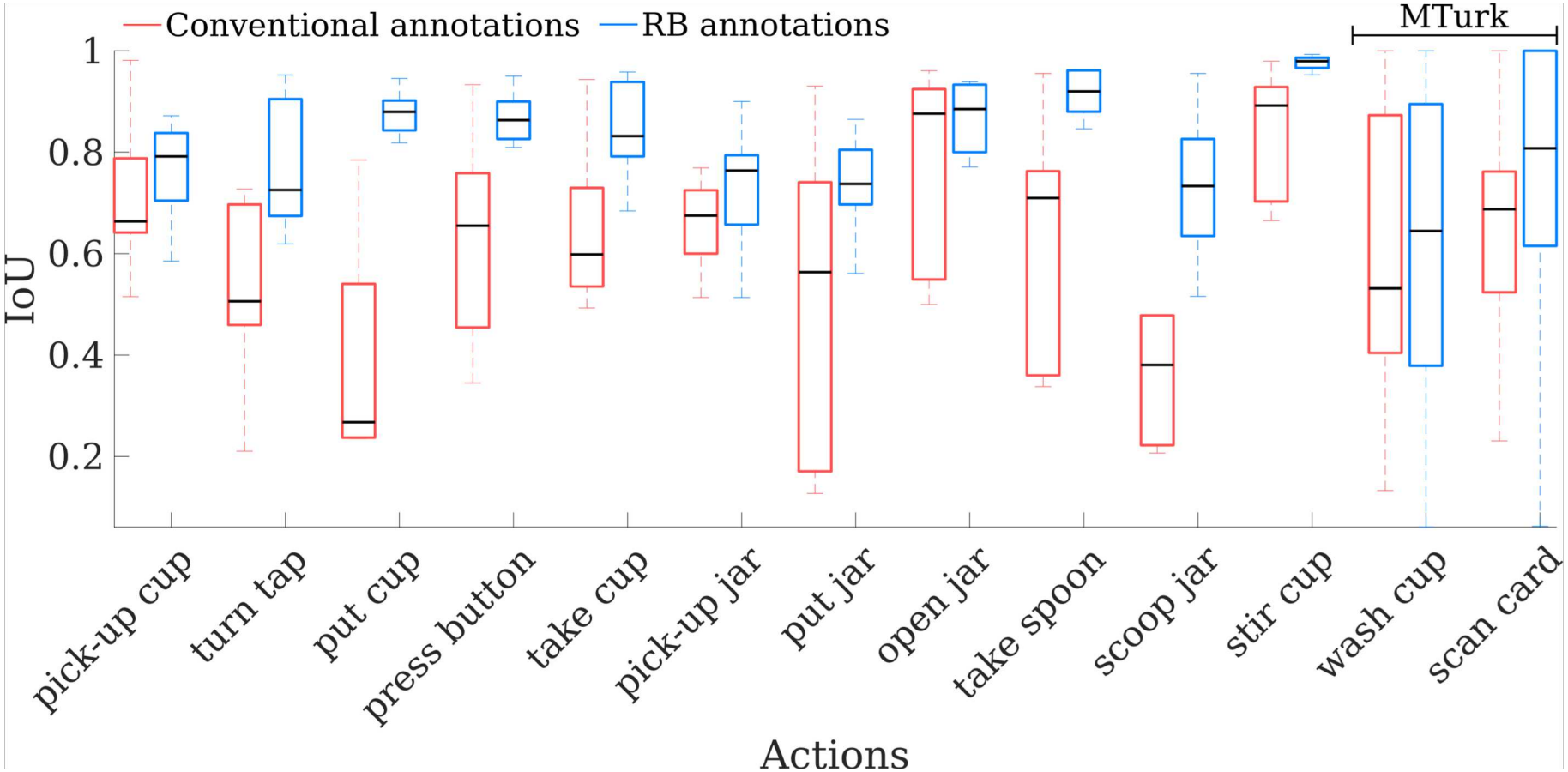
The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

Cut pepper (GTEA Gaze+)

The Rubicon Boundaries

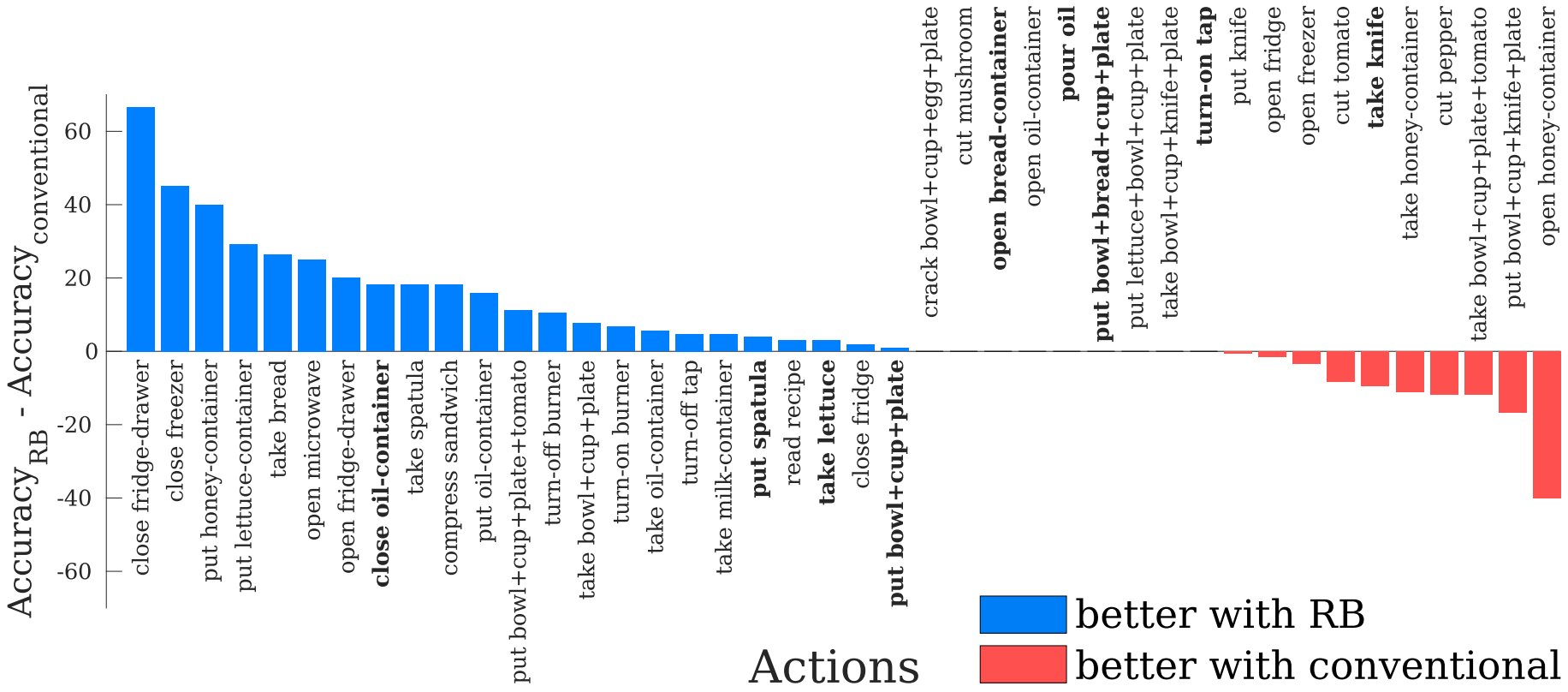
with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

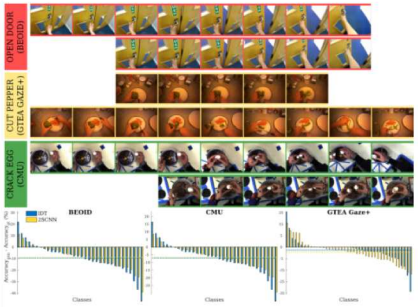
The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



More info…

Project Trespassing the Boundaries of Object Interactions



Video

Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. D Moltisanti, M Wray, W Mayol-Cuevas, D Damen. International Conference on Computer Vision (ICCV), 2017. pdf (camera ready) | arxiv

with: Davide Moltisanti
Sanja Fidler

Upcoming (CVPR 2019)…

- Learning from Single timestamps

D Moltisanti et al (2019). Action Recognition from Single Timestamp Supervision in Untrimmed Videos. *Computer Vision and Pattern Recognition (CVPR)*

