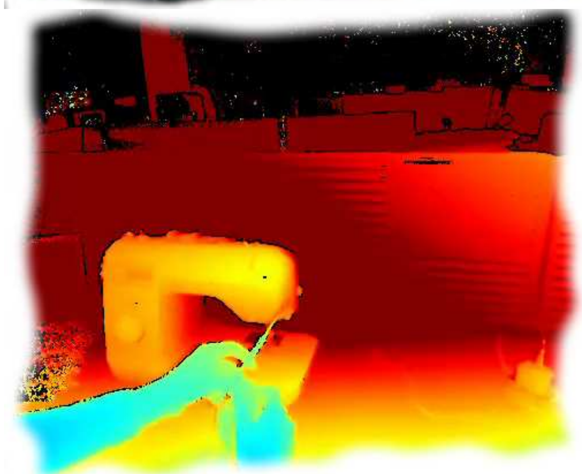
Egocentric Vision

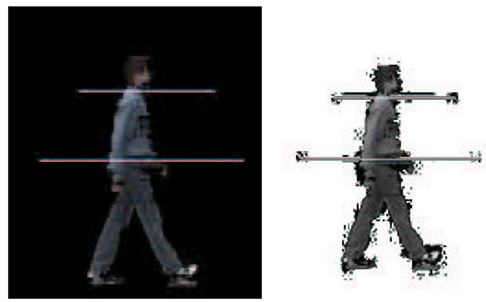
Dr Dima Damen

Department of Computer Science

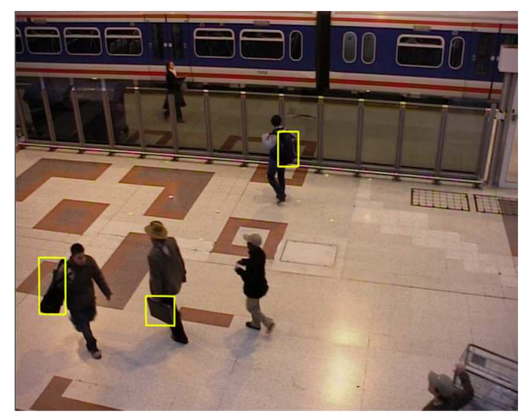


Short Bio

- 1998-2002 BSC in Computer Science
- 2002-2003 MSc in Distributed Multimedia Sys.

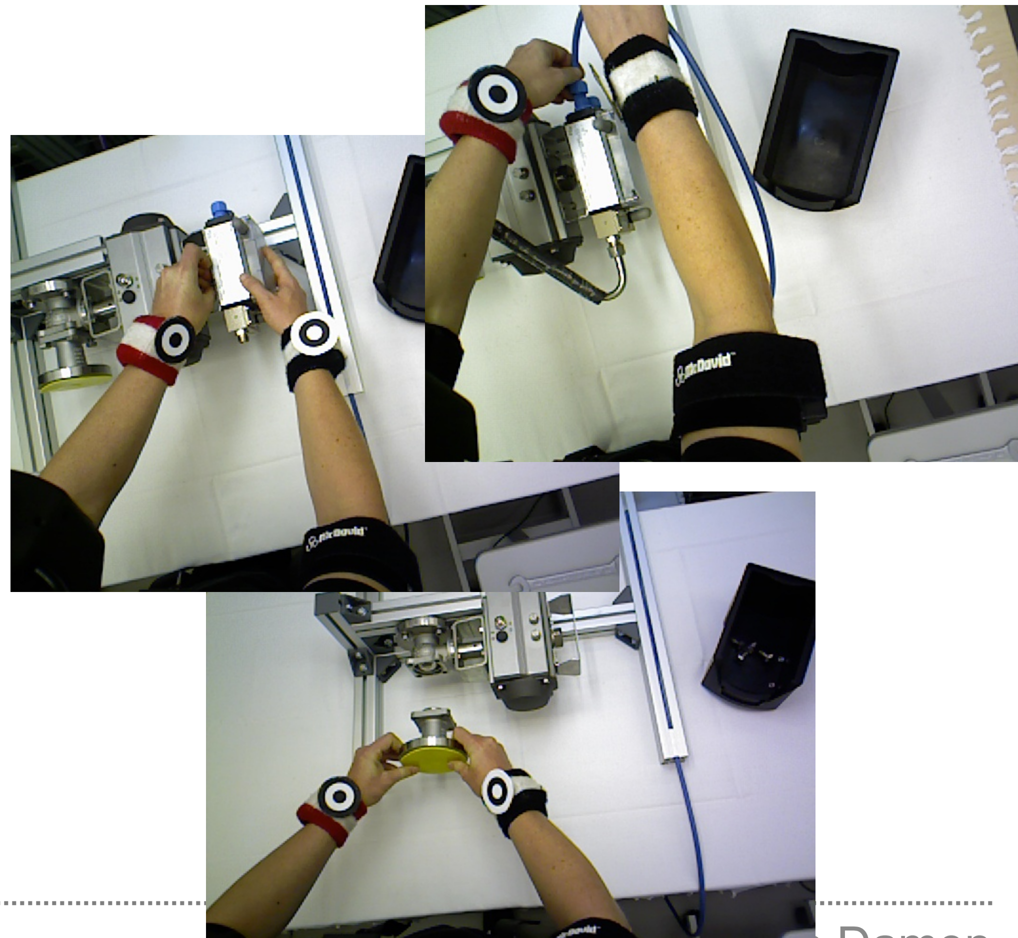


- 2006-2009 PhD in Computer Vision



Short Bio

- 2010-2012 Postdoc on EU-FP7 project



Short Bio

- 2013-2017 Assistant Prof in Computer Vision
- 2017- Associate Prof in Computer Vision

Egocentric Vision?

- Research interests: action and activity recognition
- Particularly centred around the perception of object interactions

Ego…

Ego… a person's sense of self-esteem or self-importance

Egocentric vision… the wearer serves as the central reference point in the study of interesting entities: objects, actions, interactions and intentions

Ego…

Visual Sensing – the landscape

Visual Sensing – the landscape

Least

Most

Expensive

Visual Sensing – the landscape

Most Mobile!

Least Static

Moveable

Visual Sensing – the landscape

Most
Wearable!

Hand-Held
Wireless

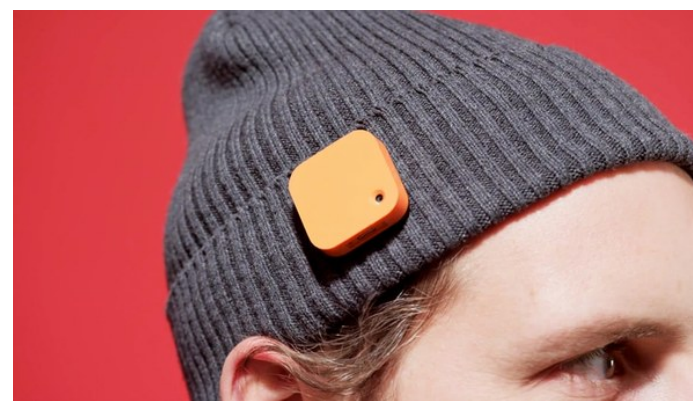
Hand-Held
Wired

Least
Static

Wearable

Wearable?

Wearable?



Wearable?

Wearable?

Wearable?

- Hat-Mounted
- Head-Mounted
- Glass-Mounted
- Shoulder-Mounted
- Chest-Mounted
- Wrist-Mounted
- Belt-Mounted
- Ankle-Mounted

But why do we care about… hardware???

- 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

Conclusions?

- Just another camera?
- Just a shaking camera?

Egocentric Vision

- The Unique Problems
  1. Camera Motion
  2. Mapping and Localisation
  3. Attention and Task-Relevance
  4. Object Interactions
  5. Multi-view Solutions
- The Unique Applications
  1. Video Summarisation
  2. Skill Determination
  3. Real-time solutions

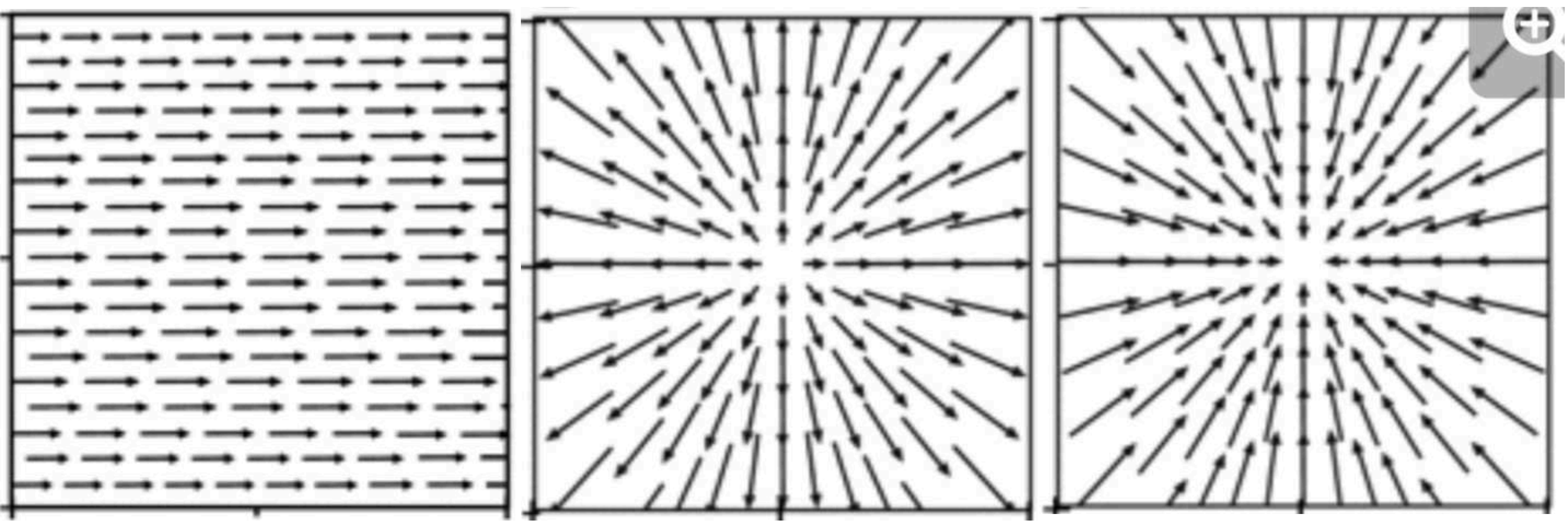
The Unique Problems

1. Camera Motion

1. Camera Motion

- Two types of motion
  - Egomotion
  - Foreground motion

Ego-motion



rotating **approaching** **receding**

Ego-motion

- Detect to:
  - Use?
  - Remove?

Hyperlapse

- https://youtu.be/sA4Za3Hv6ng

The Unique Problems

2. Mapping and Localisation

Mapping and Localisation

- https://youtu.be/ufBLu1VUQ-E

The Unique Problems

3. Attention and Task Relevance

Attention and Task Relevance

Attention and Task Relevance

- Attention in egocentric vision
  - Foreground segmentation
  - Hand-region segmentation
  - Gaze tracking

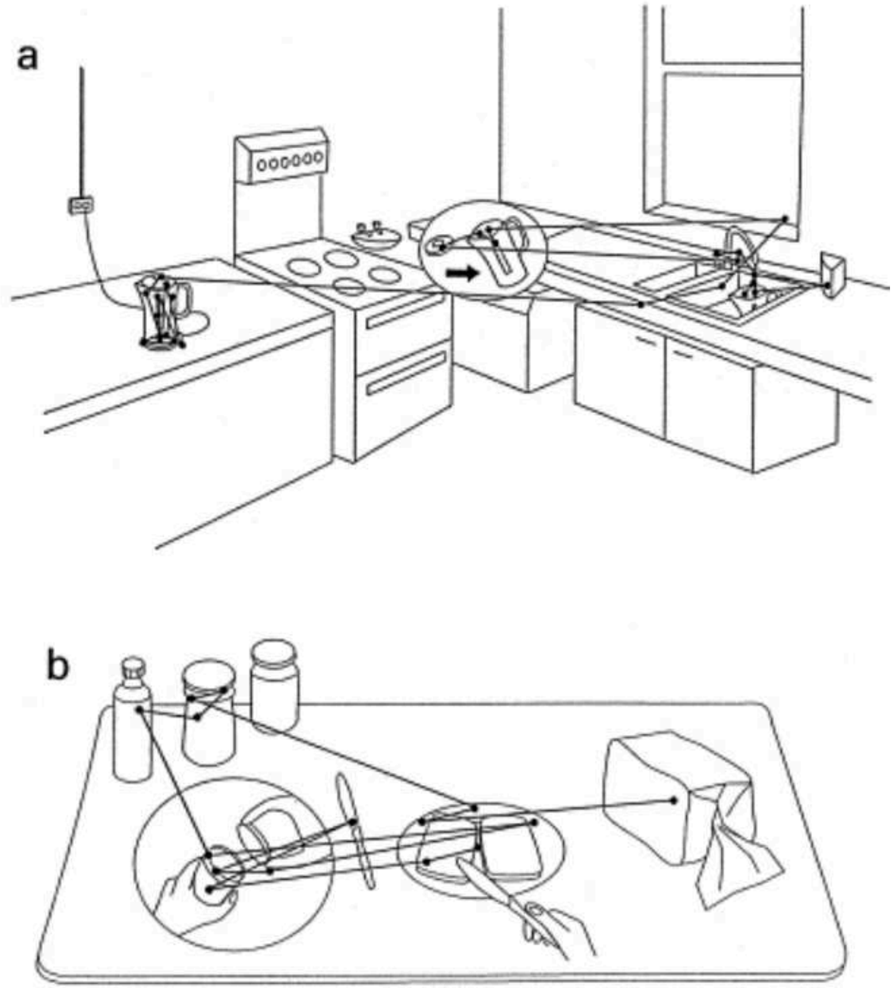
Quick introduction to human gaze

- Humans iterate between "fixations" and "saccades"
  - Fixation: short stops
  - Saccade: quick movements between fixations
- https://youtu.be/pknohrsz4Qs

Quick introduction to human gaze

Quick introduction to human gaze

Quick introduction to human gaze



Quick introduction to human gaze

- The notion of fixation/saccade has recently inspired attention models in vision

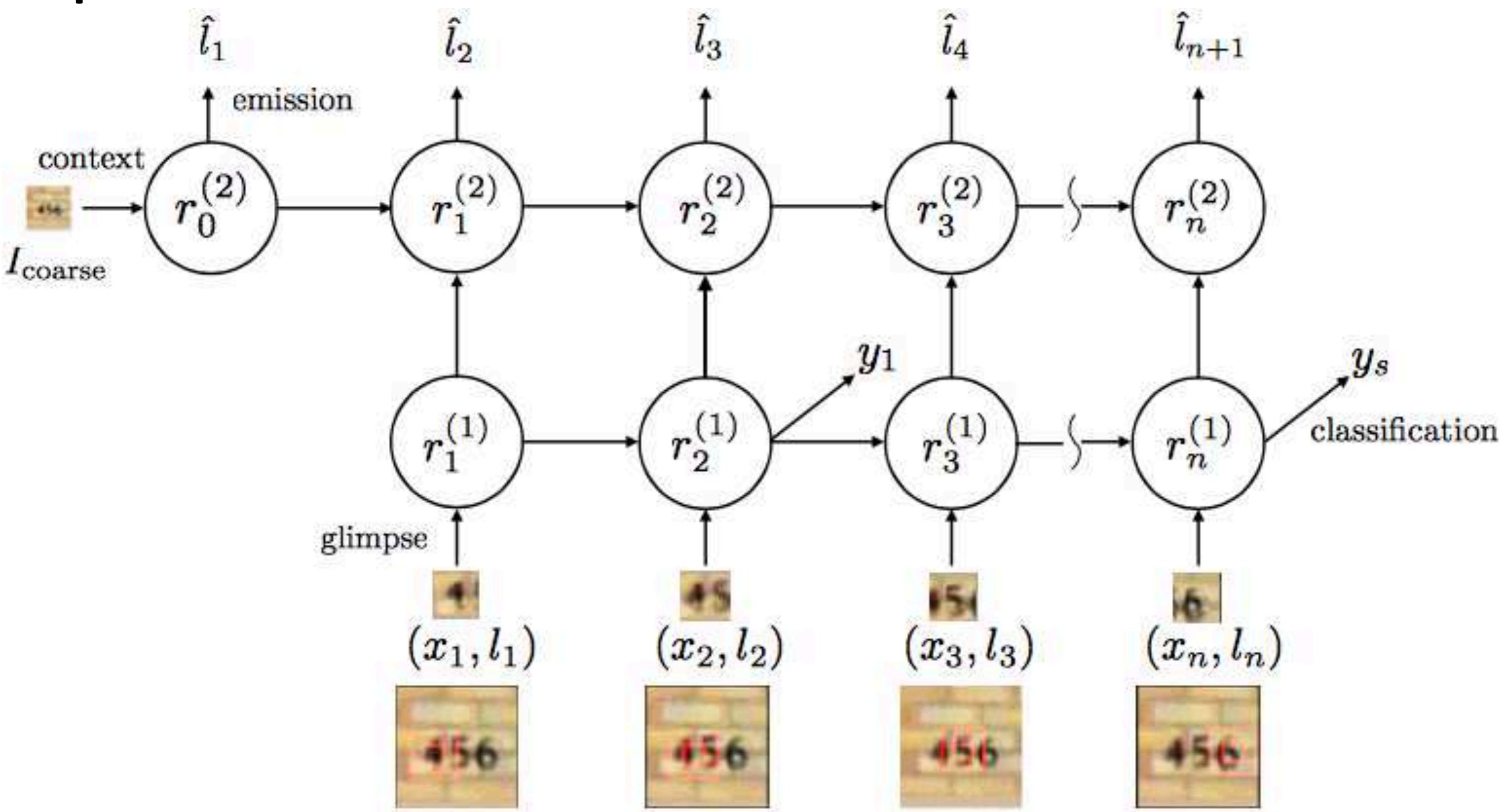


Figure: Lei Ba et al (2015). Multiple Object Recognition with Visual Attention. ICLR

The Unique Problems

3. Attention and Task Relevance

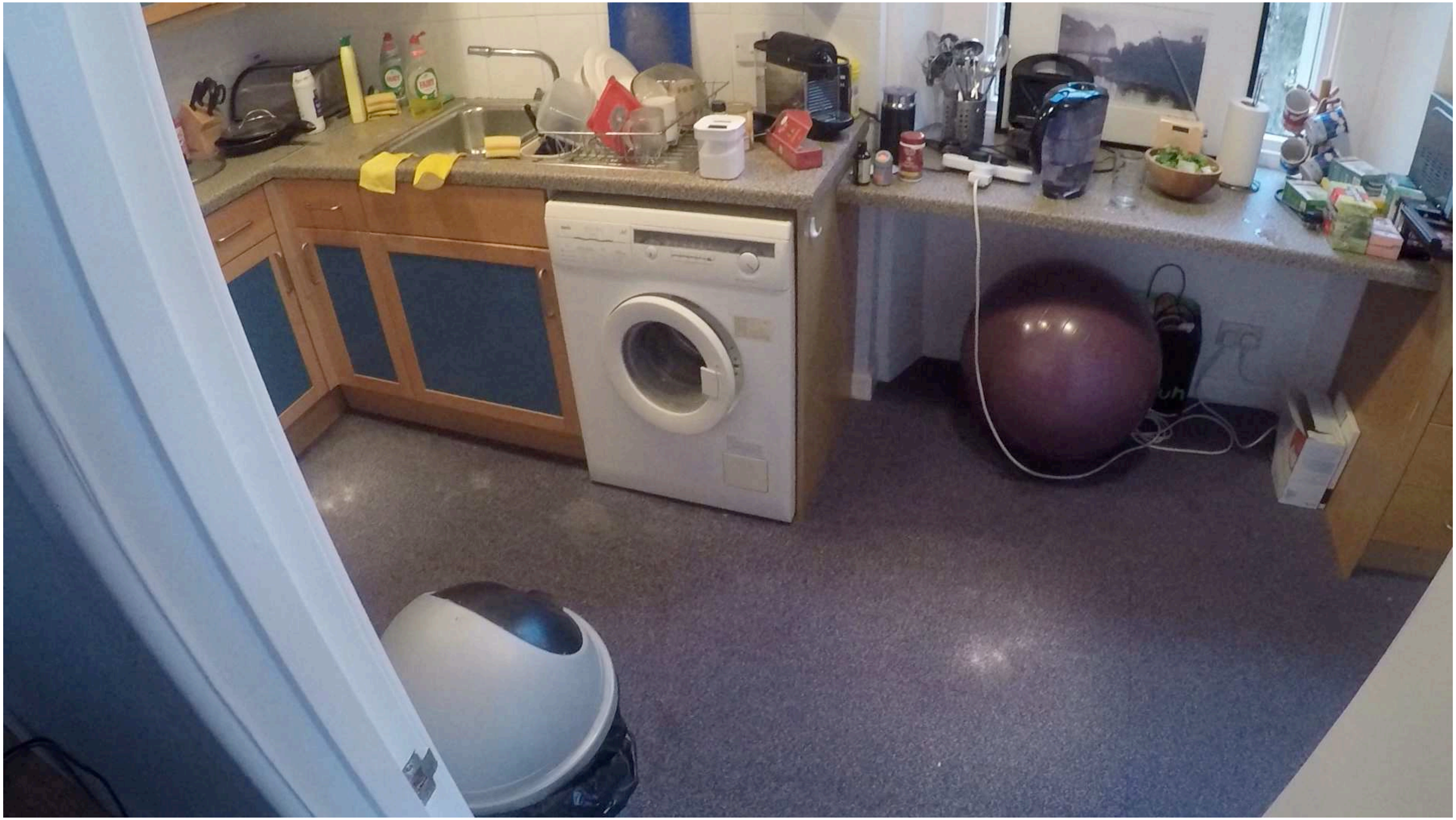
Case Study: You-Do, I-Learn

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You-Do, I-Learn

- First-person view
- Offers a unique insight into ‘used’ or ‘attended-to’ objects
- How these objects have been used

Try it yourself



with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

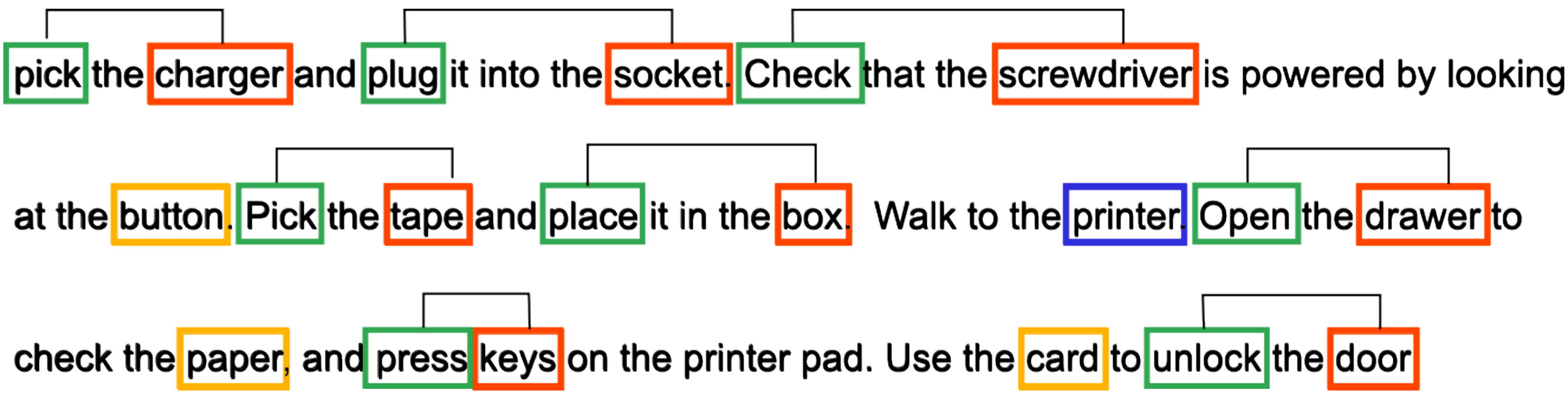
You-Do, I-Learn

- Q. How to ‘ground-truth’ objects that have been used?
- Q. How to ‘ground-truth’ how these objects have been used?

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

BEOID

- Ground-truth by written narration
- Released with dataset



with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

You Do, I Learn

- Discover used objects
- Discover how objects have been used
- Extract guidance videos
- Fully unsupervised
  - No prior knowledge of objects (number, size)
  - Static and moveable objects

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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Definition

Task-Relevant Object (TRO)

an object, or part of an object, with which a person interacts during task performance

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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Which Objects?

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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

- *Suggested* Problem Formulation…
  - Given a sequence of egocentric images $\{I_1, .., I_T\}$
  - Collected from multiple operators around a common environment
  - Automatically discover all task-relevant objects

$$\{O_k; 1 \leq k \leq K\}$$

$$O_k = \{\Omega(I_t); 1 \leq t \leq T\}$$

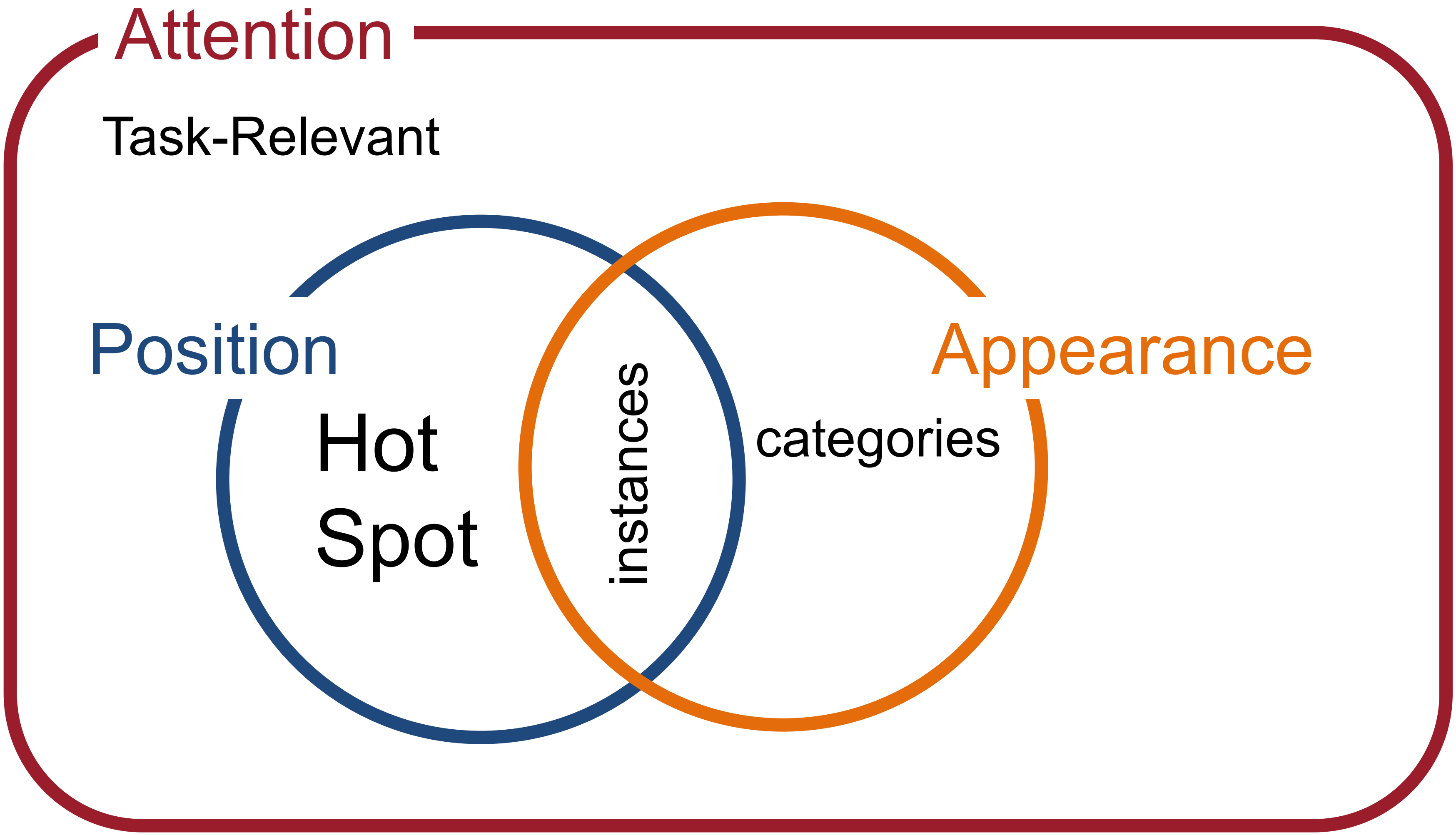
  - *Assumption:* at most one task-relevant image part is present within each image

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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

