Human-Centric Object Interactions:

A Fine-Grained Perspective from Egocentric Videos

University of BRISTOL

Dima Damen

January 13, 2021

Fine(r)-grained?

Fine(r)-grained?

- Coarse-grained: Cooking
- Fine-grained: add garlic
- Fine(r)-grained: smash garlic
  - When was the garlic smashed?
  - How was the garlic smashed?
  - Why was the garlic smashed?
  - How skilled was this person in smashing garlic?
  - Has garlic now been fully smashed?
- What information to make these decisions
  - Change in appearance
  - Motion
  - Audio
  - ??

Natural Object Interactions…

Scaling and Rescaling Egocentric Vision:
The EPIC-KITCHENS Dataset

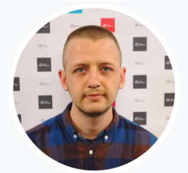

Dima Damen


Hazel Doughty


Giovanni M. Farinella


Sanja Fidler


Antonino Furnari


Evangelos Kazakos


Jian Ma


Davide Moltisanti

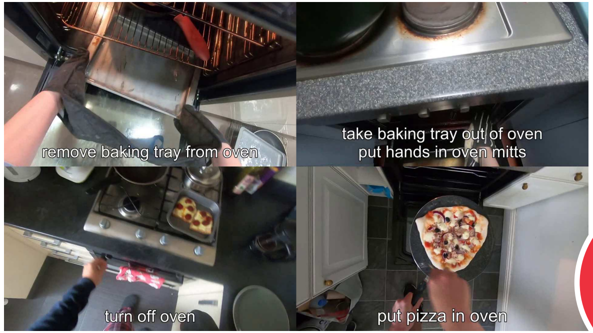

Jonathan Munro

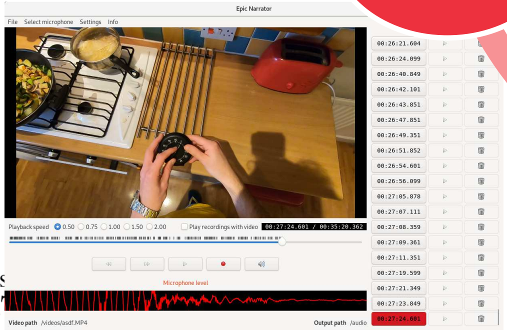

Toby Perrett


Will Price


Michael Wray

Scaling and Rescaling Egocentric Vision



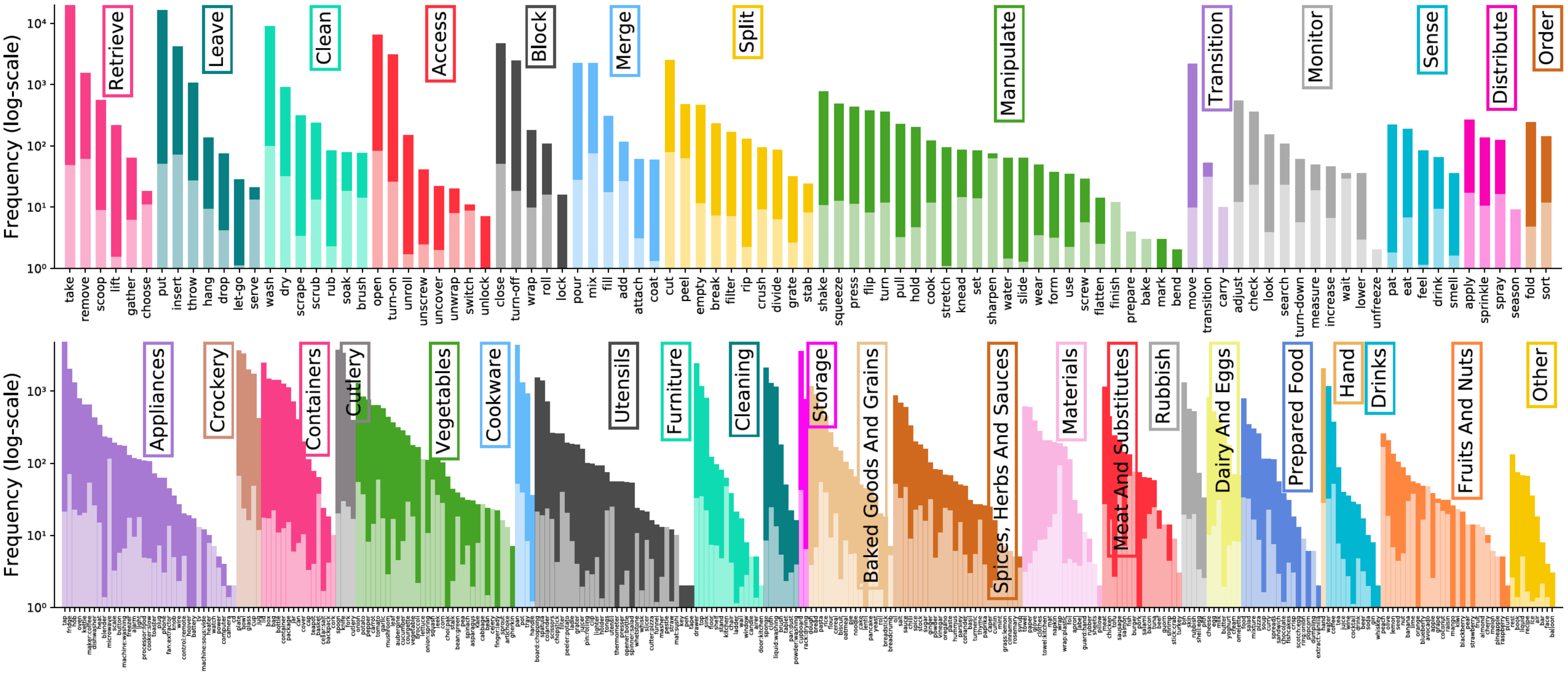


EPIC-KITCHENS-55

37 Participants

University of BRISTOL

Annotations Statistics



Open Challenges

Five currently open challenges:

- Action Recognition
- Action Detection
- Action Anticipation
- Unsupervised Domain Adaptation for Recognition
- Multi-Instance Retrieval

Action Recognition Challenge

Action Recognition Challenge

Given a trimmed action segment:

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

classify the action within.

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

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

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

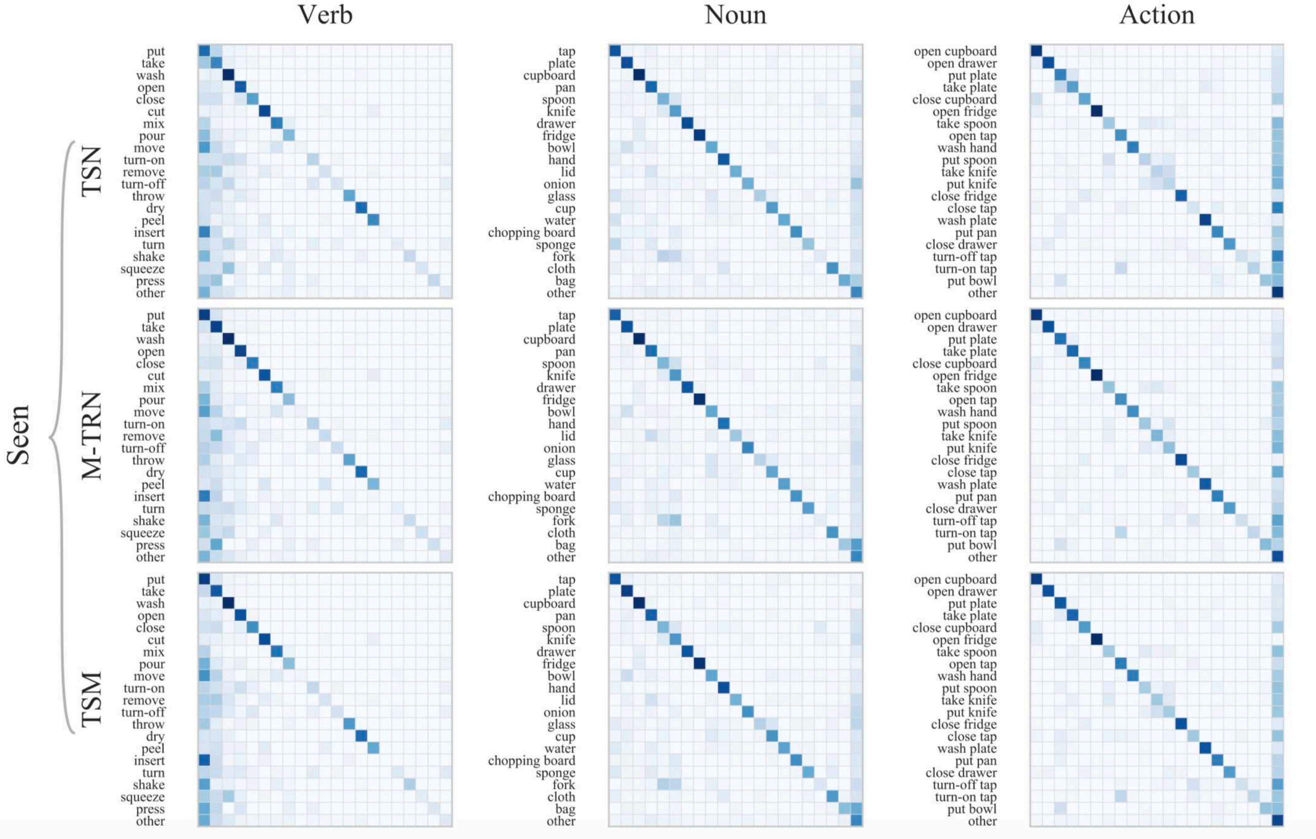
University of BRISTOL

Action Recognition Challenge

Seen Kitchens (S1)																
#	User	Entries	Date of Last Entry	Team Name	Top-1 Accuracy (%)			Top-5 Accuracy (%)			Precision (%)			Recall (%)		
					Verb	Noun	Action	Verb	Noun	Action	Verb	Noun	Action	Verb	Noun	Action
1	wasun	14	05/28/20	UTS-Baidu	70.41 (1)	52.85 (1)	42.57 (1)	90.78 (4)	76.62 (2)	63.55 (2)	60.44 (4)	47.11 (1)	24.94 (3)	45.82 (4)	50.02 (1)	26.93 (2)
2	action_banks	18	05/29/20	NUS_CVML	66.56 (6)	49.60 (4)	41.59 (2)	90.10 (5)	77.03 (1)	64.11 (1)	59.43 (7)	45.62 (3)	25.37 (1)	41.65 (8)	46.25 (4)	26.98 (1)
3	Sudhakaran	50	05/29/20	FBK_HuPBA	68.68 (3)	49.35 (5)	40.00 (3)	90.97 (3)	72.45 (5)	60.23 (4)	60.63 (3)	45.45 (4)	21.82 (6)	47.19 (2)	45.84 (5)	24.34 (4)
4	tnet	34	05/27/20	SAIC_Cambridge	69.43 (2)	49.71 (3)	40.00 (3)	91.23 (2)	73.18 (3)	60.53 (3)	60.01 (5)	45.74 (2)	24.95 (2)	47.40 (1)	46.78 (3)	25.27 (3)
5	aptx4869lm	12	01/30/20	GT-WISC-MPI	68.51 (4)	49.96 (2)	38.75 (4)	89.33 (8)	72.30 (6)	58.99 (5)	51.04 (16)	44.00 (6)	23.70 (5)	43.70 (7)	47.32 (2)	23.92 (5)
6	weiyaowang	14	05/28/20		66.67 (5)	48.48 (6)	37.12 (5)	88.90 (9)	71.36 (7)	56.21 (8)	51.86 (14)	41.26 (7)	20.97 (7)	44.33 (6)	44.92 (6)	21.48 (8)
7	TBN_Ensemble	1	07/20/19	Bristol-Oxford	66.10 (7)	47.88 (7)	36.66 (6)	91.28 (1)	72.80 (4)	58.62 (6)	60.73 (2)	44.89 (5)	24.01 (4)	46.81 (3)	43.88 (7)	22.92 (6)
8	cvg_uni_bonn	21	05/27/20	CVG Lab Uni Bonn	62.86 (8)	43.44 (10)	34.53 (7)	89.64 (6)	69.24 (8)	56.73 (7)	52.82 (13)	38.81 (11)	19.21 (10)	44.72 (5)	39.50 (10)	21.80 (7)
9	antoninofurnari	1	07/19/19		56.93 (16)	43.05 (11)	33.06 (8)	85.68 (20)	67.12 (11)	55.32 (9)	50.42 (17)	39.84 (9)	18.91 (11)	37.82 (14)	38.11 (11)	19.12 (11)
10	Wenda	12	04/25/20	Wenda Go!	61.10 (12)	43.73 (8)	31.54 (9)	89.45 (7)	68.45 (10)	52.62 (10)	55.79 (10)	41.24 (8)	20.67 (8)	40.25 (10)	40.49 (9)	19.33 (10)
11	EPIC_TSM_FUSION	1	03/30/20		62.37	41.88	29.90	88.55	66.43	49.81	59.51	39.50	18.38	34.44	36.04	15.80

Evaluating Action Recognition Models

with: Will Price



W Price, D Damen (2019). An Evaluation of Action Recognition Models on EPIC-Kitchens. Arxiv

University of BRISTOL

Evaluating Action Recognition Models

with: Will Price

	GFLOP/s		Params (M)	
Model	RGB	Flow	RGB	Flow
TSN	33.12	35.33	24.48	24.51
TRN	33.12	35.32	25.33	25.35
M-TRN	33.12	35.33	27.18	27.21
TSM	33.12	35.33	24.48	24.51

Table 3: Model parameter and FLOP/s count using a ResNet-50 backbone with 8 segments for a single video.

Models Released

W Price, D Damen (2019). An Evaluation of Action Recognition Models on EPIC-Kitchens. Arxiv

More?

http://epic-kitchens.github.io

EPIC-KITCHENS-100 2021 CHALLENGES

Challenge and Leaderboard Details with links to Codalab Leaderboards

For Challenge Results and winners on EPIC-KITCHENS-55, go to: Challenge 2020 Details.
Note that these are NEW leaderboards, and results are not directly comparable to last year's results.

EPIC-Kitchens 2021 Challenges - Dates

Aug 23rd, 2020	EPIC-Kitchens Challenges 2021 Launched alongisde EPIC@ECCV Workshop
May 28, 2021	Server Submission Deadline at 23:59:59 GMT
Jun 4, 2021	Deadline for Submission of Technical Reports
TBC	Results announcement dates will be confirmed later

Challenges Guidelines

The five challenges below and their test sets and evaluation servers are available via CodaLab. The leaderboards will decide the winners for each individual challenge. For each challenge, the CodaLab server page details submission format and evaluation metrics.

To **enter any of the five competitions**, you need to register an account for that challenge using a valid institute (university/company) email address. A single registration per research team is allowed. We perform a manual check for each submission, and expect to accept registrations within 2 working days.

For all challenges the maximum submissions per day is limited to 1, and the overall maximum number of submissions per team is limited to 50 overall, submitted once a day. This includes any failed submissions due to formats - please do not contact us to ask for increasing this limit.

To **submit** your results, follow the JSON submission format, upload your results and give time for the evaluation to complete (in the order of several minutes). **Note our new rules on declaring the supervision level, given our proposed scale, for each submission.** After the evaluation is complete, the results automatically appear on the public leaderboards but you are allowed to withdraw these at any point in time.

To **participate** in the challenge, you need to have your results on the public leaderboard, along with an informative team name (that represents your institute or the collection of institutes participating in the work), as well as brief information on your method. You are also required to submit a report (details TBC).

Make the most of the starter packs available with the challenges, and should you have any questions, please use our info email uob-epic-kitchens@bristol.ac.uk

University of BRISTOL

EPIC KITCHENS

ABOUT STATS DOWNLOADS CHALLENGES TEAM

NEWS

- **1st of July 2020: EPIC-KITCHENS-100 is now Released! Watch release webinar recording**
- **Watch the dataset's trailer and video demonstration on YouTube**

What is EPIC-KITCHENS-100?

The *extended* **largest dataset in first-person (egocentric) vision**; multi-faceted, audio-visual, **non-scripted** recordings in native environments - i.e. the wearers' homes, capturing all daily activities in the kitchen over multiple days. Annotations are collected using a novel 'Pause-and-Talk' narration interface.