with: Davide Moltisanti
Sanja Fidler

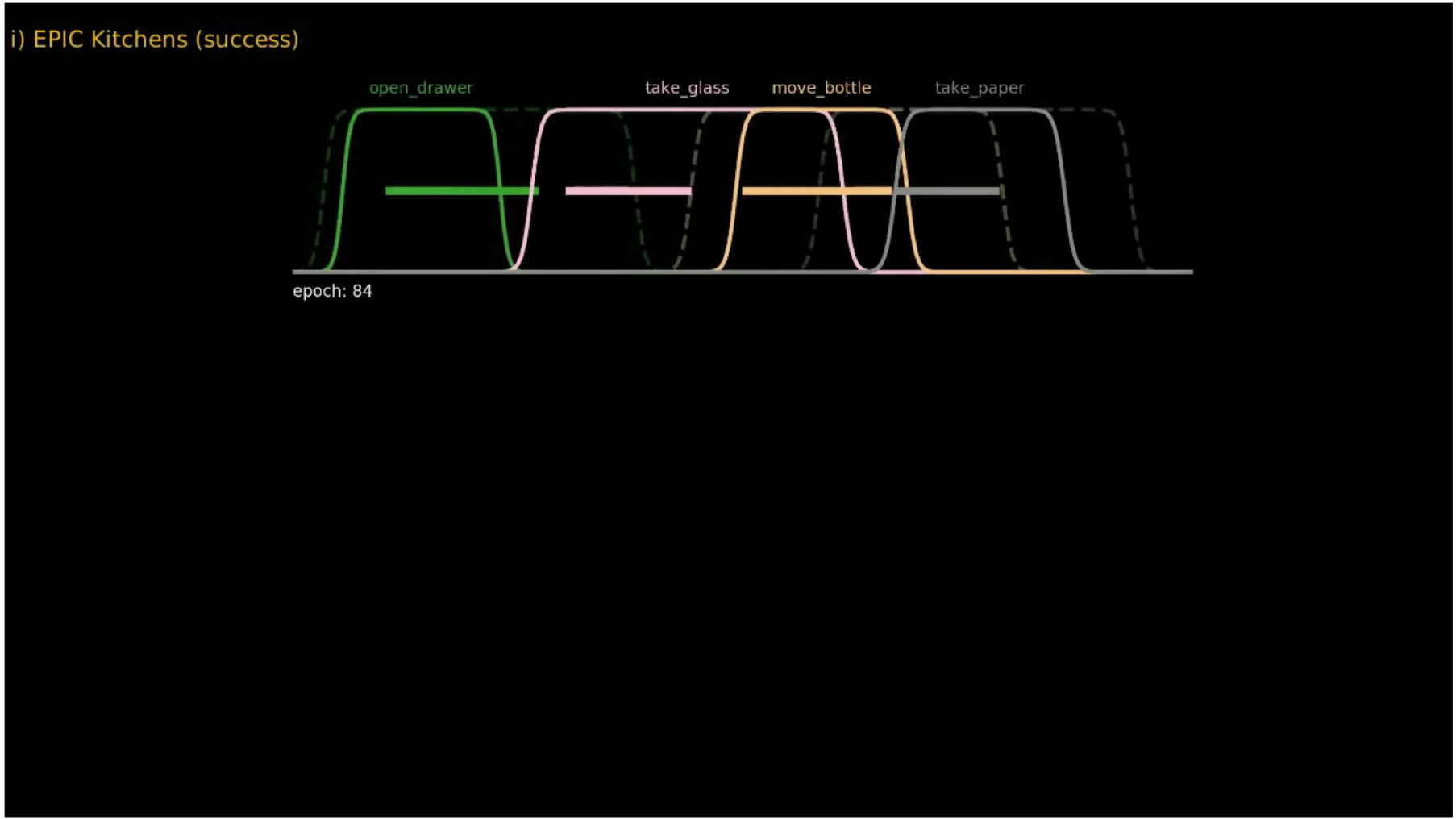
Upcoming (CVPR 2019)…

- Learning from Single timestamps

D Moltisanti et al (2019). Action Recognition from Single Timestamp Supervision in Untrimmed Videos. *Computer Vision and Pattern Recognition (CVPR)*

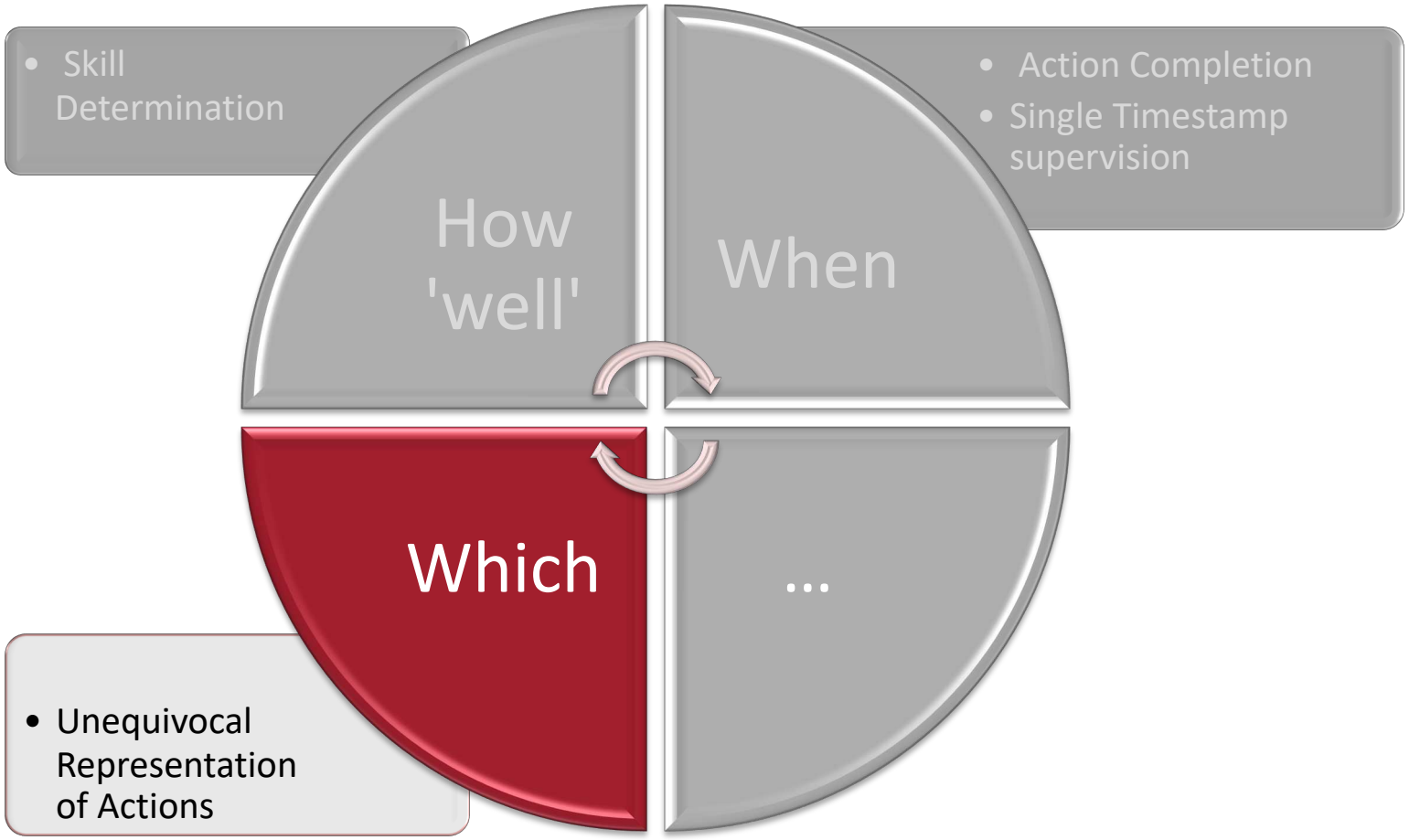
with: Davide Moltisanti
Sanja Fidler

Upcoming (CVPR 2019)...



D Moltisanti et al (2019). Action Recognition from Single Timestamp Supervision in Untrimmed Videos. *Computer Vision and Pattern Recognition (CVPR)*

Fine-Grained Object Interactions



Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

- Think of an “open” action…

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

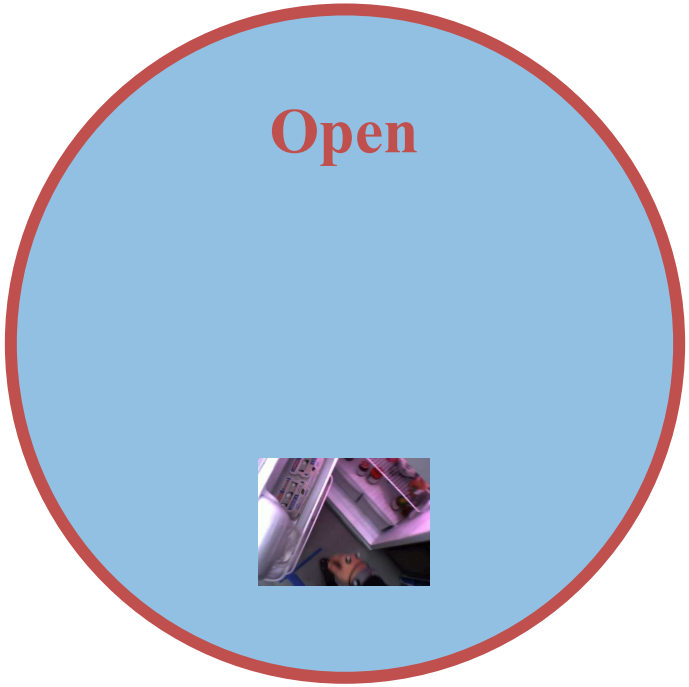
Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