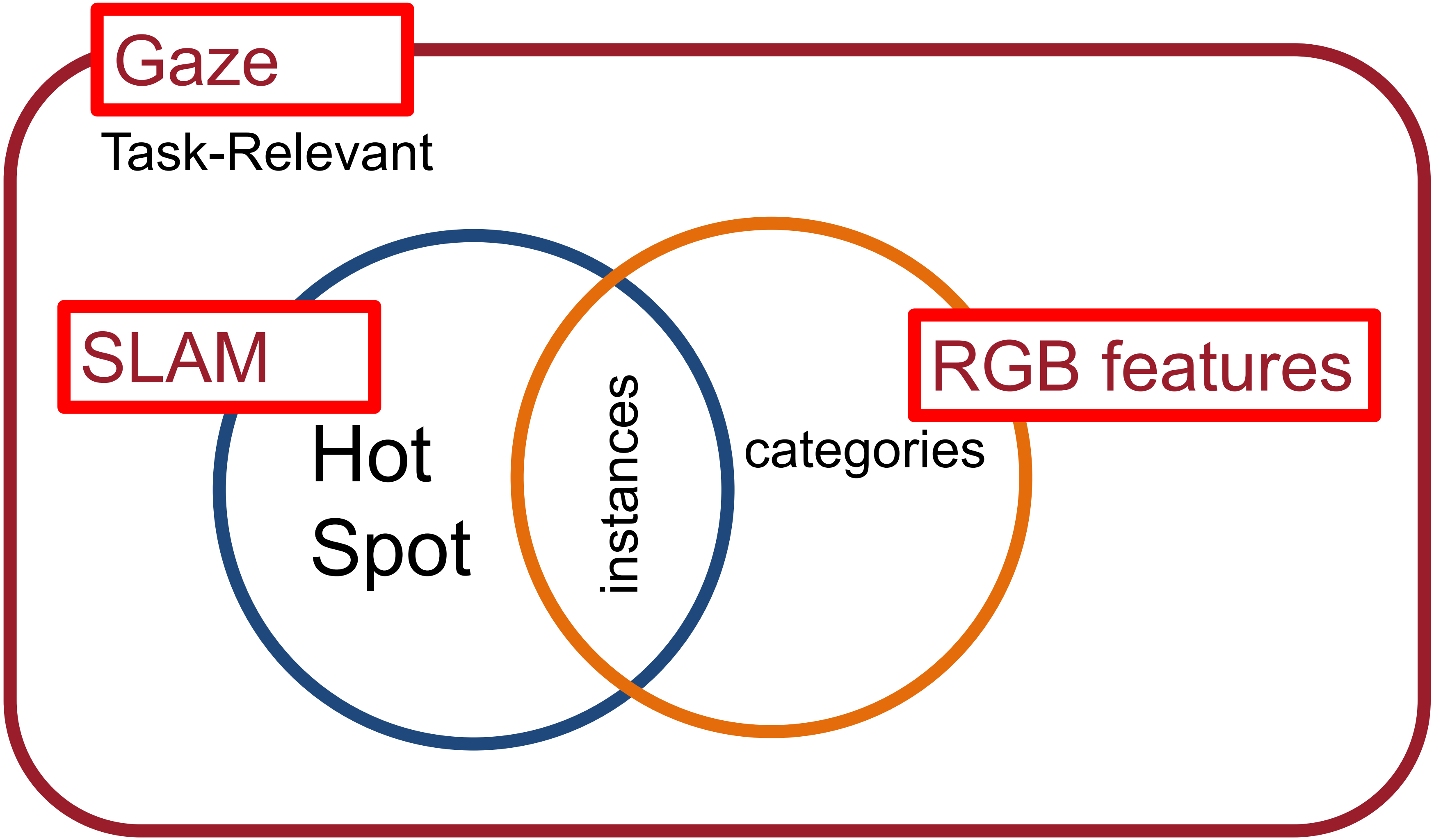
Discovering Task-Relevant Objects



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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

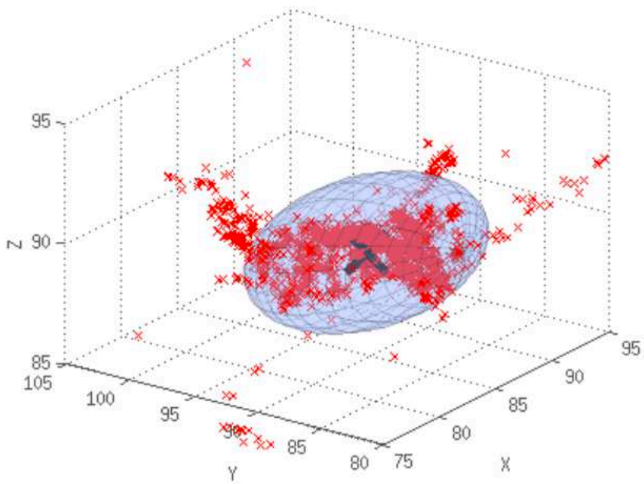
Discovering TROs

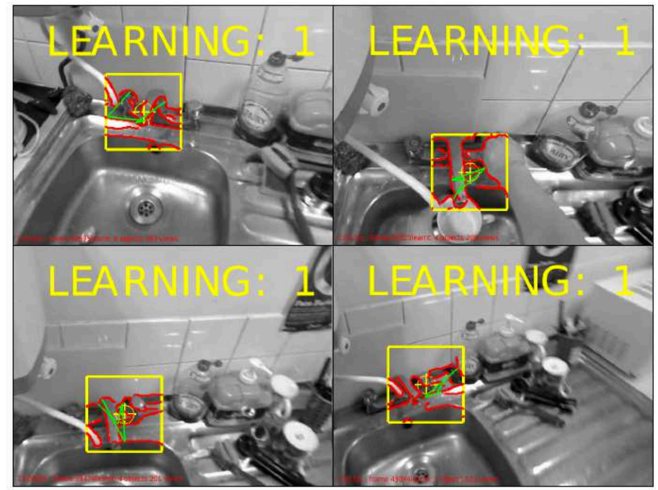
Discovering becomes a clustering task…

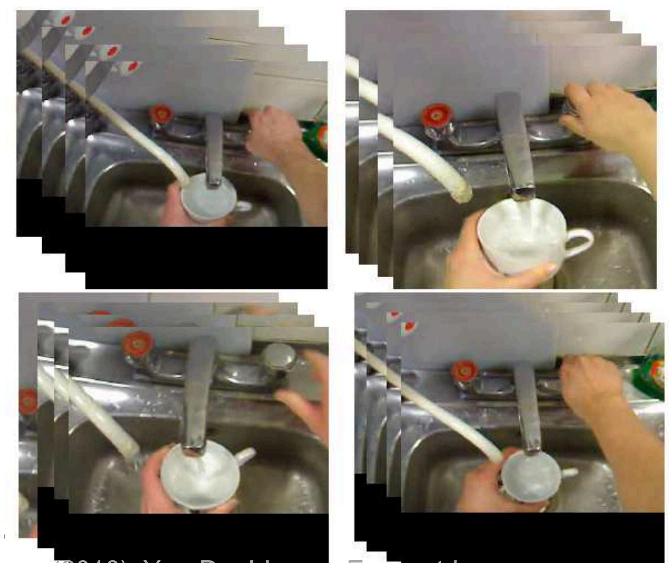
- Considers attention, position and appearance
- Unknown number of objects

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects





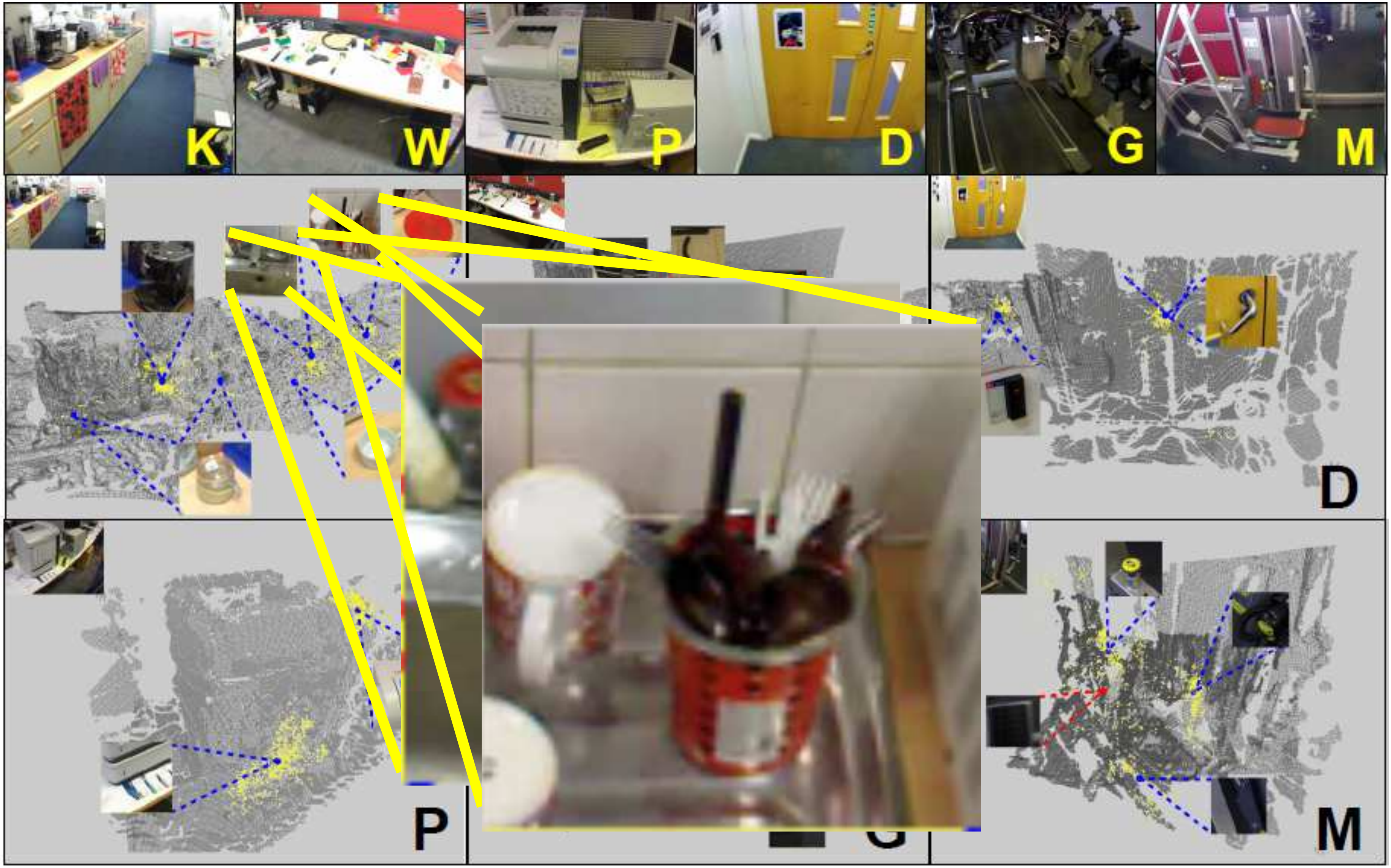


D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

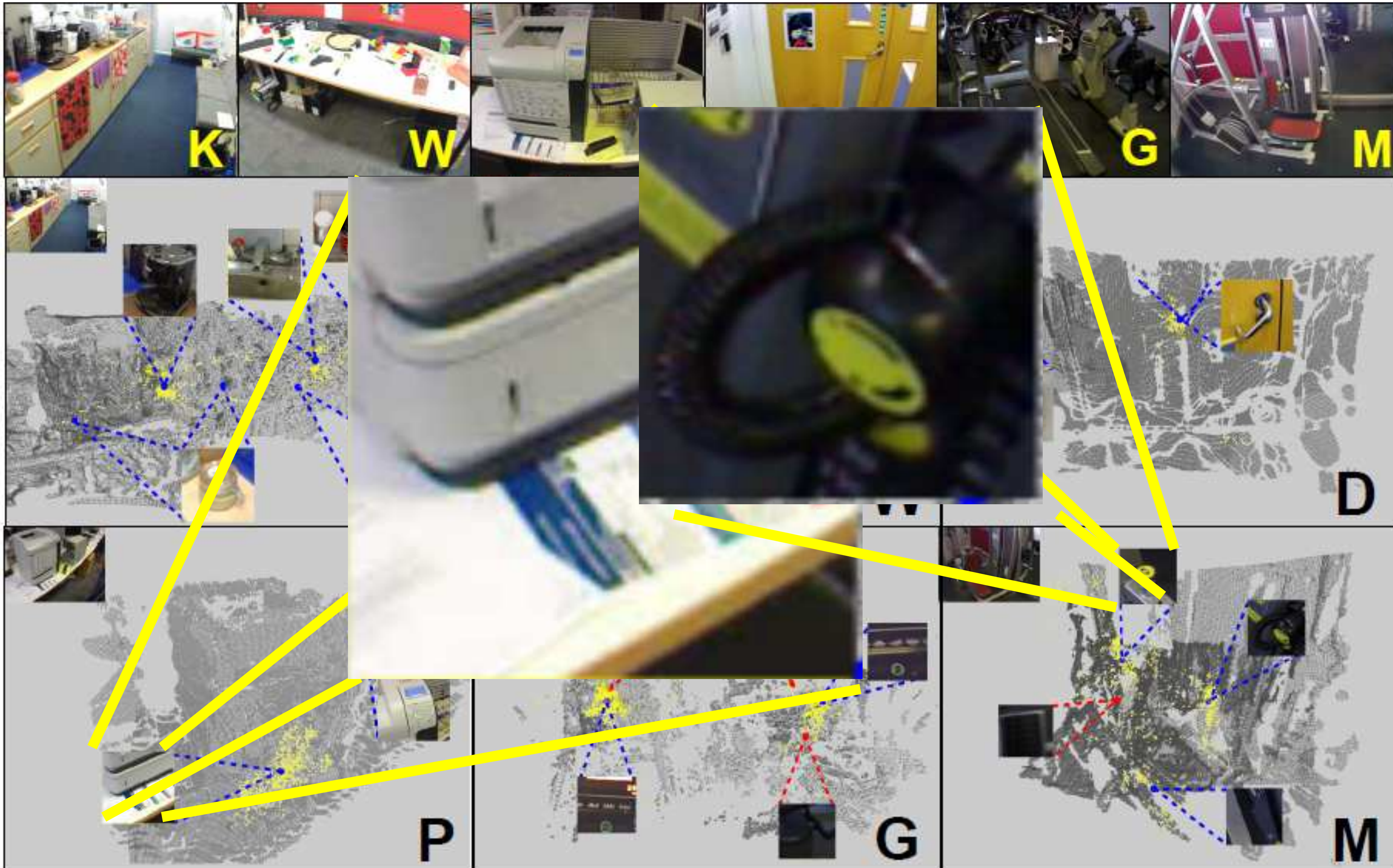
Discovering Task-Relevant Objects



Action

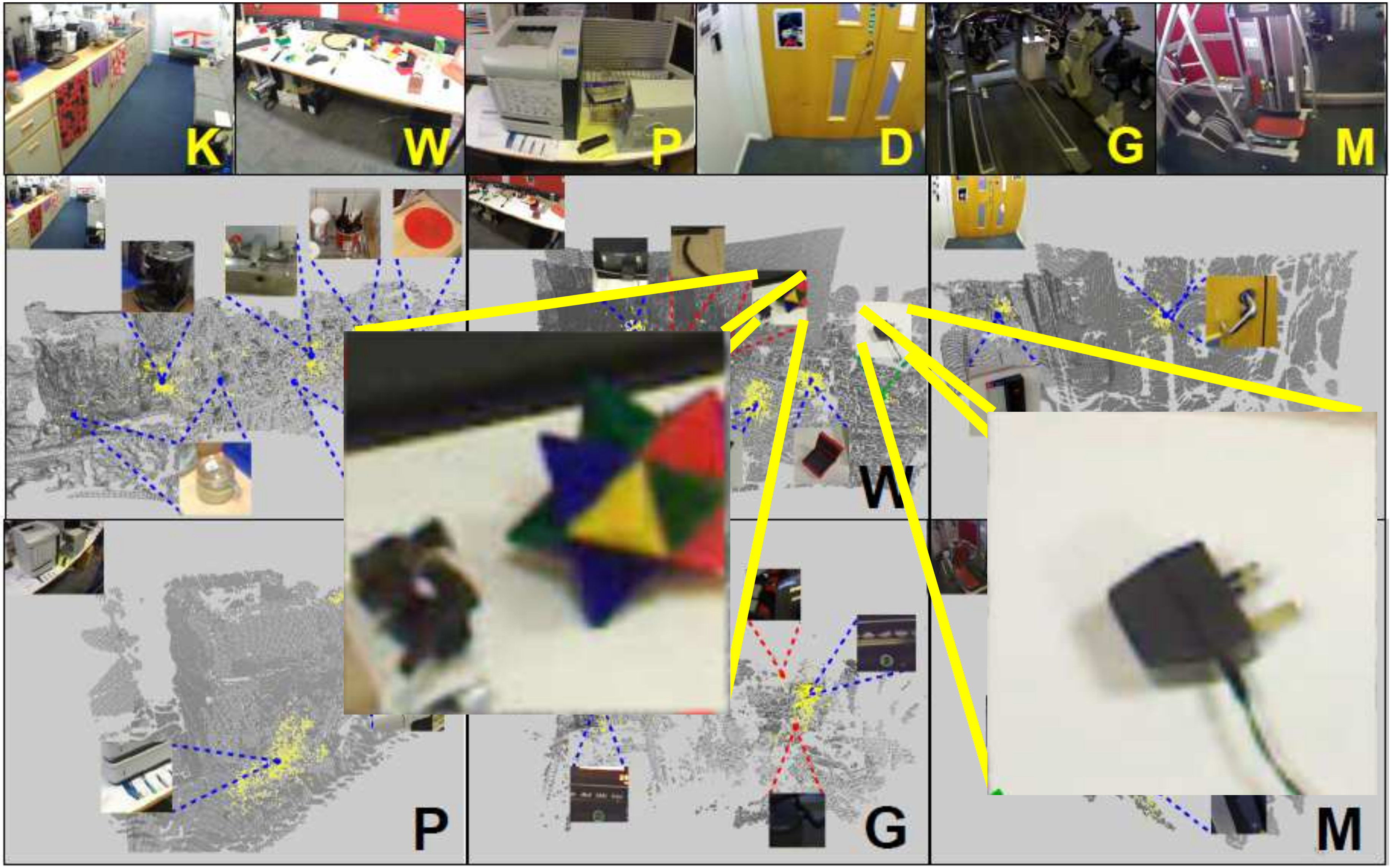
with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

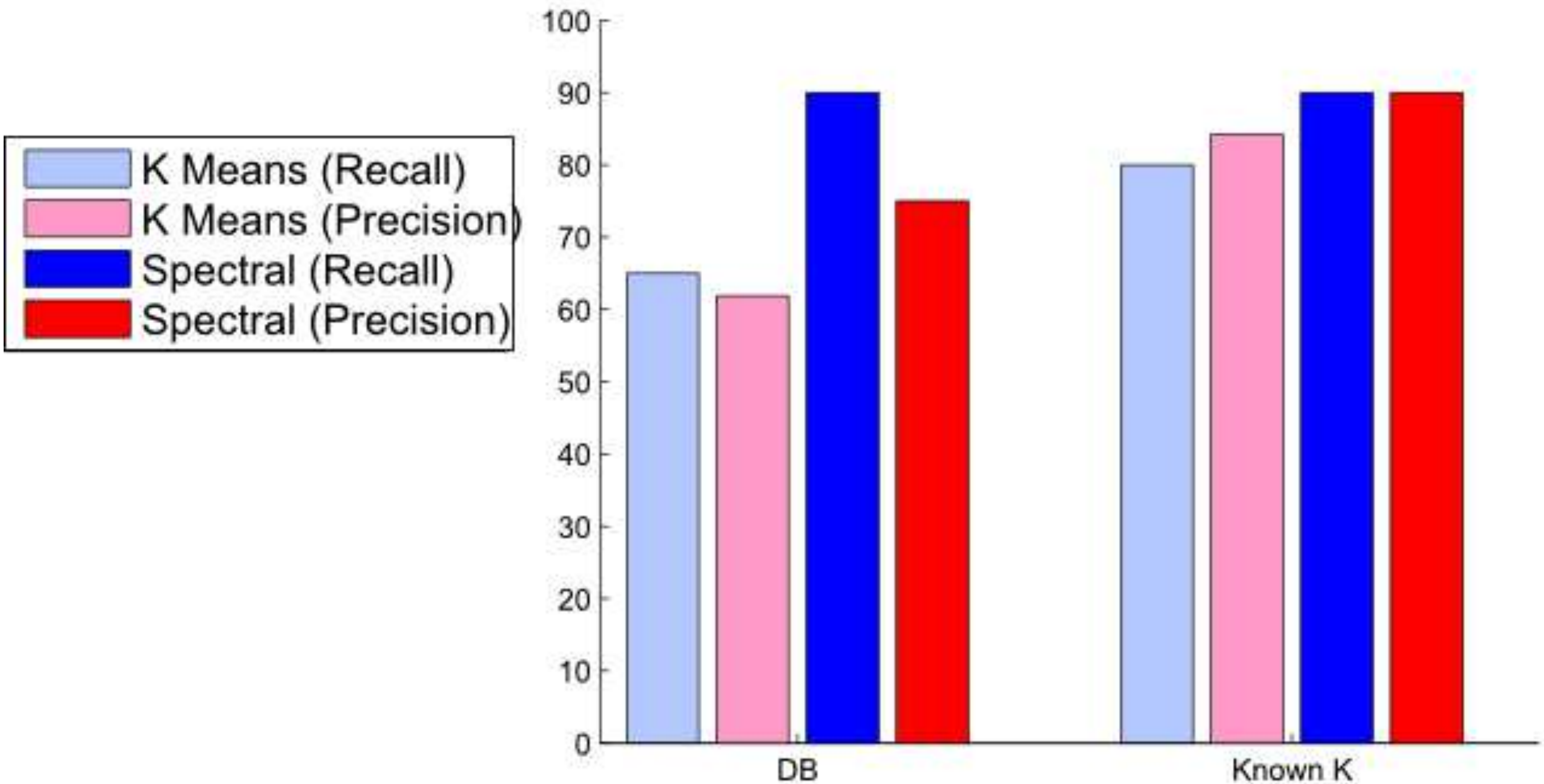
Discovering Task-Relevant Objects



Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects

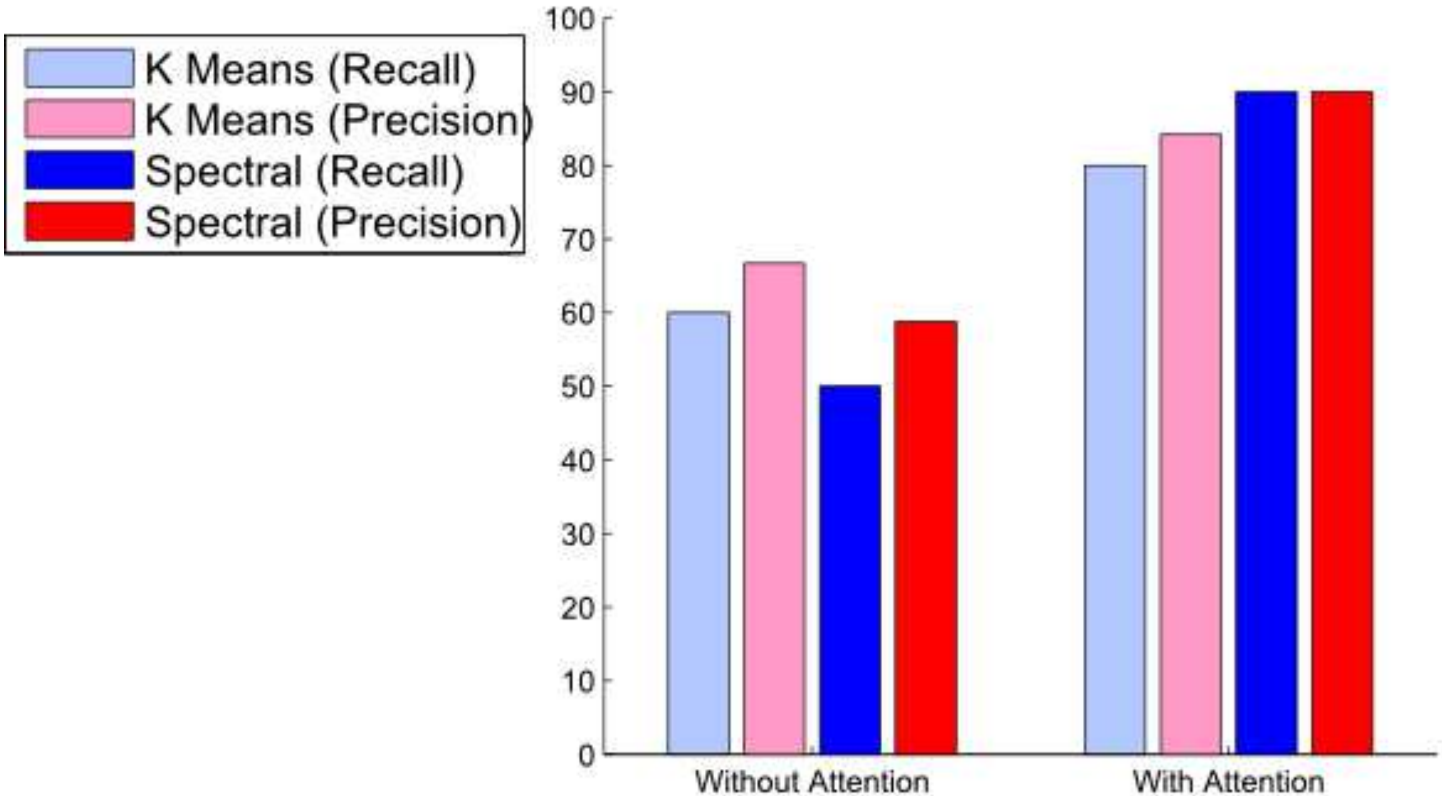


D Damen, T Leelasawassuk, W Mayol-Cuevas (2016). You-Do, I-Learn: Egocentric Unsupervised Discovery of Objects and their Modes of Interaction Towards Video-Based Guidance. *Computer Vision and Image Understanding*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Task-Relevant Objects



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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction

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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Definition

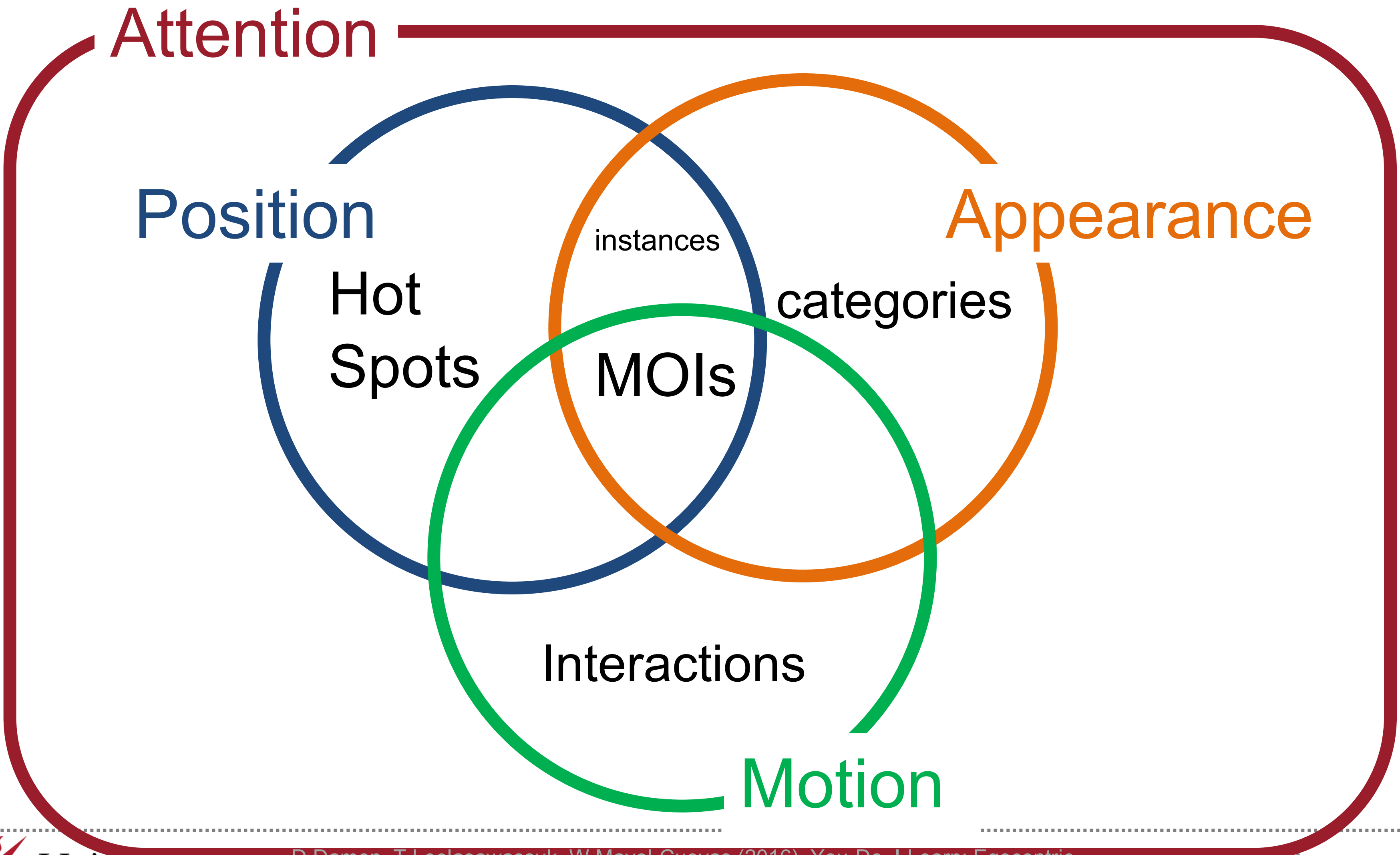
Modes of Interaction (MOI)

the different ways in which TROs are used

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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction



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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction

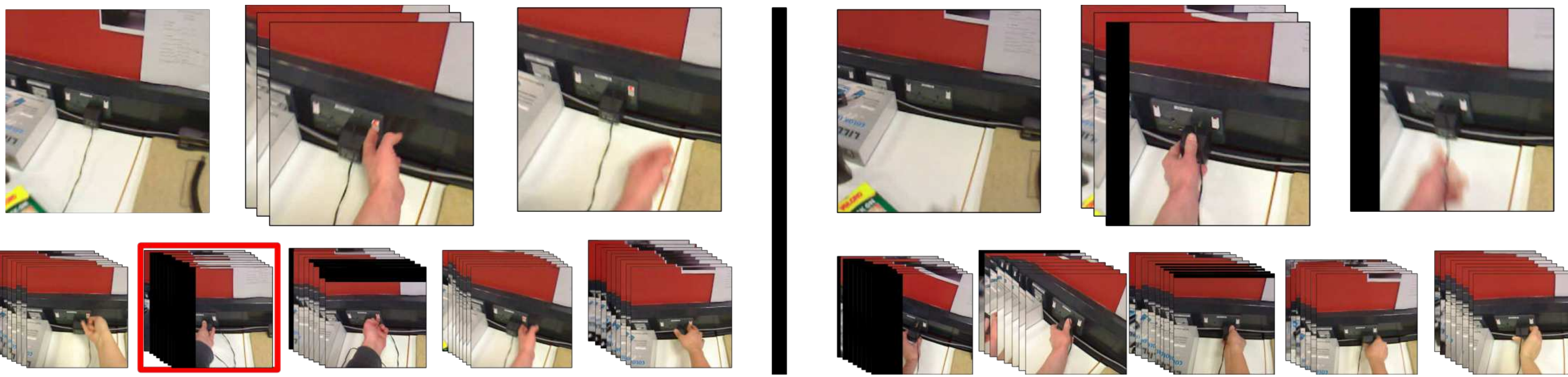
- Motion
  - Video snippets for each discovered object
  - Descriptor per snippet
  - Clustering using DB-index

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*

Action

Discovering Modes of Interaction

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk



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*

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Discovering Modes of Interaction