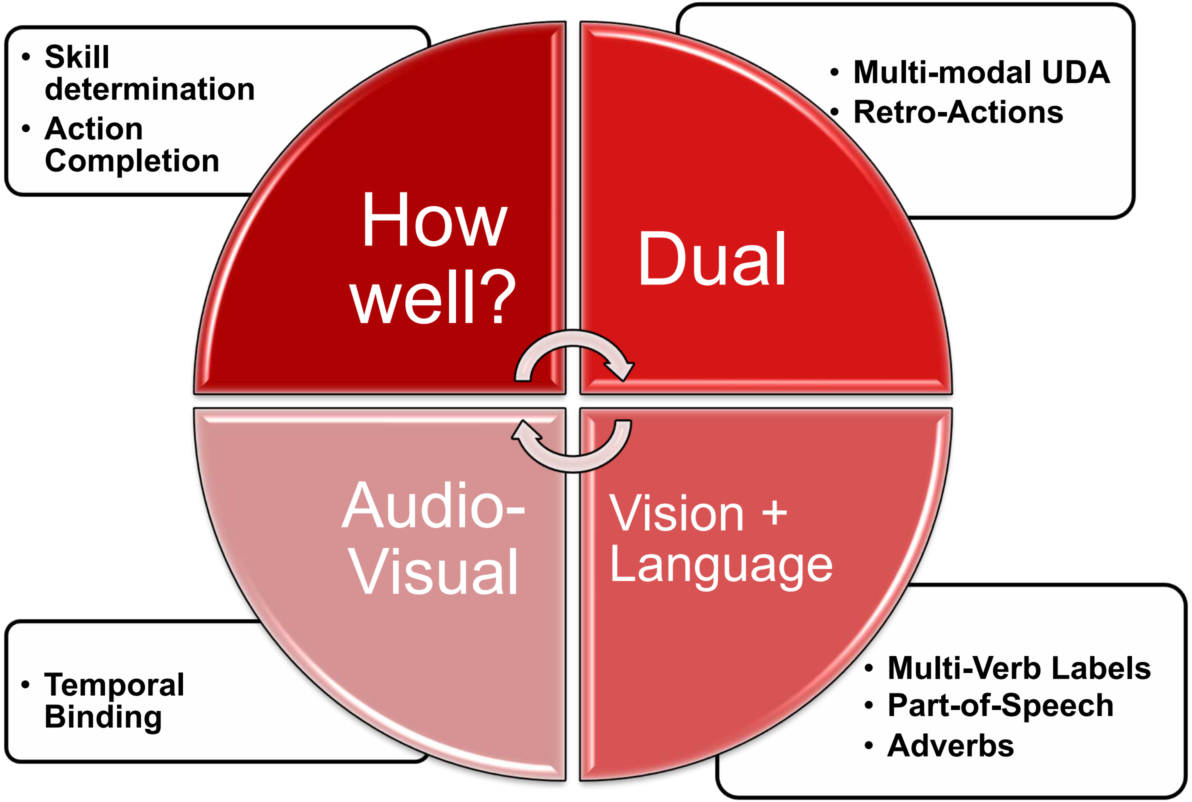
Characteristics

- 45 kitchens - 4 cities
- Head-mounted camera
- **100** hours of recording - Full HD
- 20M frames
- Multi-language narrations
- 90K action segments
- 20K unique narrations
- 97 verb classes, 300 noun classes
- 6 challenges

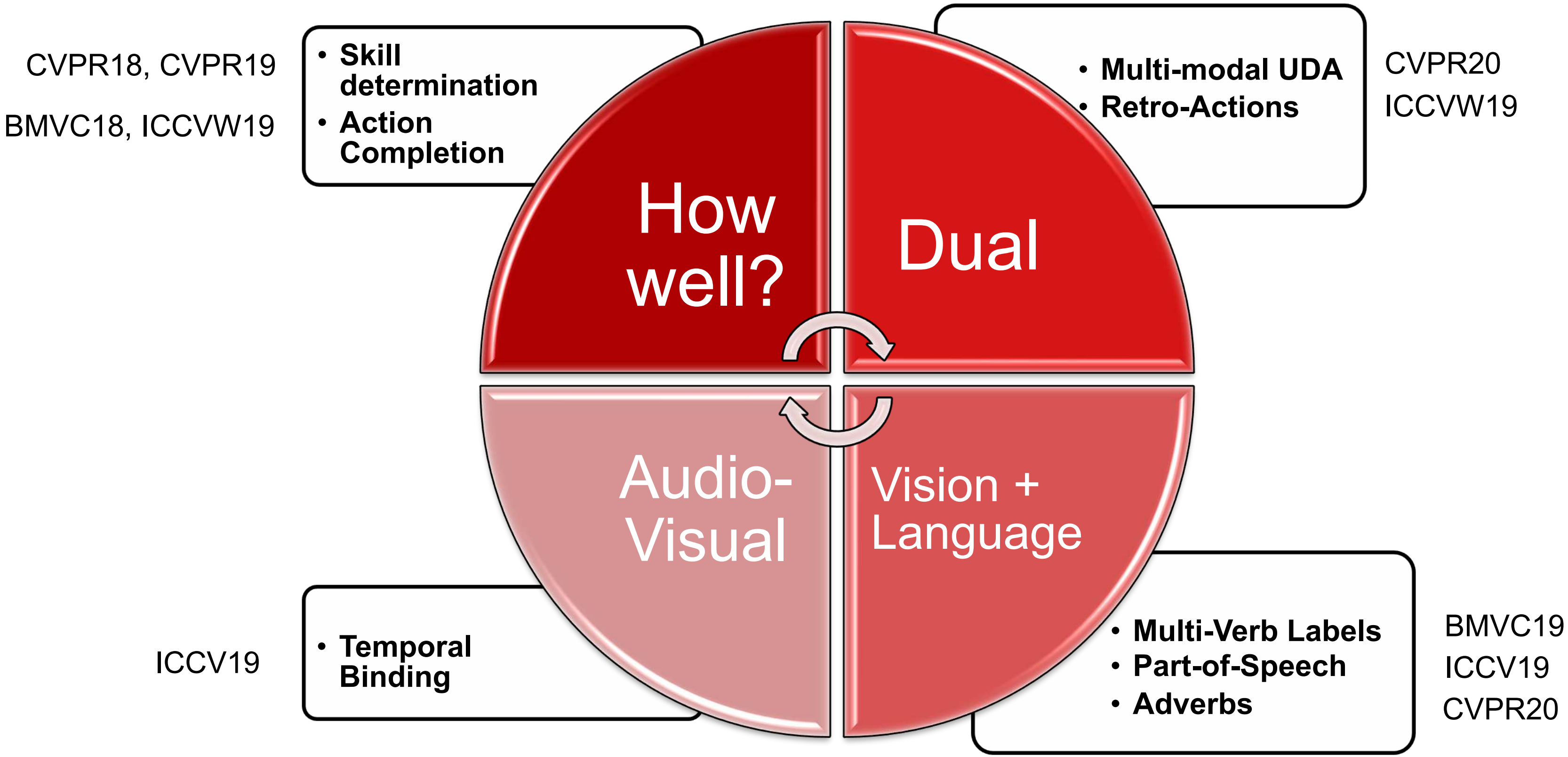
Previous versions...

- The previous version of the dataset (55 hours) was released in April 2018
- Refer to EPIC-KITCHENS-55 for details
- 2020 Challenges: Results, Tech Report
- 2019 Challenges: Results, Tech Report
- EPIC-KITCHENS-55 leaderboards remain open until the end of 2020

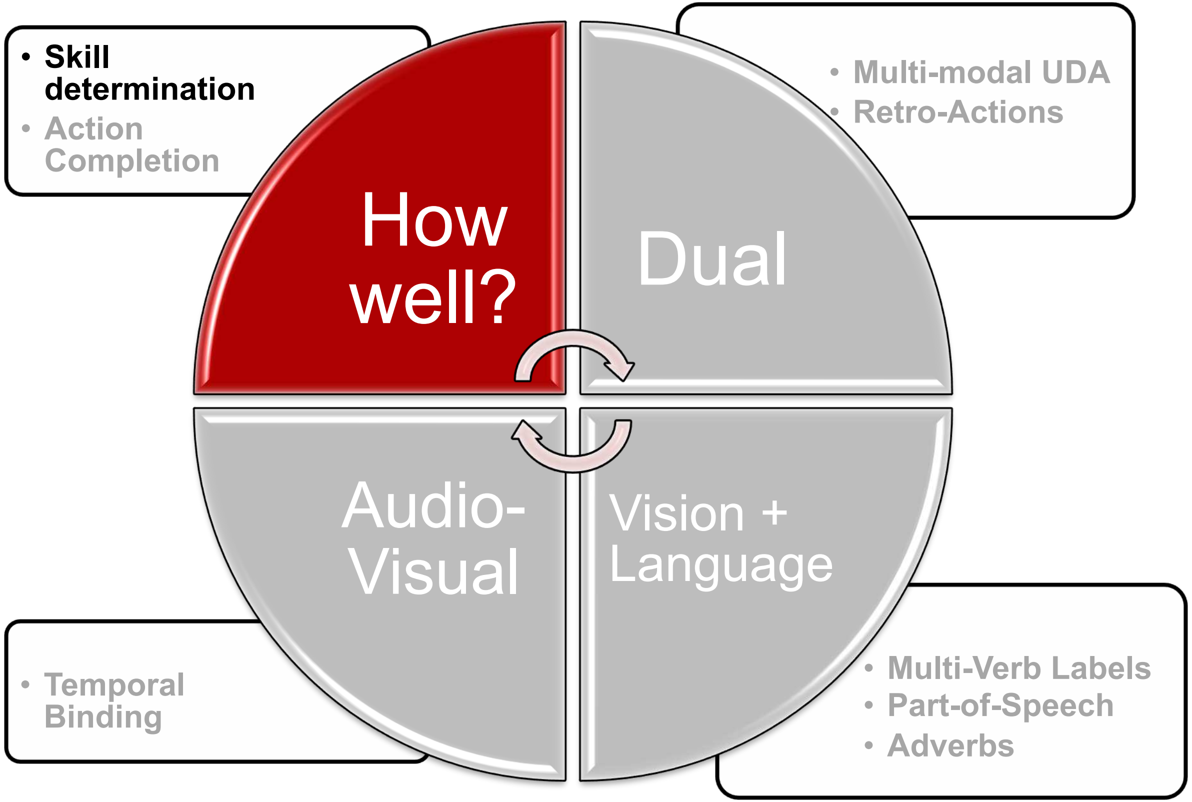
Fine(r)-grained?



Fine(r)-grained?



Fine(r)-grained?



Skill determination in video

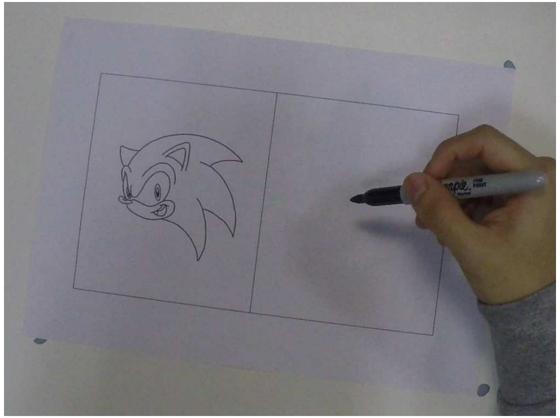
with: Hazel Doughty
Walterio Mayol-Cuevas

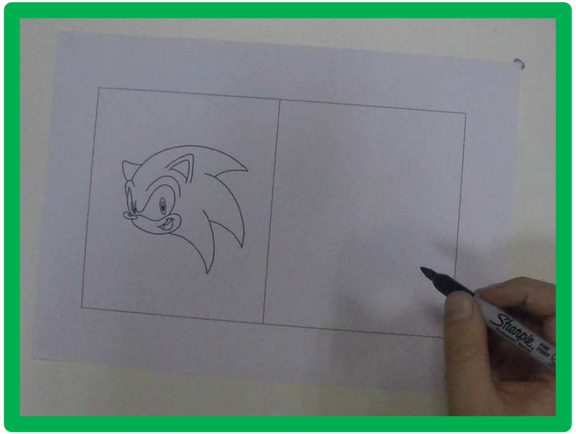
Assess relative skill for a collection of video sequences, applicable to a variety of tasks.

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

with: Hazel Doughty
Walterio Mayol-Cuevas

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

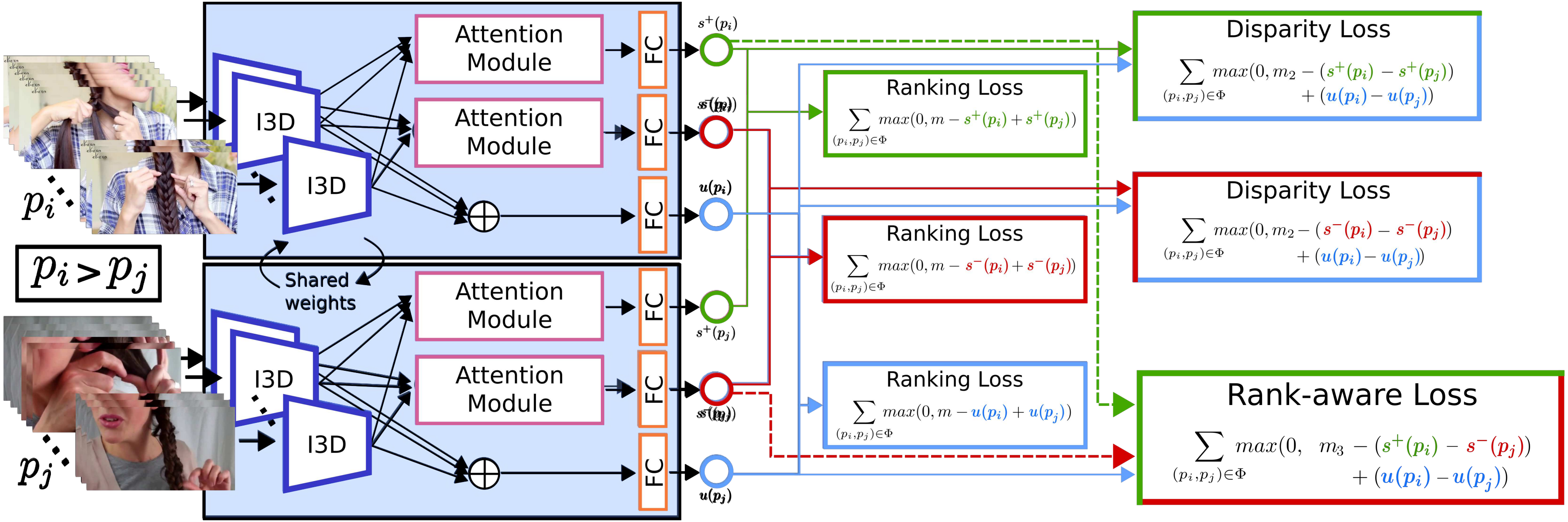




H Doughty, W Mayol-Cuevas, D Damen (2019). The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos. *Computer Vision and Pattern Recognition (CVPR)*

Skill determination in video

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)*

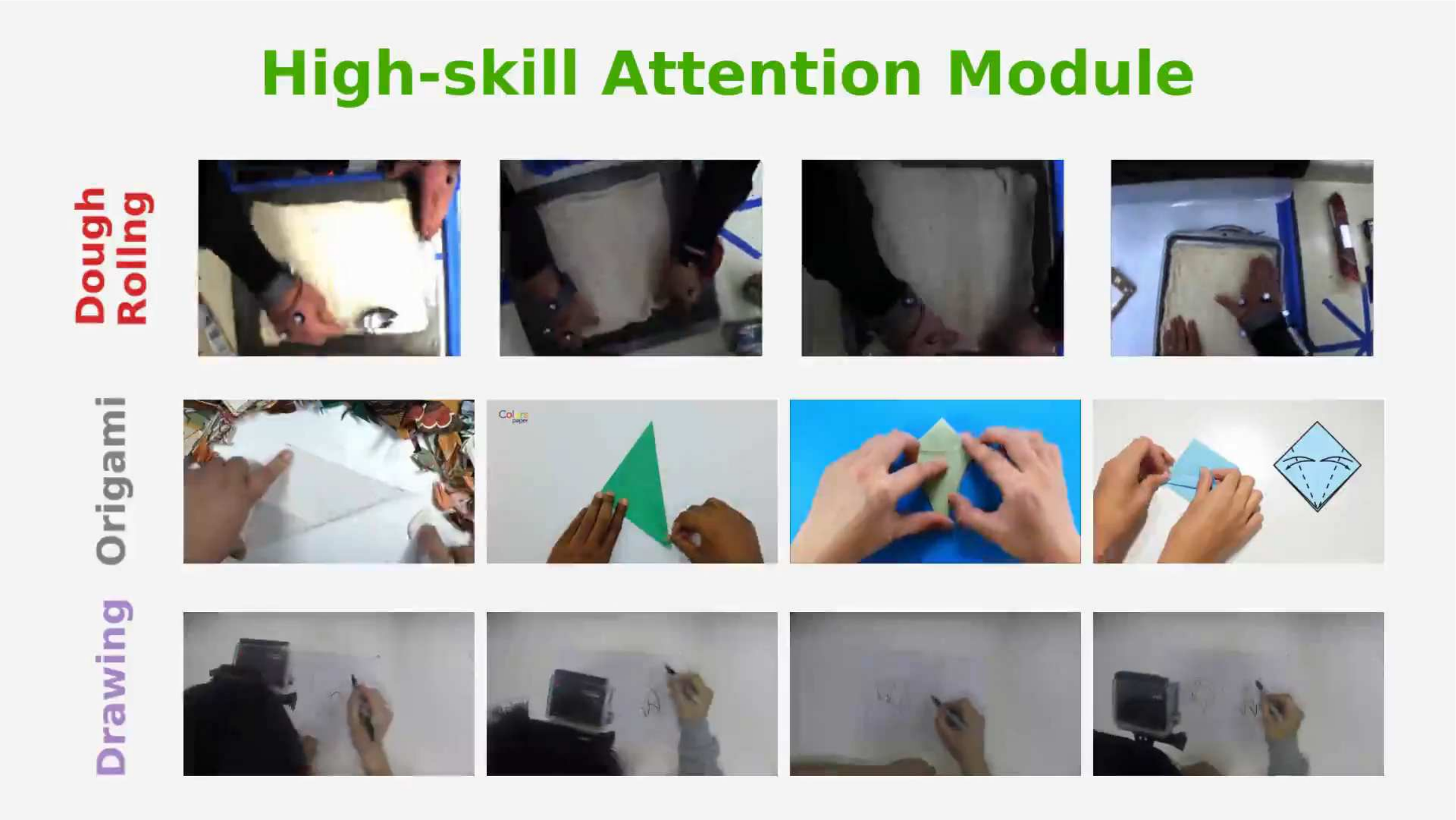
Skill determination in video

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)*

Skill determination in video

with: Hazel Doughty
Walterio Mayol-Cuevas



University of BRISTOL

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)*

Skill determination in video

with: Hazel Doughty
Walterio Mayol-Cuevas

Computer Vision and Pattern Recognition (CVPR) 2019

The Pros and Cons: Rank-aware Temporal Attention for Skill Determination in Long Videos

Hazel Doughty Walterio Mayol-Cuevas Dima Damen

University of Bristol

ABSTRACT VIDEO DOWNLOADS BIBTEX RELATED

Abstract

We present a new model to determine relative skill from long videos, through learnable temporal attention modules. Skill determination is formulated as a ranking problem, making it suitable for common and generic tasks. However, for long videos, parts of the video are irrelevant for assessing skill, and there may be variability in the skill exhibited throughout a video. We therefore propose a method which assesses the relative overall level of skill in a long video by attending to its skill-relevant parts.