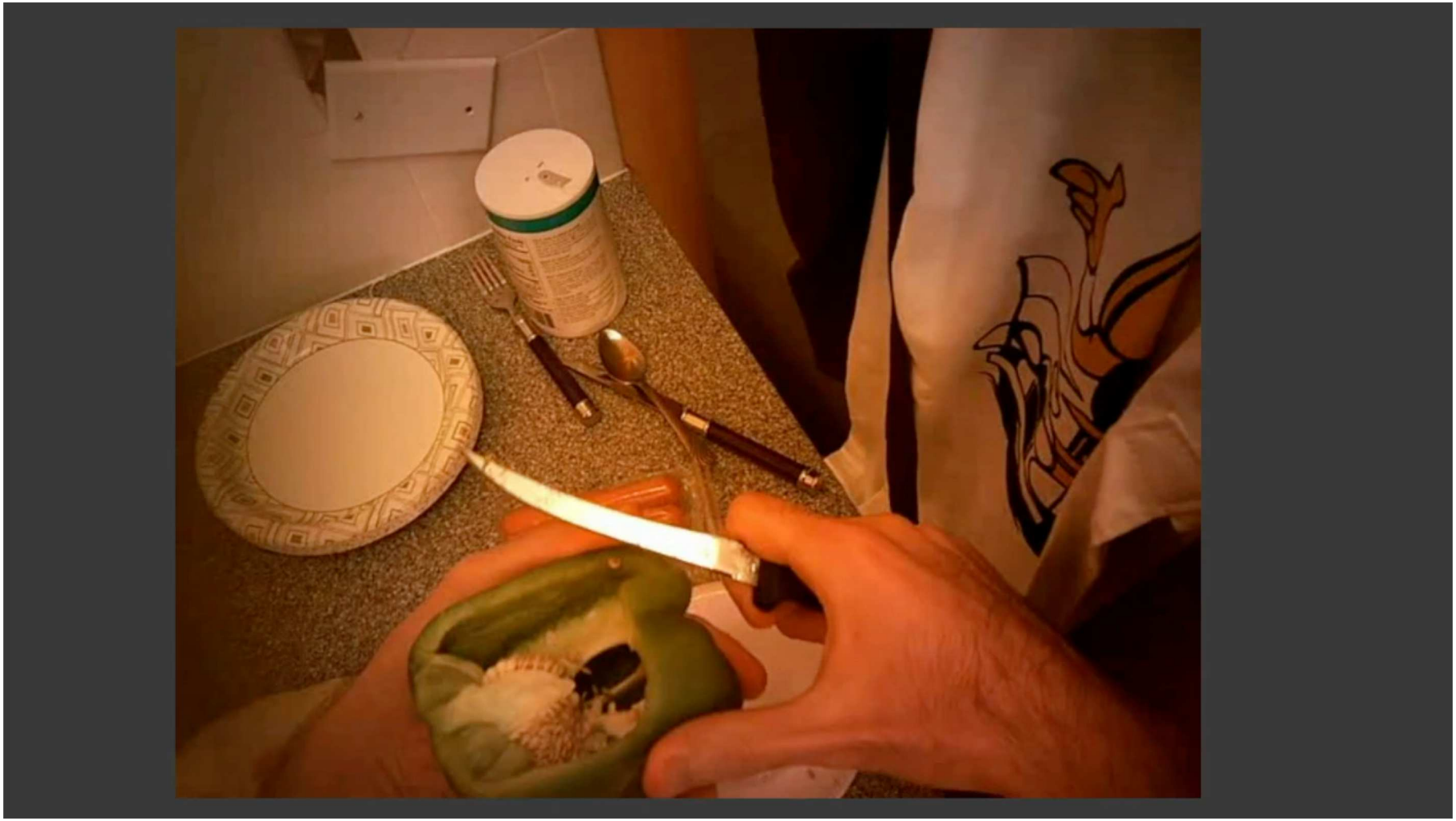
with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

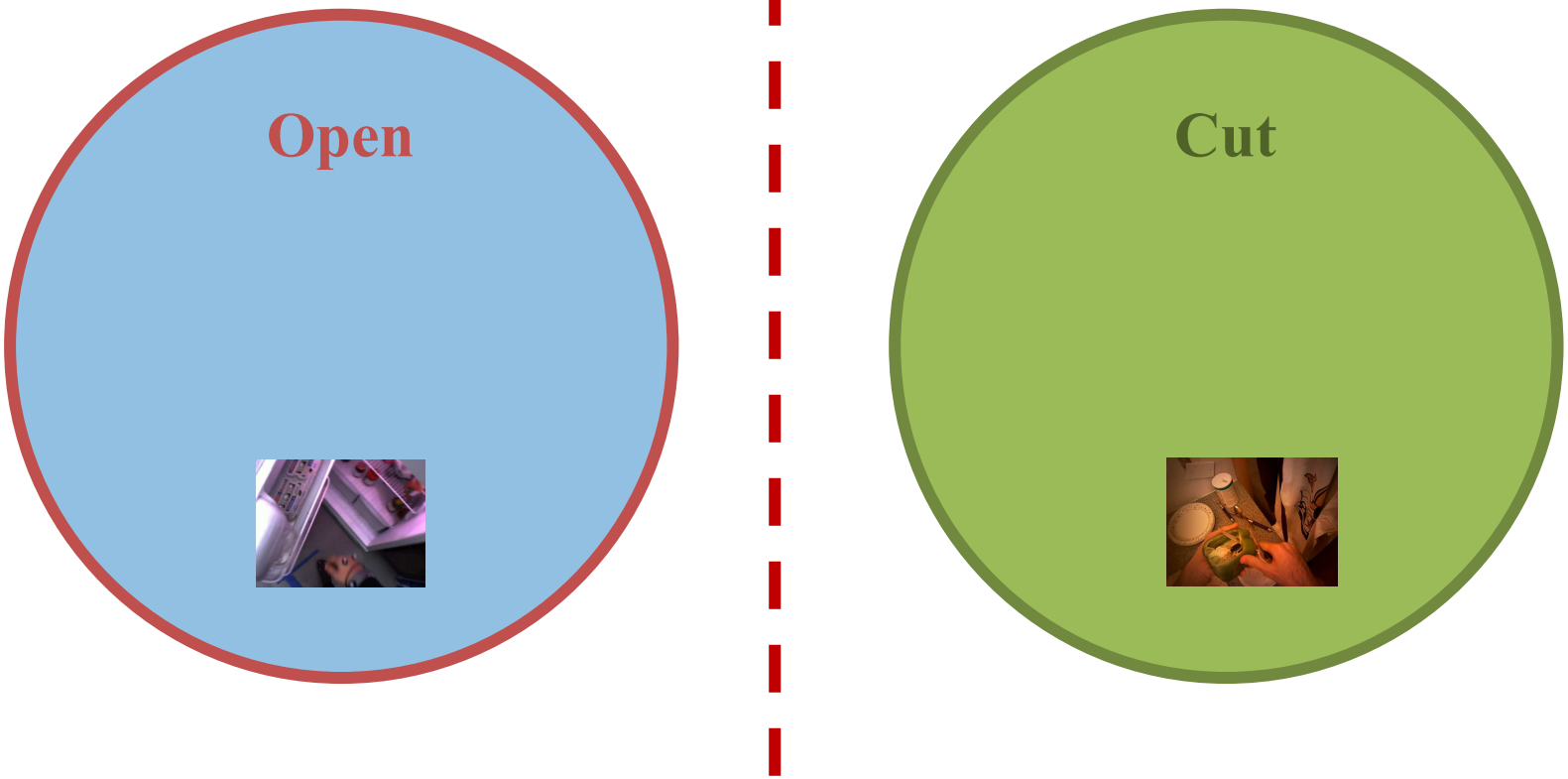
with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

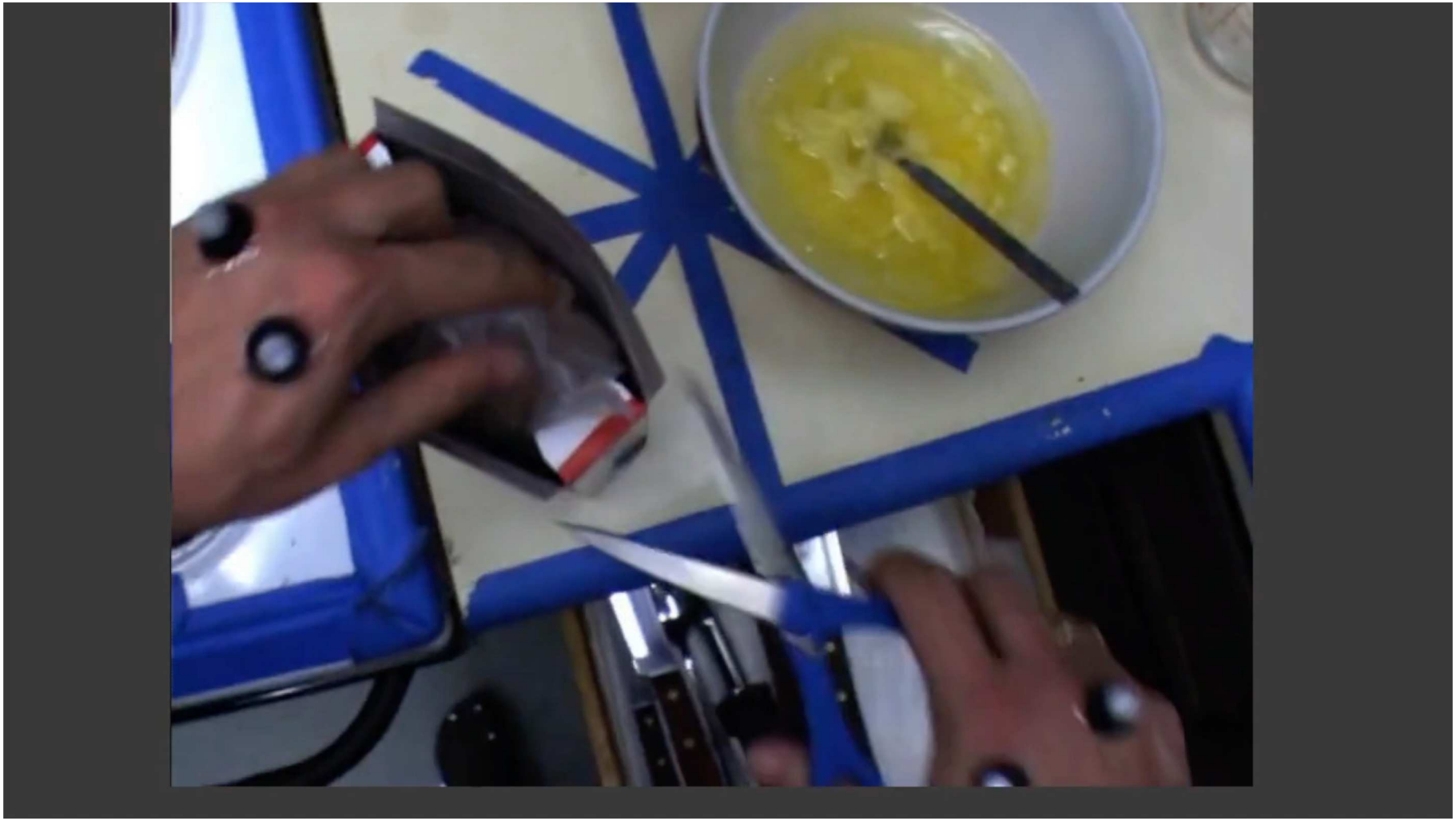
with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