Open & get sugar

Put

Pick

Open door

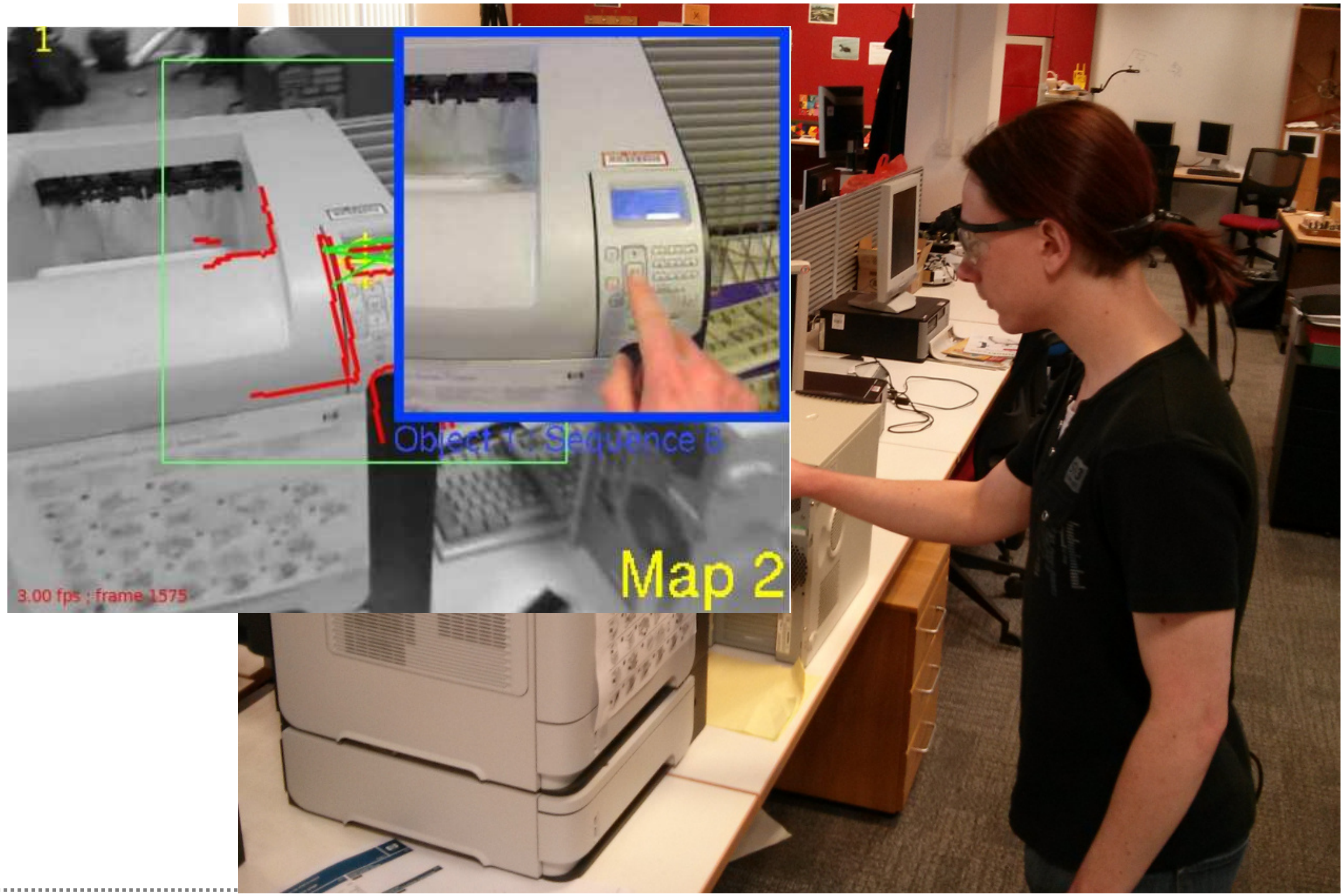
University of BRISTOL

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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

Back to…. the goal…



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*

Action

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

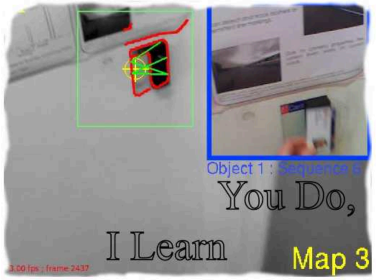
You Do, I Learn - Demonstration

University of BRISTOL

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

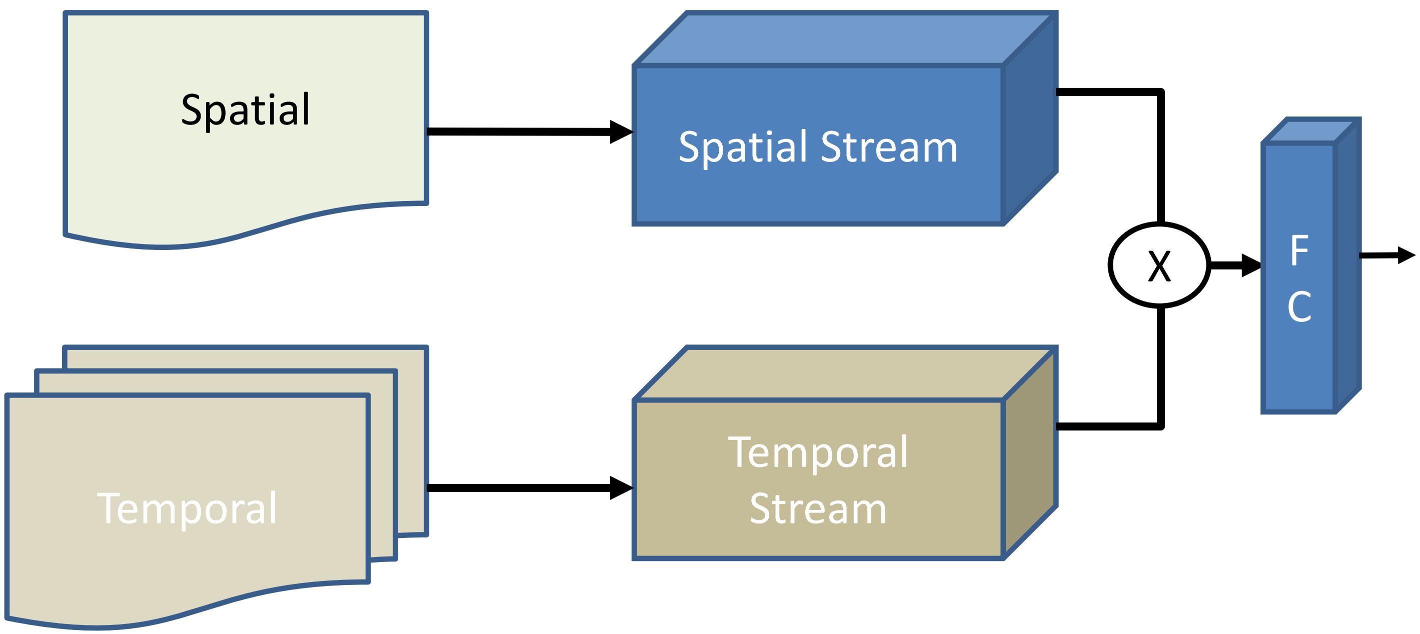
The Unique Problems

4. Object Interactions

Action Recognition – an Introduction

- CNNs for Action Recognition

Dual-Stream Neural Networks



Action Recognition – an Introduction

- CNNs for Action Recognition

  ## Dual-Stream Neural Networks

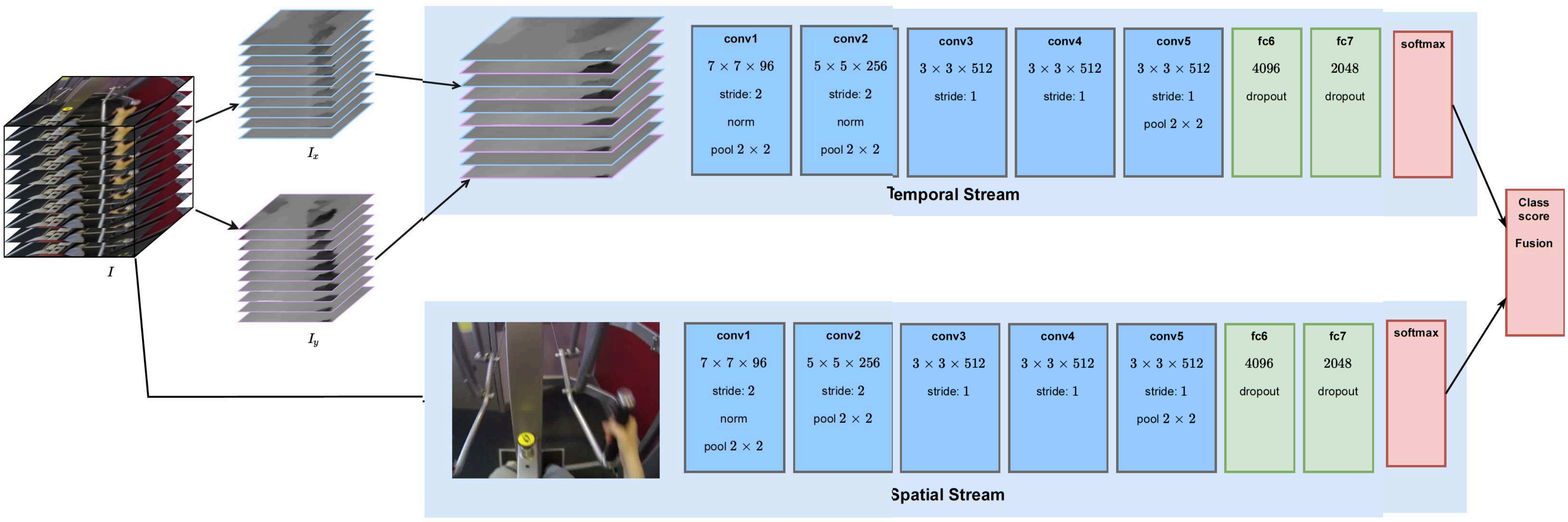


Figure by: Will Price, BSc Project, University of Bristol

Egocentric Action Recognition

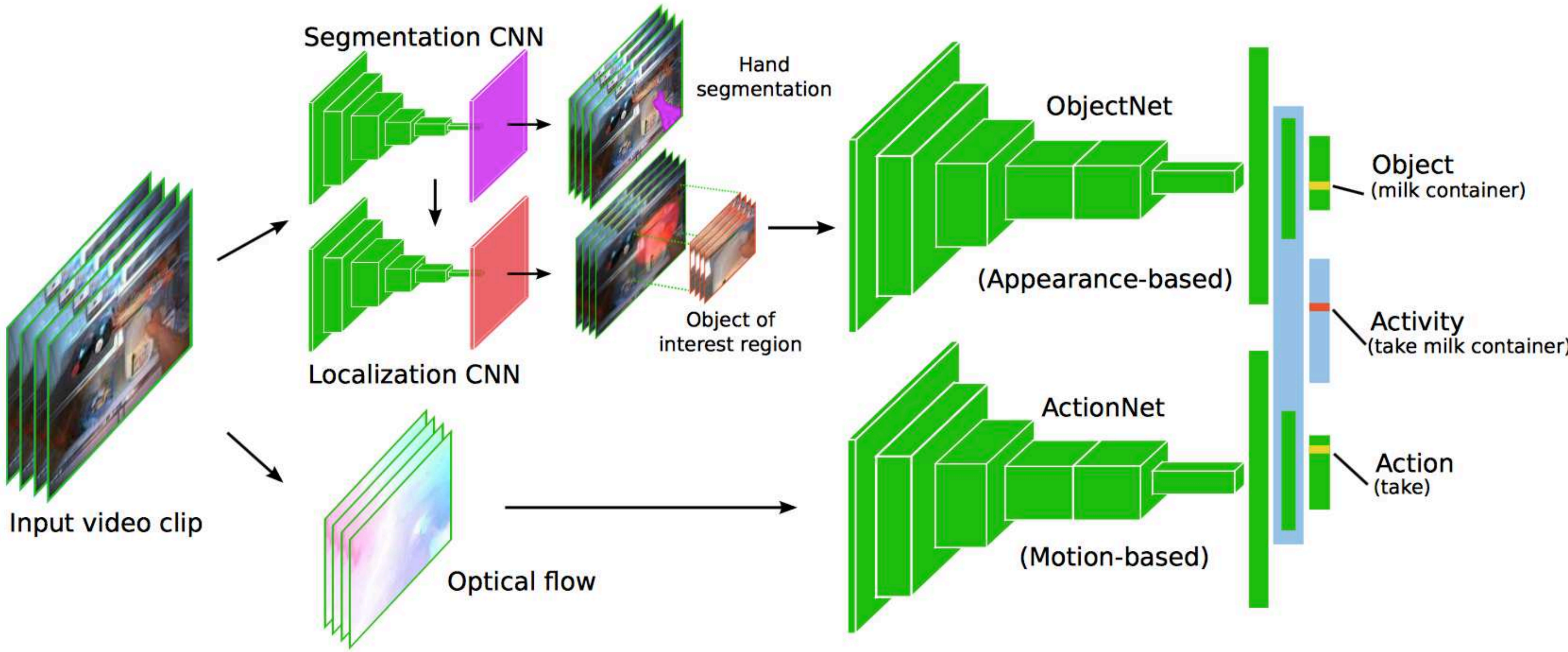
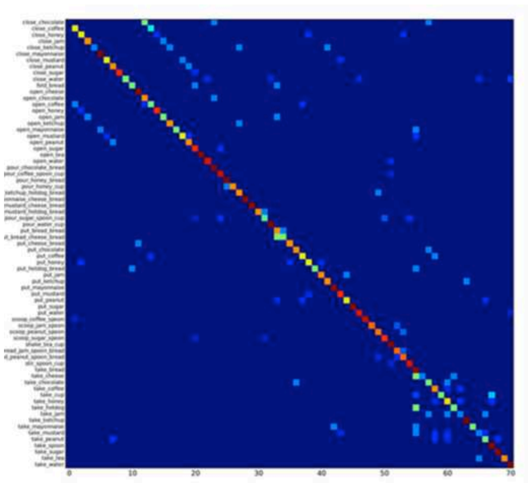
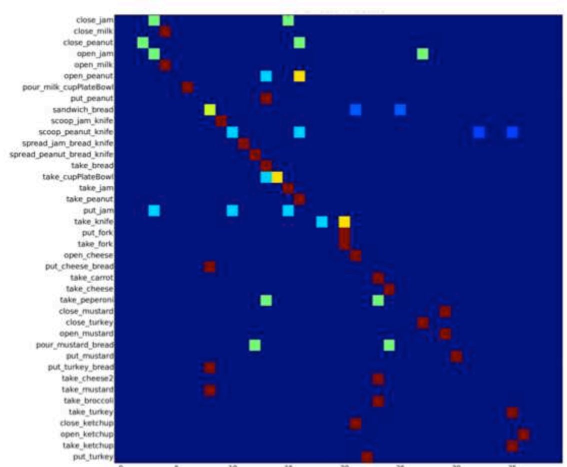


Figure from: Ma et al. Going Deeper into First-Person Activity Recognition. CVPR 2016

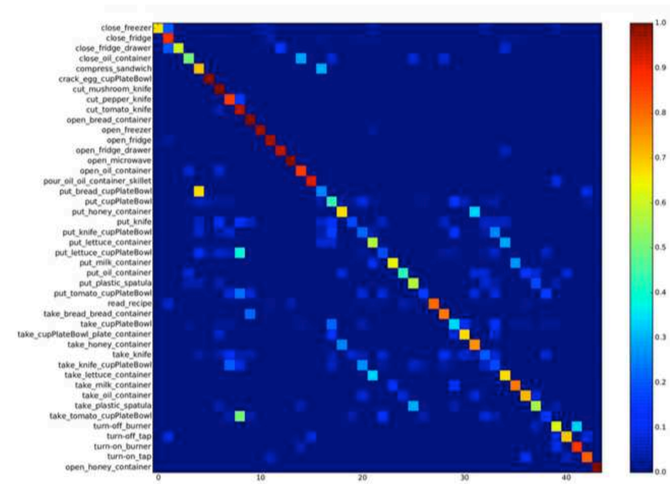
Egocentric Action Recognition



(a) GTEA 71 classes



(b) Gaze 40 classes



(c) Gaze+ 44 classes

Figure from: Ma et al. Going Deeper into First-Person Activity Recognition. CVPR 2016

Egocentric Action Recognition

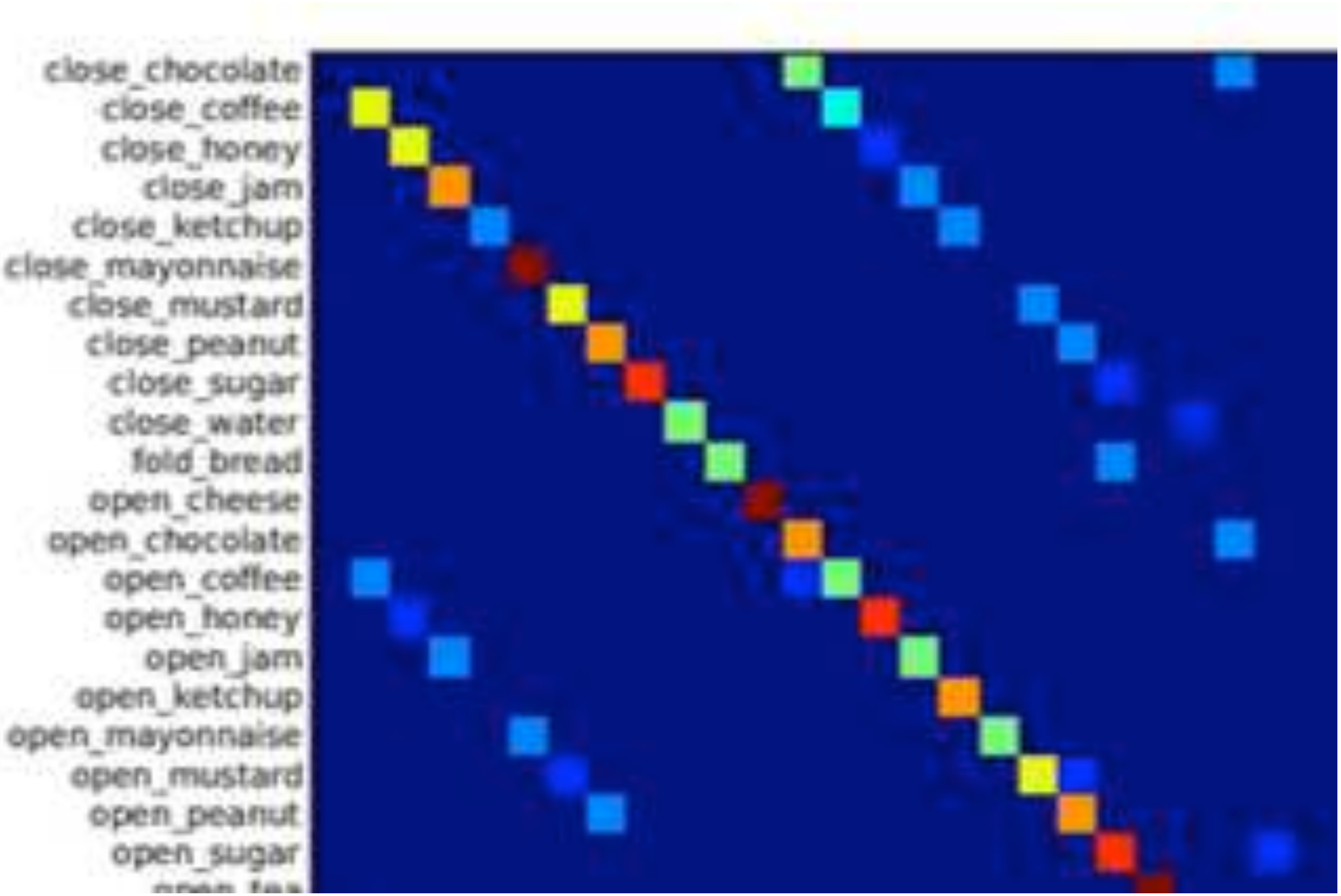


Figure from: Ma et al. Going Deeper into First-Person Activity Recognition. CVPR 2016

with: Will Price

Visualising Learnt Models

- http://youtu.be/4cZS39c7IL0

Action Recognition – an Introduction

- CNNs for Action Recognition

Dual-Stream Neural Networks

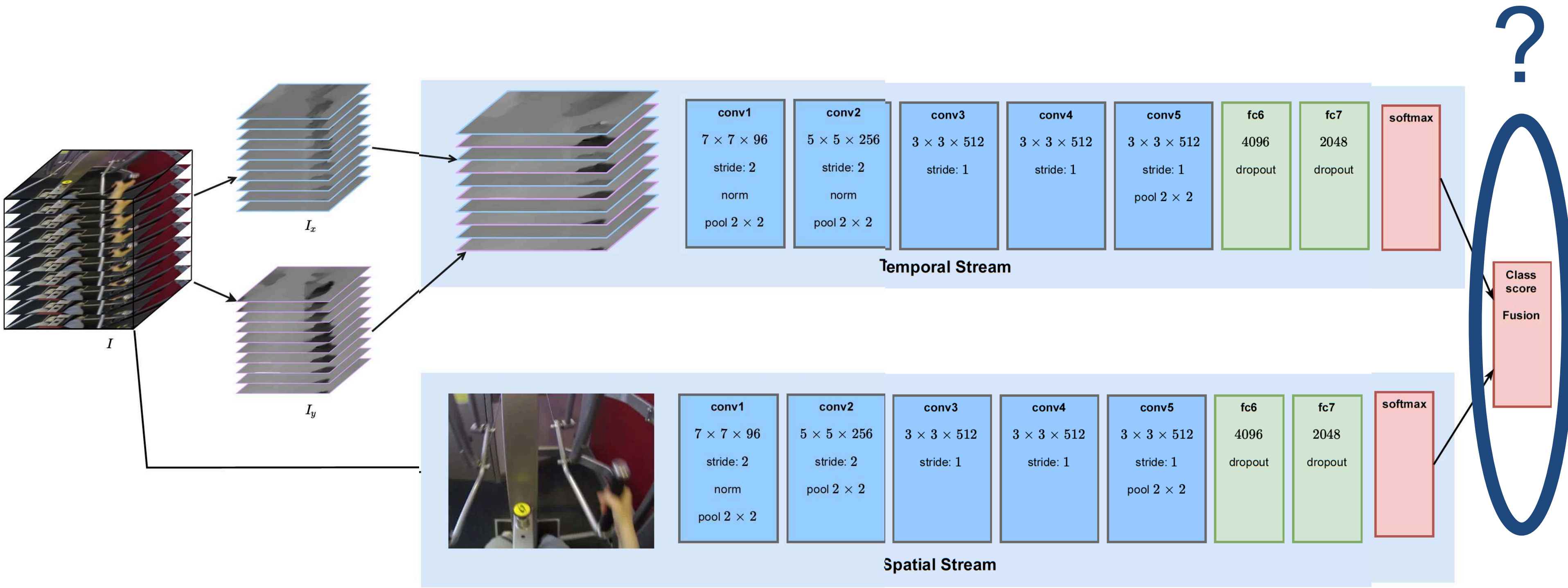
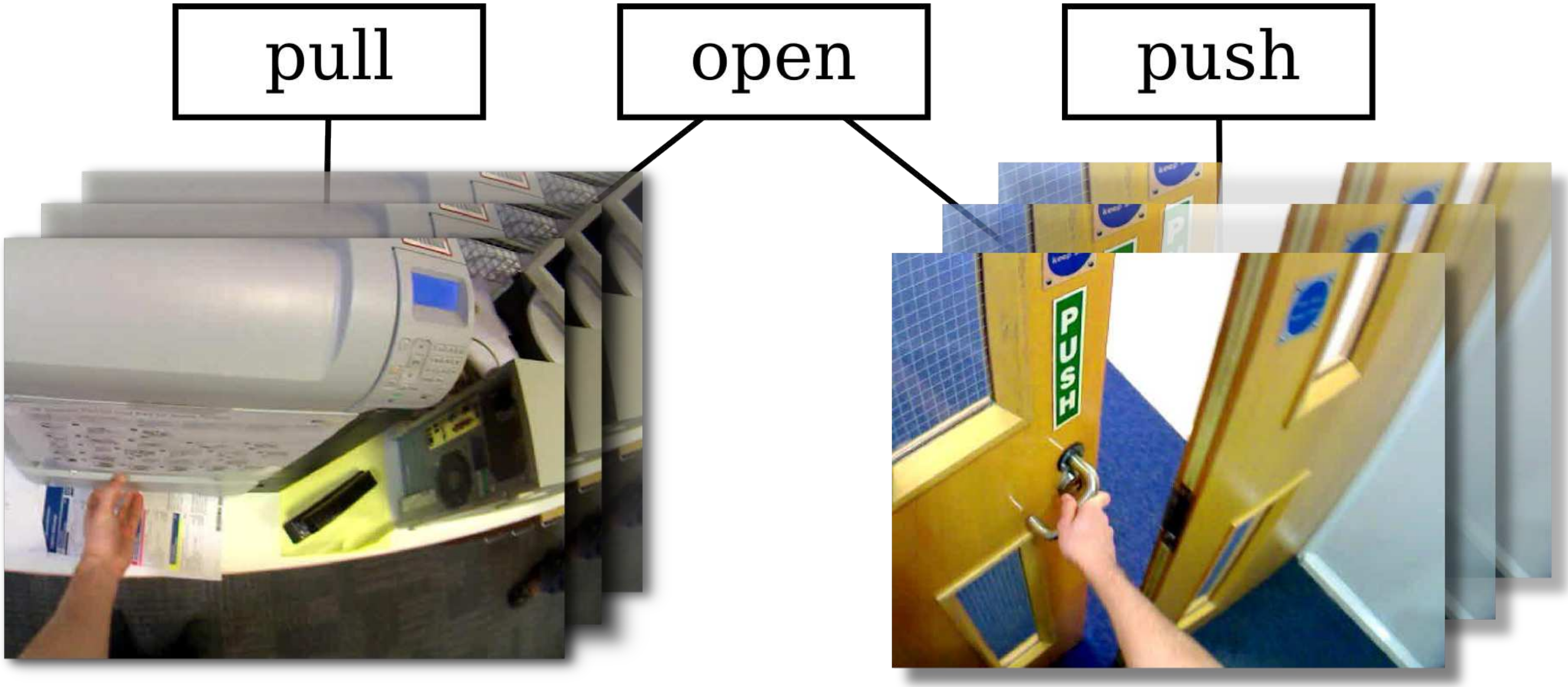