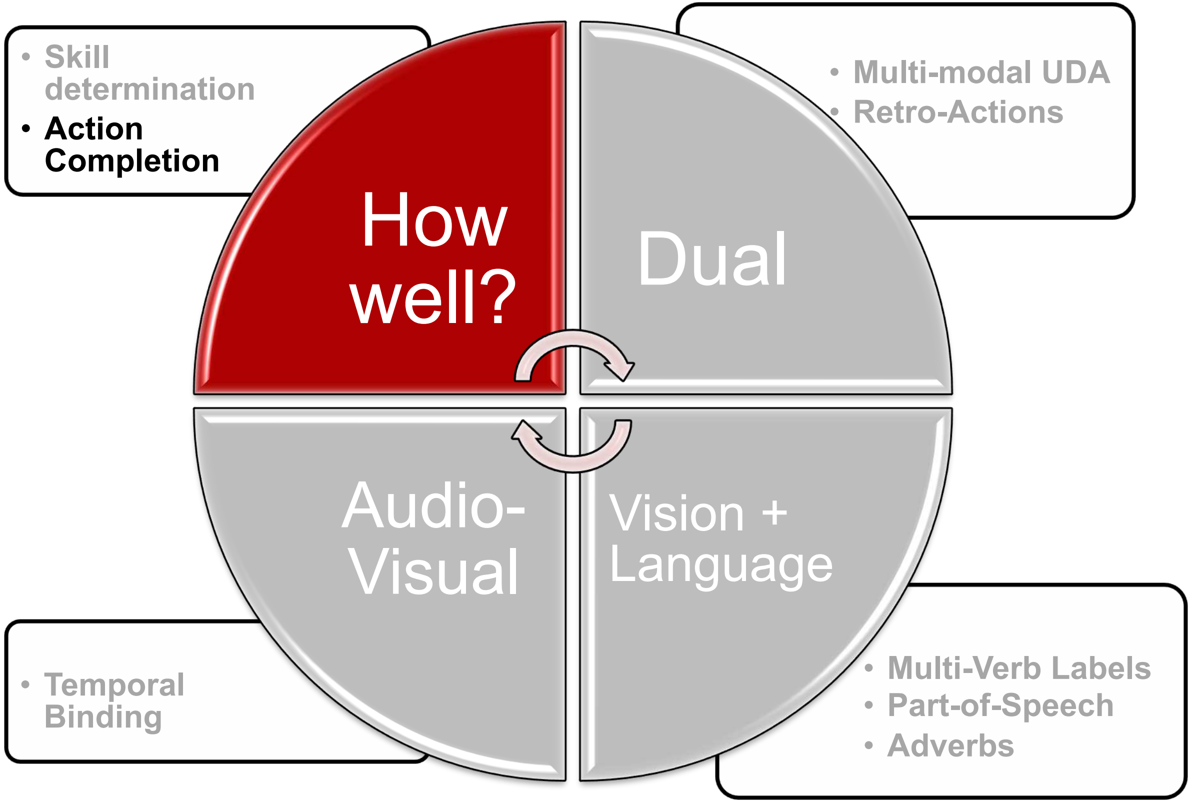
Our approach trains temporal attention modules, learned with only video-level supervision, using a novel rank-aware loss function. In addition to attending to task-relevant video parts, our proposed loss jointly trains two attention modules to separately attend to video parts which are indicative of higher (pros) and lower (cons) skill. We evaluate our approach on the EPIC-Skills dataset and additionally annotate a larger dataset from YouTube videos for skill determination with five previously unexplored tasks. Our method outperforms previous approaches and classic softmax attention on both datasets by over 4% pairwise accuracy, and as much as 12% on individual tasks. We also demonstrate our model's ability to attend to

Downloads

- Paper [PDF] [ArXiv]
- Supplementary [Video]
- Code and data [GitHub - Available Now]

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)*

Fine(r)-grained?



Action Completion Detection

with: Farnoosh Heidarivincheh
Majid Mirmehdi

Pre-V

V_R^T

C-C

R-R

R-C

C-R

Ground truth

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

Action Completion Detection

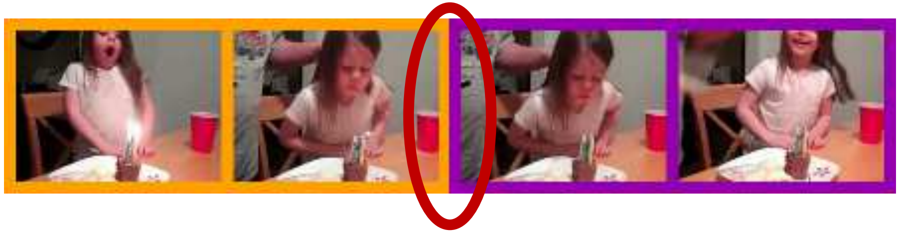
with: Farnoosh Heidarivincheh
Majid Mirmehdi

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

Action Completion Detection

with: Farnoosh Heidarivincheh
Majid Mirmehdi

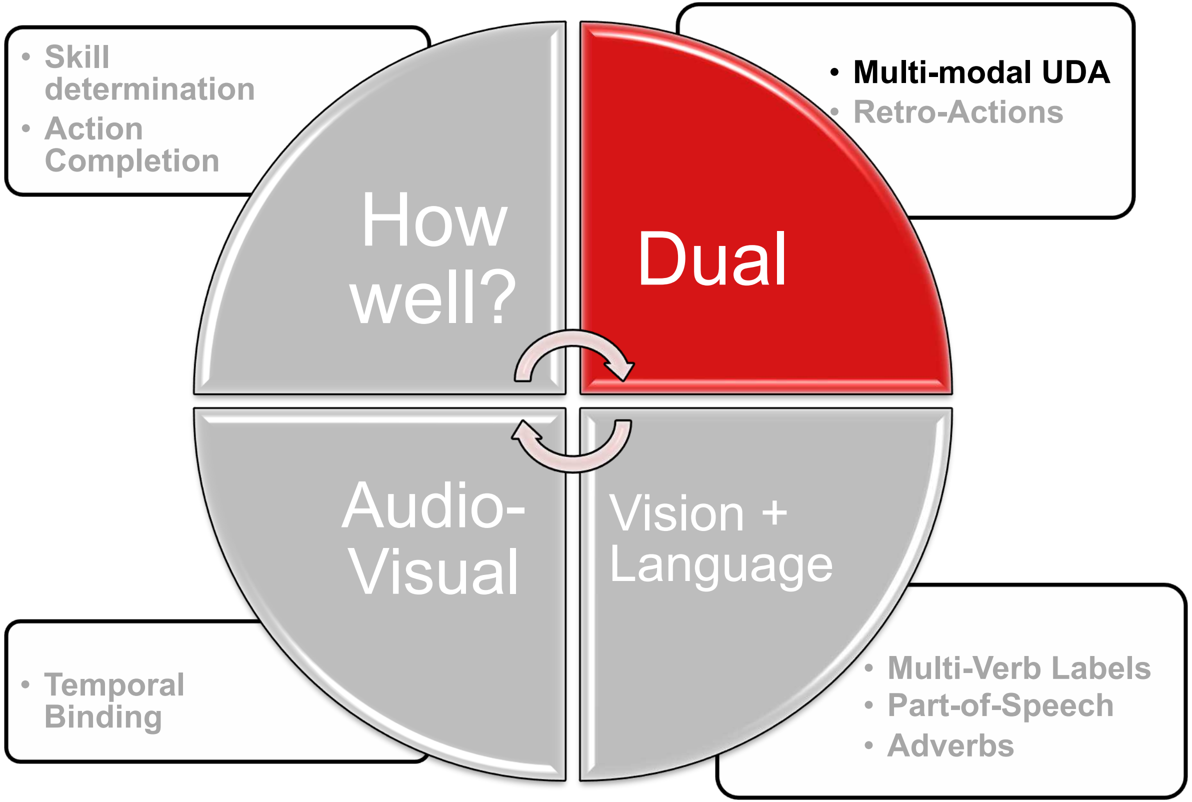
Frame-level labels: annotations are expensive, subjective and noisy.



We detect completion using only **weak labels** during training.

F Heidarivincheh, M Mirmehdi, D Damen (2019). Weakly-Supervised Completion Moment Detection using Temporal Attention. ICCV Workshop on Human Behaviour Understanding

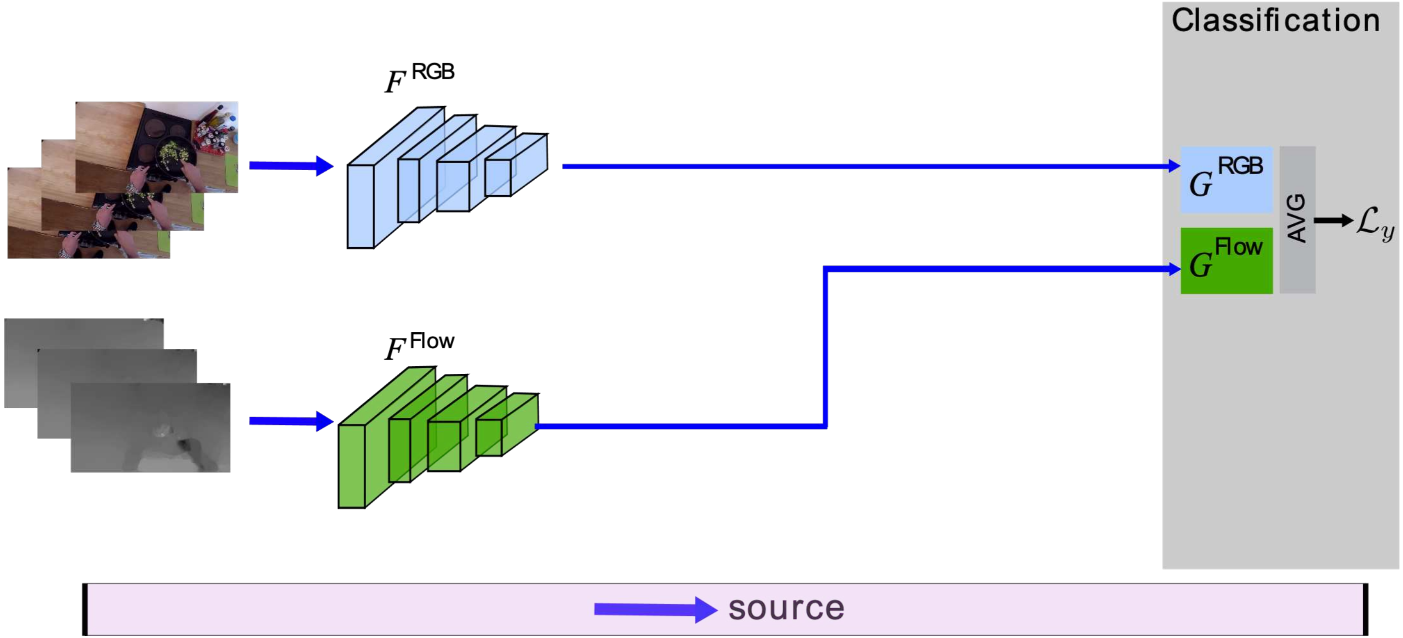
Fine(r)-grained?



University of BRISTOL

Multi-modal UDA

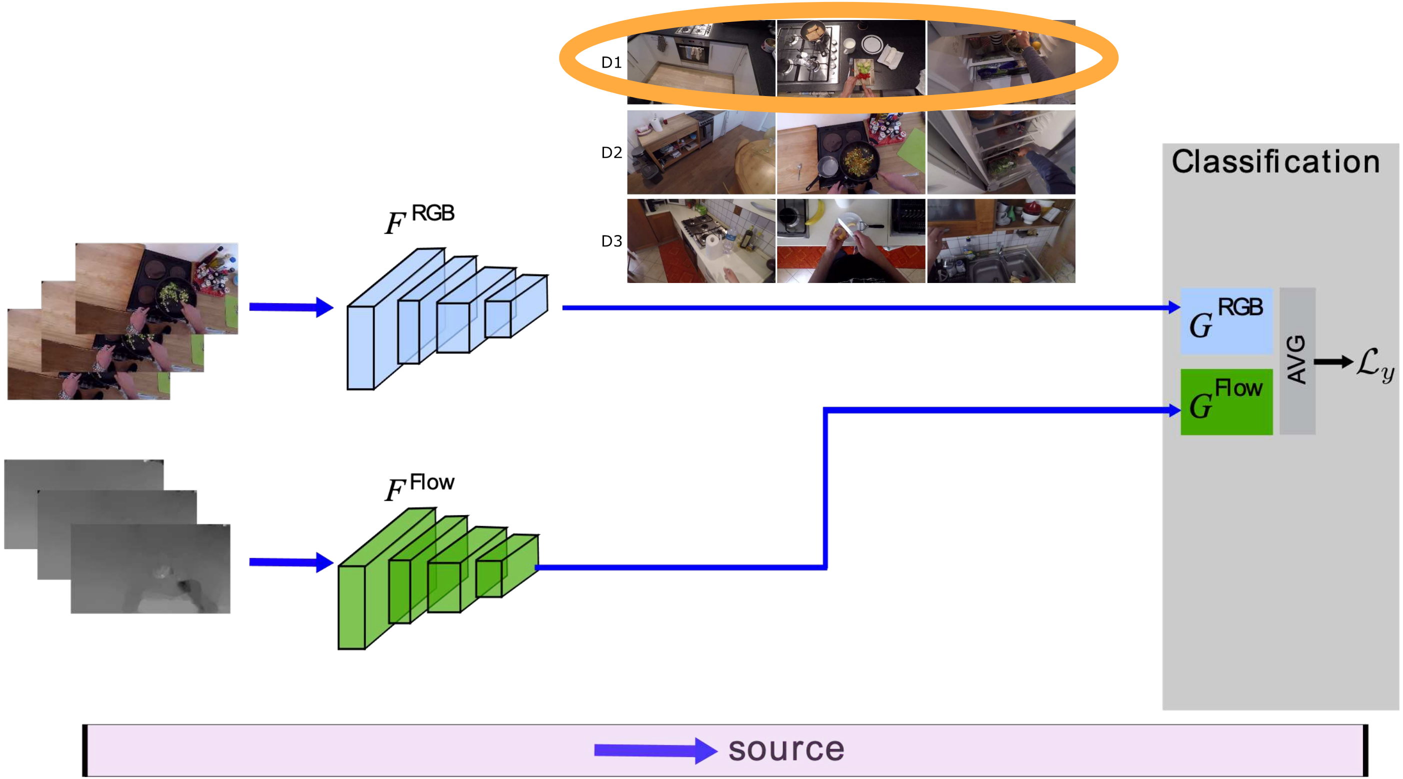
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

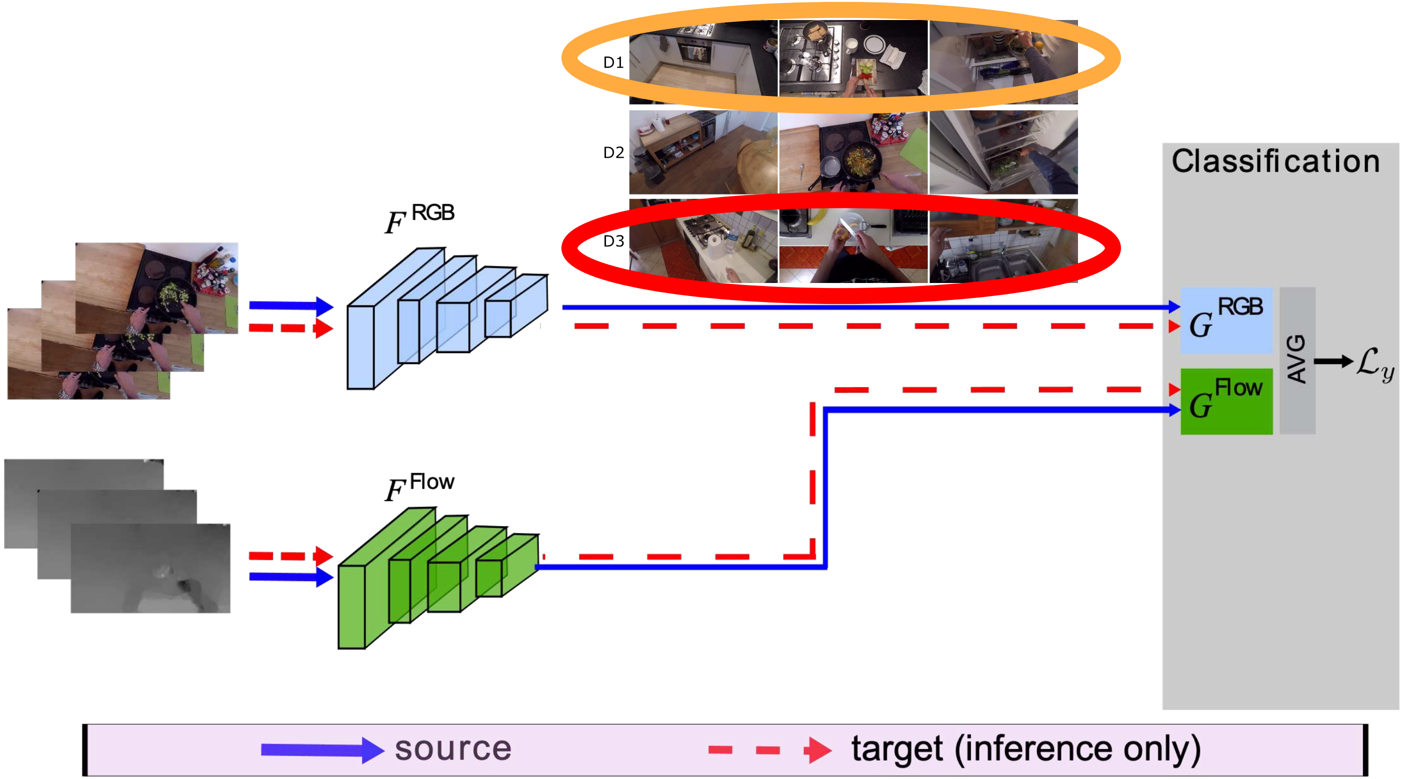
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