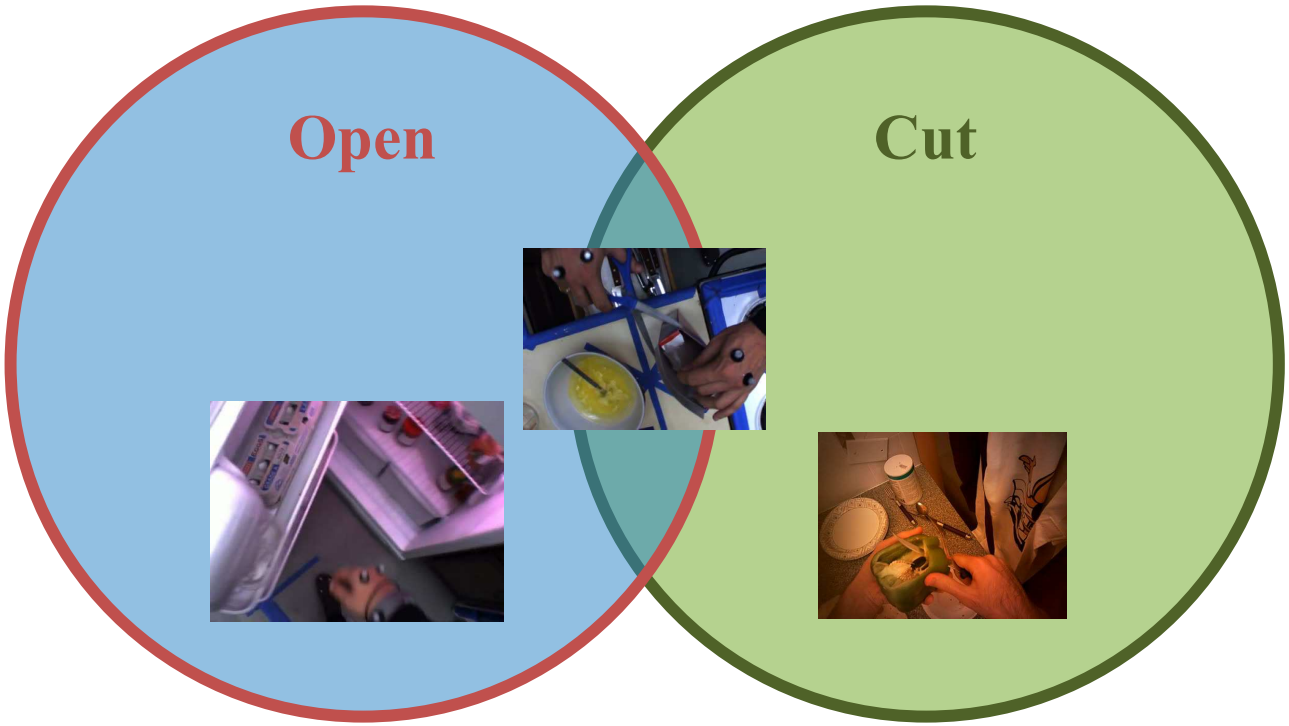
with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

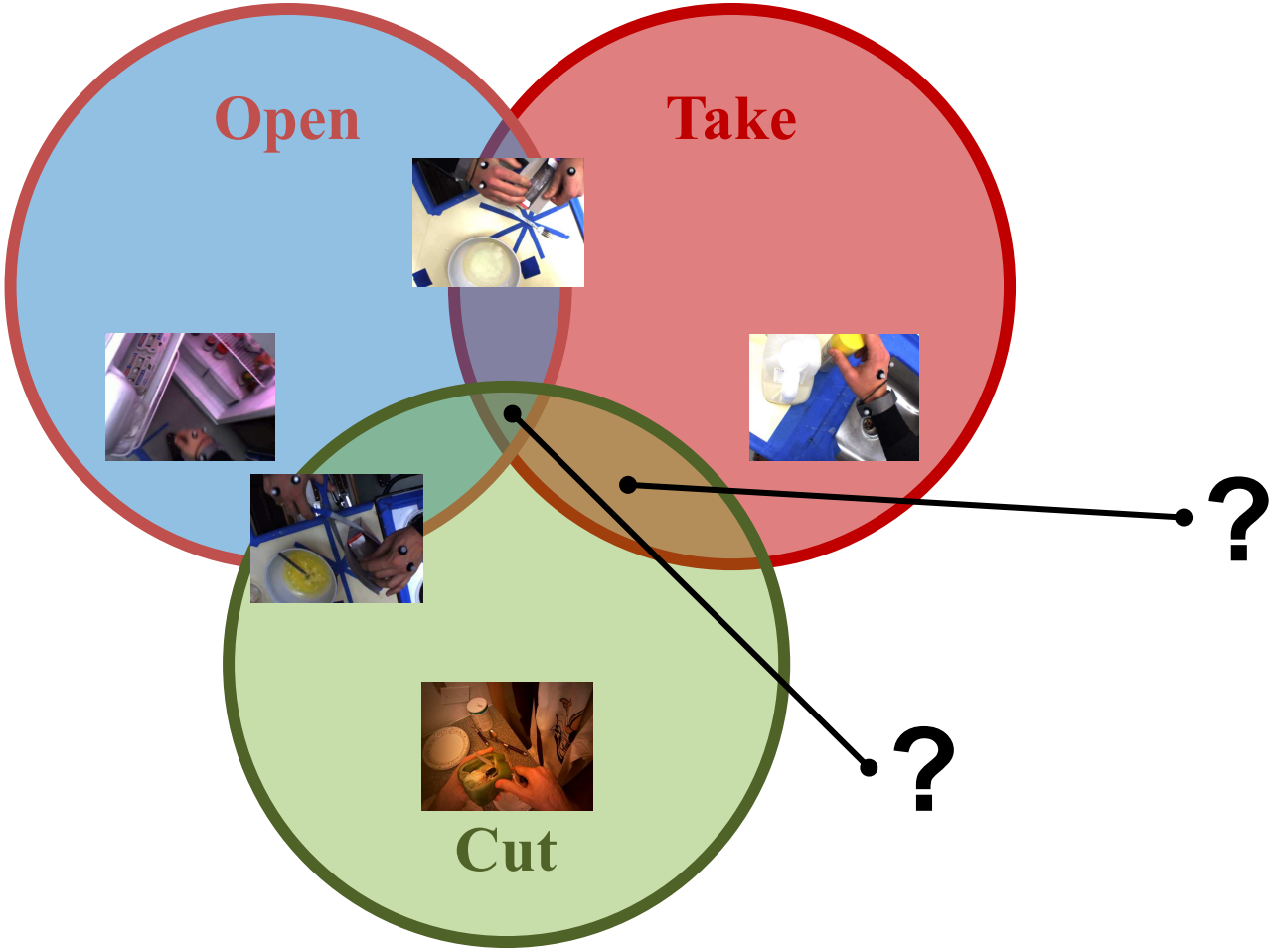
with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Object Interactions – the Dilemma

with: Michael Wray
Davide Moltisanti
Walterio Mayol-Cuevas



M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

- Action representations using a single verb is highly-ambiguous
  - Solution1: pre-selected non-overlapping verbs (SL)
    - run, walk, open, close
  - Solution2: Using nouns to disambiguate actions (V-N)
    - open-drawer, open-bottle, open-fridge
    - actions constrained to known nouns
  - Solution3: Multi-verb labels (ML, SAML)
    - open, hold, pull
    - How many verbs would be enough?

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

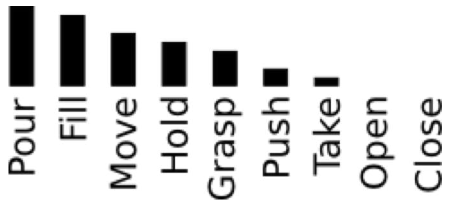
- Soft-Assigned Multi-Label
  - Multi-label using verbs only
  - Each verb assigned a value between 0 and 1
  - Object agnostic
  - Trained with Sigmoid Binary Cross Entropy

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

- Collected from AMT



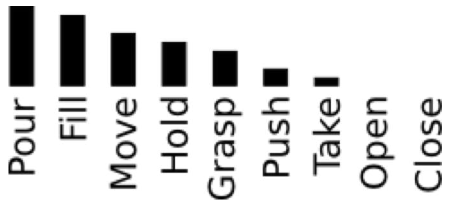
- Annotators agree:
  - Relevant Verb -> Main action
  - Irrelevant Verb -> unrelated motion
- Annotators disagree:
  - Relevant motion but not the main action

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

- Collected from AMT



SL

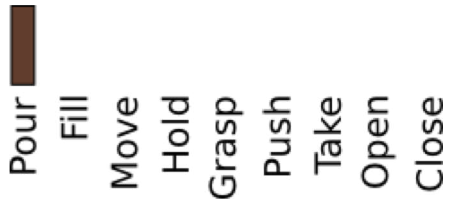
- Majority Vote.
- One-hot vector.