Figure by: Will Price, BSc Project, University of Bristol

with: Michael Wray

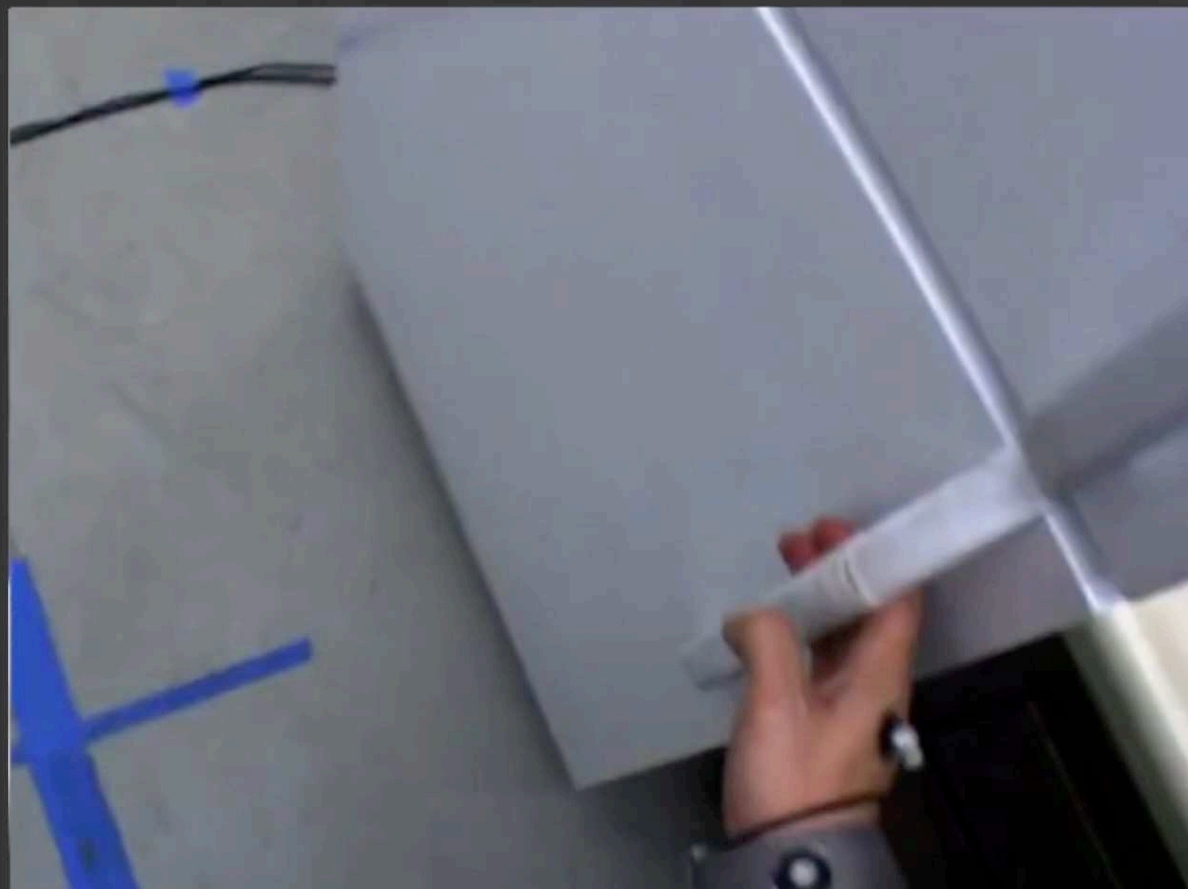
Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

with: Michael Wray

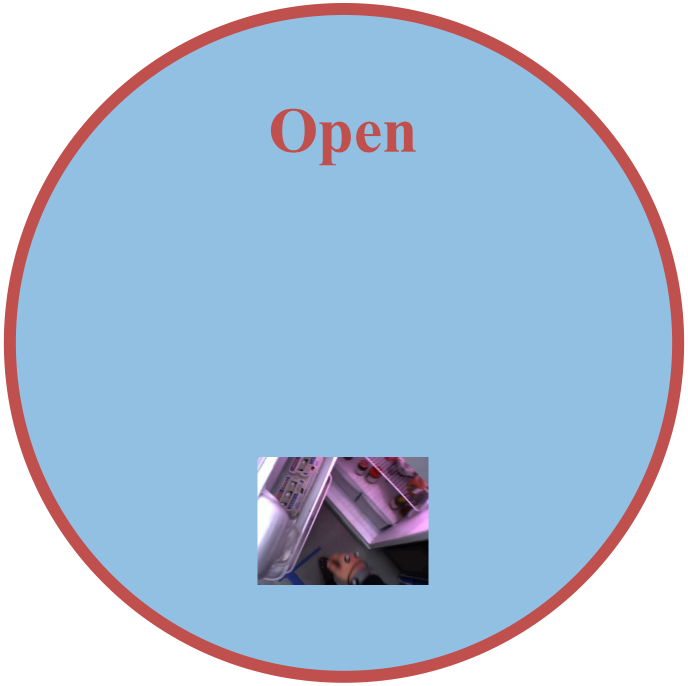
Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

with: Michael Wray

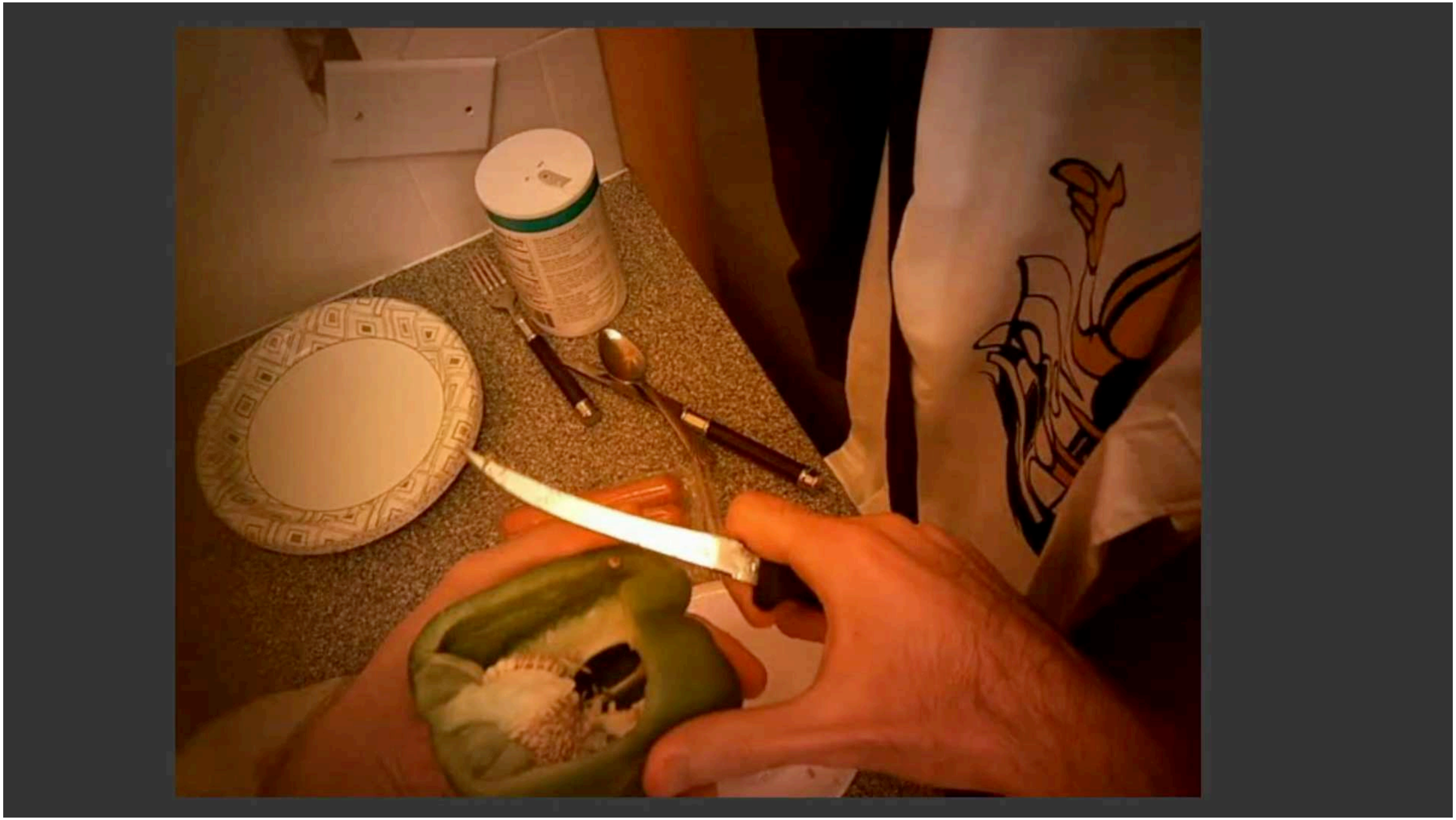
Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

with: Michael Wray

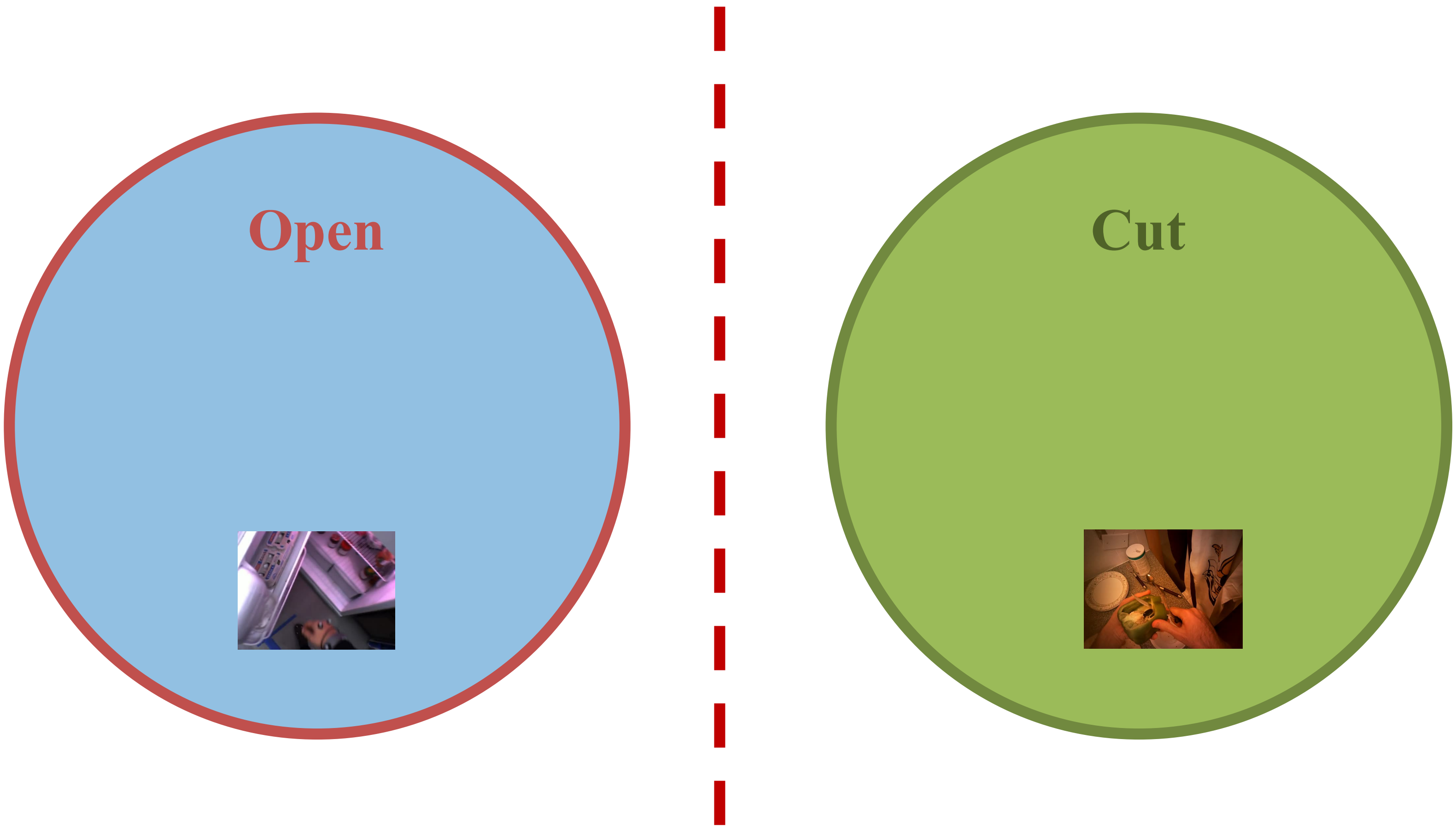
Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

with: Michael Wray

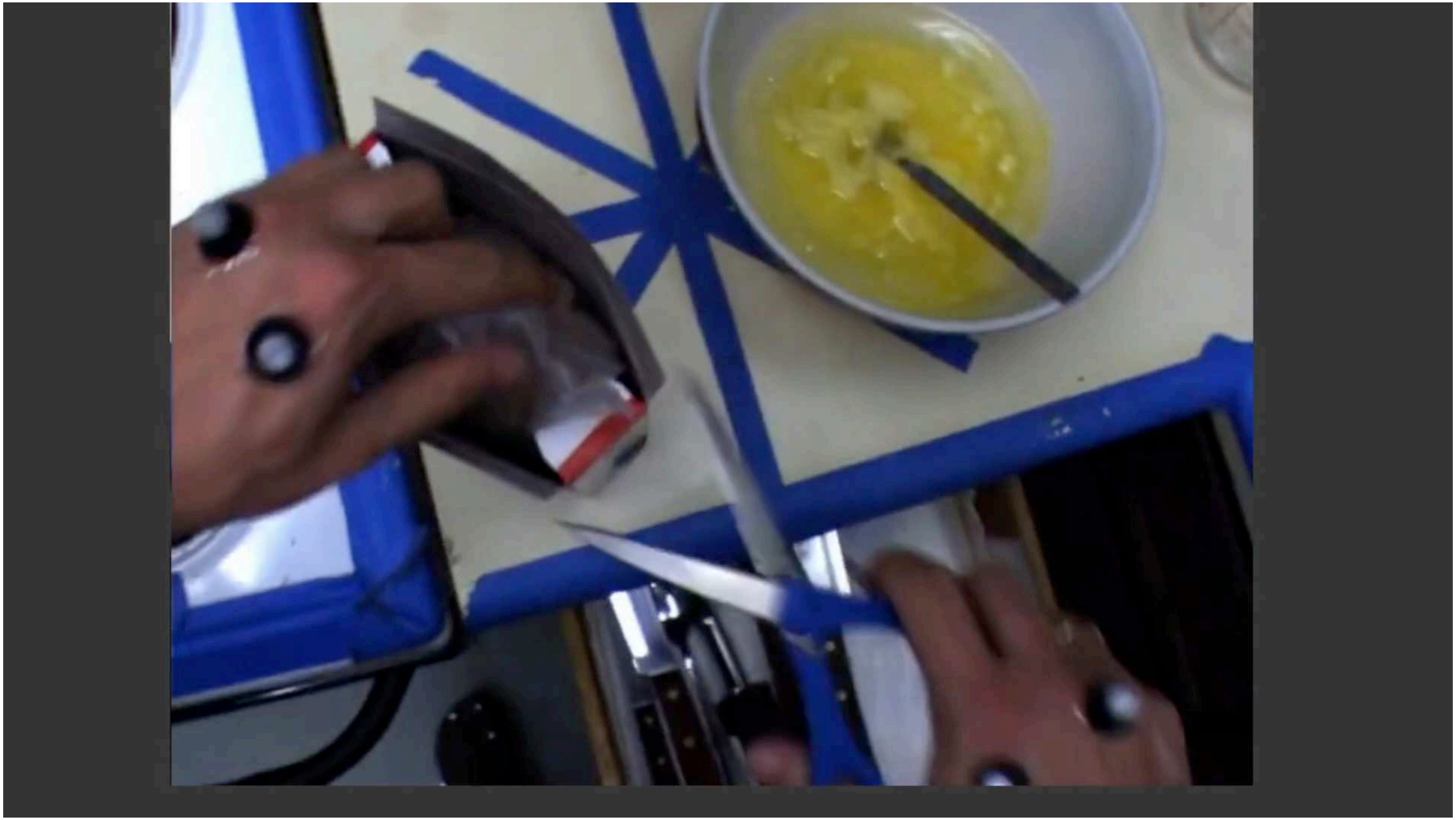
Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

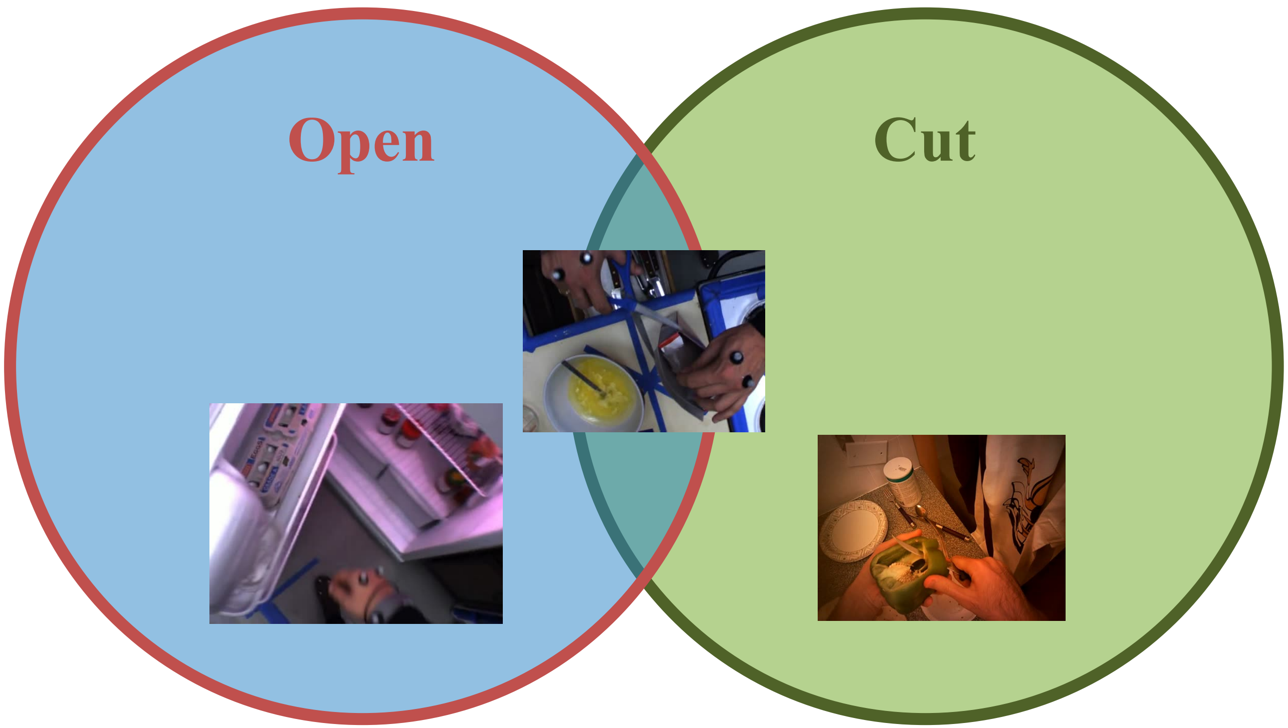
with: Michael Wray

Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Object Interactions – the Dilemma



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Object Interactions – the Dilemma

- Verbs cannot be separated into classes with hard boundaries.

- Singular classes are not enough.

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Learning Visual Actions Using Multiple Verb-Only Labels

with: Michael Wray

- 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 and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Learning Visual Actions Using Multiple Verb-Only Labels

with: Michael Wray

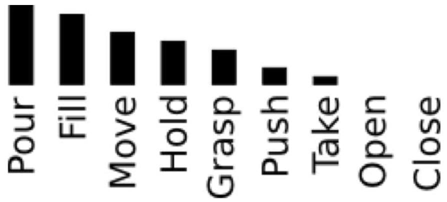
- Soft-Assigned Multi-Label
  - Multi-label using verbs only
  - Each verb assigned a value between 0 and 1
  - Object agnostic
  - Trained with Sigmoid Binary Cross Entropy

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Learning Visual Actions Using Multiple Verb-Only Labels

with: Michael Wray

- Collected from AMT



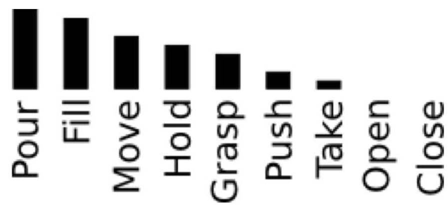
- Annotators agree:
  - Relevant Verb -> Main action
  - Irrelevant Verb -> unrelated motion
- Annotators disagree:
  - Relevant motion but not the main action

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

with: Michael Wray

Learning Visual Actions Using Multiple Verb-Only Labels

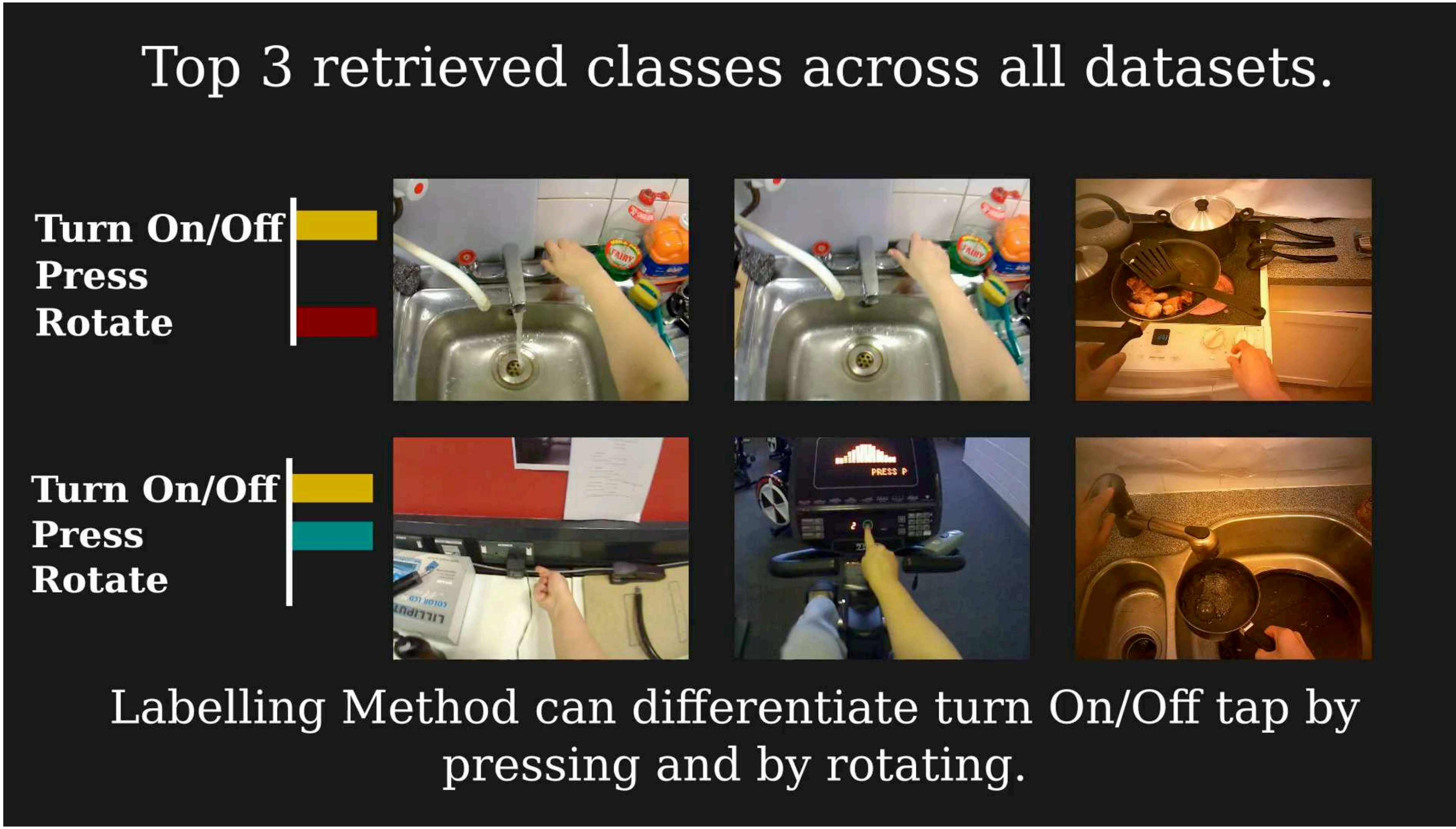
- Collected from AMT



SL	ML	SAML
• Majority Vote.	• Threshold of 0.5.	• Full Annotation.
• One-hot vector.	• Binary Vector	• Continuous Vector.

Learning Visual Actions Using Multiple Verb-Only Labels

with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC
M Wray et al (2018). Towards an Unequivocal Representation of Actions. Arxiv.

Action Recognition – an Introduction

- CNNs for Action Recognition

  Dual-Stream Neural Networks

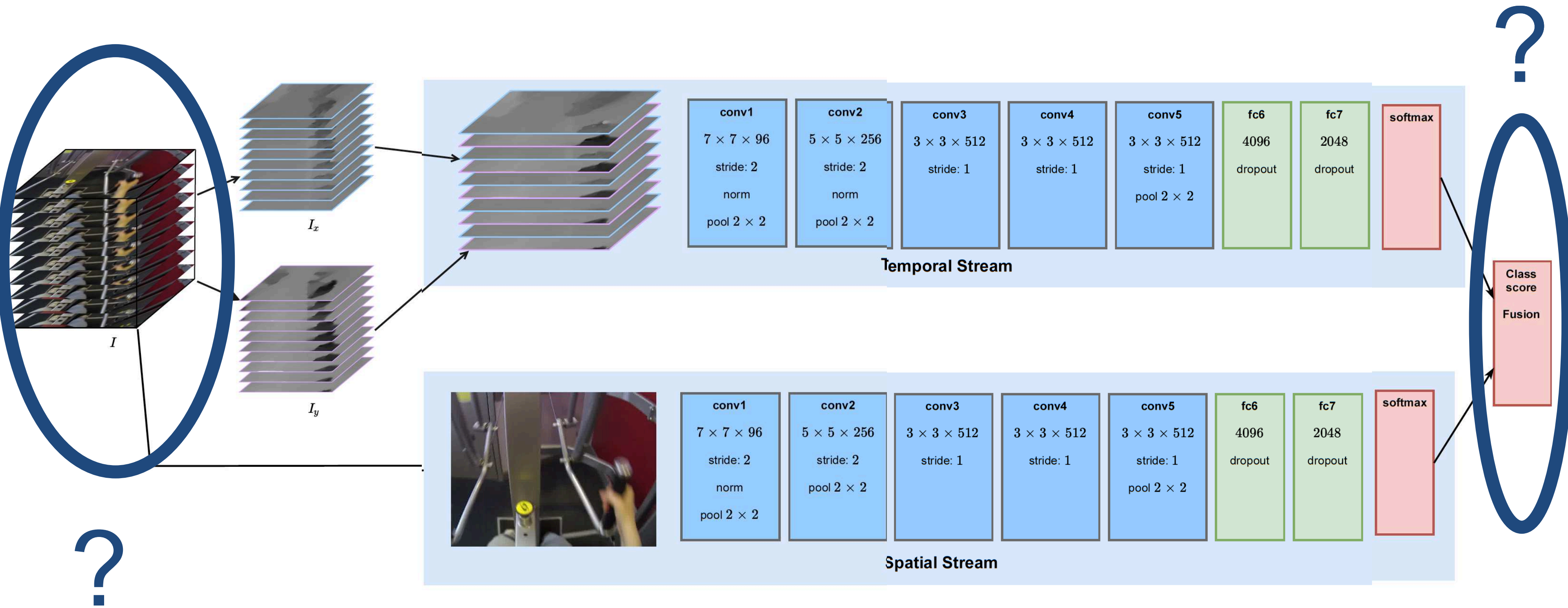
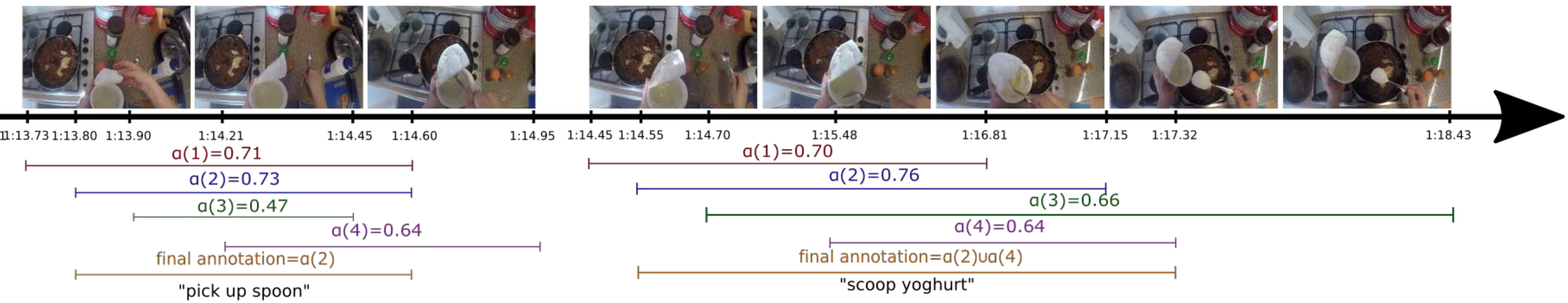


Figure by: Will Price, BSc Project, University of Bristol

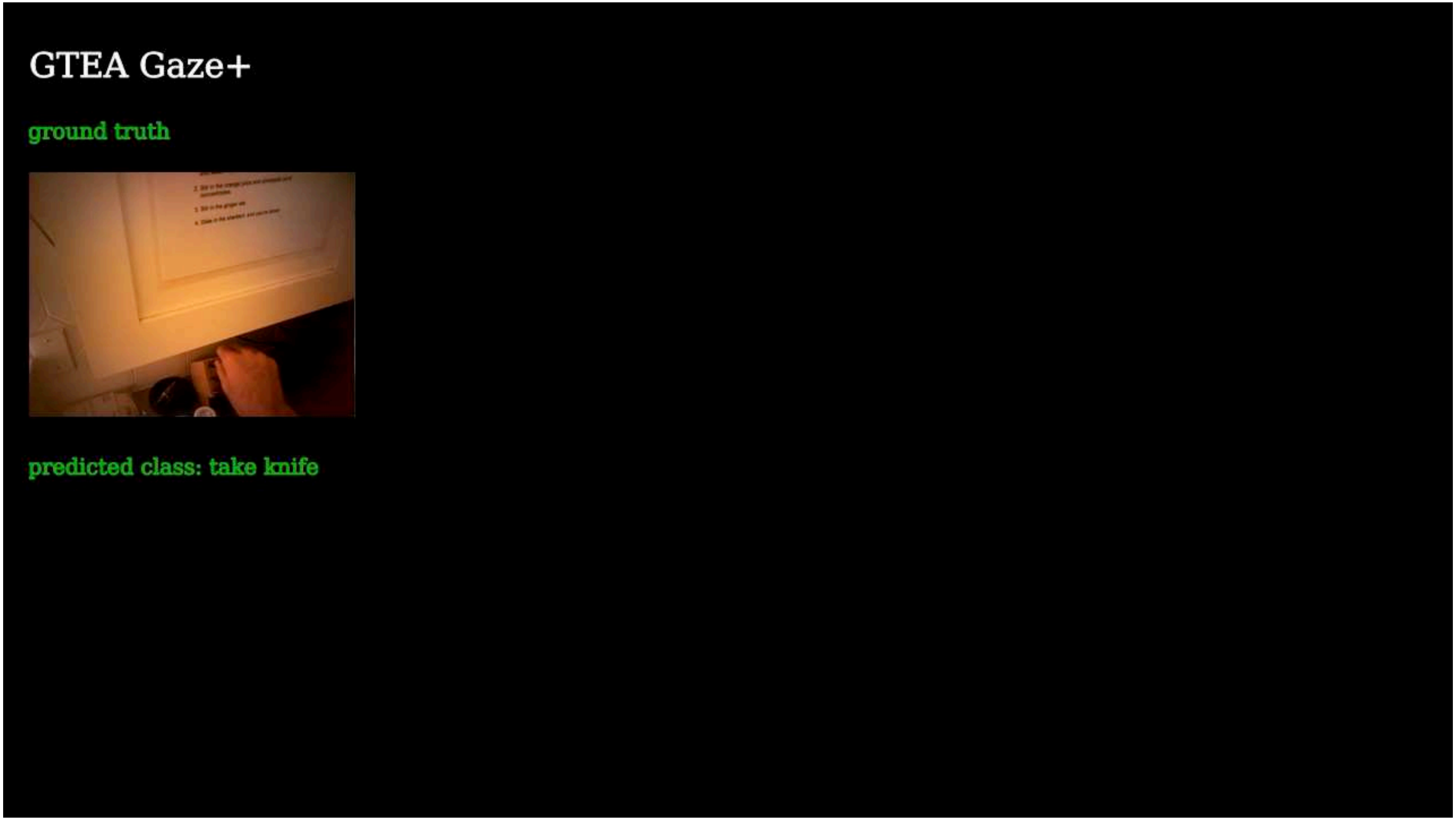
Temporal Boundaries for Object Interactions