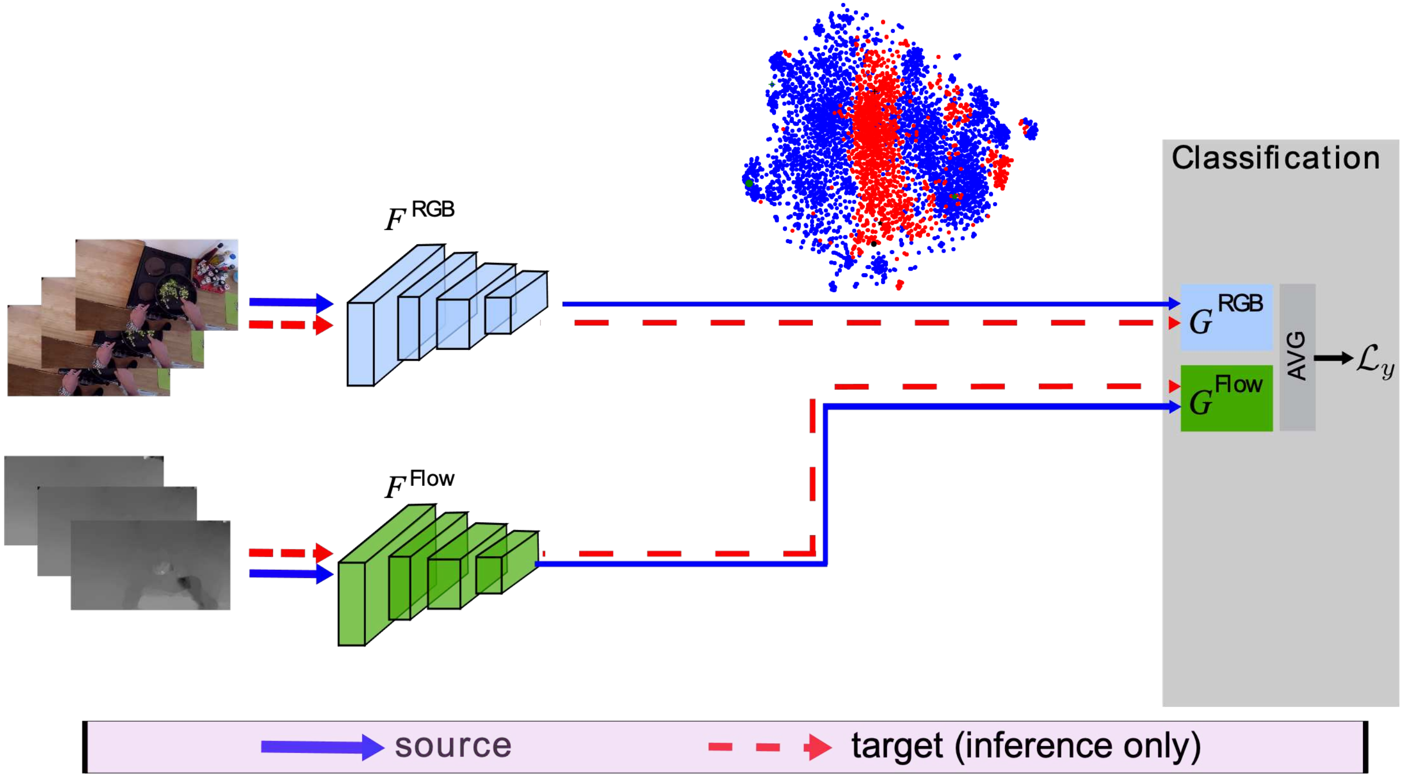
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

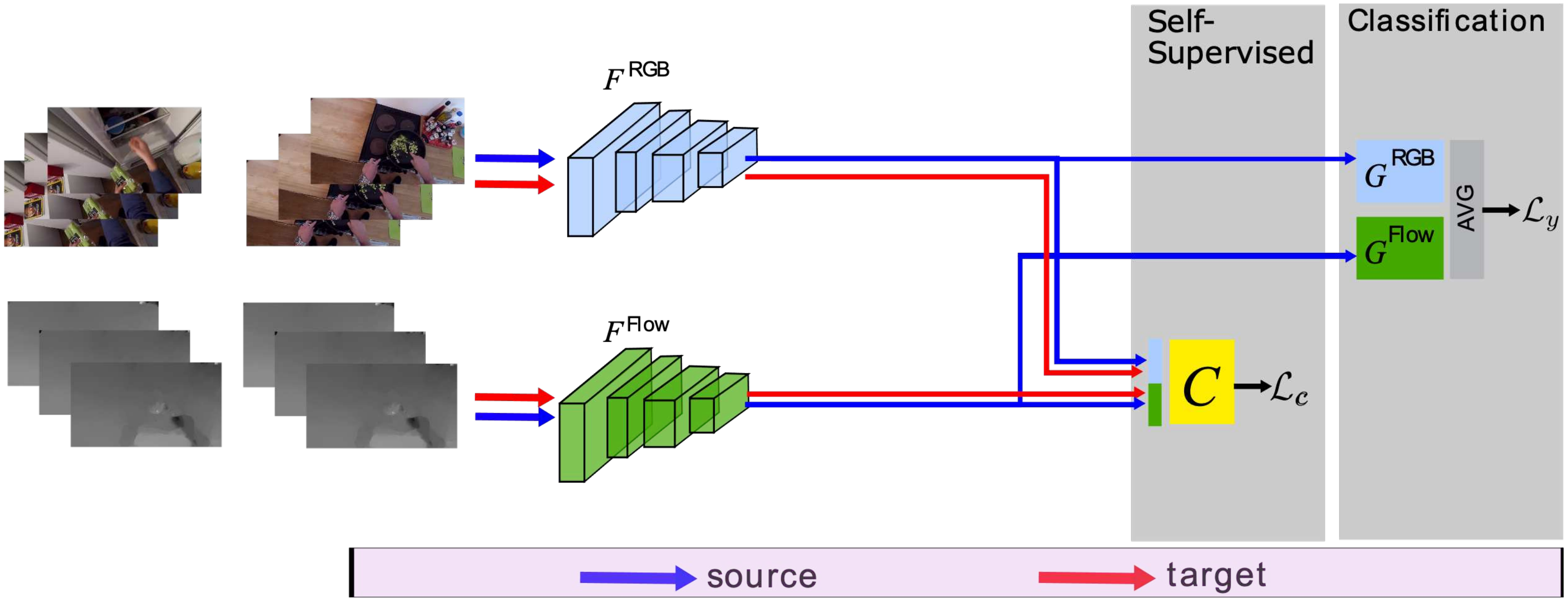
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

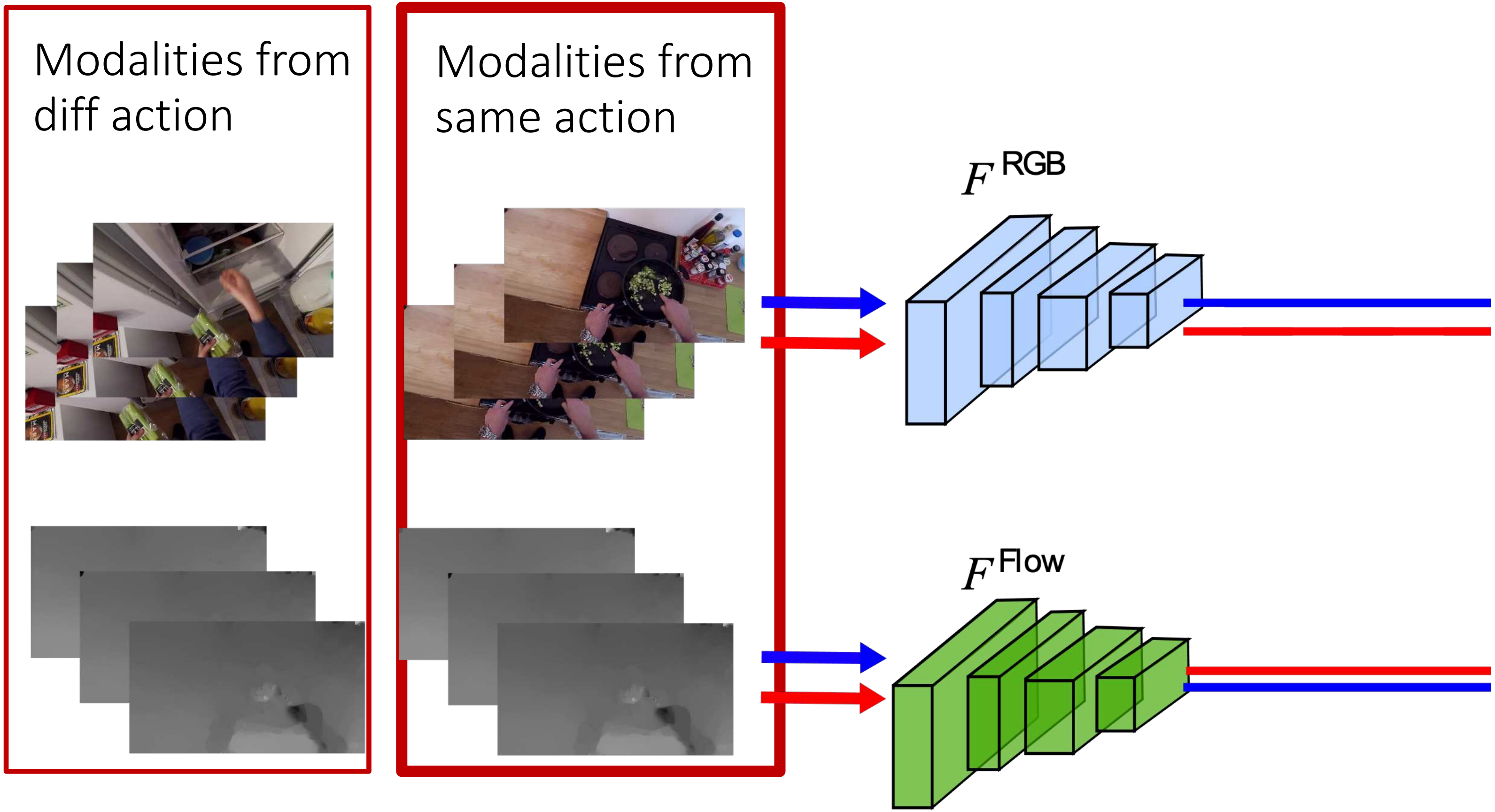
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

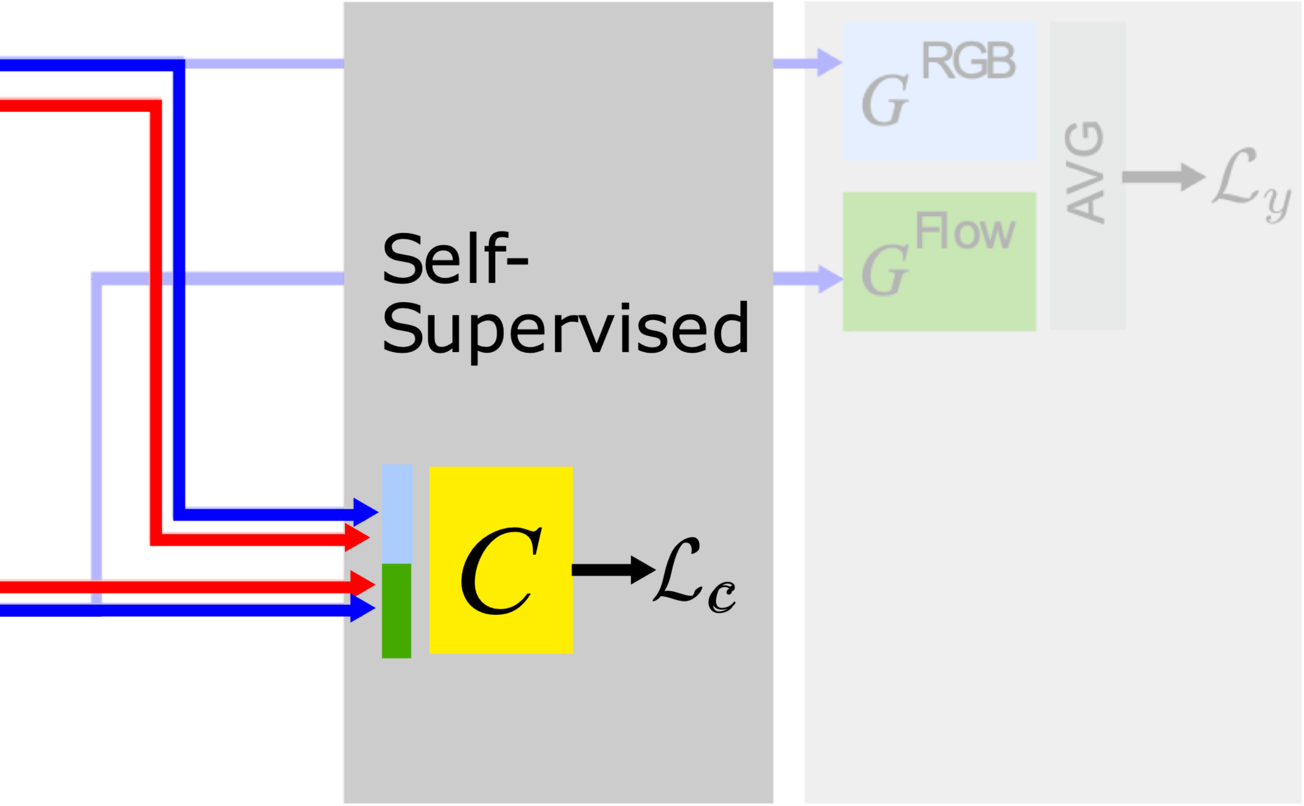
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

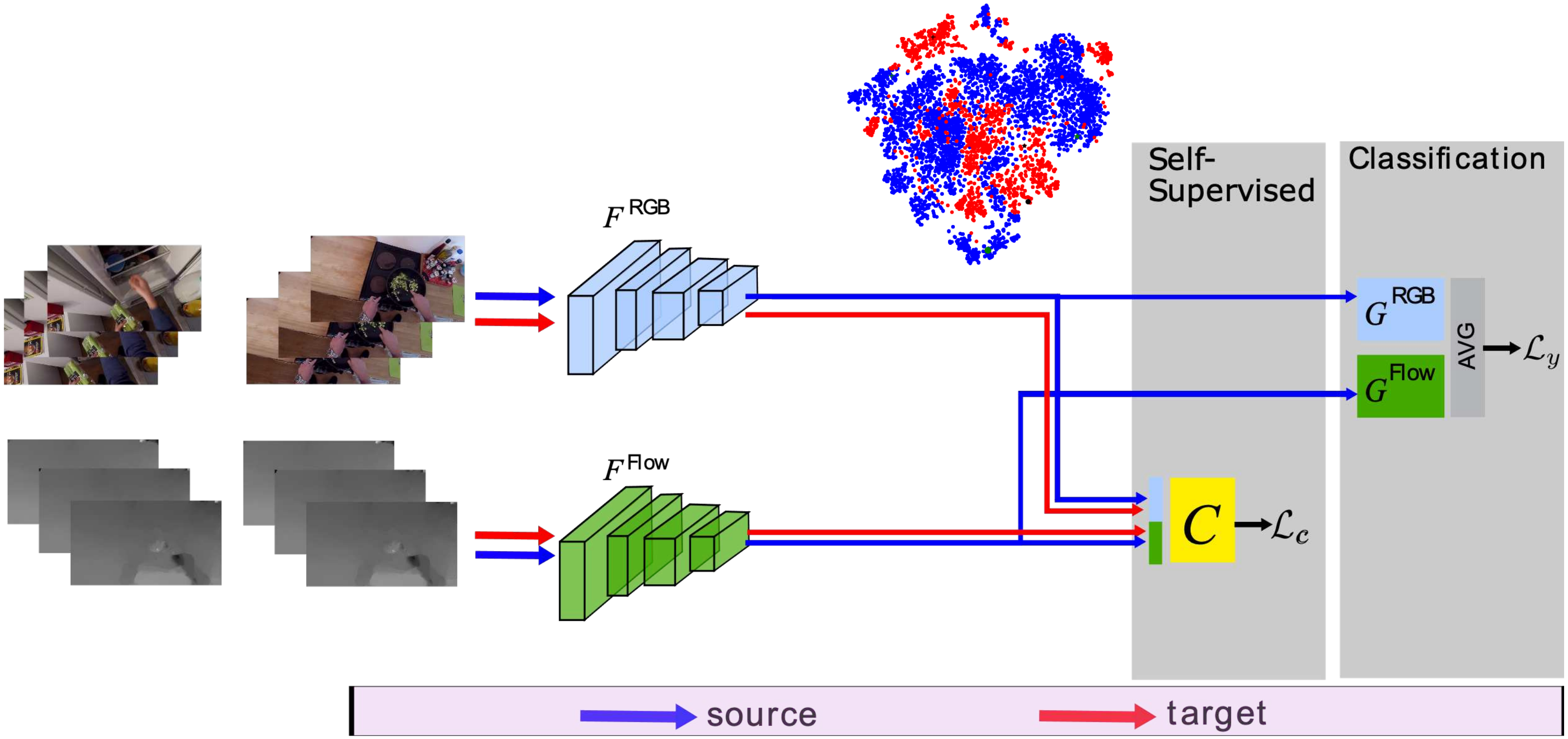
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