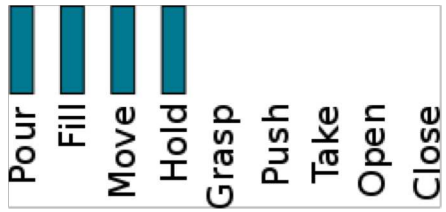
ML

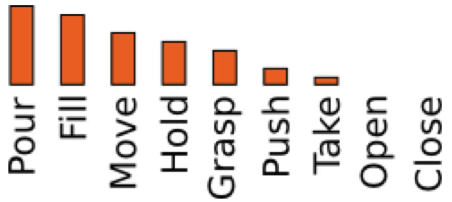
- Threshold of 0.5.
- Binary Vector

SAML

- Full Annotation.
- Continuous Vector.





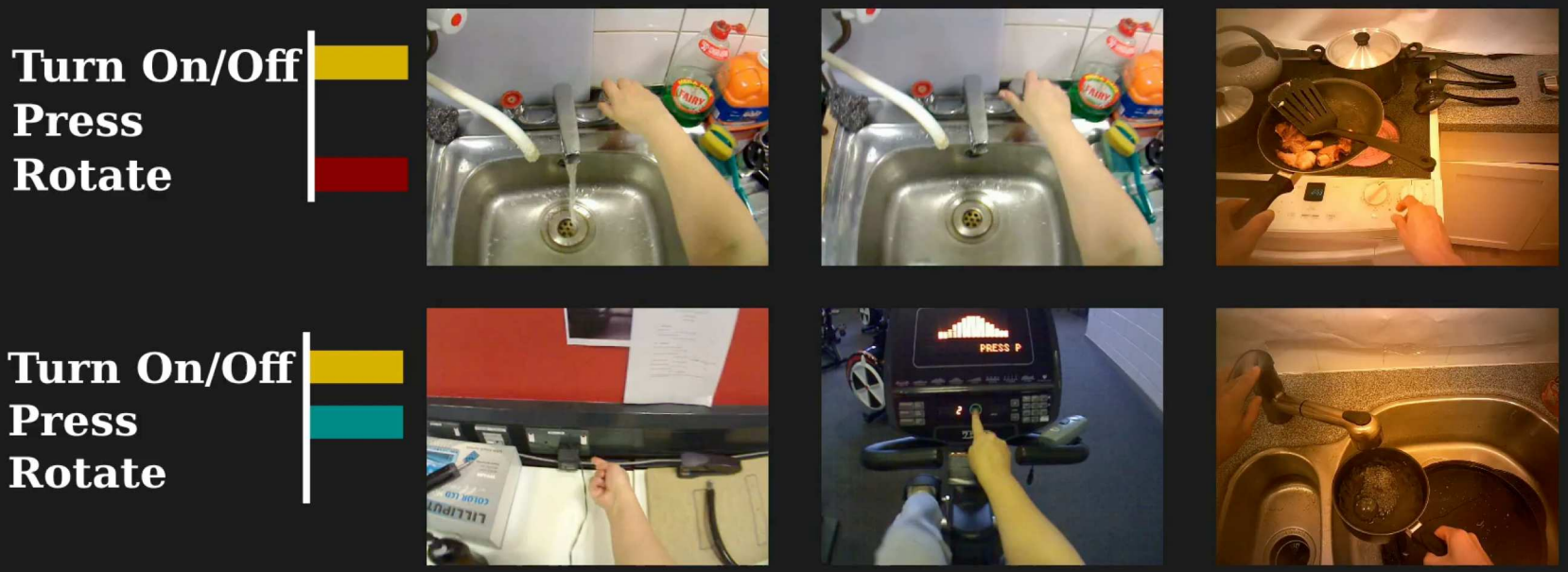


M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

Towards an Unequivocal Representation of Actions

with: Michael Wray
Davide Moltisanti

Top 3 retrieved classes across all datasets.

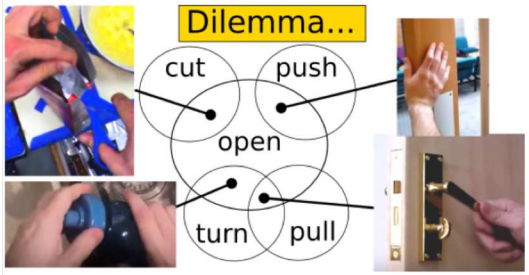


Labelling Method can differentiate turn On/Off tap by pressing and by rotating.

M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.
M Wray et al (2017). Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. Arxiv

More info…

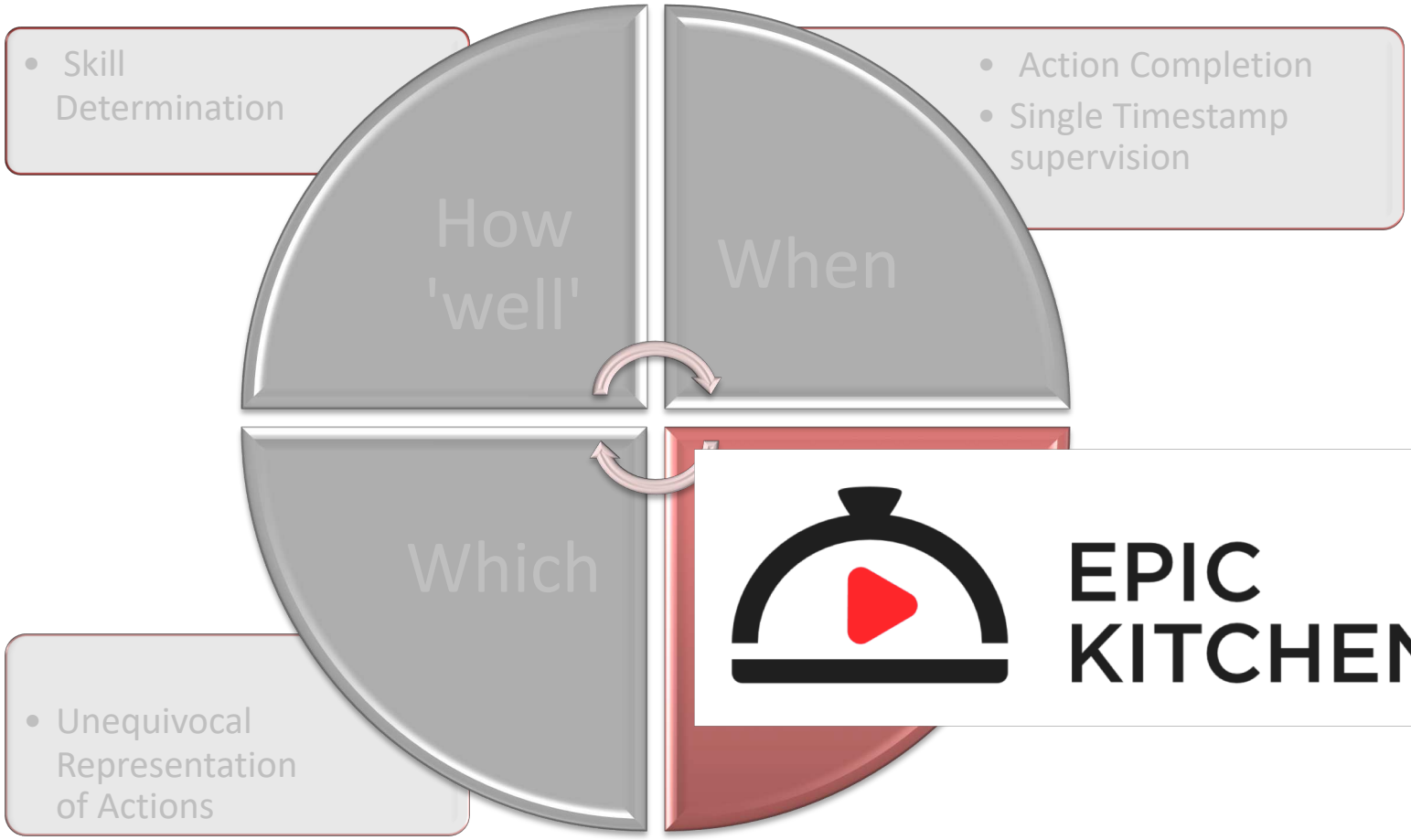
Project Towards an Unequivocal Representation of Actions