- 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**

with: Davide Moltisanti
Michael Wray

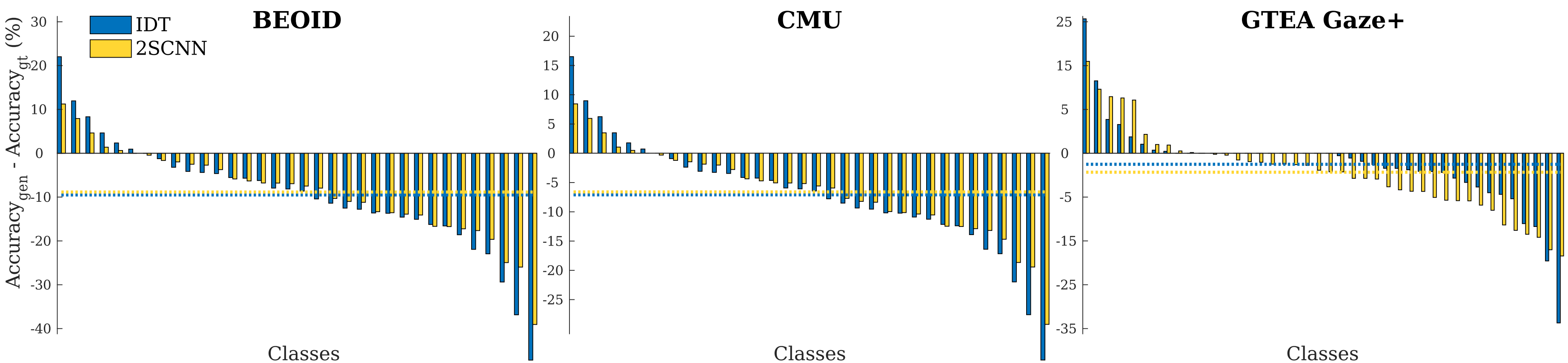
Trespassing the Boundaries



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

Trespassing the Boundaries

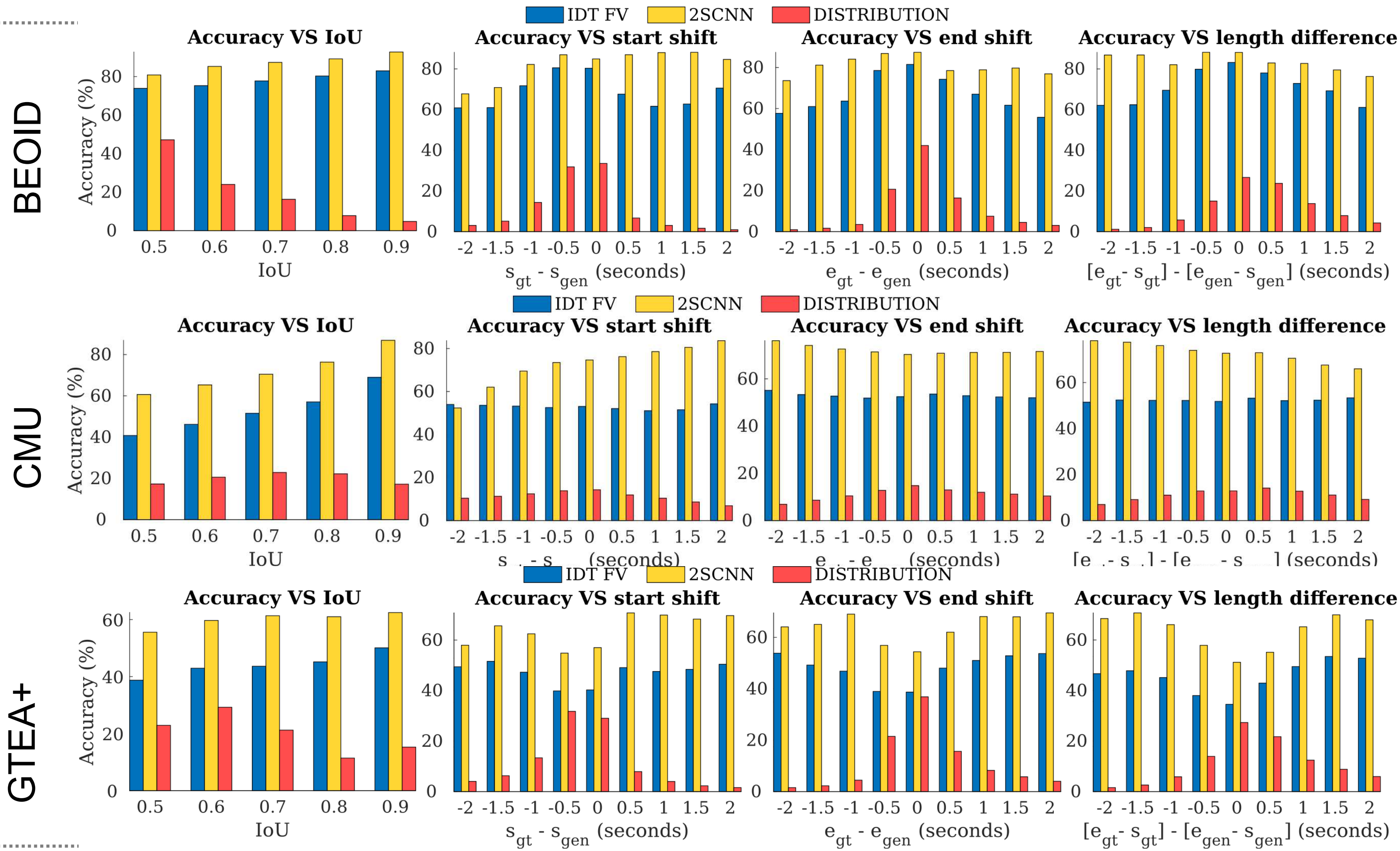


D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

Action

Trespassing the 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)*

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

The Rubicon Boundaries

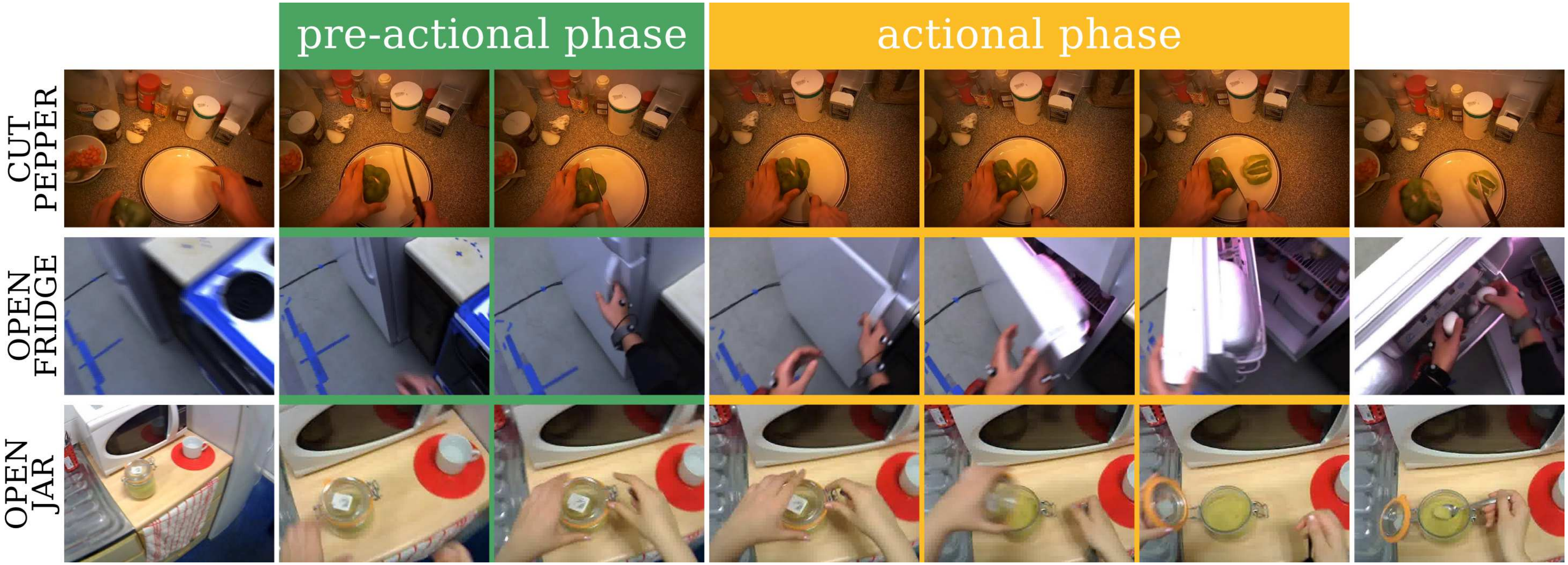
- Labelling approach proposal for temporally consistent annotations
- Decomposes an object interaction into two phases:
  - *pre-actional* phase
  - *actional* phase

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

Action

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

The Rubicon Boundaries



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

Action

The Rubicon Boundaries

with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas

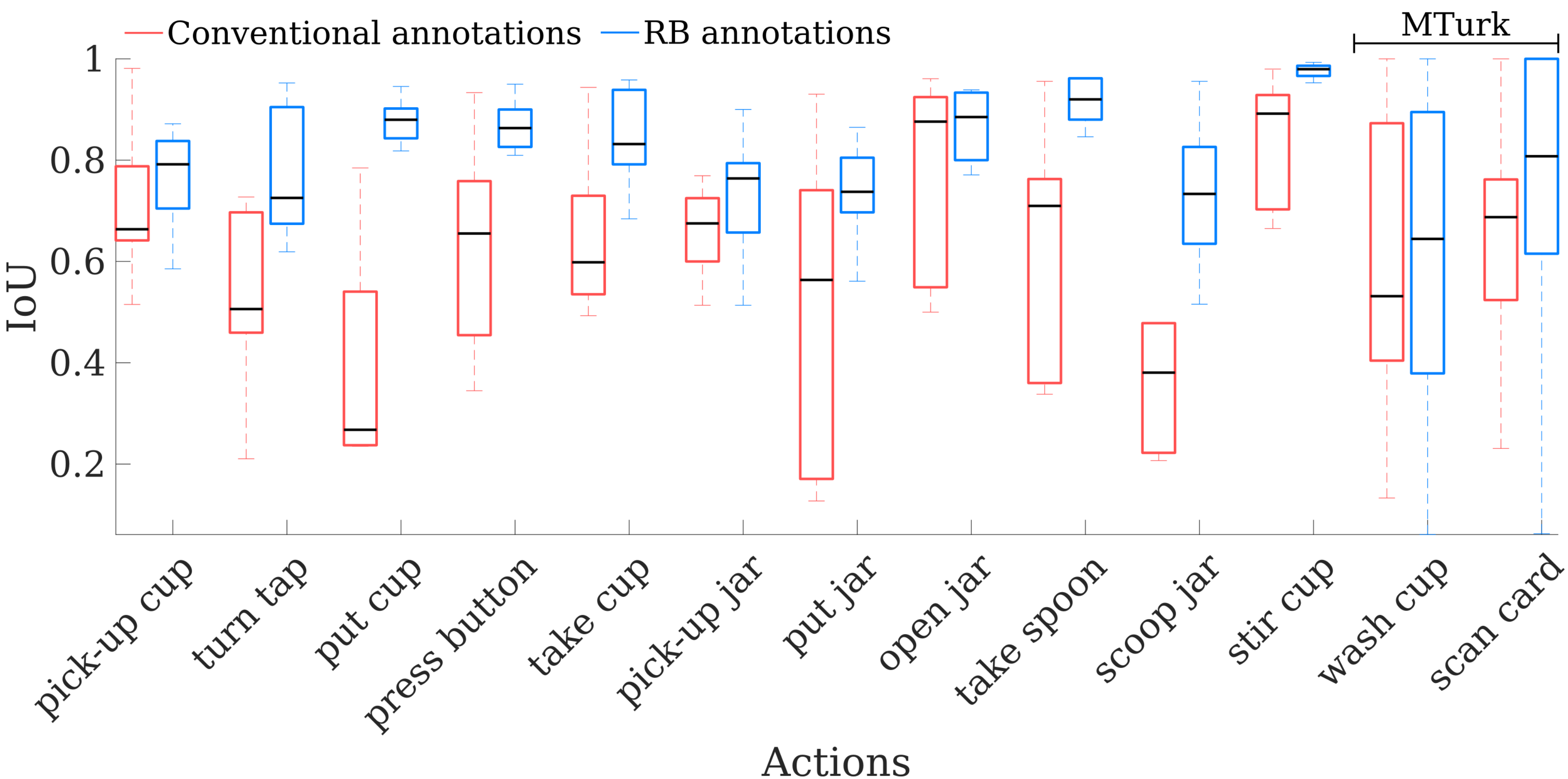
Cut pepper (GTEA Gaze+)

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

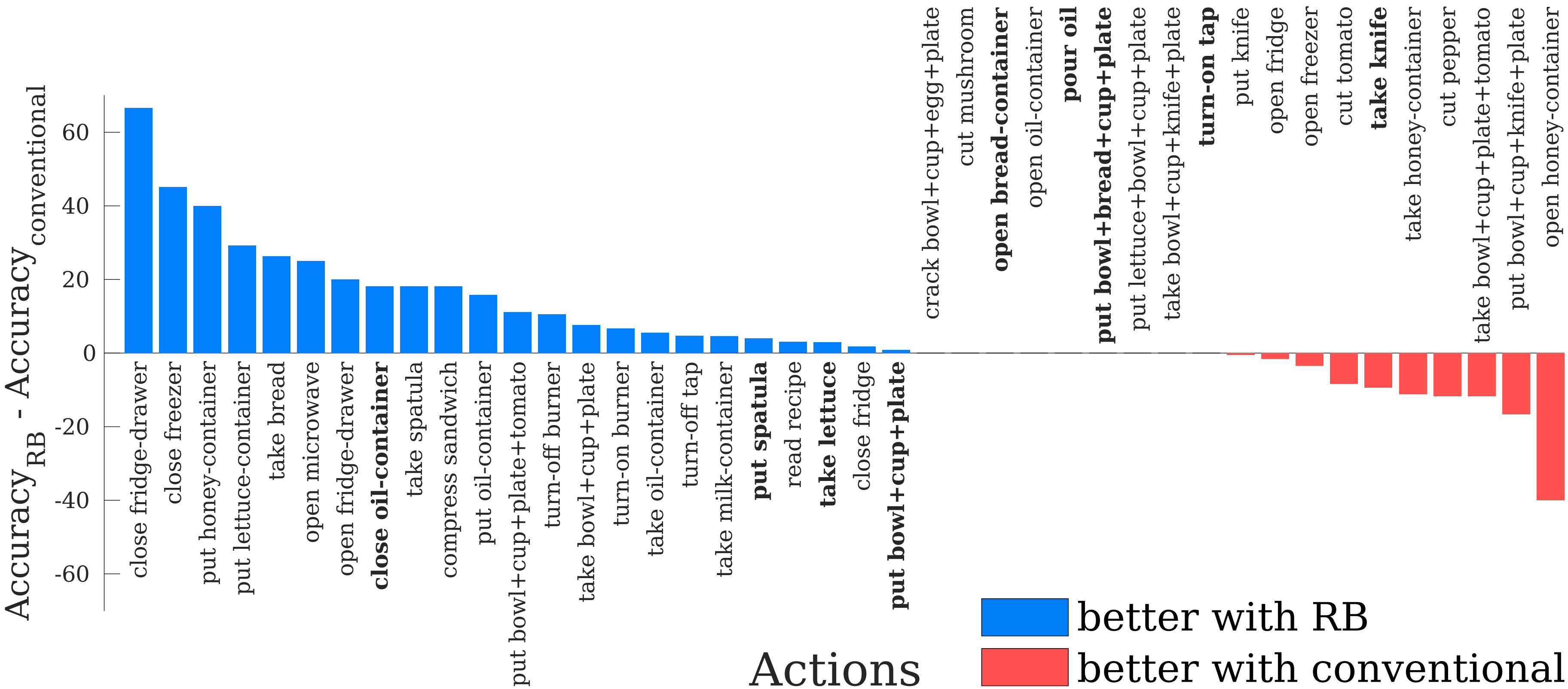
with: Davide Moltisanti
Michael Wray
Walterio Mayol-Cuevas



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

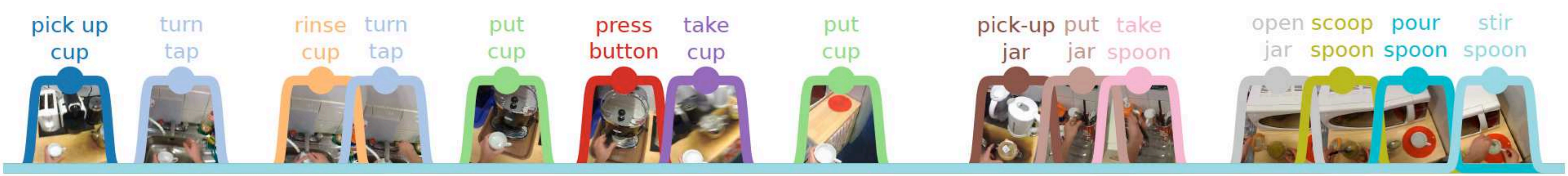


D Moltisanti et al (2017). Trespassing the Boundaries: Labeling Temporal Bounds for Object Interactions in Egocentric Video. *International Conference on Computer Vision (ICCV)*

Action Recognition from a Single Timestamp

with: Davide Moltisanti
Sanja Fidler

- Learning from Single timestamps



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

Action Recognition from a Single Timestamp

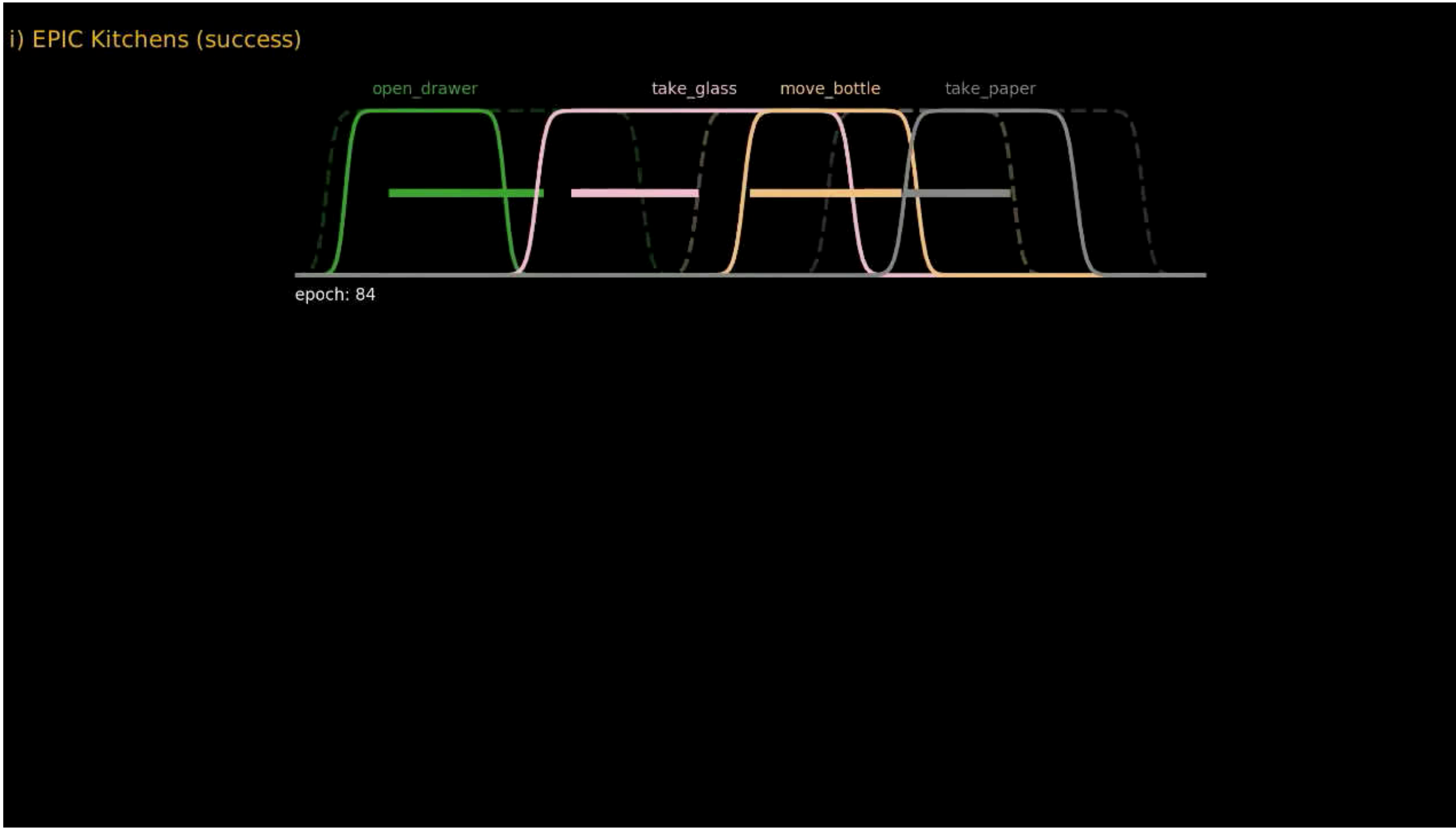
with: Davide Moltisanti
Sanja Fidler

- Learning from Single timestamps

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

Action Recognition from a Single Timestamp

with: Davide Moltisanti
Sanja Fidler

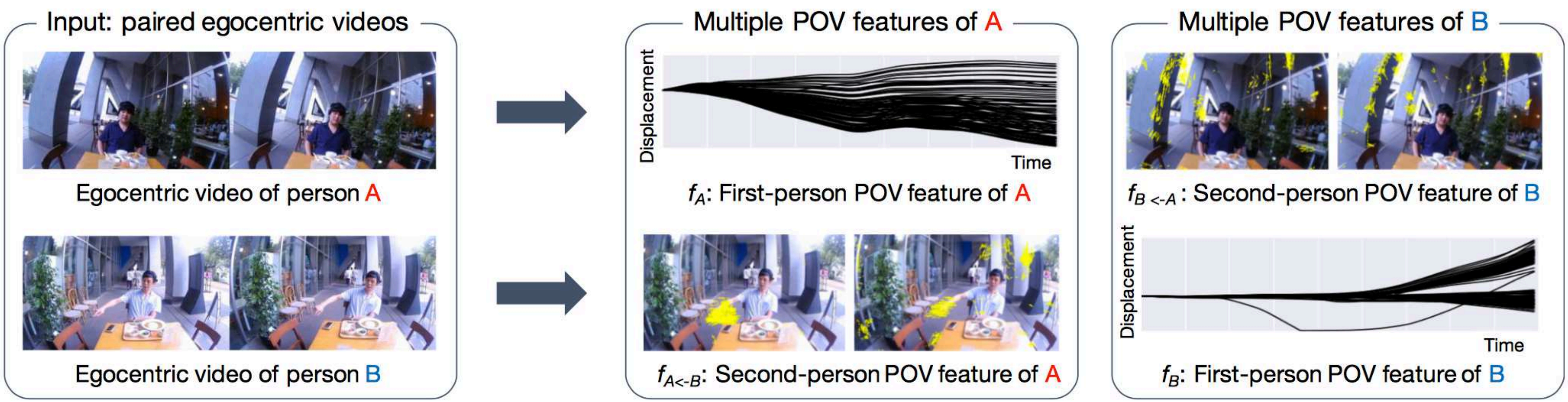


D Moltisanti et al (2019). Action Recognition from Single Timestamp Supervision in Untrimmed Videos. *Computer Vision and Pattern Recognition (CVPR)*

The Unique Problems

5. Multi-View Action Recognition

FPV with SPV



Input: paired egocentric videos
Egocentric video of person A
Egocentric video of person B
Multiple POV features of A
Displacement
Time
f_A: First-person POV feature of A
$f_{A<-B}$: Second-person POV feature of A
Multiple POV features of B
$f_{B<-A}$: Second-person POV feature of B
Displacement
Time
f_B: First-person POV feature of B

FPV with TPV (top-view)

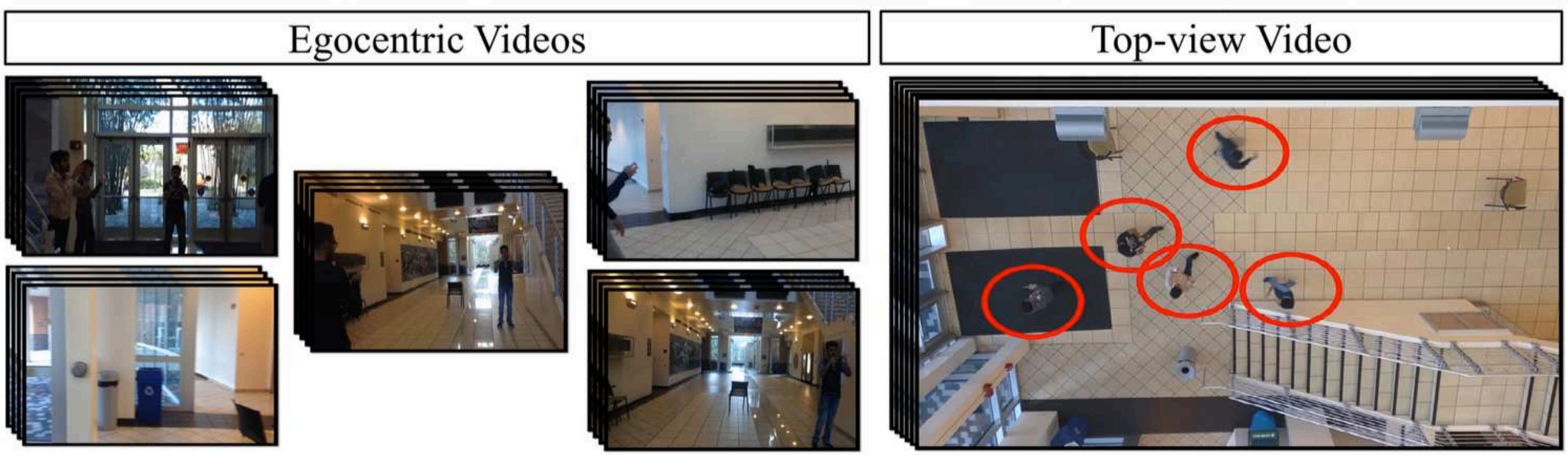


Figure from: Ardeshir and Borji (2016). Egocentric Meets Top-view CVPR

FPV with TPV (top-view)

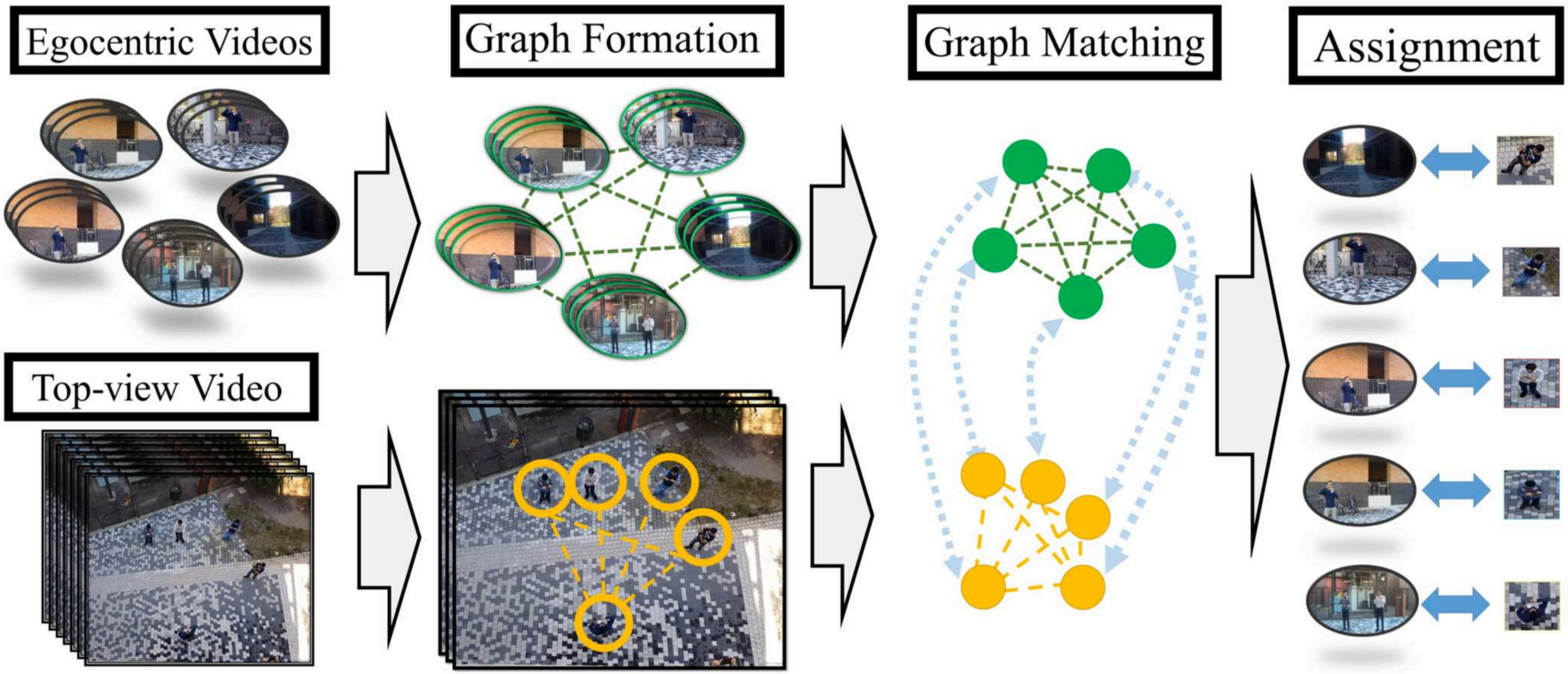


Figure from: Ardeshir and Borji (2016). Ego2Top: Matching Viewers in Egocentric and Top-view Videos. ECCV

Egocentric Vision

- The Unique Problems
  1. Camera Motion
  2. Mapping and Localisation (ref tomorrow's talk)
  3. Attention and Task-Relevance
  4. Object Interactions
  5. Multi-view Solutions
- The Unique Applications
  1. Video Summarisation
  2. Skill Determination
  3. Real-time solutions

The Unique Applications

1. Video Summarisation

Video Summarisation

- Fixations
- Highlight Detection

Egocentric Video Summarisation

- Object-Driven

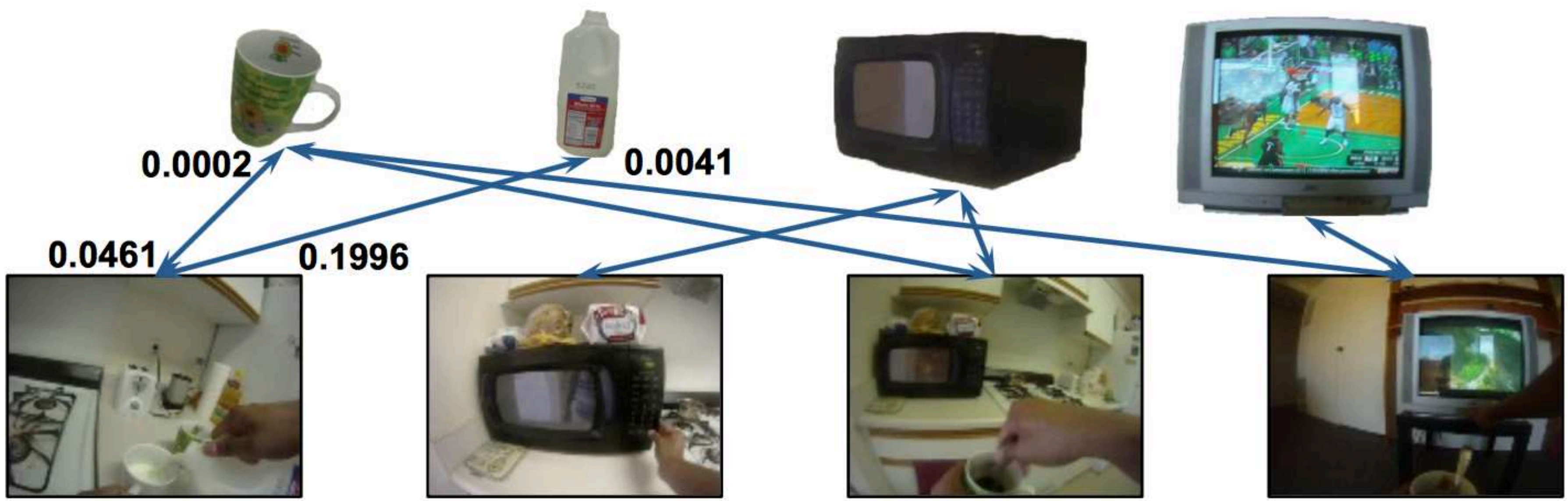


Figure from: Lu and Grauman (2013). Story-Driven Summarization for Egocentric Video. CVPR