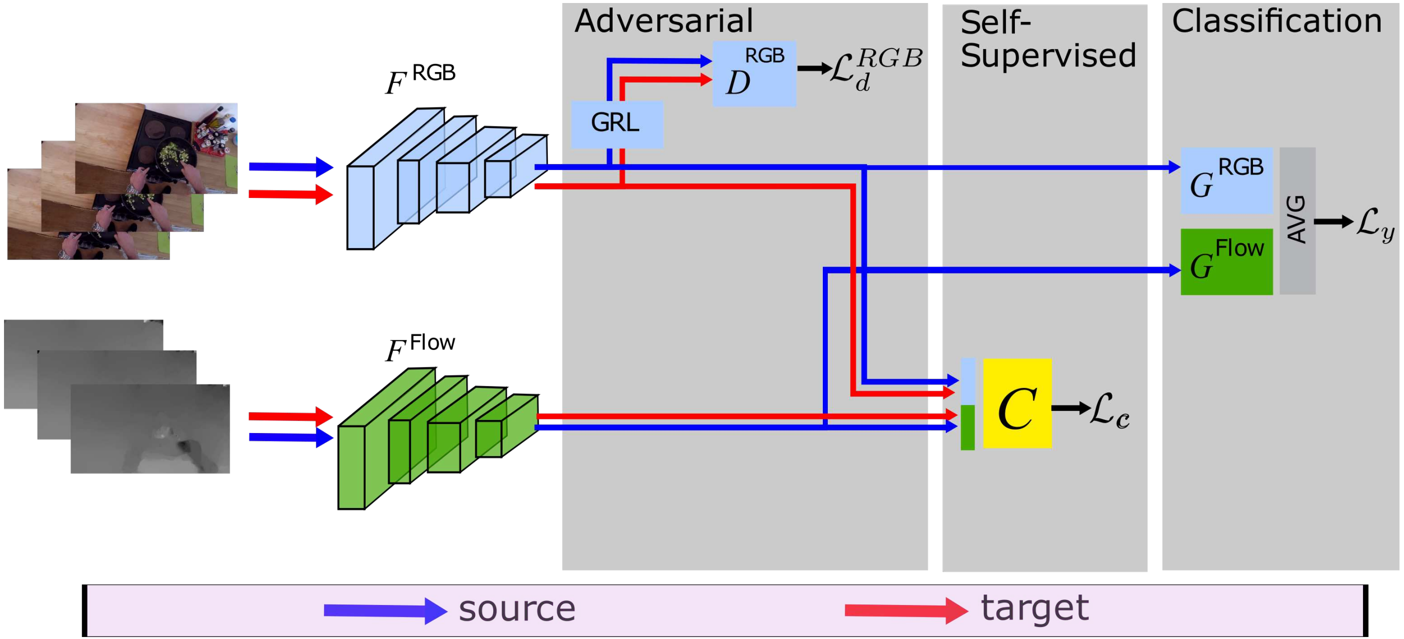
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

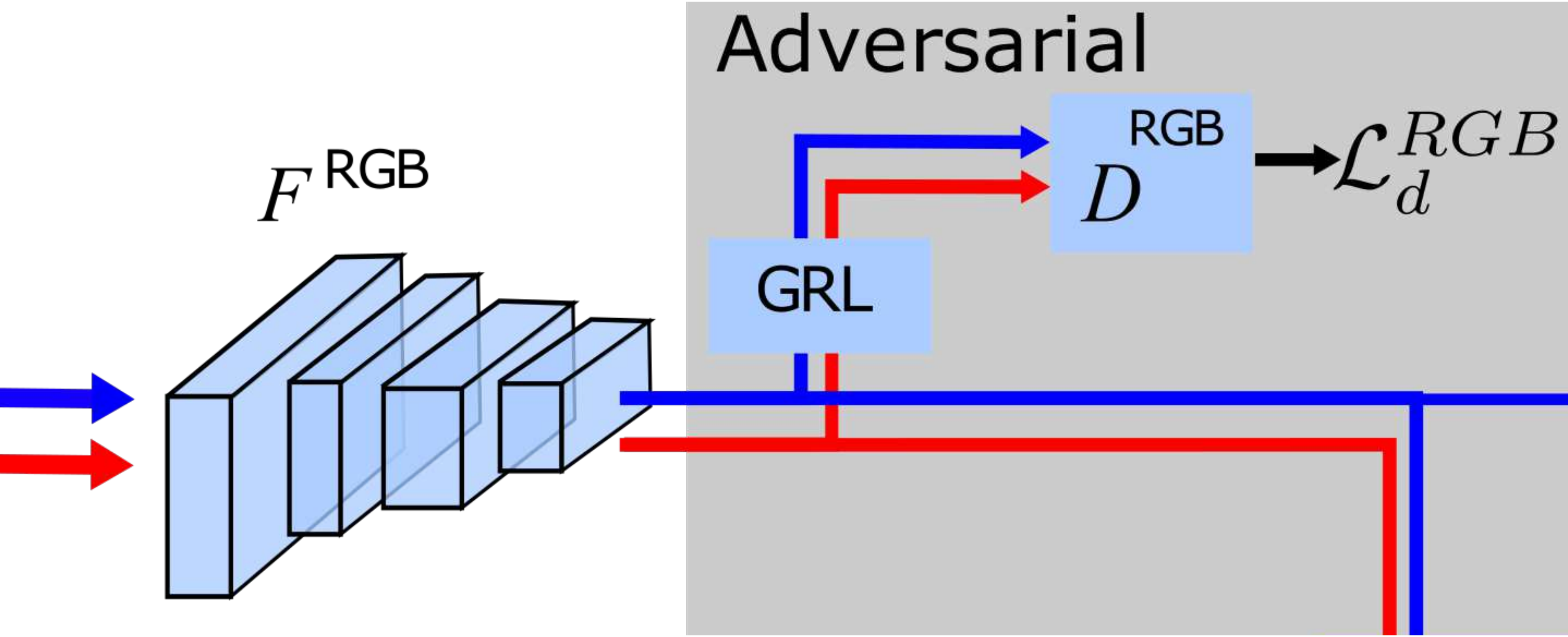
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

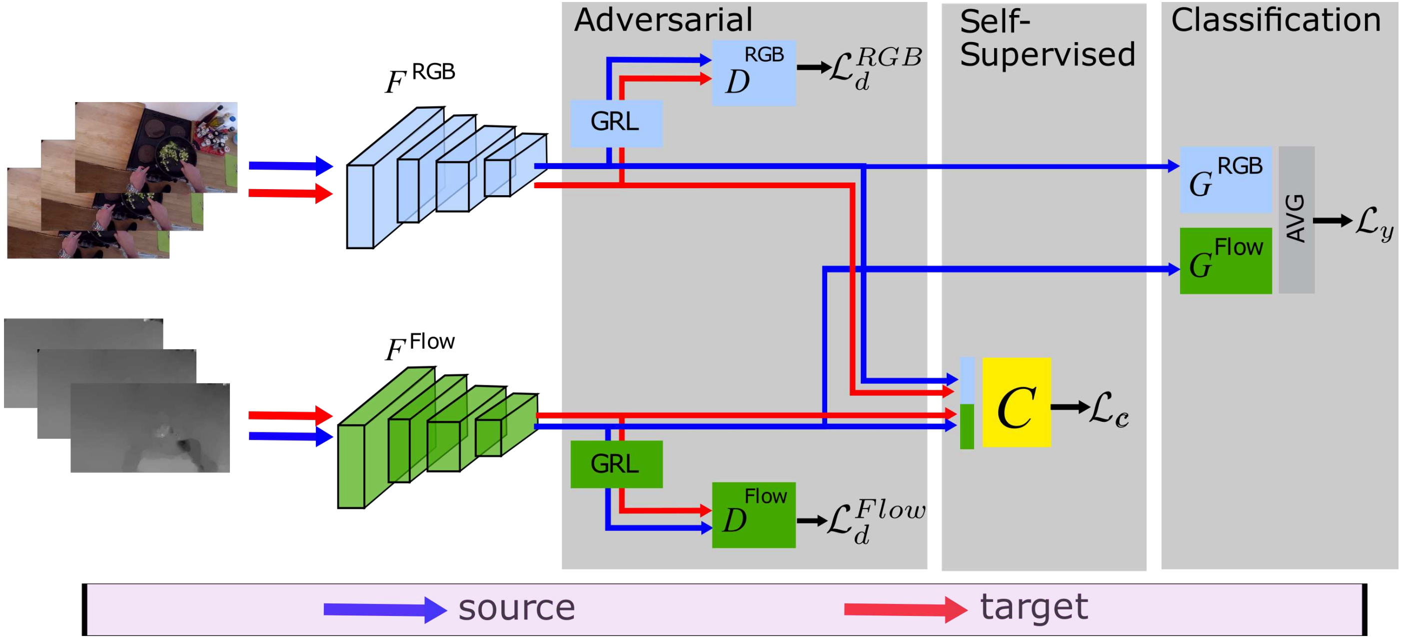
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

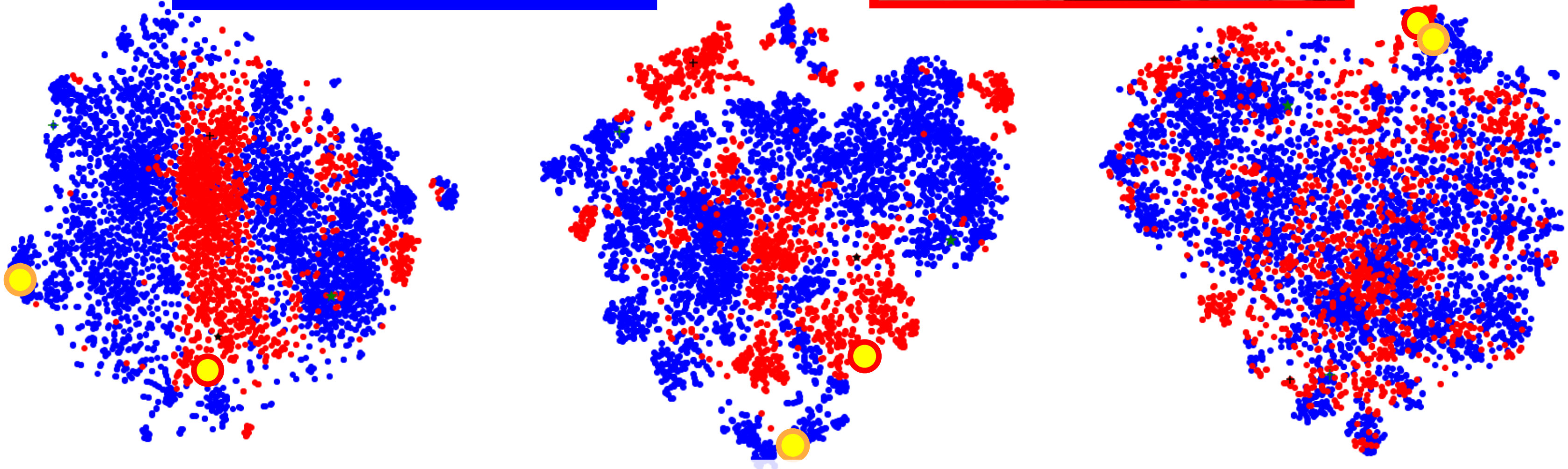
with: Jonathan Munro



J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

with: Jonathan Munro



Source-Only

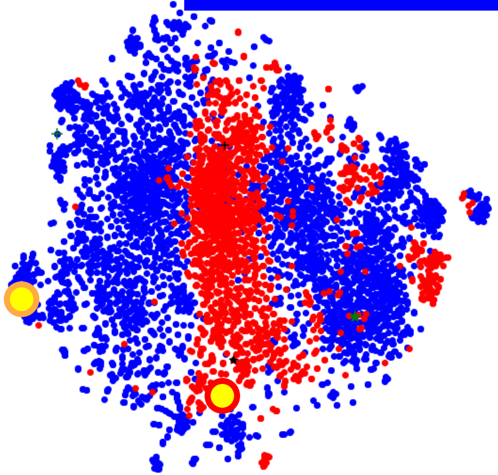
Self-Supervision

MM-SADA

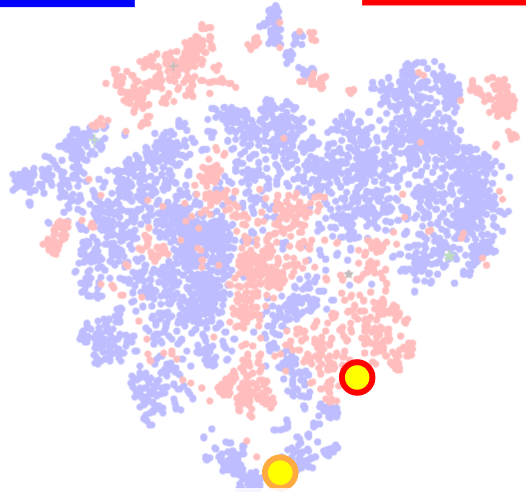
J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

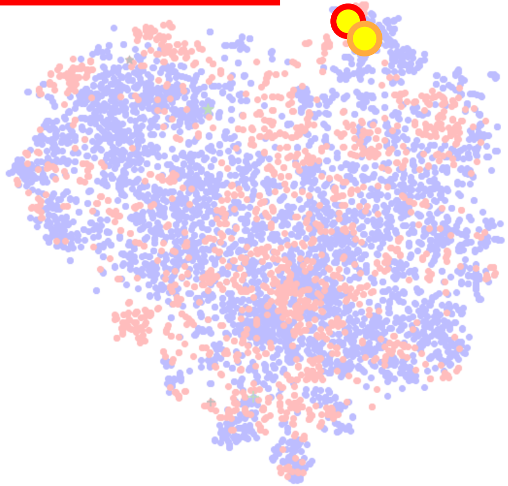
with: Jonathan Munro



Source-Only



Self-Supervision

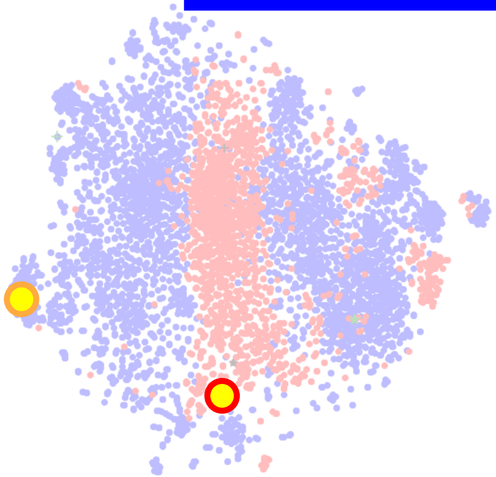


MM-SADA

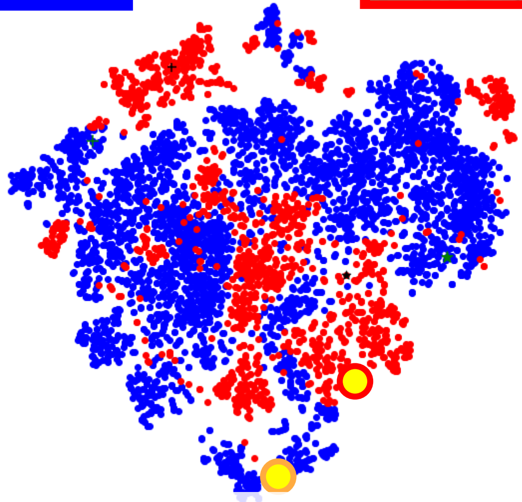
J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

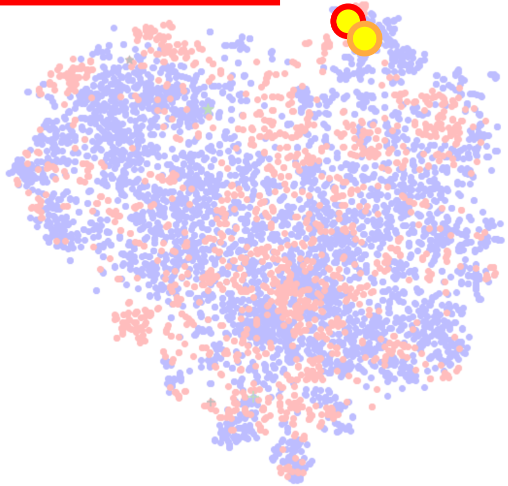
with: Jonathan Munro



Source-Only



Self-Supervision

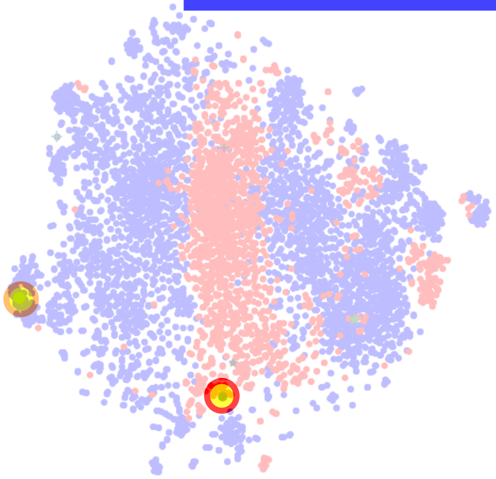


MM-SADA

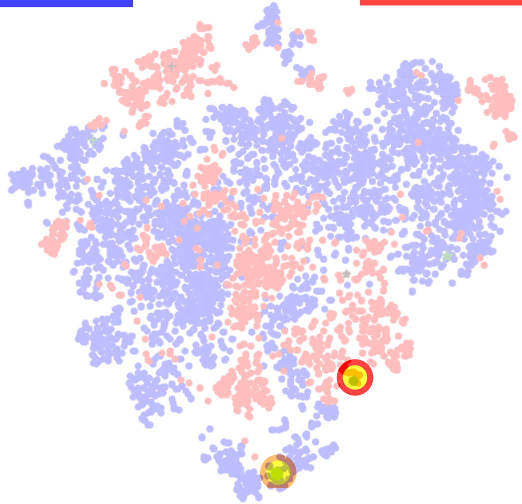
J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