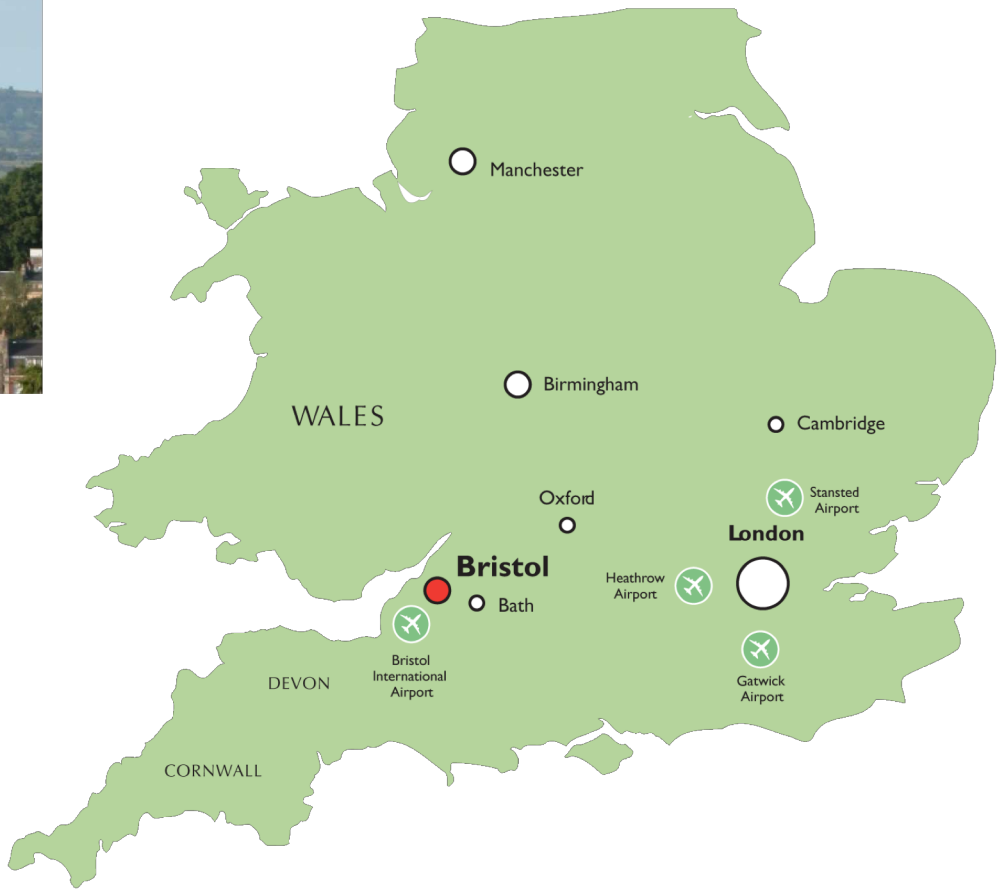
Towards an Unequivocal Representation of Actions. M Wray, D Moltisanti, D Damen. ArXiv, 2018.

Improving Classification by Improving Labelling: Introducing Probabilistic Multi-Label Object Interaction Recognition. M Wray, D Moltisanti, W Mayol-Cuevas, D Damen. ArXiv, 2017. arxiv

Fine-Grained Object Interactions



Bristol and University of Bristol



Bristol and University of Bristol

Thank you…

For further info, datasets, code, publications…

http://dimadamen.github.io

@dimadamen

http://www.linkedin.com/in/dimadamen

Join epic-community@bristol.ac.uk

- send an email to: sympa@sympa.bristol.ac.uk
- with the subject: subscribe epic-community
- and blank message content

University of BRISTOL · UNIVERSITY OF TORONTO · UNIVERSITÀ degli STUDI di CATANIA

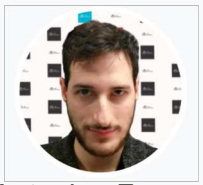
Scaling Egocentric Vision: The EPIC-KITCHENS Dataset

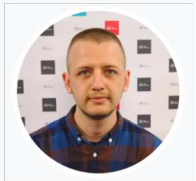

Dima Damen


Hazel Doughty


Giovanni M. Farinella


Sanja Fidler


Antonino Furnari


Evangelos Kazakos


Davide Moltisanti


Jonathan Munro


Toby Perrett


Will Price


Michael Wray

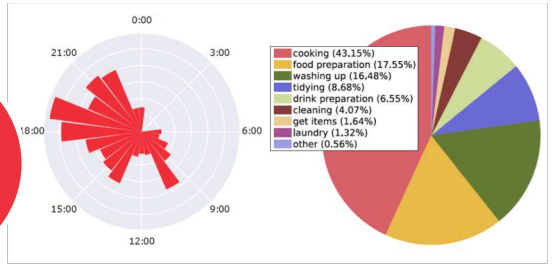
Scaling Egocentric Vision

Dataset	Ego?
EPIC-KITCHENS	✓
EGTEA Gaze+ [16]	✓
Charades-ego [41]	70% ✓
BEOID [6]	✓
GTEA Gaze+ [13]	✓
ADL [36]	✓
CMU [8]	✓
YouCook2 [56]	×
VLOG [14]	×
Charades [42]	×
Breakfast [28]	×
50 Salads [44]	×
MPII Cooking 2 [39]	×

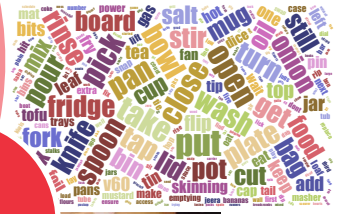
Scaling Egocentric Vision

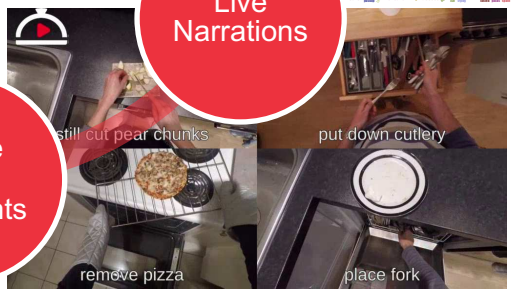
Data Collection

Native Environment, Natural Interactions



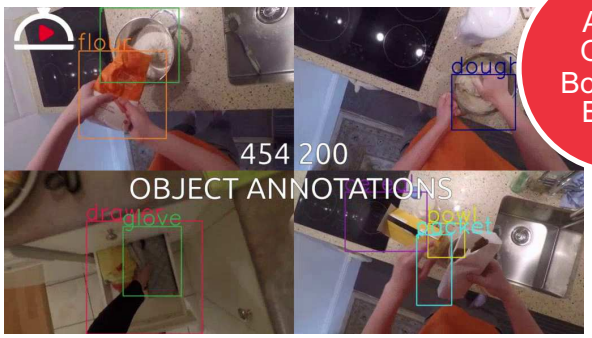
Live Narrations





Dense Action Segments

Active Object Bounding Boxes



Benchmark and Challenges

UNIVERSITY OF TORONTO
University of BRISTOL
UNIVERSITÀ degli STUDI di CATANIA

EPIC KITCHENS

Data Collection

- Head-Mounted Go-Pro, adjustable mounting
- Recording starts immediately before entering the kitchen
- Only stopped before leaving the kitchen

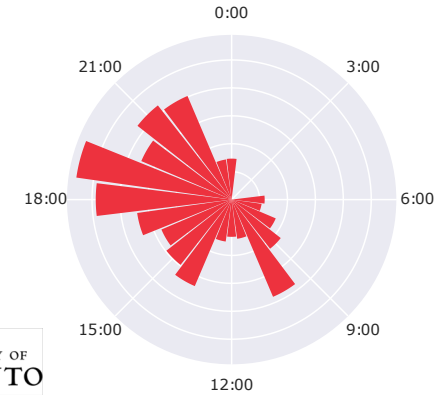
UNIVERSITY OF TORONTO

University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA

Data Collection

- 32 kitchens
- Single-person environments
- 4 cities
- May – Nov 2017 – 55 hours
- 10 nationalities
- 3 days - all kitchen activities