Egocentric Video Summarisation

- Object-Driven

Figure from: Lu and Grauman (2013). Story-Driven Summarization for Egocentric Video. CVPR

Egocentric Video Summarisation

- Fixation-Driven with Constraints

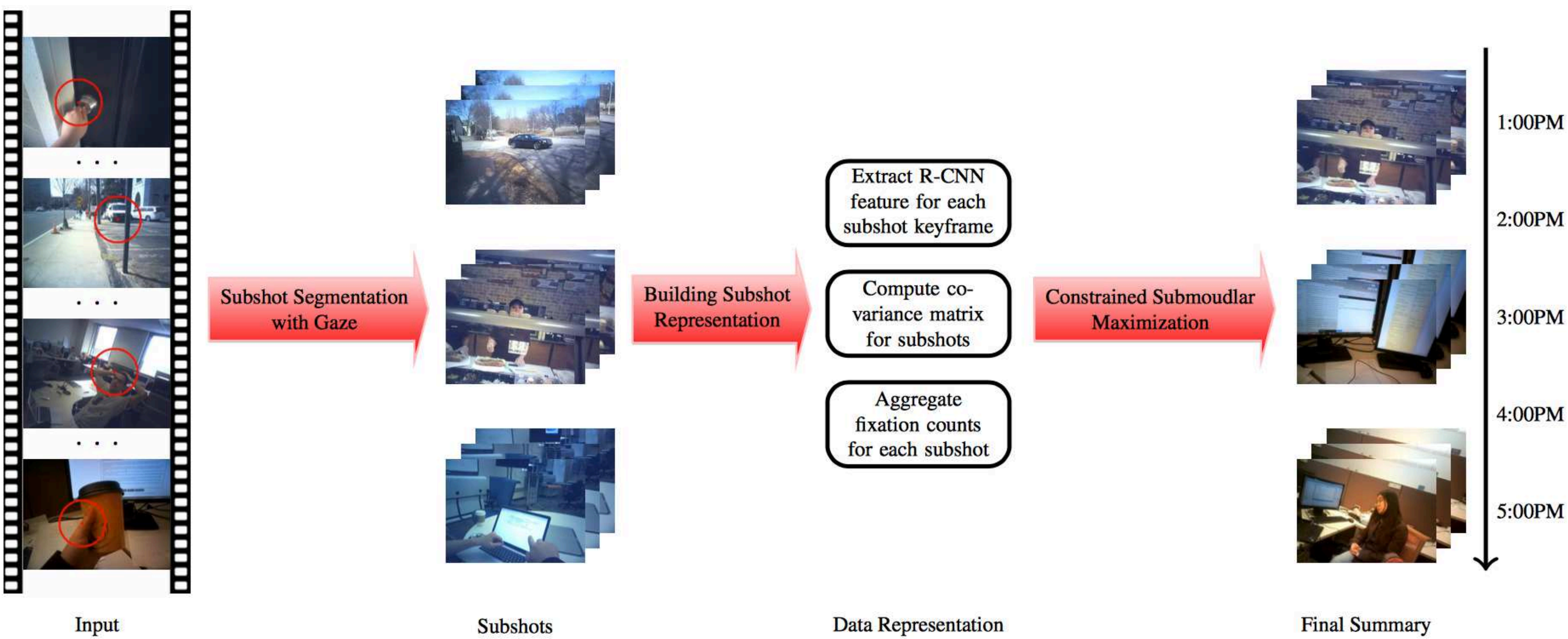


Figure from: Xu et al (2015). Gaze-enabled Egocentric Video Summarization via Constrained Submodular Maximization . CVPR

Egocentric Video Summarisation

- Fixations from IMUs

The Unique Applications

2. Skill Determination

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

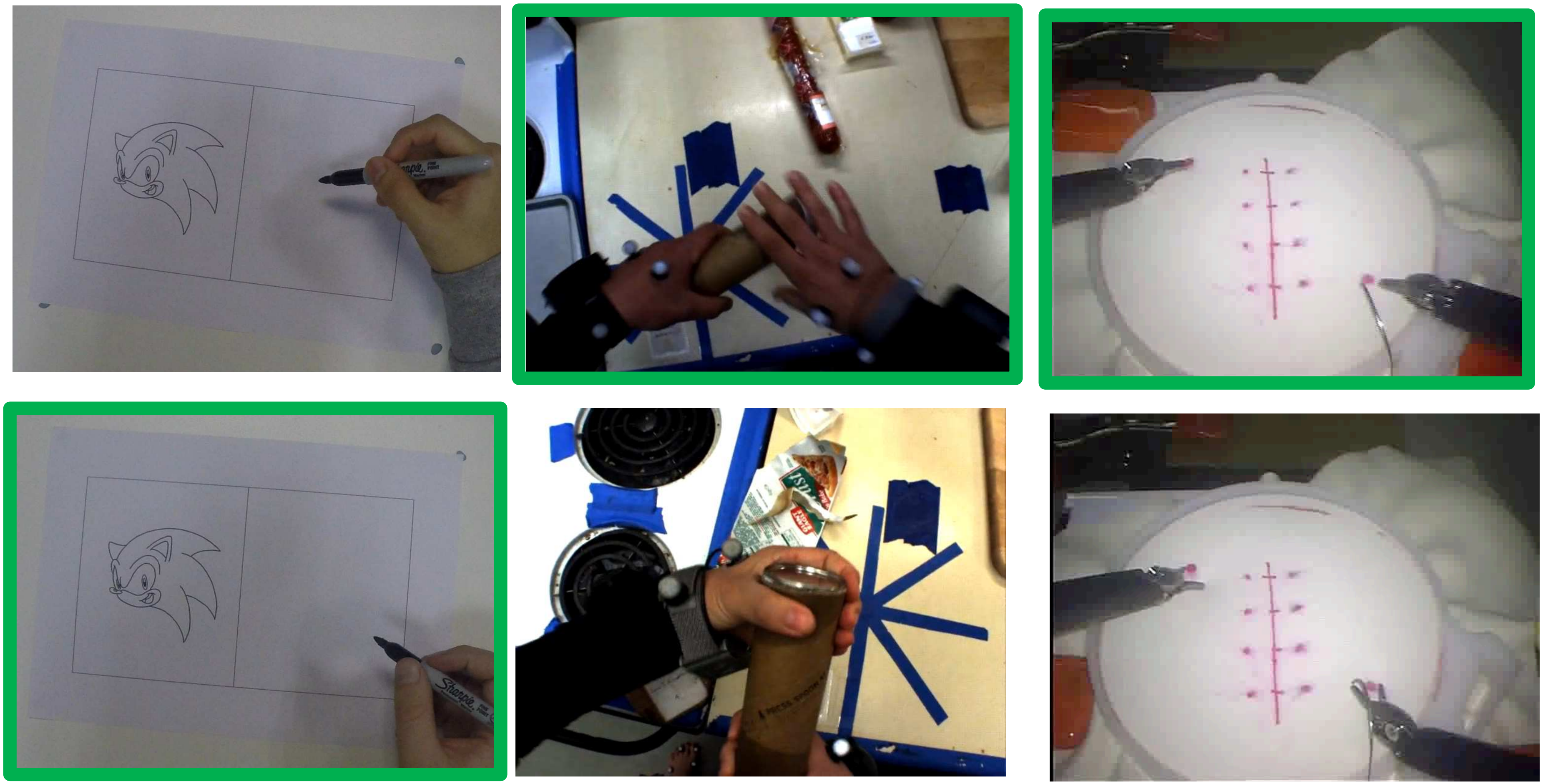
with: Hazel Doughty
Walterio Mayol-Cuevas

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

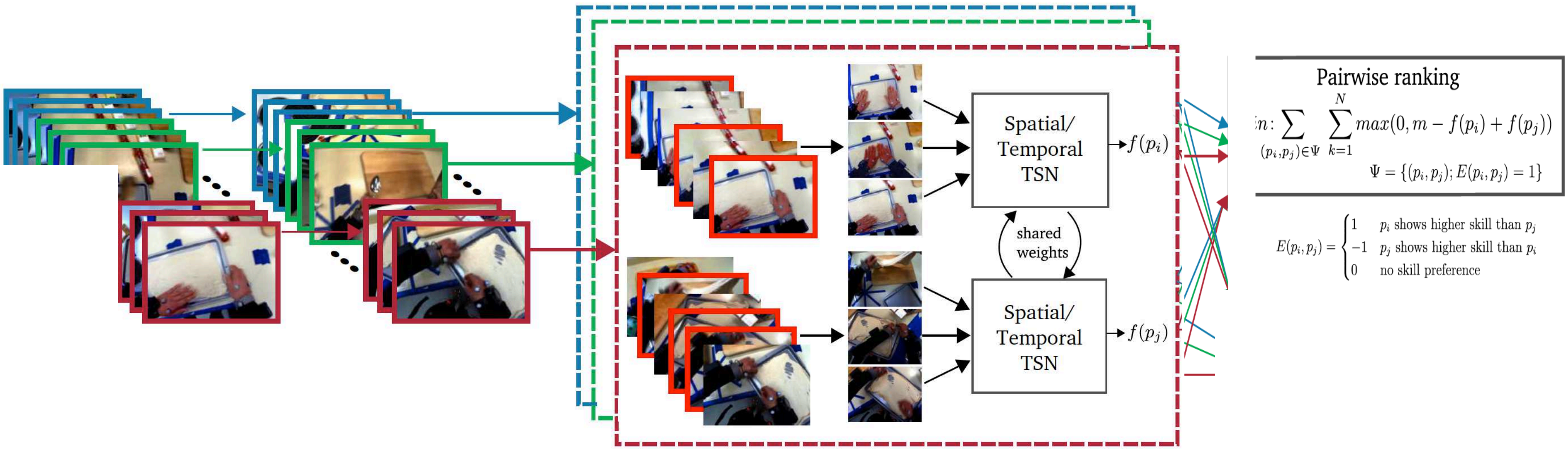
with: Hazel Doughty
Walterio Mayol-Cuevas

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



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

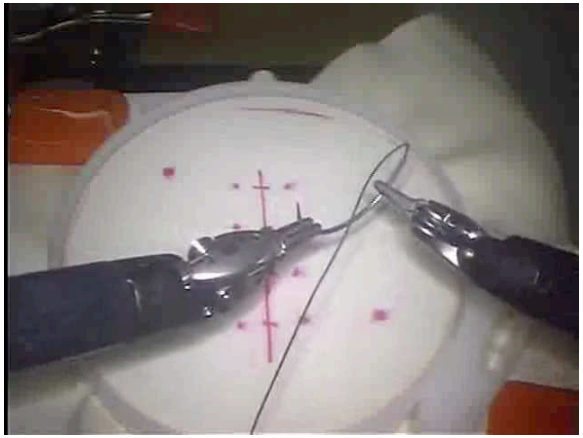


Video 1 ~ Video 2

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

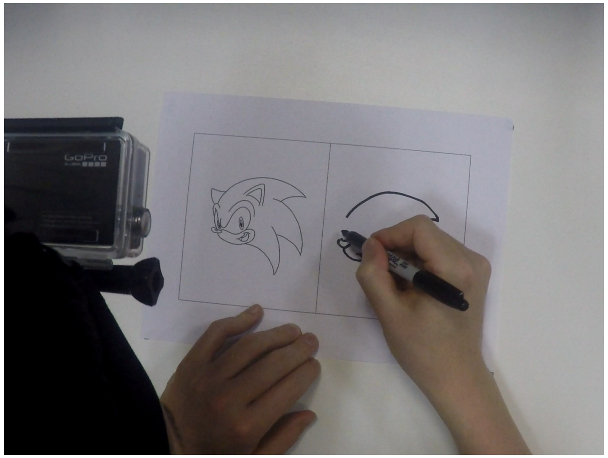
with: Hazel Doughty
Walterio Mayol-Cuevas

Surgery[1]



EPIC-SKILLS 2018

Drawing



Dough-Rolling[2]

Chopstick Using

[1] Gao, Yixin, et al. "The JHU-ISI gesture and skill assessment dataset (JIGSAWS): A surgical activity working set for human motion modeling." Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2014.
[2] De la Torre, Fernando, et al. "Guide to the carnegie mellon university multimodal activity (CMU-MMAC) database." *Robotics Institute* (2008): 135.

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

with: Hazel Doughty
Walterio Mayol-Cuevas

$$L_{rank1} = \sum_{(p_i,p_j)\in\Psi} max(0, m - f(p_i) + f(p_j)) \quad (3)$$

$$L_{rank2} = \sum_{(p_i,p_j)\in\Psi} \sum_{k=1}^{N} max(0, m - f_k(p_i) + f_k(p_j)) \quad (5)$$

$$L_{sim} = \sum_{(p_i,p_j)\in\Phi} \sum_{k=1}^{N} max(0, |f(p_i) - f(p_j)| - m) \quad (7)$$

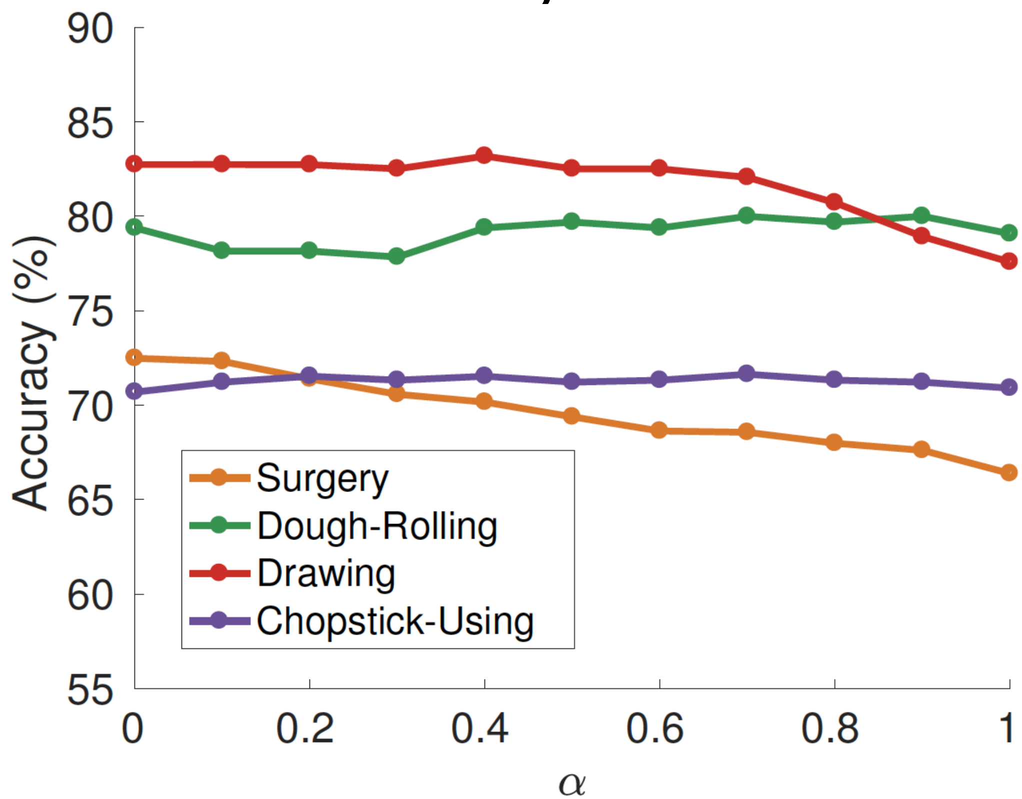
$$L_{rank3} = \beta L_{rank2} + (1-\beta) L_{sim} \quad (8)$$

Method	Surgery			Dough-Rolling			Drawing			Chopstick-Using		
	S	T	TS	S	T	TS	S	T	TS	S	T	TS
Siamese TSN with margin loss	64.7	72.8	69.1	77.6	79.4	78.5	75.6	77.4	78.0	67.2	67.9	68.8
+ splits	64.4	**73.3**	69.0	79.1	**80.4**	78.5	74.9	81.8	79.1	67.2	69.9	68.8
+ similarity loss	**66.4**	72.5	**70.2**	**79.5**	79.5	**79.4**	**77.6**	**82.7**	**83.2**	**70.8**	**70.6**	**71.5**

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

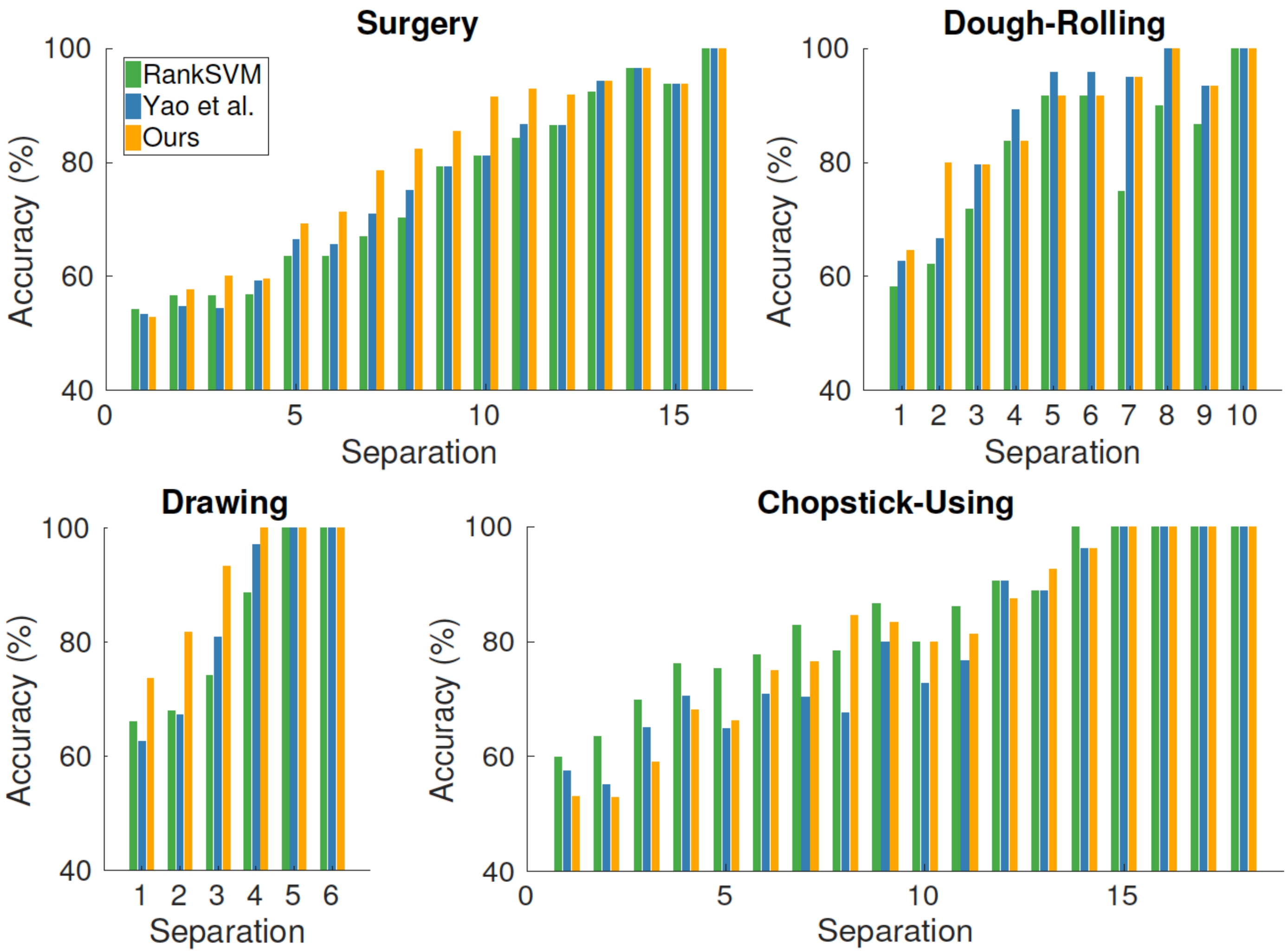
with: Hazel Doughty
Walterio Mayol-Cuevas

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



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

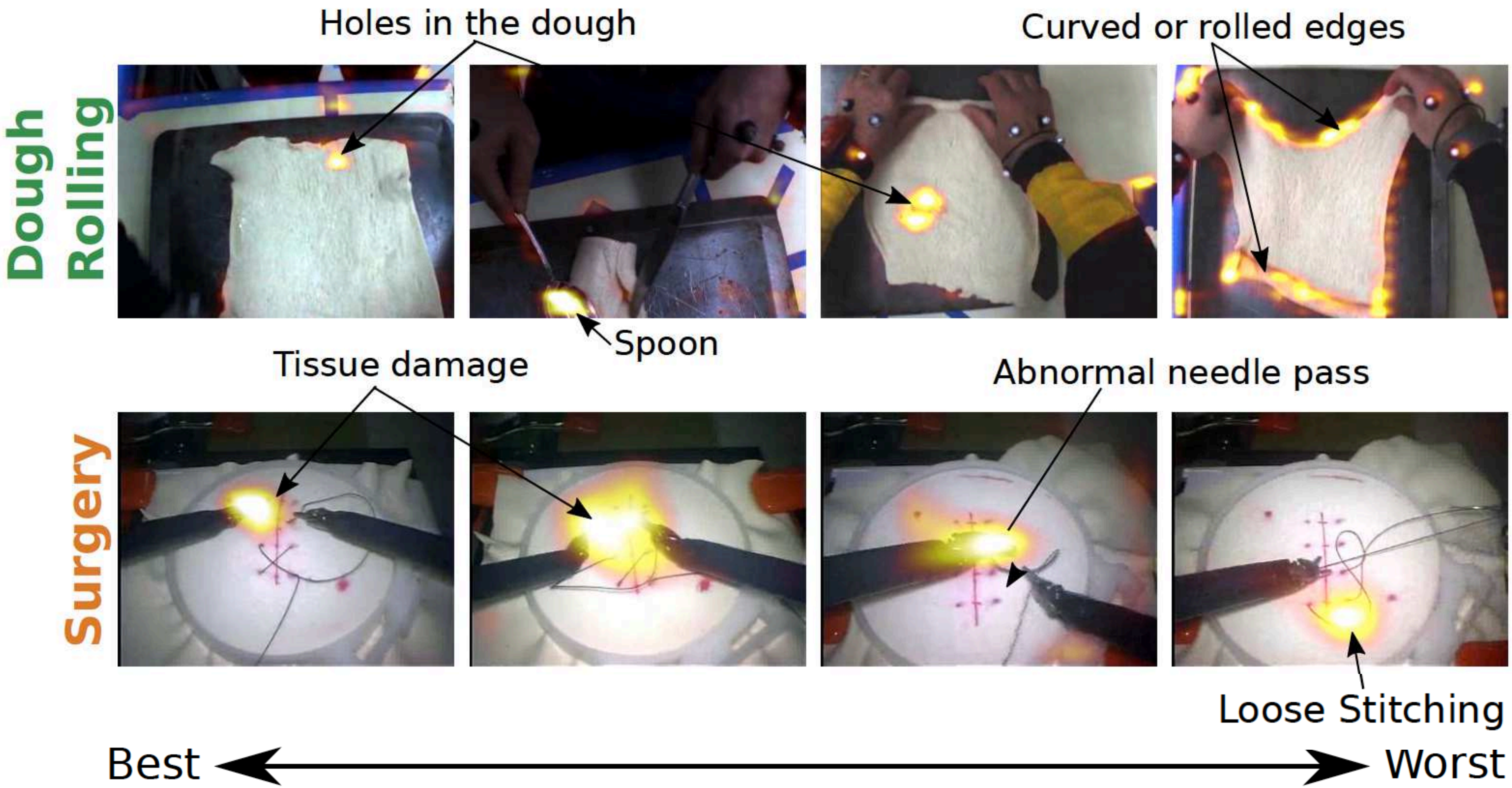
with: Hazel Doughty
Walterio Mayol-Cuevas



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

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

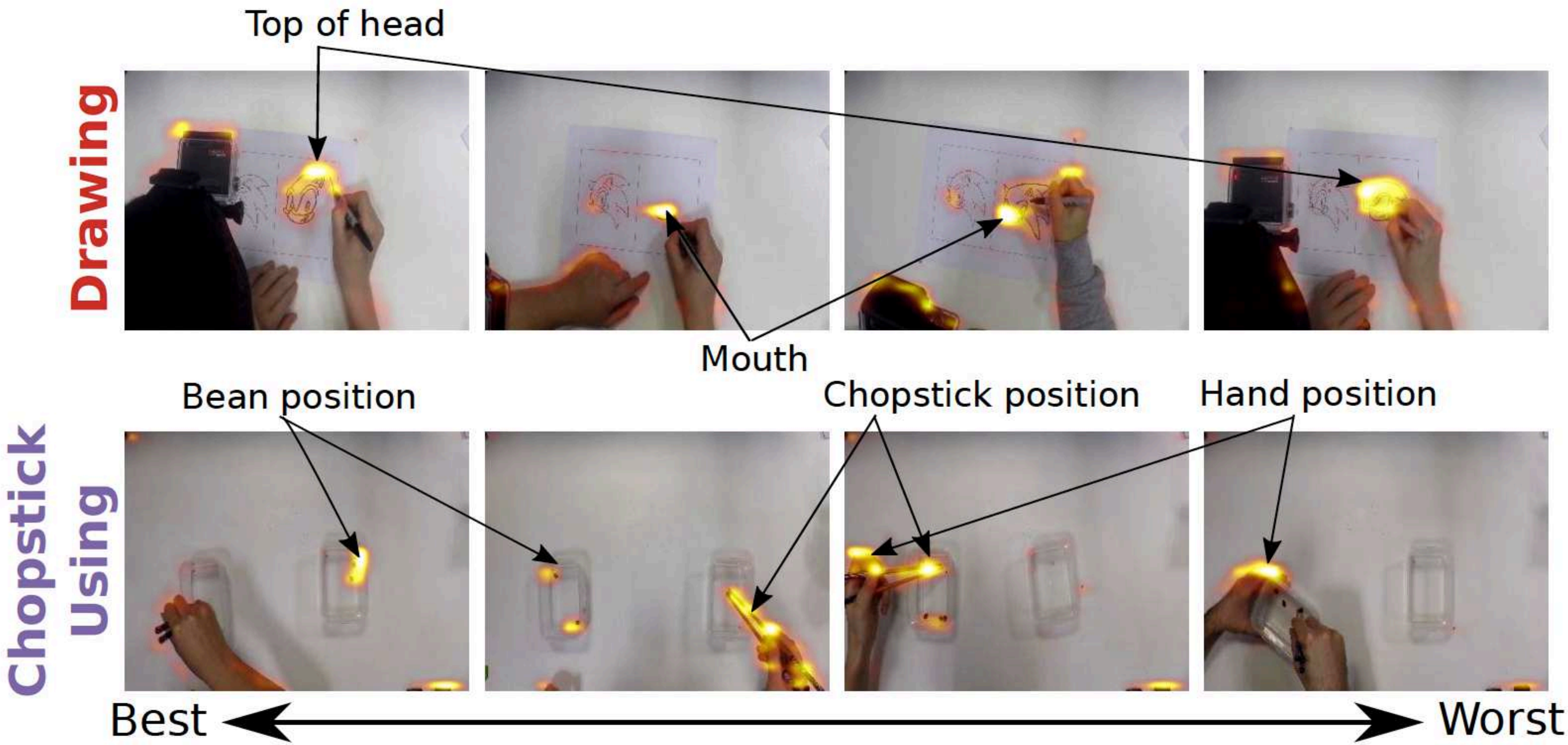
with: Hazel Doughty
Walterio Mayol-Cuevas



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

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

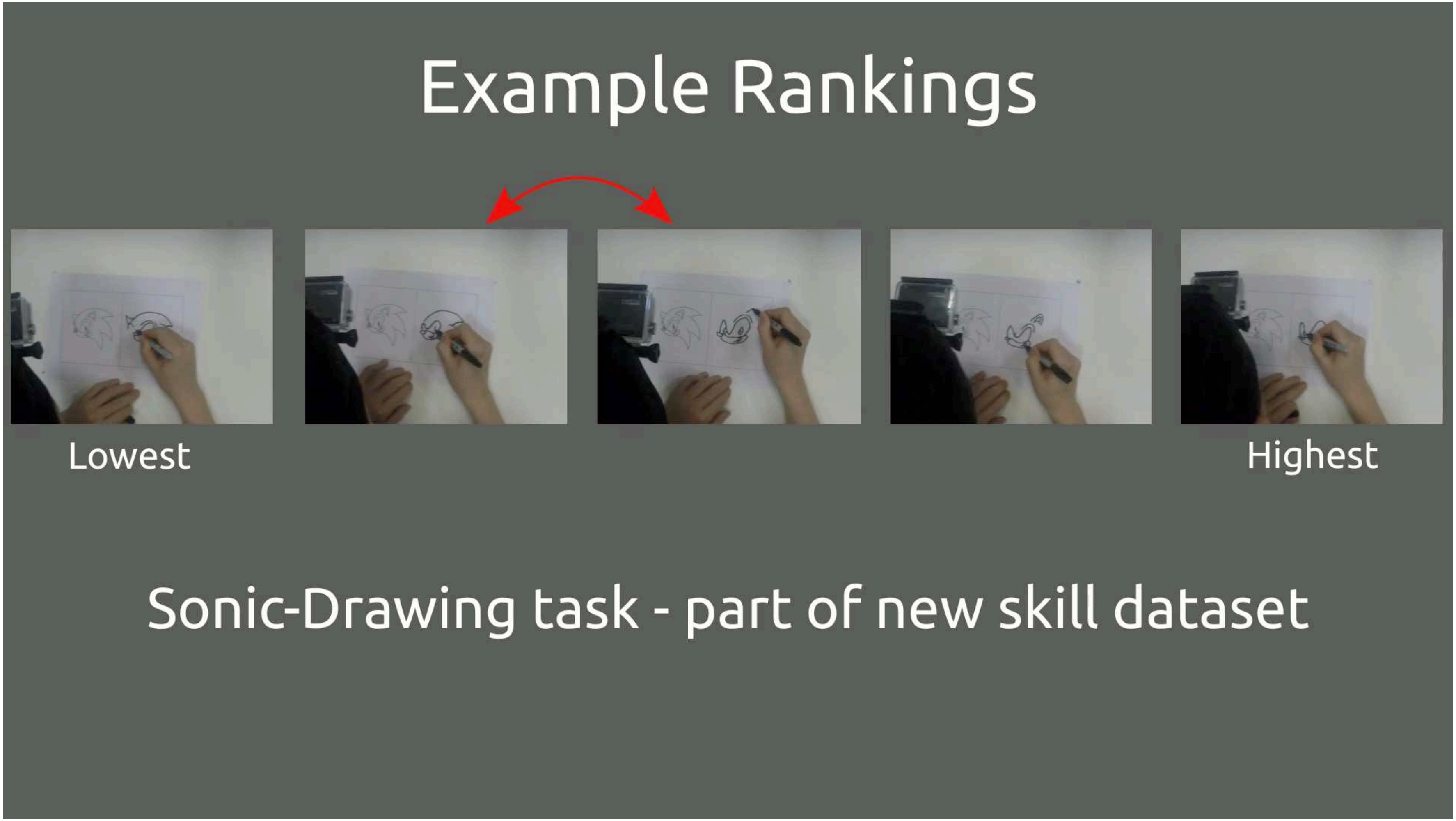
with: Hazel Doughty
Walterio Mayol-Cuevas