Multi-modal UDA

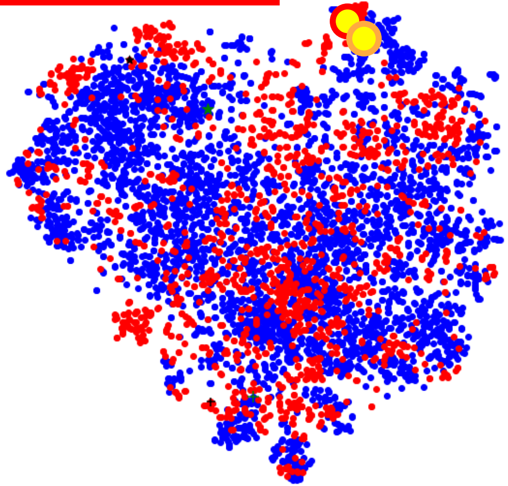
with: Jonathan Munro

Source-Only



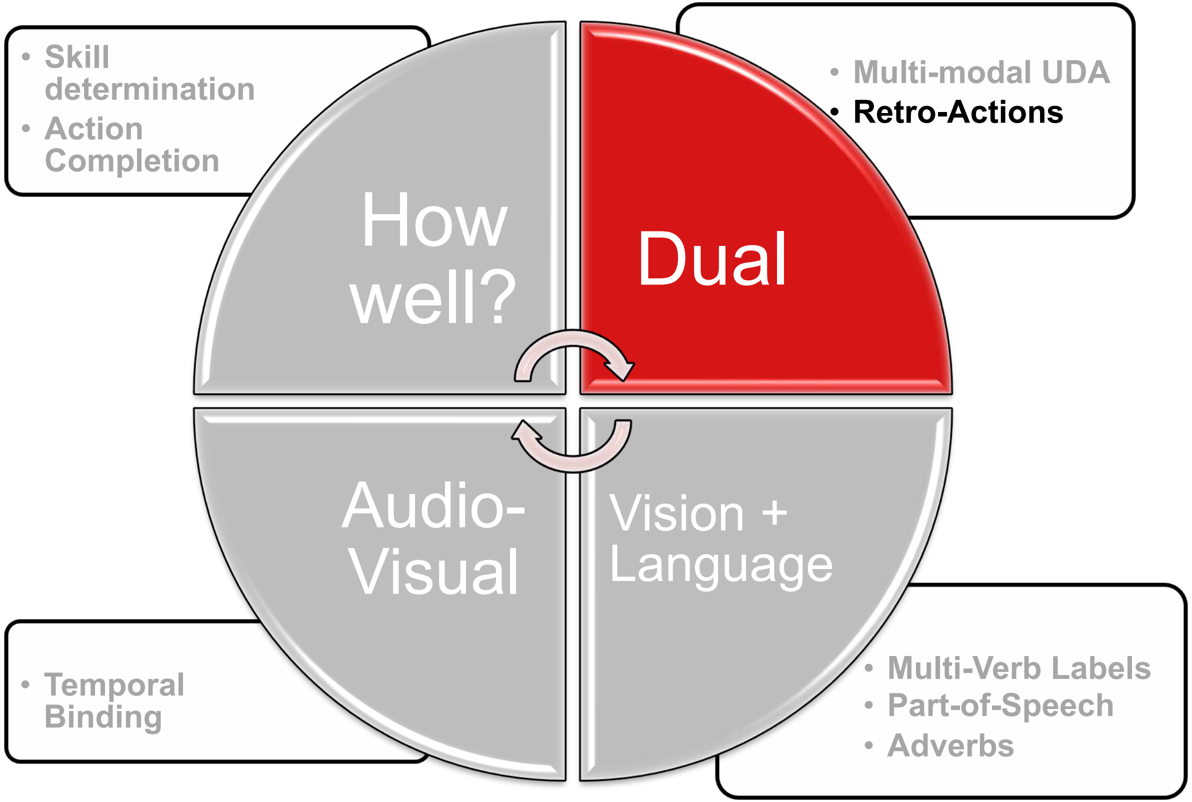
Self-Supervision



MM-SADA

J Munro, D Damen (2020). Multi-Modal Domain Adaptation for Fine-Grained Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*

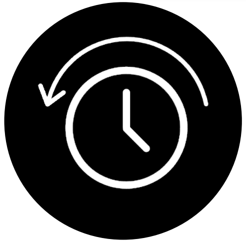
Fine(r)-grained?



University of BRISTOL

Retro-Actions

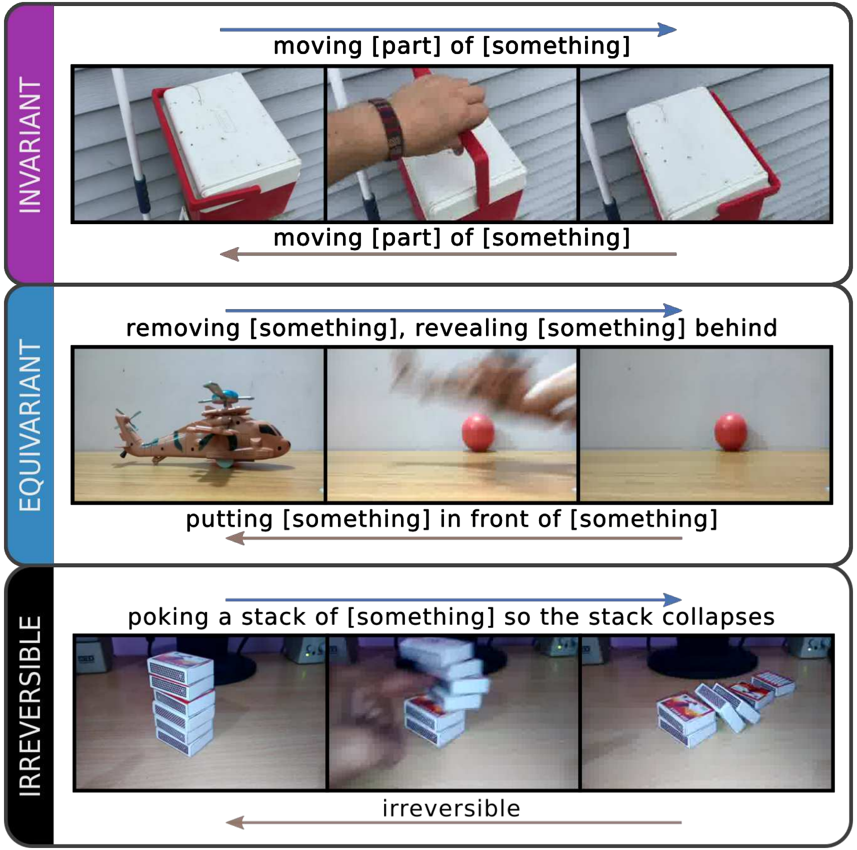
with: Will Price



W Price, D Damen (2019). Retro-Actions: Learning ‘Close’ by Time-Reversing ‘Open’ Videos. ICCV MDALC Workshop

Retro-Actions

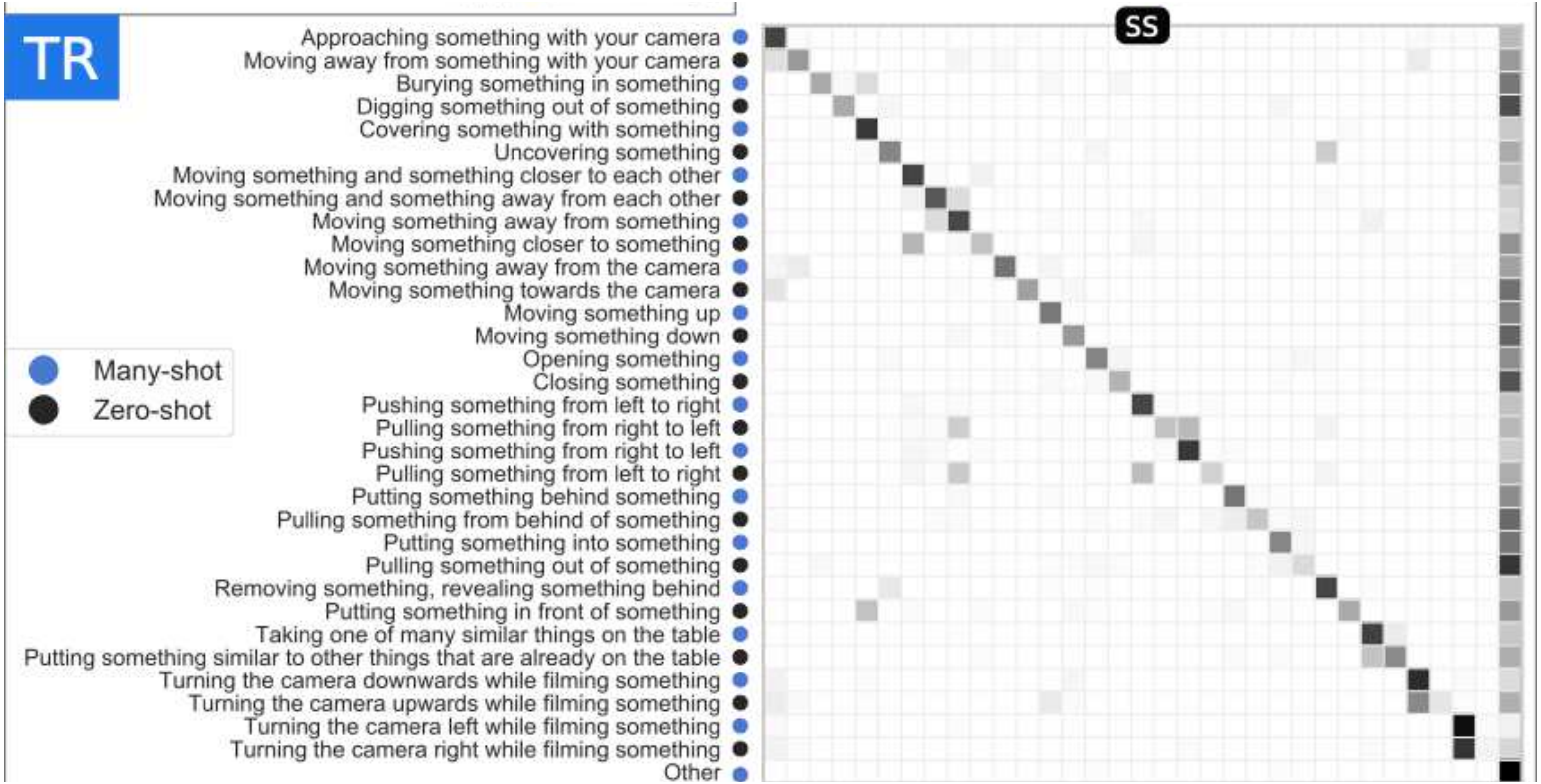
with: Will Price



W Price, D Damen (2019). Retro-Actions: Learning 'Close' by Time-Reversing 'Open' Videos. ICCV MDALC Workshop

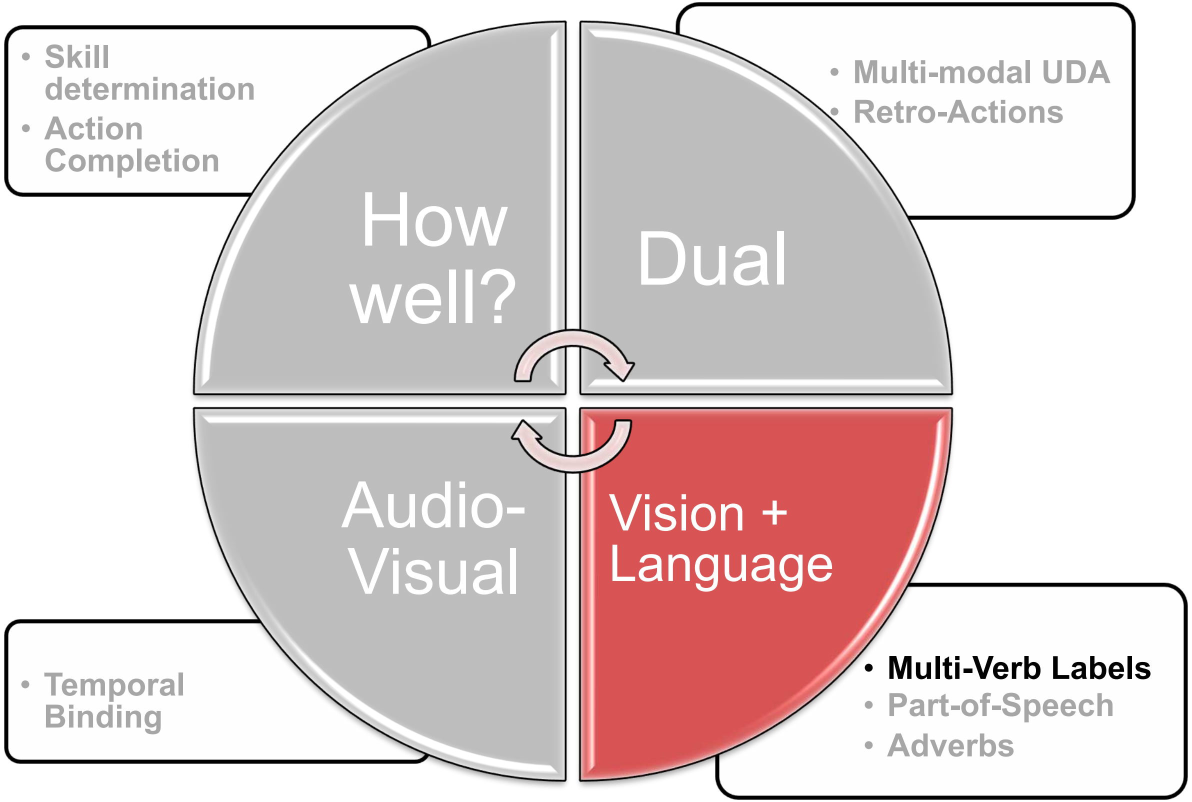
Retro-Actions

with: Will Price



W Price, D Damen (2019). Retro-Actions: Learning 'Close' by Time-Reversing 'Open' Videos. ICCV MDALC Workshop

Fine(r)-grained?