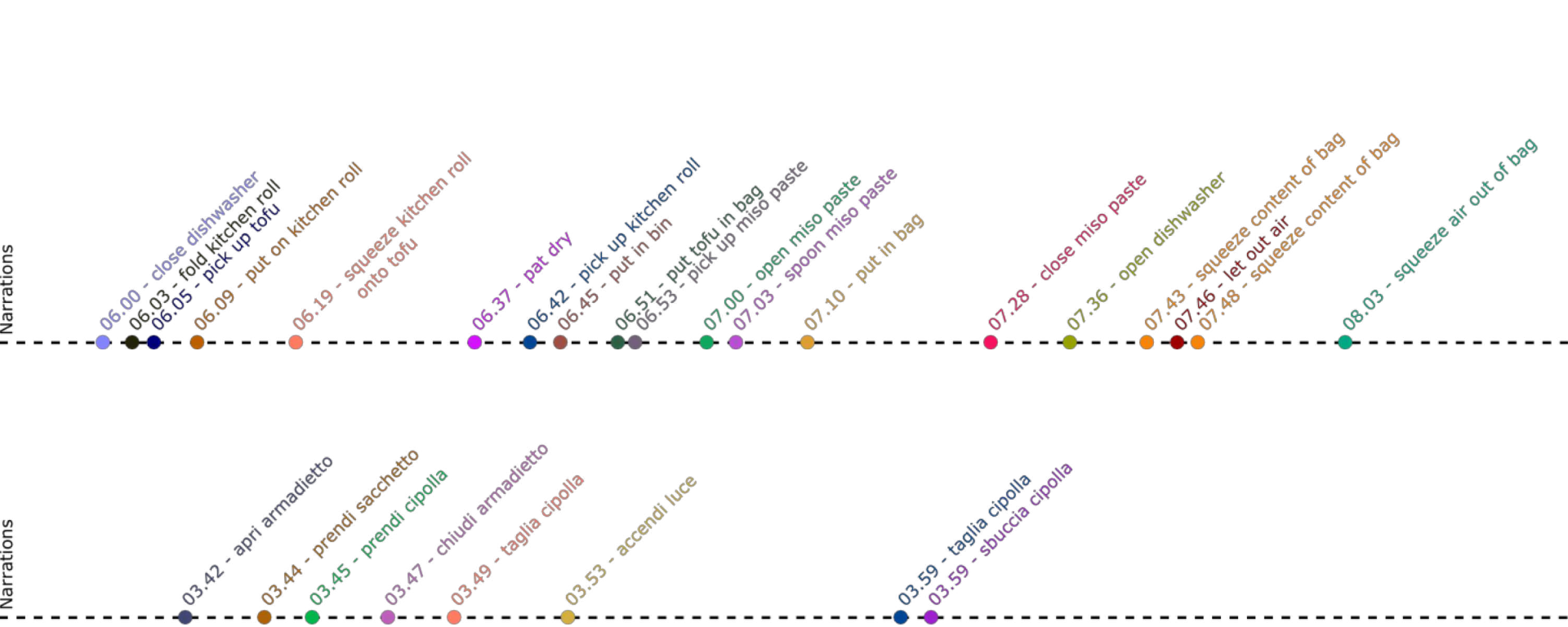


UNIVERSITY OF TORONTO

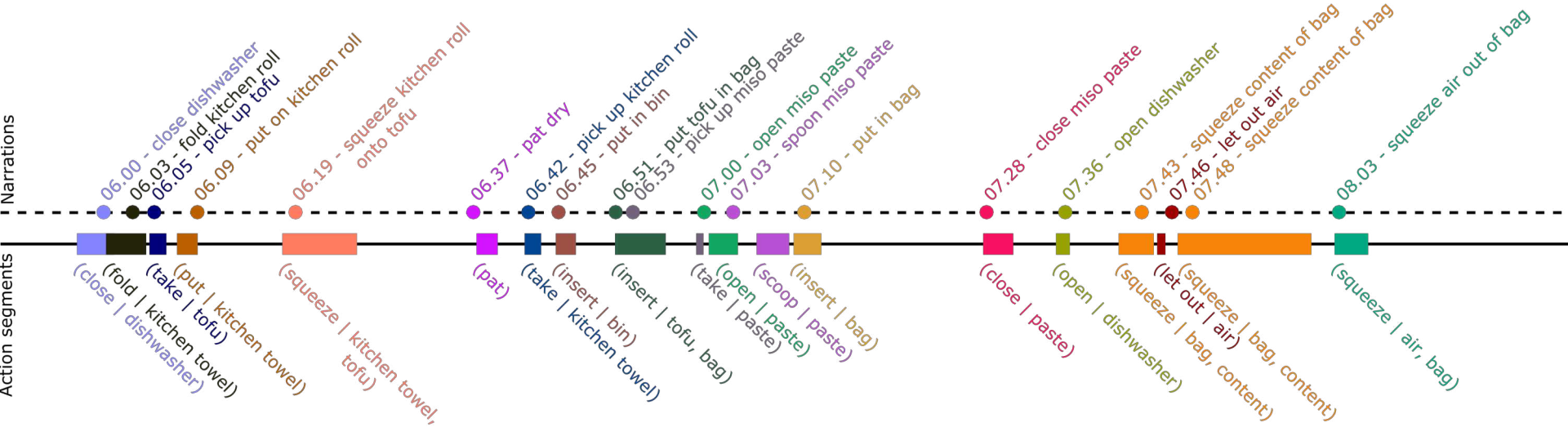
University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA

Annotations (1) - Narrations



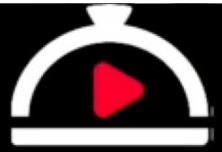
Annotations (2) – Action Segments



UNIVERSITY OF TORONTO

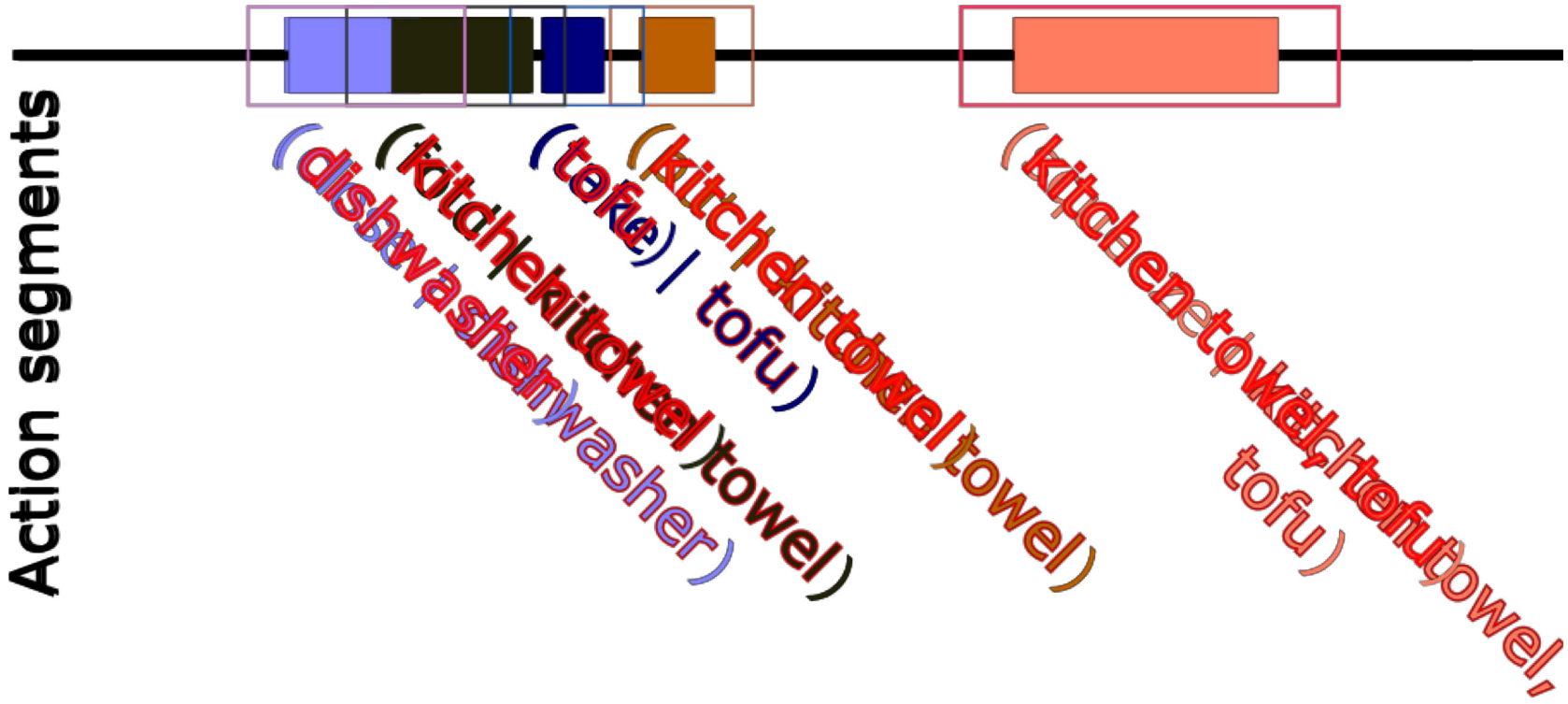
University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA



39 000
ACTION SEGMENTS

Annotations (3) – Object Bounding Boxes




454 200
OBJECT ANNOTATIONS

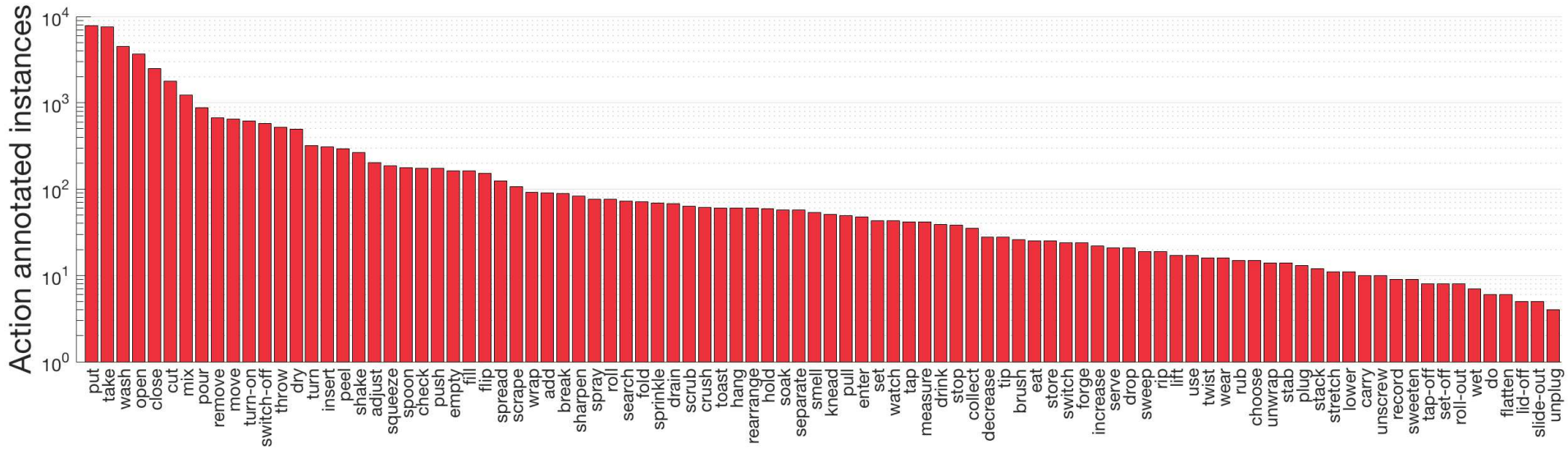
Annotations (4) – Verb and Noun Classes

take, grab, pick, get, fetch, pick-up, ...