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

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

with: Hazel Doughty
Walterio Mayol-Cuevas

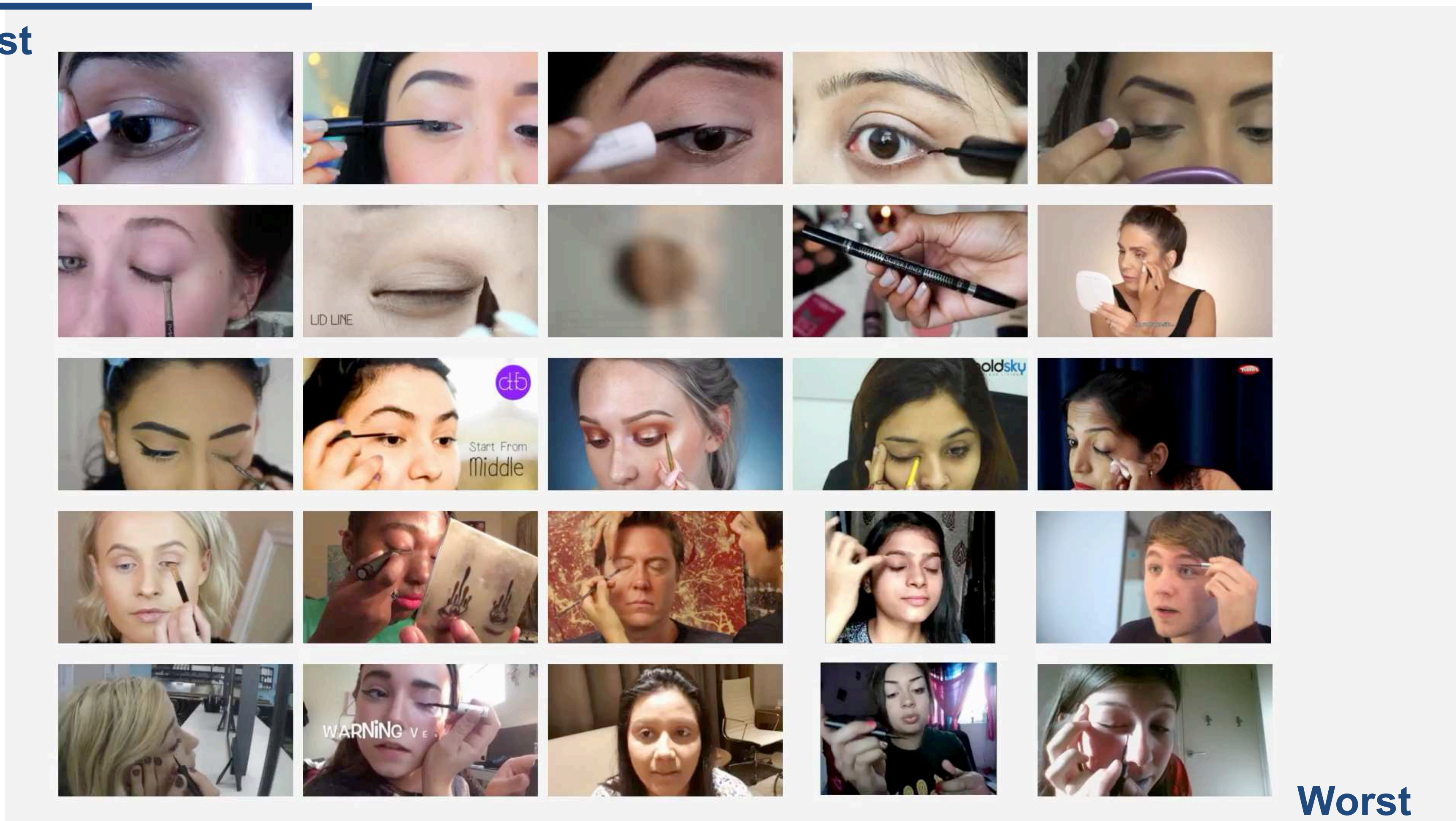


H Doughty, D Damen, W Mayol-Cuevas (2018). Who's Better? Who's Best? Pairwise Deep Ranking for Skill Determination. CVPR

The Pros and Cons: Rank-Aware Temporal Attention

with: Hazel Doughty
Walterio Mayol-Cuevas

Best

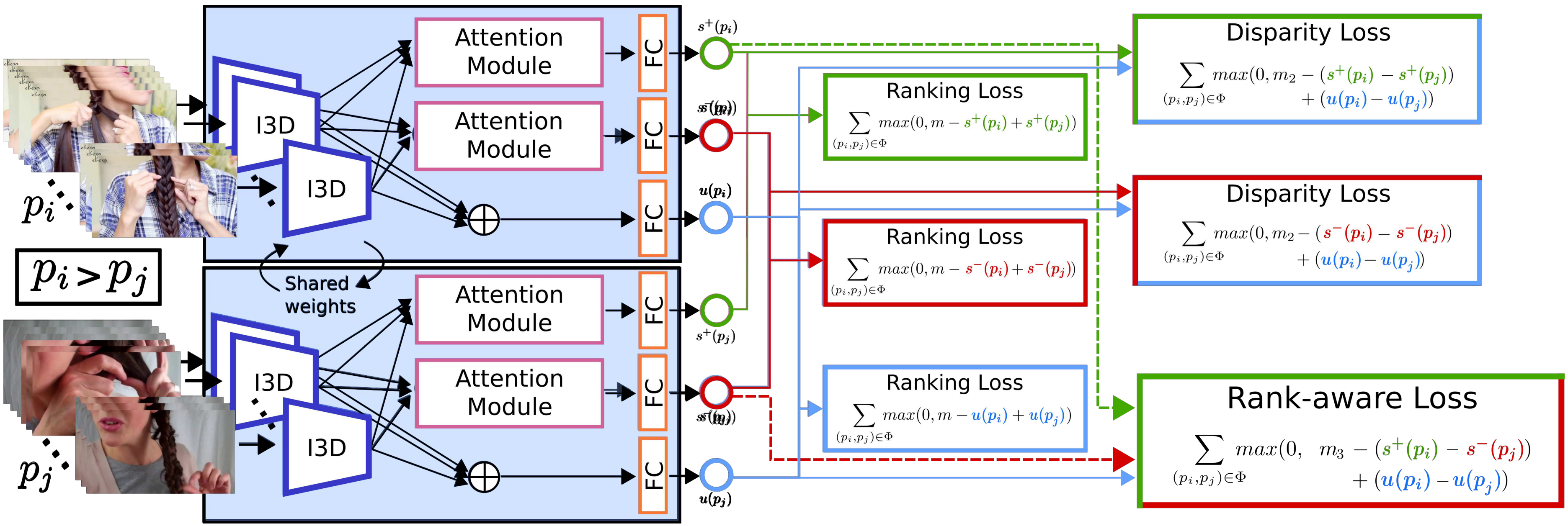


Worst

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



Novel Rank-Aware Loss

$$L^+_{rank} = \sum_{(p_i,p_j)\in P} max(0, m - s^+(p_i) + s^+(p_j)) \quad (8)$$

$$L^-_{rank} = \sum_{(p_i,p_j)\in P} max(0, m - s^-(p_i) + s^-(p_j)) \quad (9)$$

$$L^u_{rank} = \sum_{(p_i,p_j)\in P} max(0, m - u(p_i) + u(p_j)) \quad (10)$$

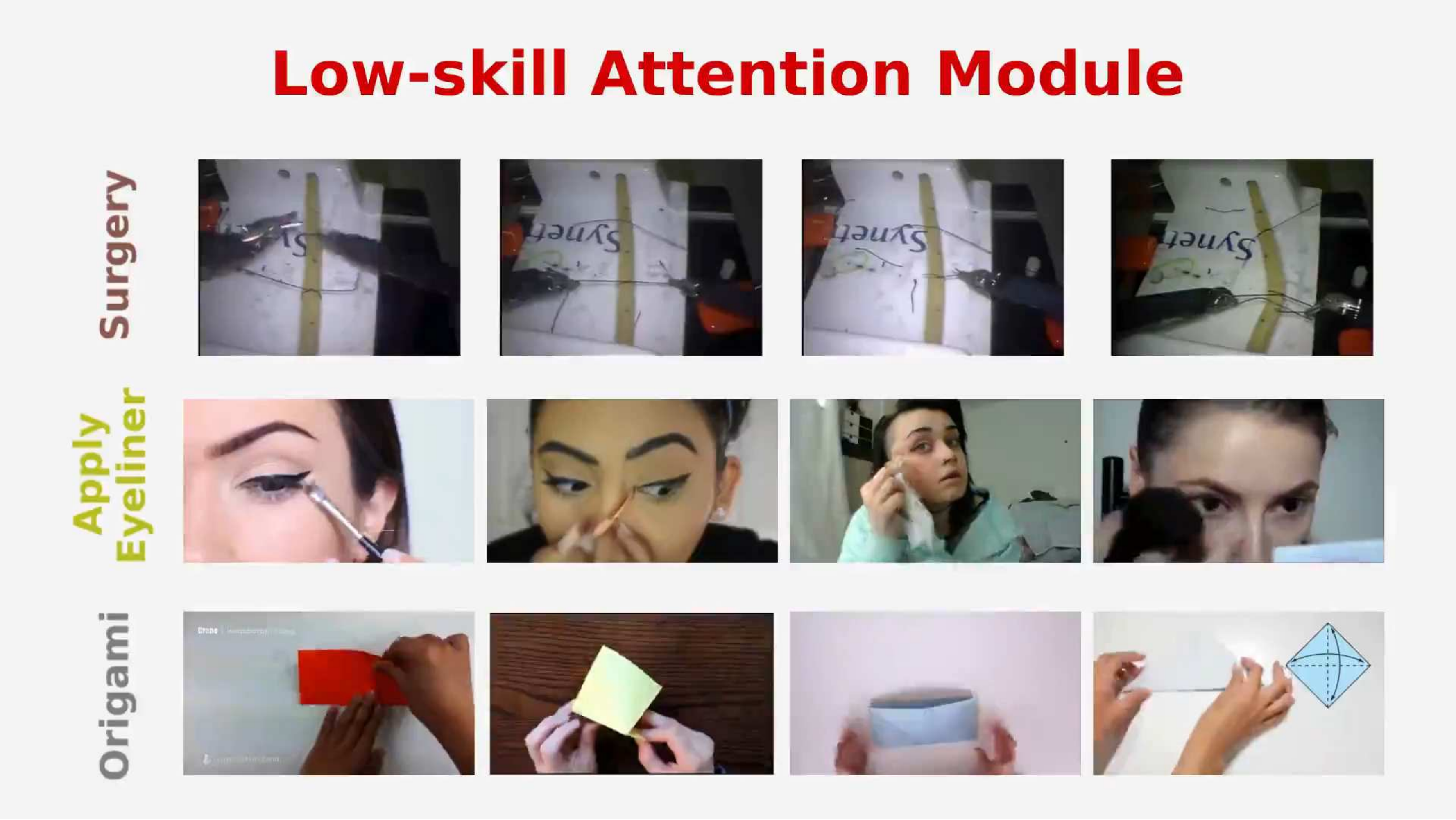
$$\begin{aligned} L^+_{disp} = \sum_{(p_i,p_j)\in P} max(0, m_2 - (s^+(p_i) - s^+(p_j)) \\ + (u(p_i) - u(p_j))) \quad (11) \end{aligned}$$

$$\begin{aligned} L_{rAware} = \sum_{(p_i,p_j)\in P} max(0, m_3 - (s^+(p_i) - s^-(p_j)) \\ + (u(p_i) - u(p_j))) \quad (12) \end{aligned}$$

$$L_R = \sum_{i=\{+,-,u\}} L^i_{rank} + \sum_{i=\{+,-\}} L^i_{disp} + L_{rAware} \quad (13)$$

The Pros and Cons: Rank-Aware Temporal Attention

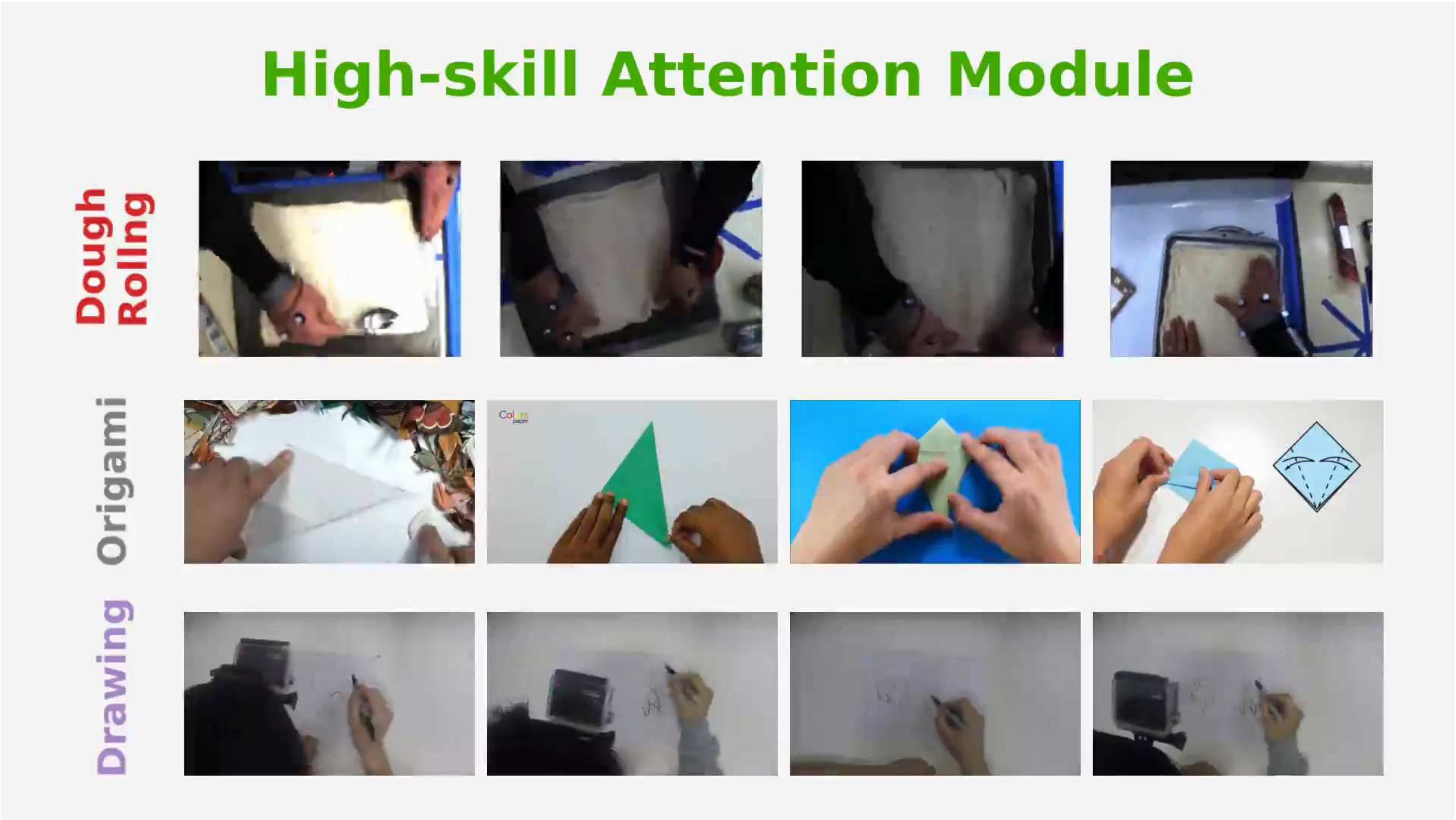
with: Hazel Doughty
Walterio Mayol-Cuevas



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)*

The Pros and Cons: Rank-Aware Temporal Attention

with: Hazel Doughty
Walterio Mayol-Cuevas

Method	EPIC Skills	YouTubeSkill
Who's Better [4]	76.0	75.8
Last Segment	76.8	61.0
Uniform Weighting	78.8	73.6
Softmax Attention	74.5	72.3
STPN [18]	74.3	70.0
Ours (Rank Aware Attention)	**80.3**	**81.2**

Table 2. Results of our method in comparison to baseline. Our final method outperforms every baseline on both datasets.

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 Unique Applications

3. Real-time Solutions

Wearable (Systems)!

- On-the-cloud processing
- On-the-mobile processing
- Onboard processing!

Connecting-to-the-cloud

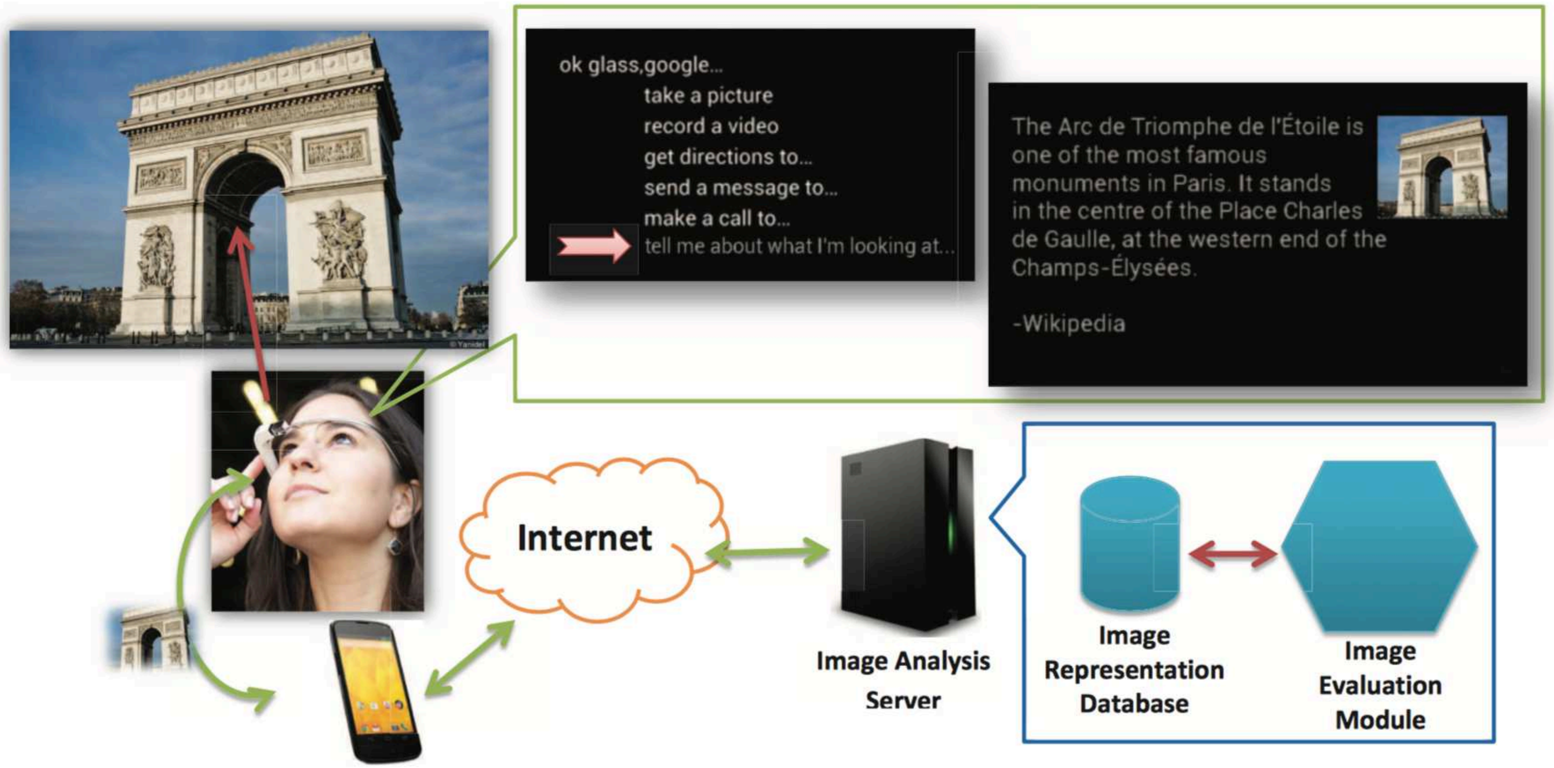


Figure 1. System overview. The user asks the device to inform her about her current view of Arc de Triomphe, and the system responds with the most relevant description in its database.

Altwaijry et al (2014). Recognizing locations with Google Glass: A case study. WACV

with: Walterio Mayol-Cuevas
Teesid Leelasawassuk

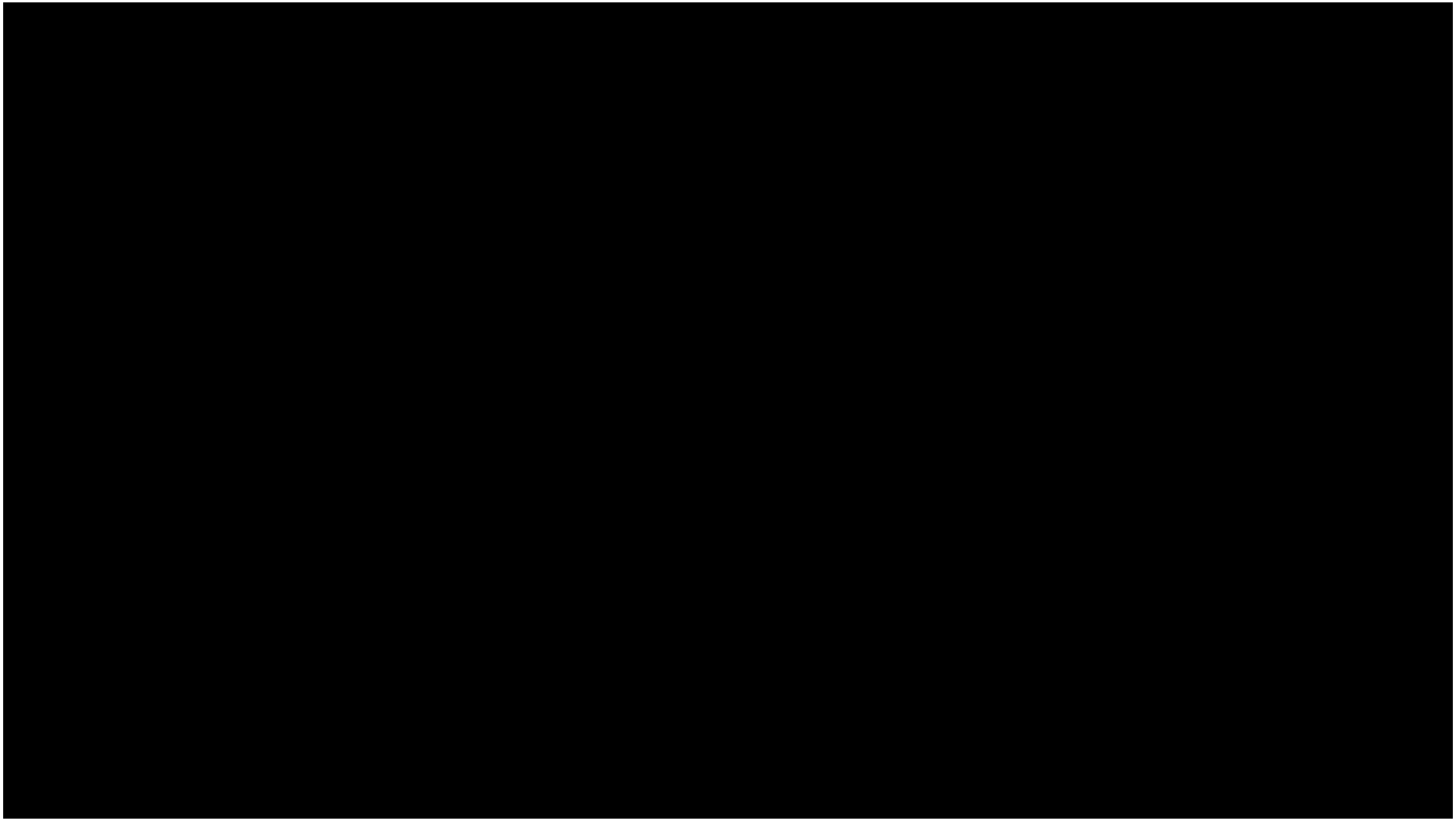
You Do, I Learn – Google Glass Prototype

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*

The need for large-scaled datasets…

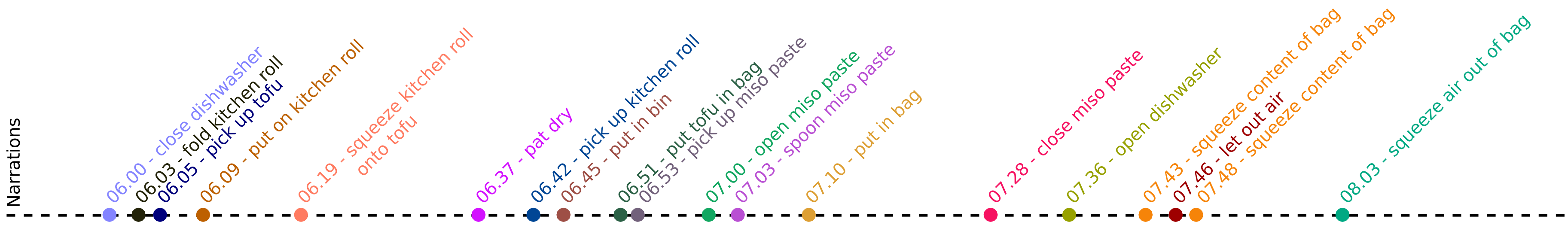
EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



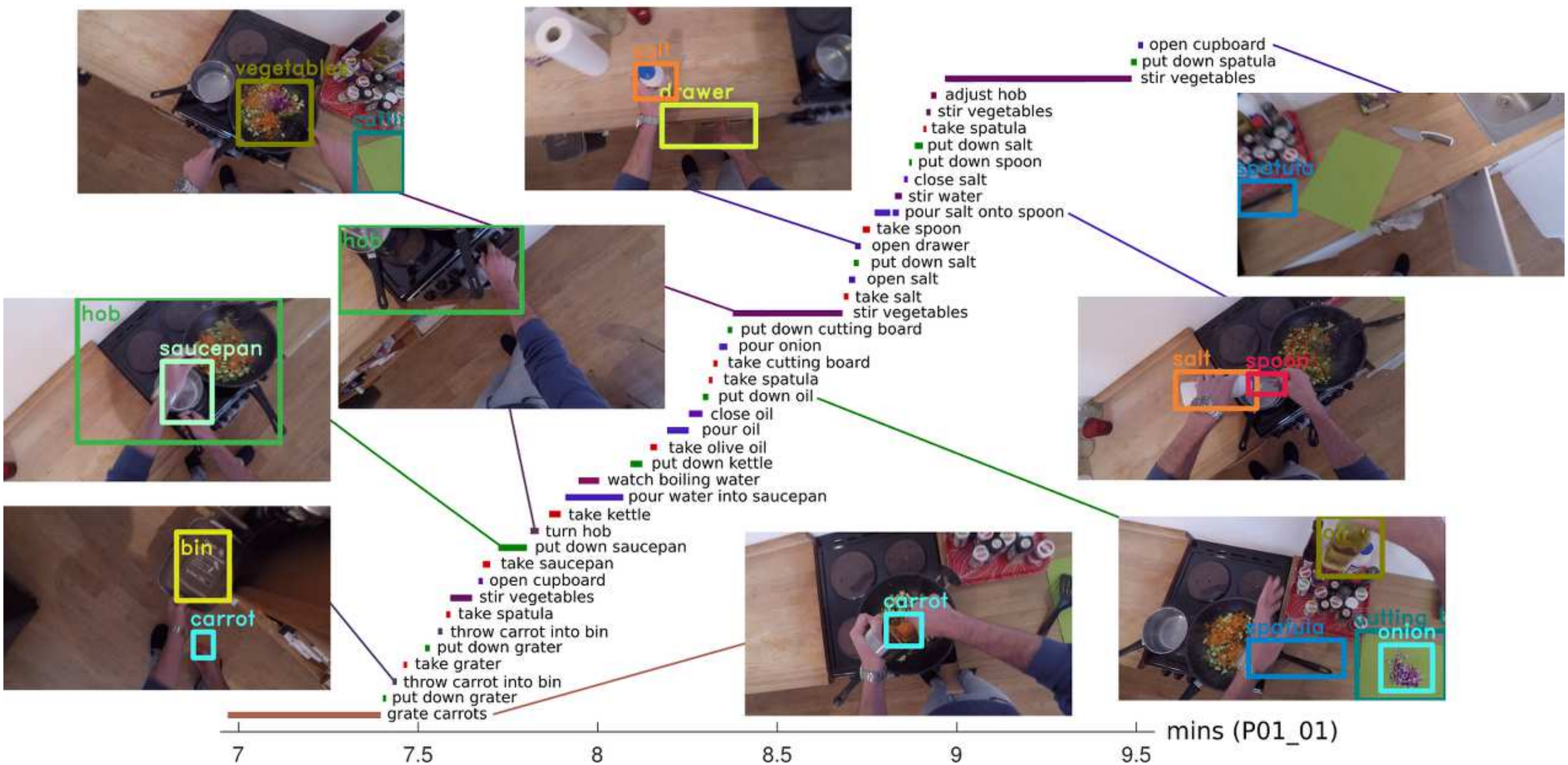
EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