The *Verbs* Dilemma

with: Michael Wray

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

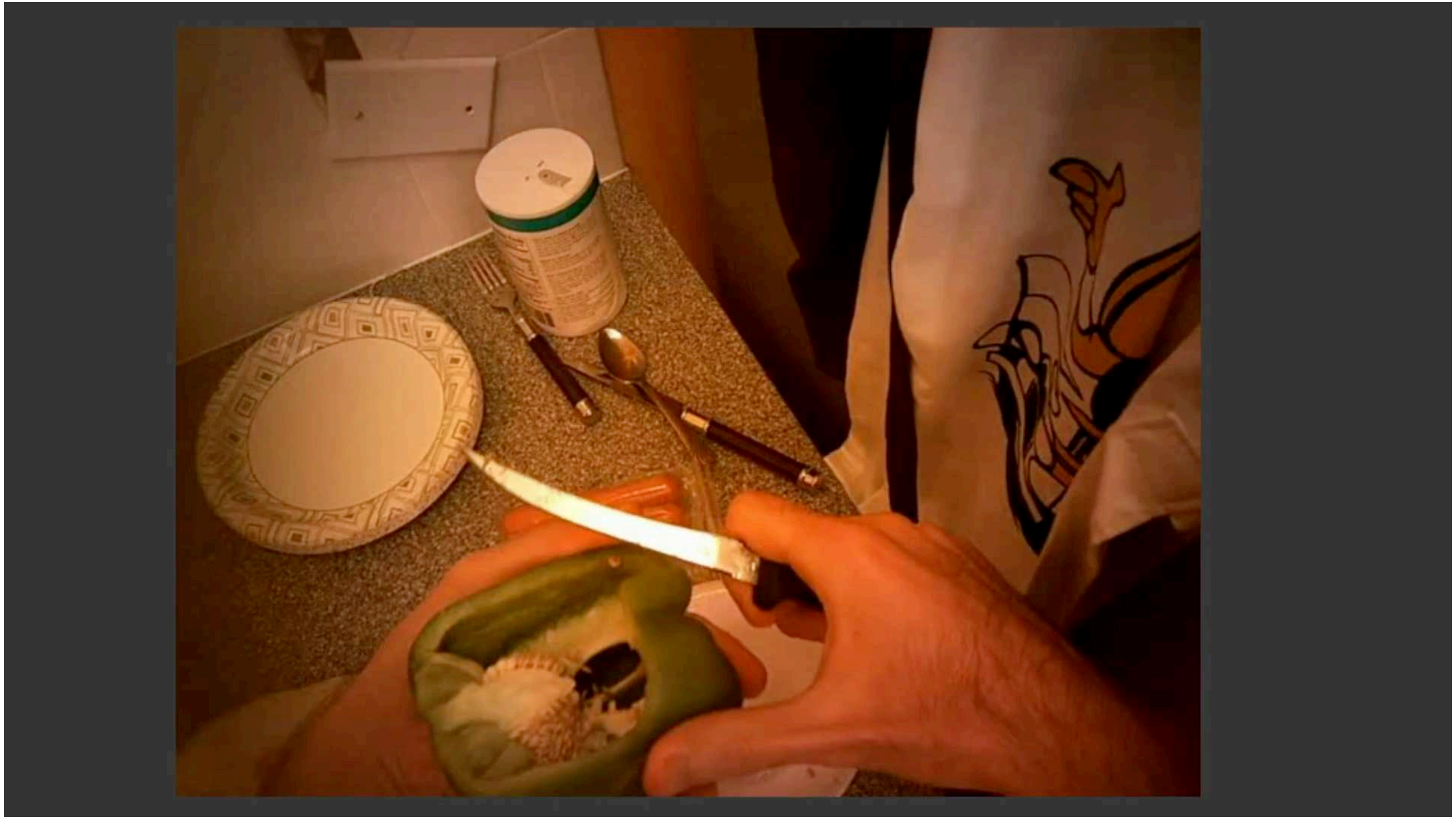
with: Michael Wray

Open

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

with: Michael Wray

Open

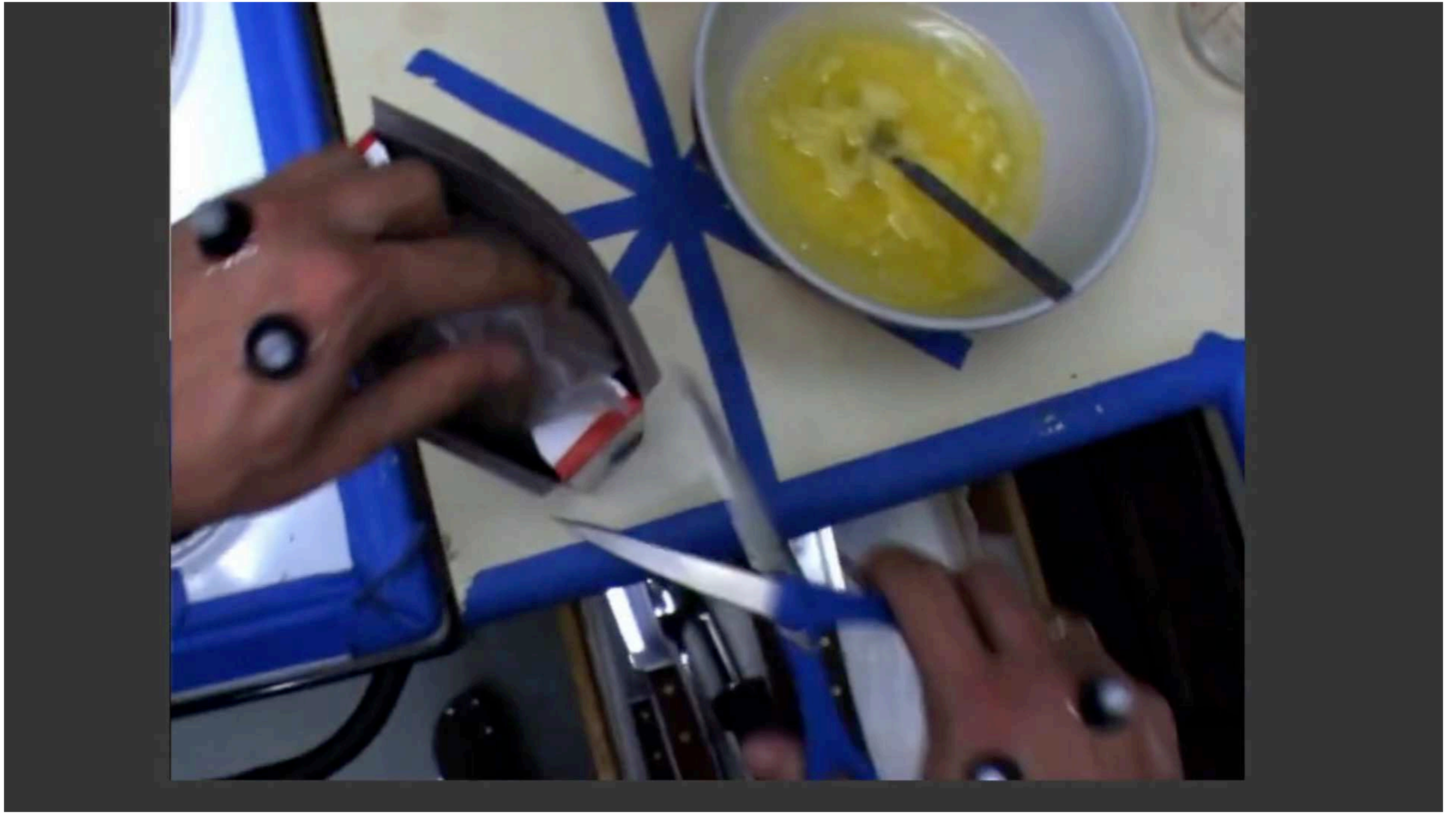
Cut

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

University of BRISTOL

The *Verbs* Dilemma

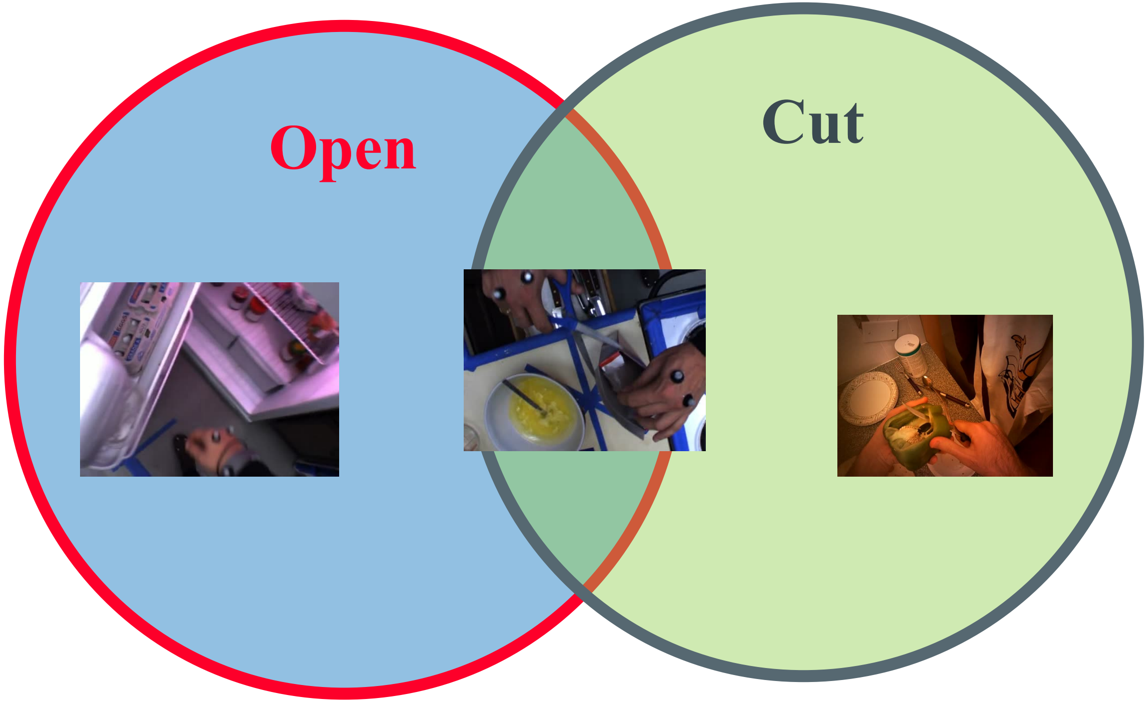
with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

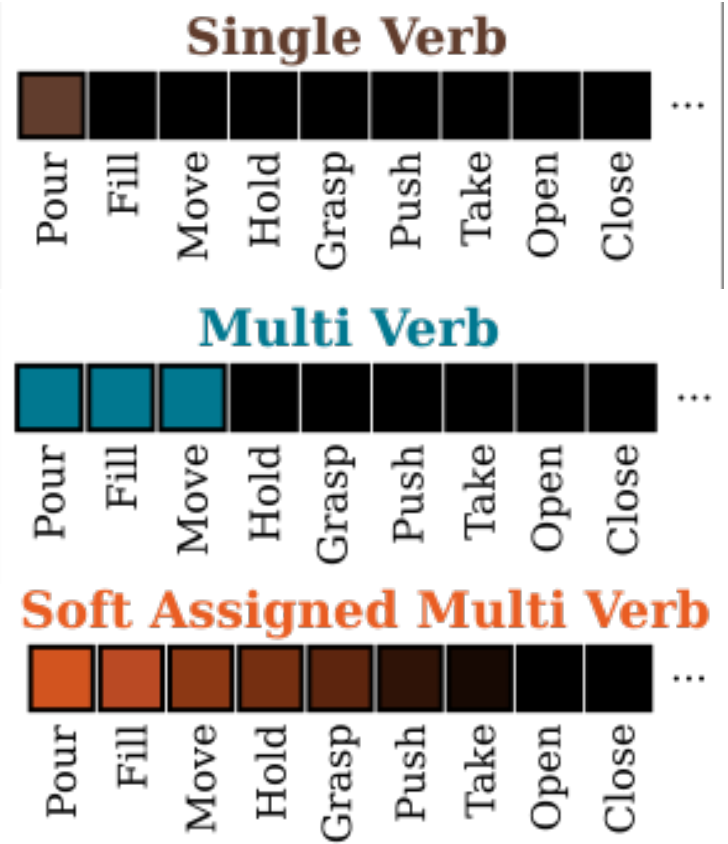
with: Michael Wray

- 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

M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

The *Verbs* Dilemma

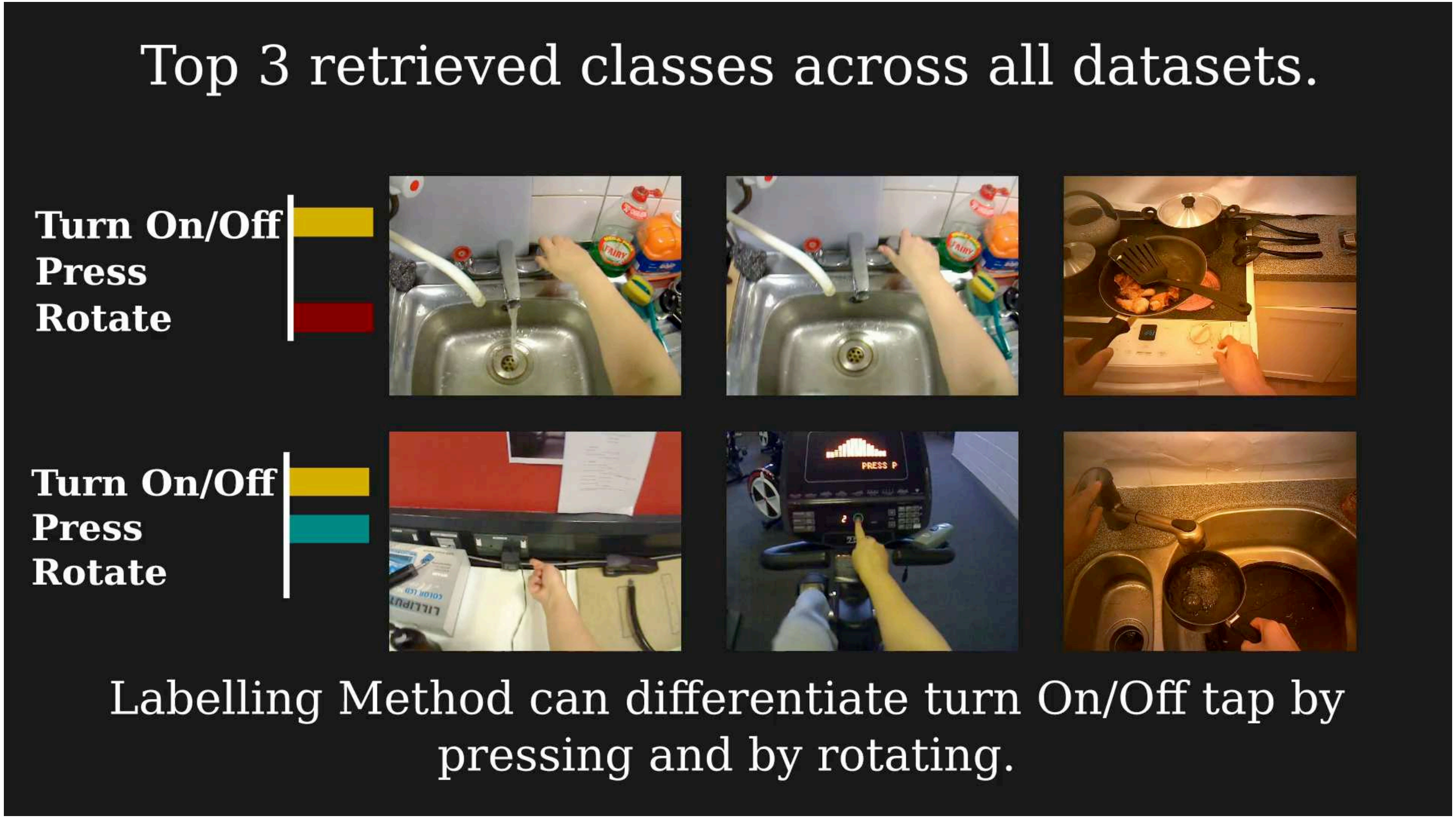
with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

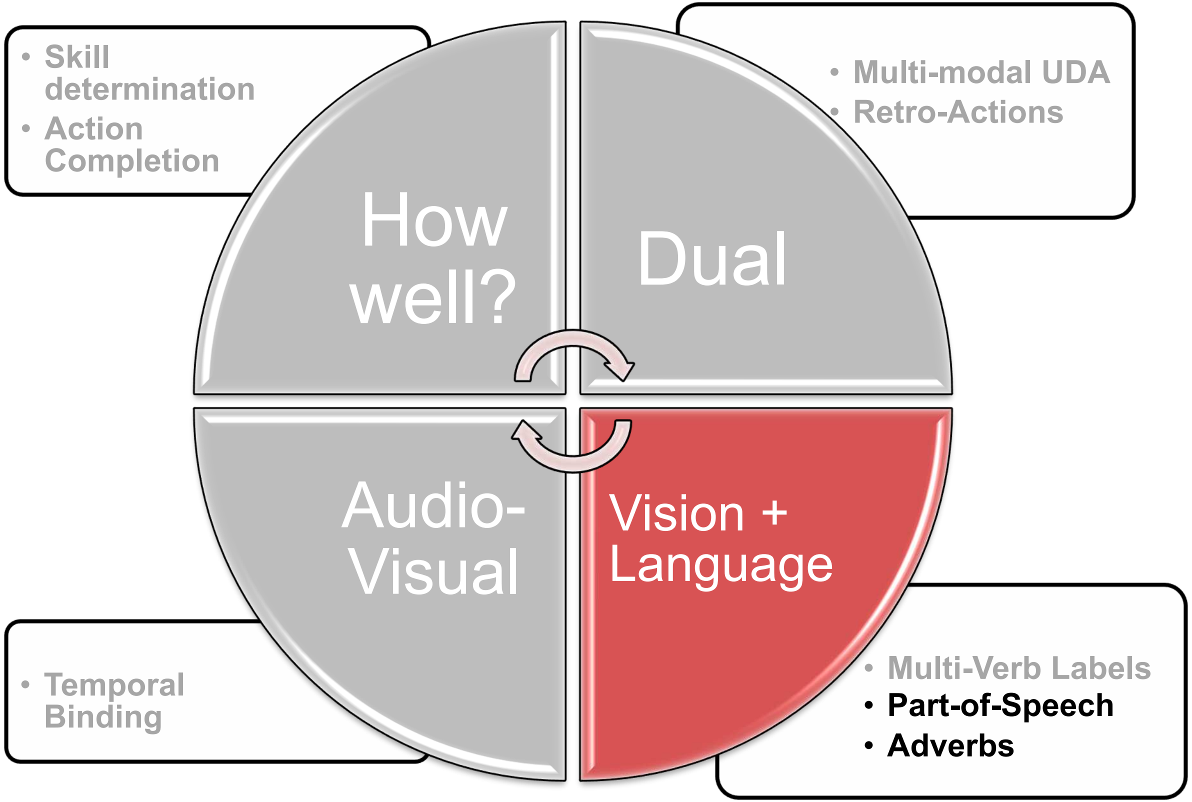
The *Verbs* Dilemma

with: Michael Wray



M Wray and D Damen (2019). Learning Visual Actions Using Multiple Verb-Only Labels. BMVC

Fine(r)-grained?