- 125 verb classes
- 331 noun classes

EPIC KITCHENS

Annotations Statistics

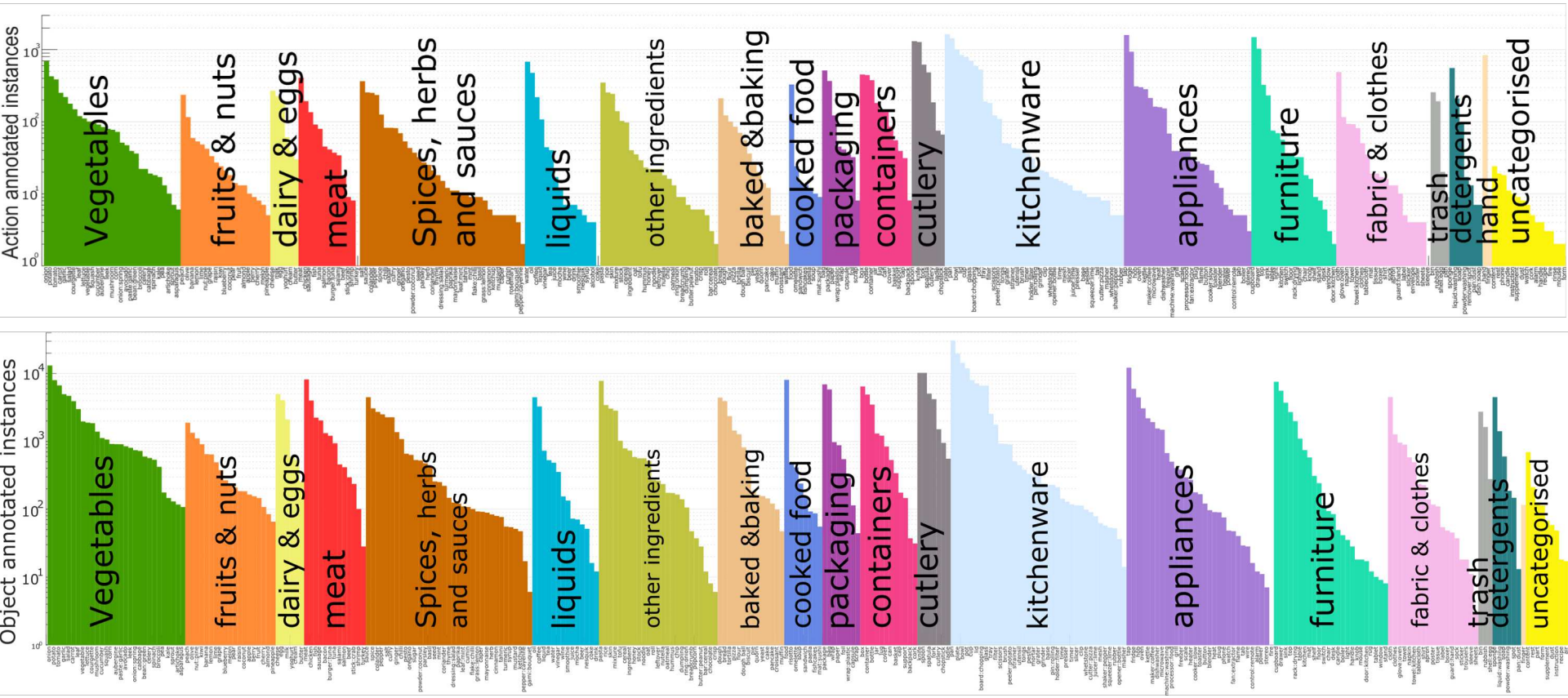


UNIVERSITY OF TORONTO

University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA

Annotations Statistics



Train/Test Splits

- 20% - Seen Test Set
  - 28 Kitchens
- 7% - Unseen Test Set
  - 4 Kitchens

Table 4: Statistics of test splits: seen (S1) and unseen (S2) kitchens

	#Subjects	#Sequences	Duration (s)	%	Narrated Segments	Action Segments	Bounding Boxes
Train/Val	28	272	141731		28,587	28,561	326,388
S1 Test	28	106	39084	20%	8,069	8,064	97,872
S2 Test	4	54	13231	7%	2,939	2,939	29,995

Dataset Release

FHD video:

- 1920x1080 px
- 60FPS

RGB frames:

- 456x256 px
- 60FPS

TVL$_1$ optical flow (u, v) frames:

- 456x256 px
- 30FPS

UNIVERSITY OF TORONTO

University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA

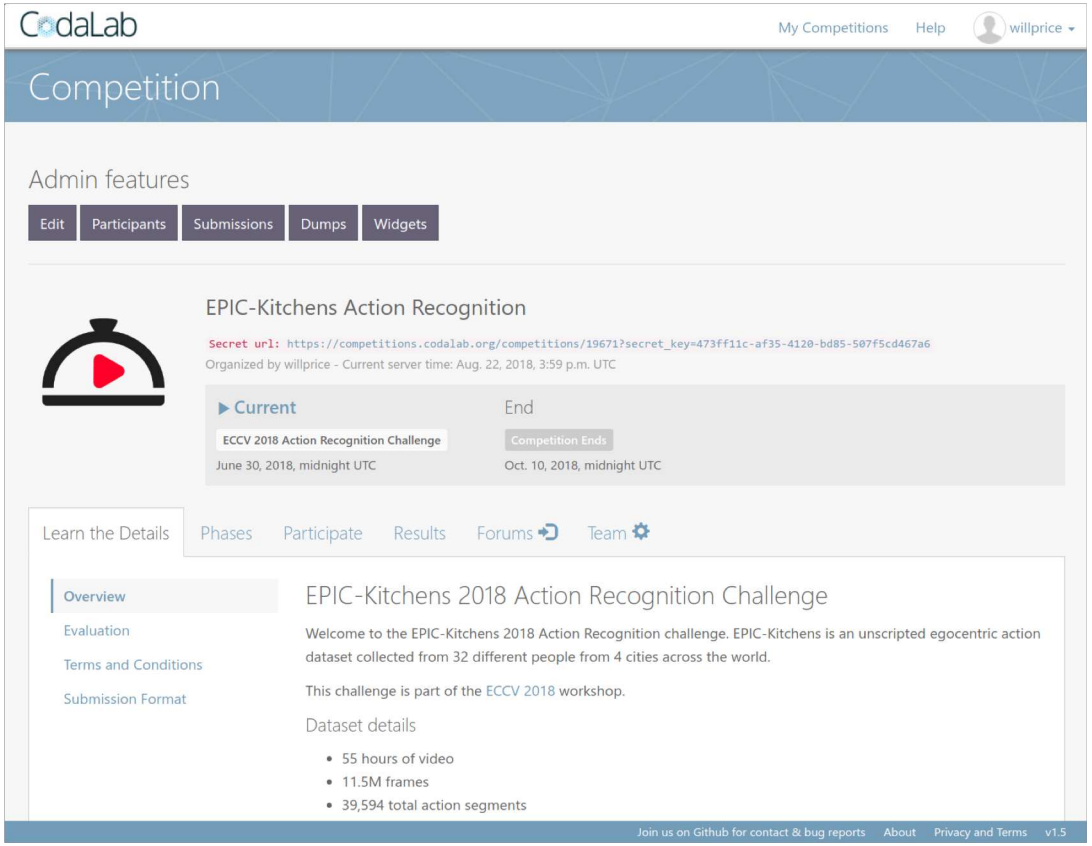
Open Challenges

1. Object Detection Challenge
2. Action Recognition Challenge
3. Action Anticipation Challenge

Open Challenges

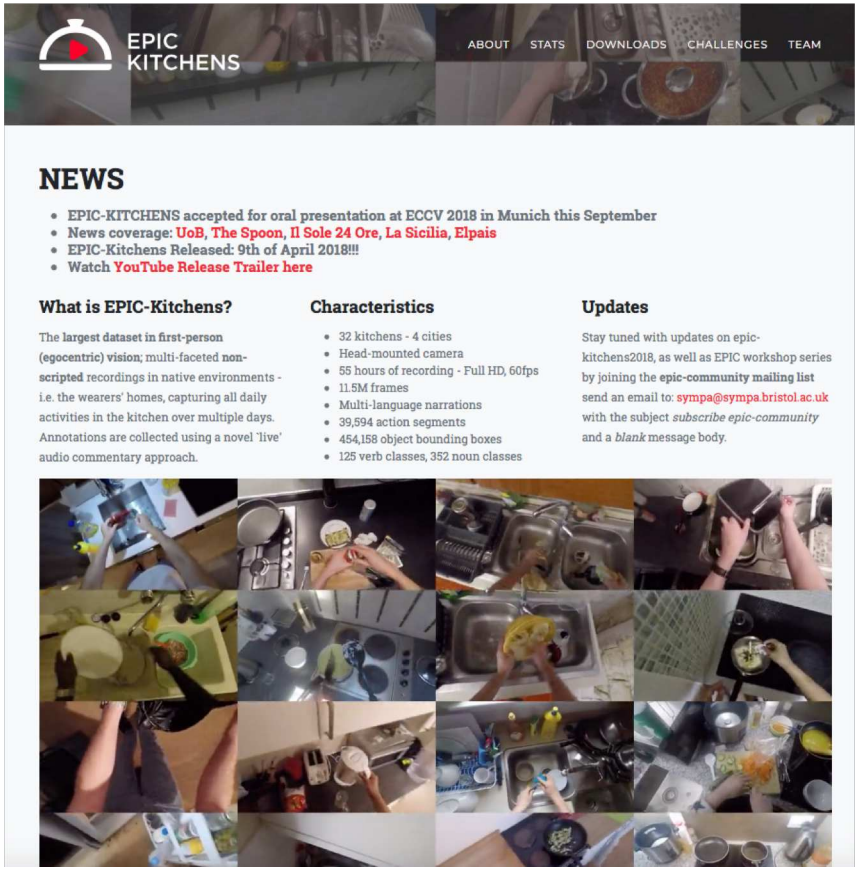
- Challenges open on CodaLab – 9 Sep

- First Challenge Results in CVPR 2019
- EPIC@CVPR2019
- ActivityNet@CVPR2019



More?

http://epic-kitchens.github.io



UNIVERSITY OF TORONTO

University of BRISTOL

UNIVERSITÀ degli STUDI di CATANIA