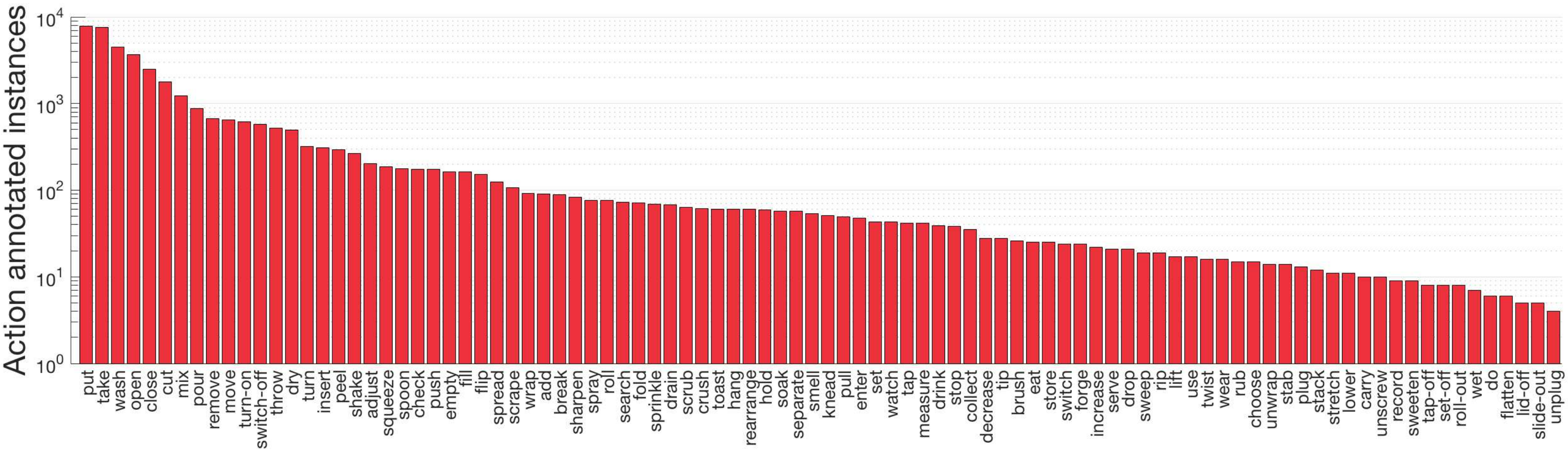
EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



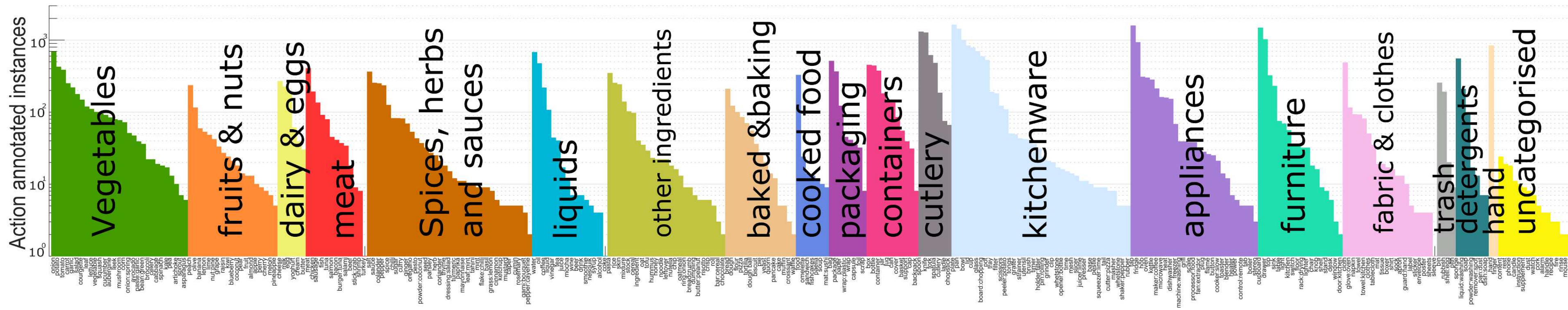
EPIC KITCHENS

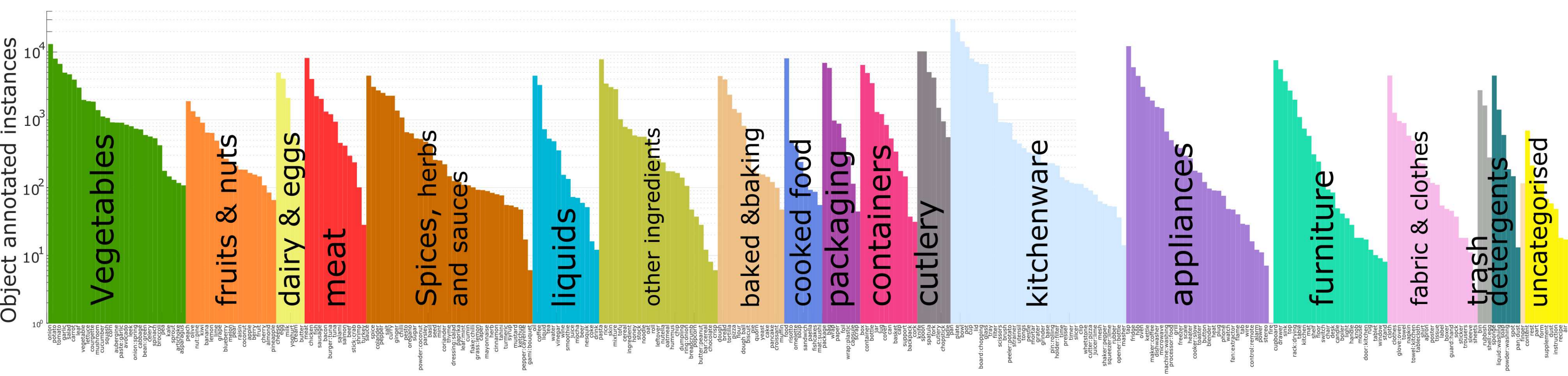
with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray



EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray





D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

TABLE 1: Comparative overview of relevant datasets. *action classes with $>$ 50 samples

Dataset	Ego?	Non-Scripted?	Native Env?	Year	Frames	Sequ-ences	Action Segments	Action Classes	Object BBs	Object Classes	Partici-pants	No. Env.s
EPIC-KITCHENS	✓	✓	✓	2018	11.5M	432	39,596	149*	454,158	323	32	32
EGTEA Gaze+ [19]	✓	×	×	2018	2.4M	86	10,325	106	0	0	32	1
BEOID [21]	✓	×	×	2014	0.1M	58	1,488	34	0	0	5	1
GTEA Gaze+ [20]	✓	×	×	2012	0.4M	35	3,371	42	0	0	13	1
ADL [23]	✓	×	✓	2012	1.0M	20	436	32	137,780	42	20	20
CMU [22]	✓	×	×	2009	0.2M	16	516	31	0	0	16	1
VLOG [15]	×	✓	✓	2017	37.2M	114K	0	0	0	0	10.7K	N/A
Charades [16]	×	×	✓	2016	7.4M	9,848	67,000	157	0	0	N/A	267
Breakfast [24]	×	✓	✓	2014	3.0M	433	3078	50	0	0	52	18
50 Salads [25]	×	×	×	2013	0.6M	50	2967	52	0	0	25	1
MPII Cooking 2 [26]	×	×	×	2012	2.9M	273	14,105	88	0	0	30	1

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

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		28,588	28,561	326,298
S1 Test	28	106	39084	20%	8,069	8,064	97,865
S2 Test	4	54	13231	7%	2,939	2,939	29,995

		15 Most Frequent Object Classes															**Totals**		
	mAP	pan	plate	bowl	onion	tap	pot	knife	spoon	meat	food	potato	cup	pasta	cupboard	lid	few-shot	many-shot	all
S1	IoU > 0.05	74.00	72.61	71.50	60.72	84.44	69.97	44.03	40.93	29.65	58.52	62.82	53.30	78.39	51.95	62.77	9.71	49.80	38.23
	IoU > 0.5	67.60	66.21	65.98	39.96	73.80	64.71	28.80	23.89	20.75	49.85	55.48	42.99	69.75	29.20	58.48	6.98	36.50	28.06
	IoU > 0.75	21.94	44.60	39.48	3.52	25.83	19.67	3.42	2.59	5.27	15.78	13.18	8.00	24.53	4.05	26.51	0.36	8.73	6.50
S2	IoU > 0.05	75.94	87.36	72.72	47.61	78.14	75.92	55.51	41.28	71.59	38.61	N/A	44.62	80.58	53.88	58.40	6.00	51.71	40.61
	IoU > 0.5	62.88	84.86	68.61	32.18	59.75	62.86	39.60	27.52	53.54	35.47	N/A	39.19	76.27	32.54	49.36	5.32	36.27	28.57
	IoU > 0.75	14.56	62.82	38.44	2.25	4.89	14.91	3.85	1.51	9.56	8.10	N/A	7.60	43.30	5.61	25.48	0.18	9.05	7.04

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

TABLE 6: Baseline results for the action recognition challenge

		Top-1 Accuracy			Top-5 Accuracy			Avg Class Precision			Avg Class Recall		
		VERB	NOUN	ACTION	VERB	NOUN	ACTION	VERB	NOUN	ACTION	VERB	NOUN	ACTION
S1	RGB	45.25	35.78	18.91	**86.07**	**62.80**	**39.39**	54.94	40.41	07.01	**23.31**	**30.03**	**05.29**
	FLOW	43.27	17.92	09.10	79.89	39.63	21.91	**64.58**	24.51	01.52	15.35	09.72	01.28
	FUSION	**47.36**	**36.05**	**19.44**	84.27	61.05	35.45	63.12	**44.24**	**07.33**	21.95	29.25	05.22
S2	RGB	35.96	21.74	09.96	**74.70**	44.95	**24.59**	45.40	22.14	02.06	11.79	16.75	**01.91**
	FLOW	**40.56**	14.91	07.28	73.66	33.87	18.29	44.83	22.99	00.92	**14.16**	08.79	00.94
	FUSION	39.67	**22.33**	**10.84**	74.53	**45.23**	23.52	**59.60**	**23.65**	**02.09**	13.37	**16.84**	01.84

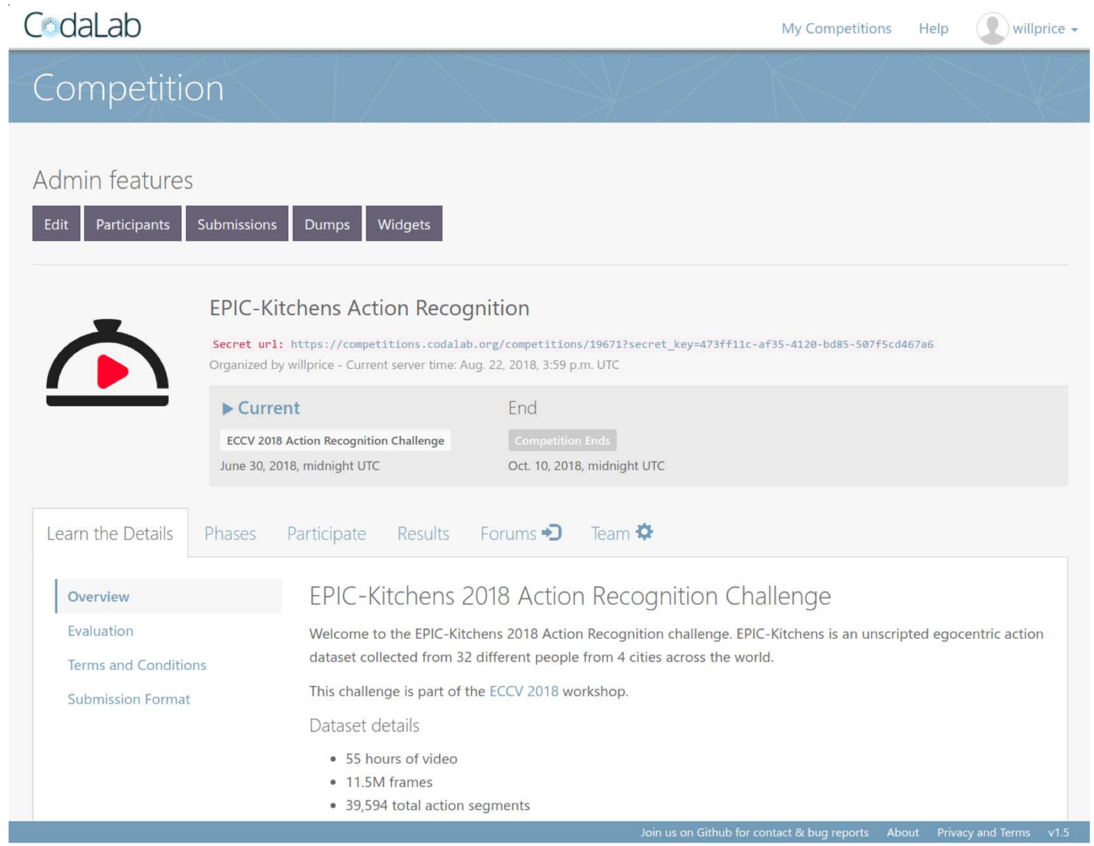
TABLE 7: Sample baseline action recognition per-class metrics (using fusion)

		15 Most Frequent Verb Classes														
		put	take	wash	open	close	cut	mix	pour	move	turn-on	remove	turn-off	throw	dry	peel
S1	RECALL	65.32	51.01	80.45	60.98	27.13	74.27	52.63	24.87	00.00	05.63	01.58	03.67	10.11	29.73	26.09
	PRECISION	35.62	41.24	63.17	72.67	72.46	69.38	69.52	66.20	-	53.33	66.67	50.00	56.25	88.00	54.55
S2	RECALL	64.16	48.03	87.76	42.06	15.10	45.69	35.85	06.06	00.00	00.00	00.81	00.00	00.00	00.00	00.00
	PRECISION	30.19	30.46	67.79	57.31	61.54	85.48	65.52	40.00	-	00.00	100.0	-	-	-	00.00

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

- Challenges open on CodaLab – 9 Sep

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



EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

Given a trimmed action segment:

$$(t_{\text{start}}, t_{\text{stop}})$$

classify the action within.

$$\hat{y}_{\text{verb}} = \text{open}$$

$$\hat{y}_{\text{noun}} = \text{oven}$$

$$\hat{y}_{\text{action}} = (\text{open}, \text{oven})$$

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

25 Teams

230 submissions

EPIC KITCHENS

with: Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, Michael Wray

Seen Kitchens (S1)

#	User	Entries	Date of Last Entry	Team Name	Top-1 Accuracy (%) Verb	Top-1 Noun	Top-1 Action	Top-5 Accuracy (%) Verb	Top-5 Noun	Top-5 Action	Precision (%) Verb	Precision Noun	Precision Action	Recall (%) Verb	Recall Noun	Recall Action
1	wasun	16	05/30/19	Baidu-UTS	69.80 (1)	52.27 (1)	41.37 (1)	90.95 (2)	76.71 (1)	63.59 (1)	63.55 (1)	46.86 (1)	25.13 (1)	46.94 (1)	49.17 (1)	26.39 (1)
2	TBN_Ensemble	2	05/30/19	Bristol-Oxford	66.10 (2)	47.88 (2)	36.66 (2)	91.28 (1)	72.80 (2)	58.62 (2)	60.73 (4)	44.89 (2)	24.01 (2)	46.81 (2)	43.88 (3)	22.92 (2)
3	deeptigp	9	10/30/18		64.14 (4)	47.65 (3)	35.75 (3)	87.64 (5)	70.66 (4)	54.65 (6)	43.64 (14)	40.52 (5)	18.95 (9)	38.31 (9)	45.29 (2)	21.13 (5)
4	sudhakaran	42	05/29/19	FBK-HUPBA	63.34 (5)	44.75 (6)	35.54 (4)	89.01 (4)	69.88 (5)	57.18 (4)	63.21 (2)	42.26 (4)	19.76 (7)	37.77 (10)	41.28 (5)	21.19 (4)
5	antoninofurnari	5	05/05/19	DMI-UNICT	58.99 (9)	45.00 (5)	35.14 (5)	86.70 (8)	69.08 (6)	57.62 (3)	52.23 (8)	40.06 (6)	19.40 (8)	42.12 (4)	39.32 (9)	20.28 (8)
6	TBN_Single_Model	1	03/22/19	Bristol-Oxford	64.74 (3)	46.03 (4)	34.80 (6)	90.69 (3)	71.33 (3)	56.64 (5)	55.66 (6)	43.65 (3)	22.06 (3)	45.55 (3)	42.30 (4)	21.30 (3)
7	zhe2325138	12	05/30/19	NTU CML Mira	61.65 (7)	43.63 (8)	30.55 (7)	87.09 (7)	68.65 (7)	40.11 (14)	48.63 (9)	39.62 (8)	16.92 (11)	33.41 (12)	40.57 (7)	16.68 (9)
8	pengx2019	2	05/31/19		57.49 (10)	43.00 (9)	30.50 (8)	86.45 (10)	66.57 (8)	48.27 (7)	45.92 (12)	39.34 (9)	20.18 (5)	40.89 (5)	40.76 (6)	20.87 (7)
9	largefish	7	05/05/19		57.47 (11)	43.00 (9)	30.50 (8)	86.08 (13)	66.51 (9)	48.20 (8)	45.91 (13)	39.34 (9)	20.18 (6)	40.88 (6)	40.76 (6)	20.87 (7)
10	mxtx0509	1	05/30/19		57.25 (12)	42.51 (10)	30.20 (9)	86.65 (9)	66.36 (10)	47.83 (9)	46.02 (11)	39.05 (10)	20.31 (4)	40.81 (7)	40.56 (8)	20.92 (6)
11	Casia-Airia	4	05/31/19	CA	63.29 (6)	44.03 (7)	29.18 (10)	86.24 (12)	65.99 (11)	46.94 (10)	56.35 (5)	39.95 (7)	12.86 (12)	34.83 (11)	38.76 (10)	10.53 (11)
12	zolfagha	8	05/29/19		60.90 (8)	38.19 (11)	28.05 (11)	87.52 (6)	63.19 (13)	47.83 (9)	54.85 (7)	35.48 (11)	17.29 (10)	38.46 (8)	33.95 (11)	15.63 (10)
13	Nour	47	05/26/19	RML-Ryerson University	52.75 (13)	35.19 (14)	24.15 (12)	86.26 (11)	61.72 (15)	44.96 (11)	42.80 (16)	29.32 (16)	7.19 (18)	31.44 (13)	30.67 (13)	10.06 (13)
14	EPIC_TSN_Fusion	1	09/04/18		48.23 (14)	36.71 (13)	20.54 (13)	84.09 (17)	62.32 (14)	39.79 (15)	47.26 (10)	35.42 (12)	11.57 (13)	22.33 (17)	30.53 (14)	9.78 (14)
15	masterchef	1	10/02/18	Inria / Facebook	43.51 (18)	32.94 (15)	20.19 (14)	84.38 (16)	61.66 (16)	43.57 (12)	28.42 (20)	27.99 (18)	7.62 (17)	24.18 (15)	26.83 (16)	8.85 (17)
16	EPIC_TSN_RGB	1	09/04/18		45.68 (16)	36.80 (12)	19.86 (15)	85.56 (14)	64.19 (12)	41.89 (13)	61.64 (3)	34.32 (13)	11.02 (15)	23.81 (16)	31.62 (12)	9.76 (15)
17	nachwa	34	05/23/19	UGA	47.41 (15)	28.31 (19)	19.76 (16)	81.33 (19)	53.77 (19)	36.98 (17)	31.20 (18)	21.21 (20)	9.83 (16)	20.43 (18)	22.48 (18)	10.23 (12)
18	wsy2019	12	05/30/19		44.74 (17)	31.81 (16)	17.99 (17)	84.39 (15)	58.08 (17)	37.06 (16)	43.33 (15)	31.68 (14)	11.18 (14)	24.47 (14)	26.94 (15)	9.11 (16)
19	EPIC_2SCNN_RGB	2	09/06/18		40.44 (21)	30.46 (17)	13.67 (18)	83.04 (18)	57.05 (18)	33.25 (18)	34.74 (17)	28.23 (17)	6.66 (19)	15.90 (20)	23.23 (17)	5.47 (18)
20	EPIC_2SCNN_Fusion	1	09/06/18		42.16 (20)	29.14 (18)	13.23 (19)	80.58 (20)	53.70 (20)	30.36 (19)	29.39 (19)	30.73 (15)	5.92 (20)	14.83 (22)	21.10 (19)	4.93 (19)

Unseen Kitchens (S2)

#	User	Entries	Date of Last Entry	Team Name	Top-1 Accuracy (%) Verb	Top-1 Noun	Top-1 Action	Top-5 Accuracy (%) Verb	Top-5 Noun	Top-5 Action	Precision (%) Verb	Precision Noun	Precision Action	Recall (%) Verb	Recall Noun	Recall Action
1	wasun	16	05/30/19	Baidu-UTS	59.68 (1)	34.14 (1)	25.06 (1)	82.69 (1)	62.38 (1)	45.95 (1)	37.20 (1)	29.14 (1)	15.44 (2)	29.81 (1)	30.48 (1)	18.67 (1)
2	deeptigp	9	10/30/18		55.24 (2)	33.87 (2)	23.93 (2)	80.23 (3)	58.25 (2)	40.15 (3)	25.71 (10)	28.19 (2)	15.72 (1)	25.69 (4)	29.51 (2)	17.06 (2)
3	TBN_Ensemble	2	05/30/19	Bristol-Oxford	54.46 (3)	30.39 (3)	21.99 (3)	81.22 (2)	55.68 (3)	40.59 (2)	32.56 (2)	21.67 (5)	9.83 (6)	27.60 (3)	25.58 (3)	13.52 (5)
4	antoninofurnari	5	05/05/19	DMI-UNICT	47.35 (9)	28.64 (4)	21.37 (4)	73.75 (13)	51.01 (6)	39.47 (4)	26.88 (9)	22.09 (4)	10.53 (4)	22.12 (8)	23.31 (5)	13.98 (4)
5	sudhakaran	42	05/29/19	FBK-HUPBA	49.37 (7)	27.11 (7)	20.25 (5)	77.50 (7)	51.96 (5)	37.56 (5)	31.09 (5)	21.06 (7)	9.18 (8)	18.73 (12)	21.88 (8)	14.23 (3)
6	TBN_Single_Model	1	03/22/19	Bristol-Oxford	52.68 (6)	27.86 (6)	19.05 (6)	79.92 (4)	53.77 (4)	36.53 (6)	31.43 (3)	21.47 (6)	11.99 (3)	28.21 (2)	23.53 (4)	12.68 (6)
7	Casia-Airia	4	05/31/19	CA	53.36 (4)	28.37 (5)	18.47 (7)	77.47 (8)	50.15 (7)	33.63 (8)	31.23 (4)	22.12 (3)	7.24 (11)	21.29 (10)	22.56 (6)	9.40 (11)
8	pengx2019	2	05/31/19		44.66 (12)	25.91 (9)	16.56 (8)	73.75 (13)	49.74 (9)	29.77 (11)	28.58 (7)	18.88 (10)	10.37 (5)	23.09 (5)	21.91 (7)	11.67 (7)
9	largefish	7	05/05/19		44.62 (13)	25.91 (9)	16.56 (8)	73.40 (14)	49.85 (8)	29.74 (12)	28.53 (8)	18.88 (9)	10.37 (5)	23.08 (6)	21.91 (7)	11.67 (7)
10	mxtx0509	1	05/30/19		44.93 (11)	26.05 (8)	16.42 (9)	73.92 (12)	49.23 (12)	29.22 (13)	29.03 (6)	18.98 (8)	9.49 (7)	22.89 (7)	21.83 (9)	11.24 (8)
11	zhe2325138	12	05/30/19	NTU CML Mira	52.78 (5)	24.62 (11)	16.35 (10)	79.72 (5)	49.61 (10)	22.81 (16)	23.31 (12)	17.91 (12)	9.00 (9)	22.02 (9)	19.29 (11)	9.62 (10)
12	Nour	47	05/26/19	RML-Ryerson University	45.51 (10)	24.41 (12)	16.08 (11)	77.47 (8)	50.15 (7)	34.14 (7)	21.70 (14)	17.12 (13)	5.16 (15)	18.33 (13)	18.82 (12)	9.27 (12)
13	zolfagha	8	05/29/19		47.59 (8)	24.68 (10)	15.47 (12)	77.53 (6)	49.30 (11)	31.31 (10)	24.98 (11)	18.44 (11)	8.91 (10)	20.18 (11)	21.18 (10)	10.53 (9)
14	masterchef	1	10/02/18	Inria / Facebook	39.30 (16)	22.43 (14)	14.10 (13)	76.41 (9)	47.35 (13)	32.43 (9)	20.42 (15)	15.96 (14)	4.83 (16)	16.95 (14)	17.72 (13)	8.46 (13)
15	EPIC_TSN_Fusion	1	09/04/18		39.40 (15)	22.70 (13)	10.89 (14)	74.29 (11)	45.72 (14)	25.26 (15)	22.54 (13)	15.33 (15)	6.21 (12)	13.06 (16)	17.52 (14)	6.49 (14)
16	EPIC_TSN_RGB	1	09/04/18		34.89 (19)	21.82 (15)	10.11 (15)	74.56 (10)	45.34 (15)	25.33 (14)	19.48 (17)	14.67 (17)	5.32 (14)	11.22 (18)	17.24 (15)	6.34 (15)
17	nachwa	34	05/23/19	UGA	34.35 (20)	17.48 (18)	9.08 (16)	69.24 (19)	37.56 (18)	19.46 (19)	15.09 (21)	10.71 (20)	3.68 (17)	11.00 (19)	12.55 (17)	4.77 (16)
18	wsy2019	12	05/30/19		33.56 (21)	16.56 (19)	8.09 (17)	72.38 (16)	37.08 (19)	18.61 (21)	16.89 (19)	11.12 (19)	5.73 (13)	10.54 (20)	11.80 (19)	4.67 (17)
19	EPIC_2SCNN_Fusion	1	09/06/18		36.16 (18)	18.03 (16)	7.31 (18)	71.97 (17)	38.41 (17)	19.49 (18)	18.11 (18)	15.31 (16)	3.19 (19)	10.52 (21)	12.55 (16)	3.00 (19)
20	EPIC_2SCNN_RGB	2	09/06/18		33.12 (22)	17.58 (17)	6.79 (19)	73.23 (15)	40.46 (16)	20.42 (17)	16.06 (20)	11.97 (18)	3.39 (18)	9.44 (23)	12.53 (18)	3.01 (18)

EPIC KITCHENS

with: Hazel Doughty
Giovanni Maria Farinella
Sanja Fidler
Antonino Furnari
Evangelos Kazakos
Davide Moltisanti
Jonathan Munro
Toby Perrett
Will Price
Michael Wray

D Damen et al (2018). Scaling Egocentric Vision: The EPIC-KITCHENS Dataset. Arxiv. https://epic-kitchens.github.io

Interested in More?

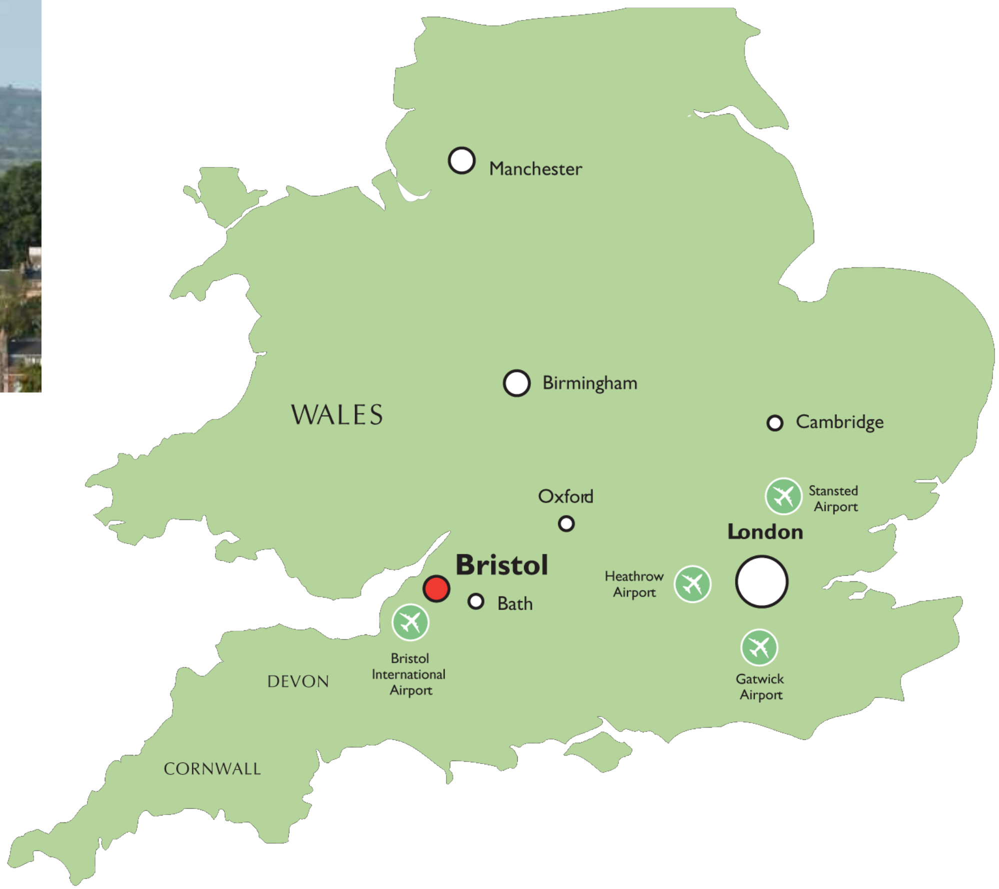
- Egocentric Perception, Interaction and Computing (EPIC) Workshop Series
  - ECCV 2016 (Amsterdam)
  - ICCV 2017 (Venice)
  - ECCV 2018 (Munich)
  - CVPR 2019 (Long Beach)
  - ICCV 2019 (Seoul) – 5th edition!
    - Paper deadline: 26 July 2019
    - Abstract submission: 5 August 2019 (ongoing work)

Interested in More?

- Subscribe to the newly introduced mailing list: epic-community@bristol.ac.uk

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

Bristol and University of Bristol



Bristol and University of Bristol

Thank you…

For further info, datasets, code, publications…

http://www.cs.bris.ac.uk/~damen

@dimadamen

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