Fine-Grained Action Retrieval

with: Michael Wray
Gabriela Csurka
Diane Larlus

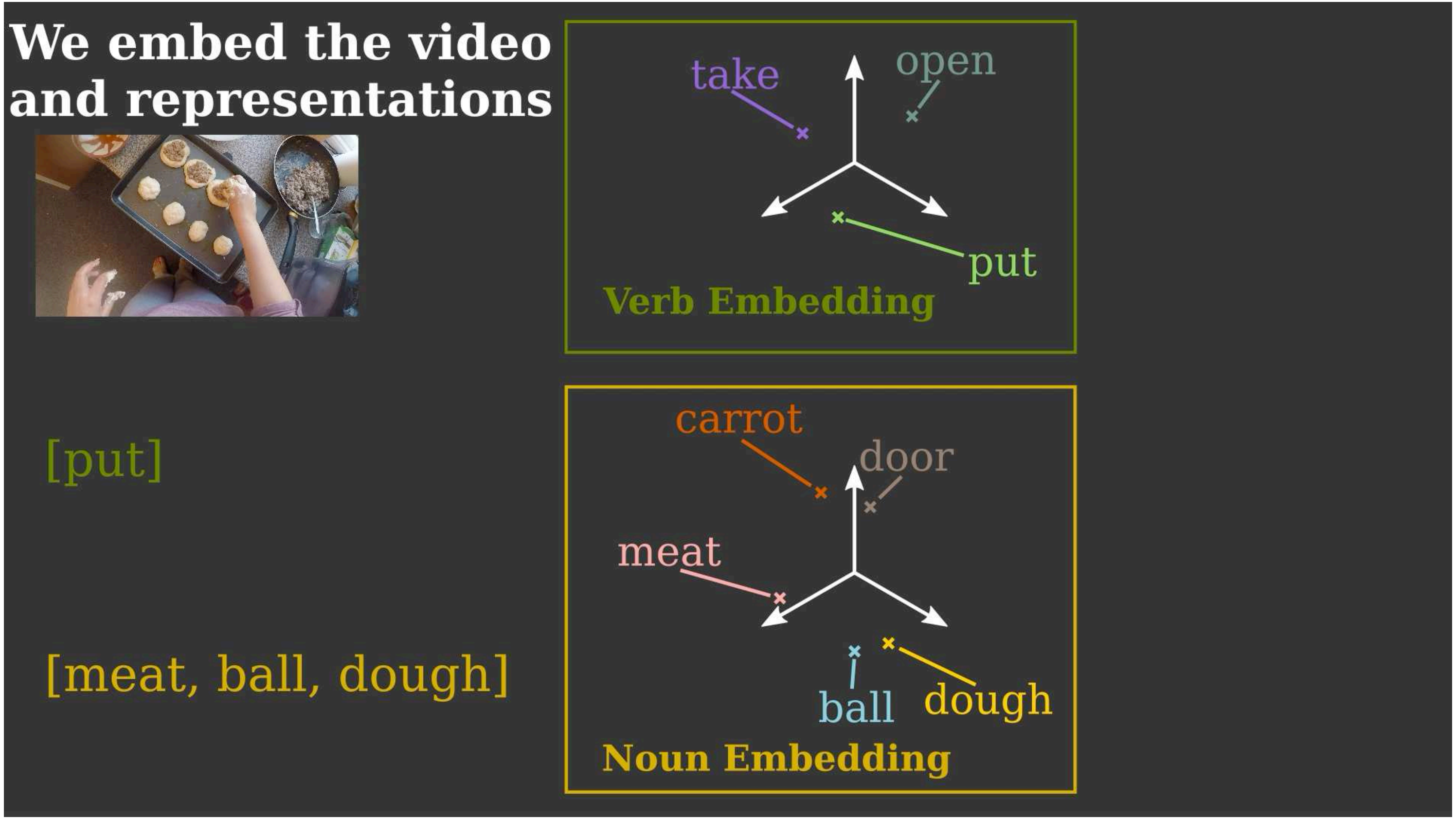
In this work we focus on
Fine-Grained Action Retrieval

I put meat on a ball of dough ⟷

M Wray, D Larlus, G Csurka, D Damen (2019). Fine-Grained Action Retrieval through Multiple Parts-of-Speech Embeddings. ICCV

Fine-Grained Action Retrieval

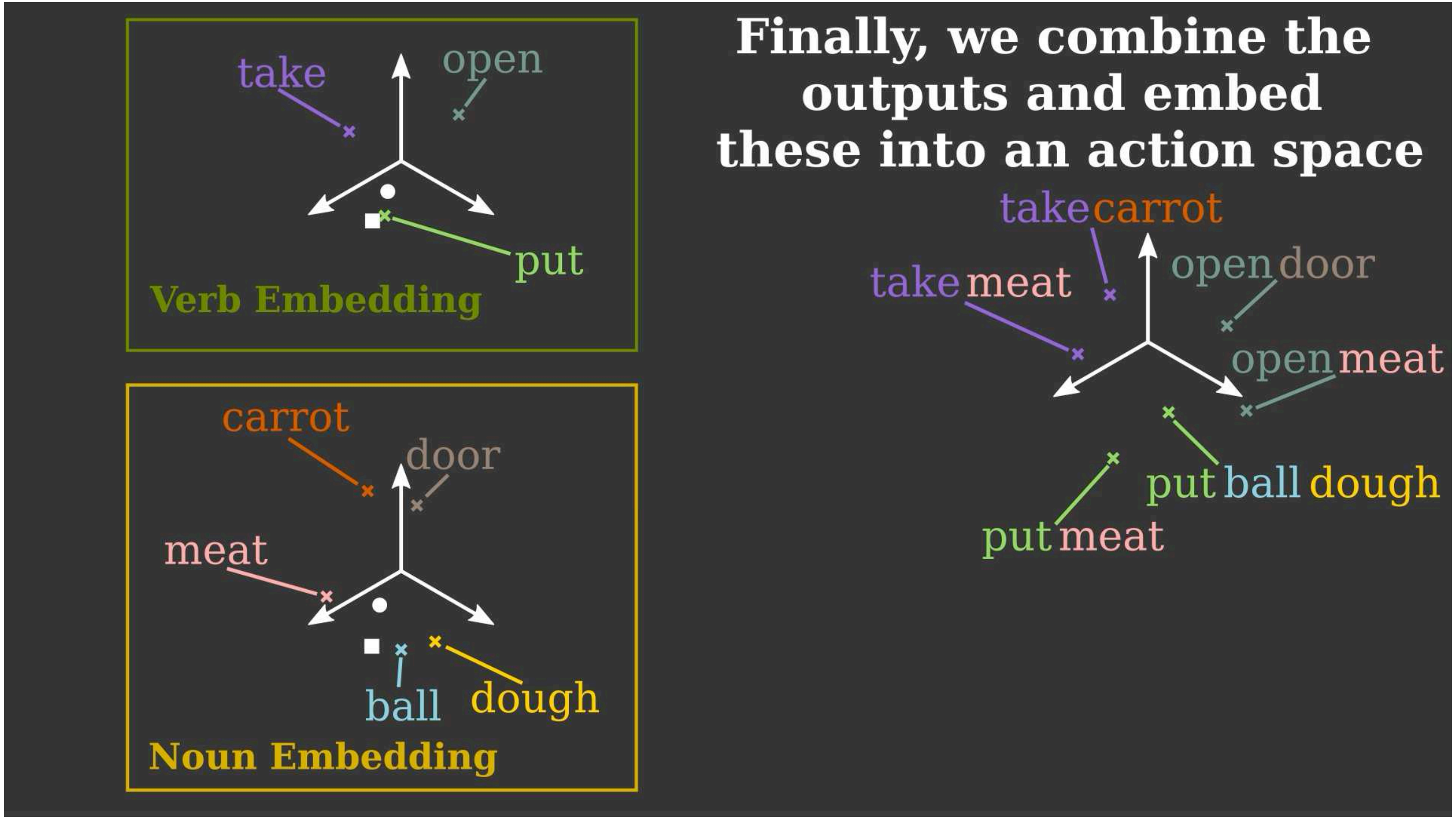
with: Michael Wray
Gabriela Csurka
Diane Larlus



M Wray, D Larlus, G Csurka, D Damen (2019). Fine-Grained Action Retrieval through Multiple Parts-of-Speech Embeddings. ICCV

Fine-Grained Action Retrieval

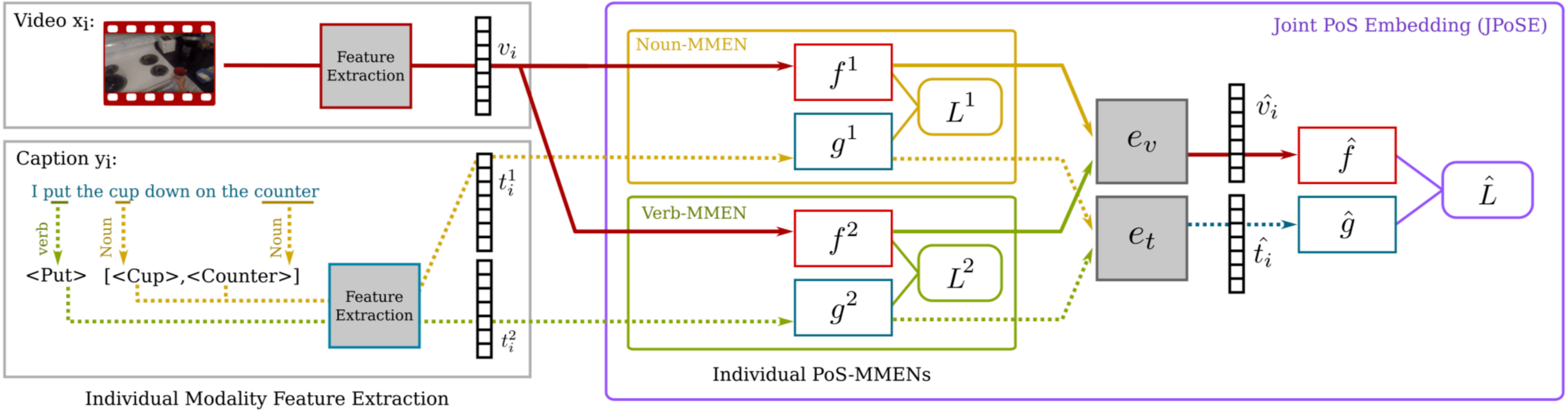
with: Michael Wray
Gabriela Csurka
Diane Larlus



M Wray, D Larlus, G Csurka, D Damen (2019). Fine-Grained Action Retrieval through Multiple Parts-of-Speech Embeddings. ICCV

Fine-Grained Action Retrieval

with: Michael Wray
Gabriela Csurka
Diane Larlus



M Wray, D Larlus, G Csurka, D Damen (2019). Fine-Grained Action Retrieval through Multiple Parts-of-Speech Embeddings. ICCV

Fine-Grained Action Retrieval

with: Michael Wray
Gabriela Csurka
Diane Larlus

Maximum activation examples for a neuron in a noun PoS Embedding (Cutting Board) - Figure 4

M Wray, D Larlus, G Csurka, D Damen (2019). Fine-Grained Action Retrieval through Multiple Parts-of-Speech Embeddings. ICCV

Action Modifiers: Learning from Adverbs

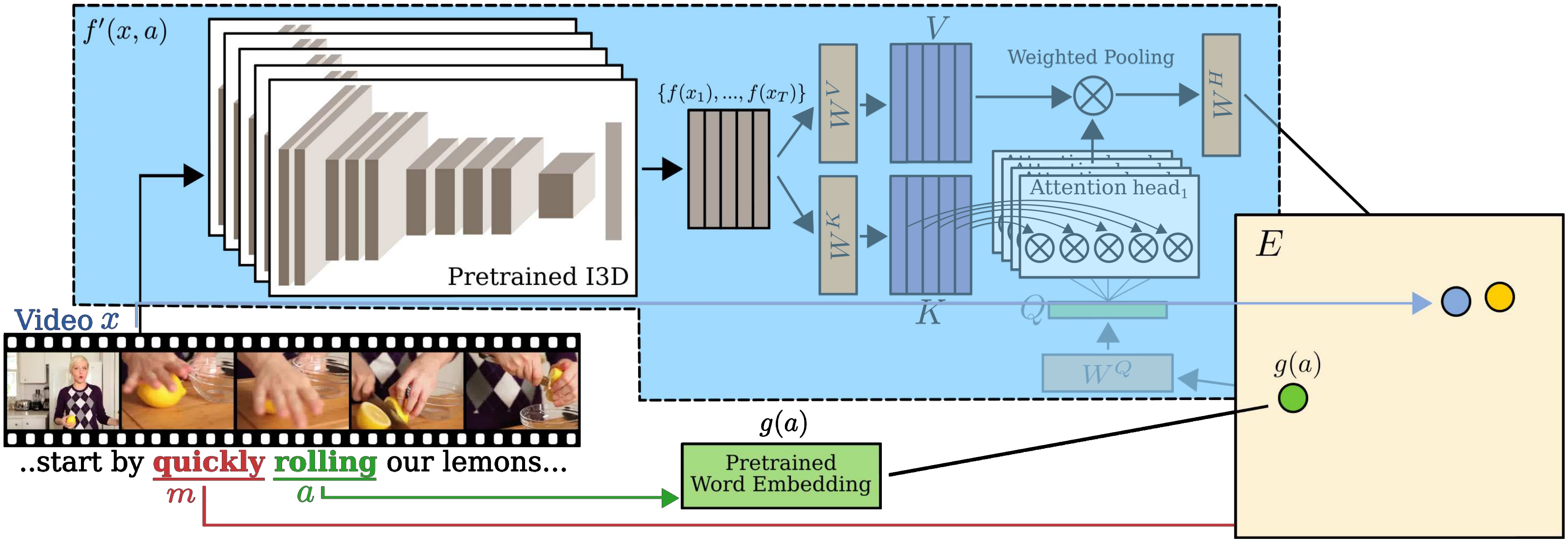
with: Hazel Doughty
Ivan Laptev
Walterio Mayol-Cuevas

University of BRISTOL

H Doughty, I Laptev, W Mayol-Cuevas, D Damen (2020). Action Modifiers: Learning from Adverbs in Instructional Videos. Computer Vision and Pattern Recognition (CVPR)

Action Modifiers: Learning from Adverbs

with: Hazel Doughty
Ivan Laptev
Walterio Mayol-Cuevas



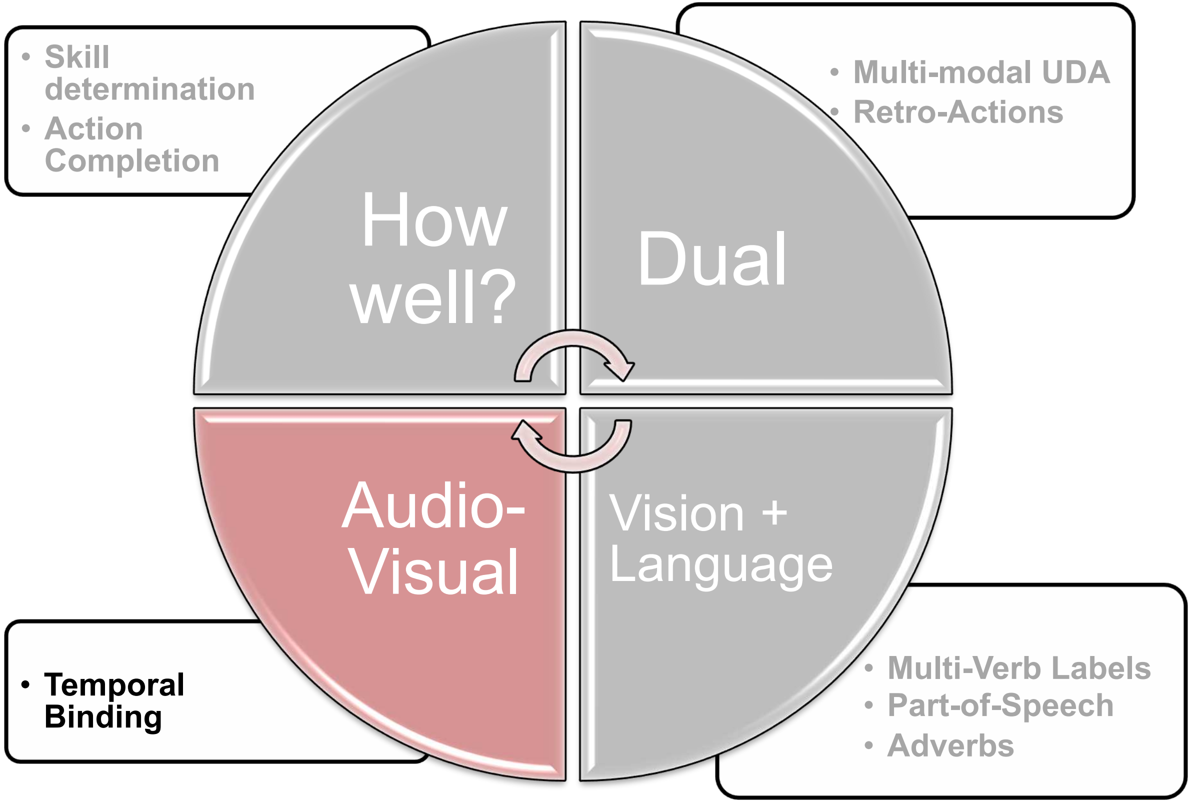
H Doughty, I Laptev, W Mayol-Cuevas, D Damen (2020). Action Modifiers: Learning from Adverbs in Instructional Videos. Computer Vision and Pattern Recognition (CVPR)

with: Hazel Doughty
Ivan Laptev
Walterio Mayol-Cuevas

... we're going to mix these up real **quick**...

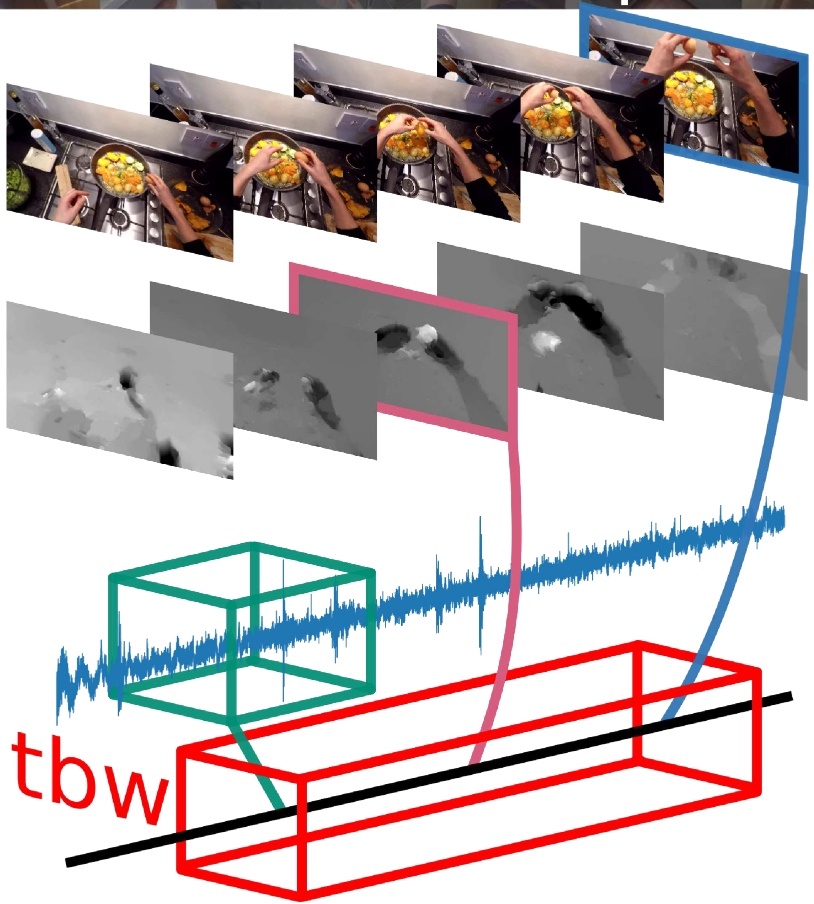
H Doughty, I Laptev, W Mayol-Cuevas, D Damen (2020). Action Modifiers: Learning from Adverbs in Instructional Videos. Computer Vision and Pattern Recognition (CVPR)

Fine(r)-grained?



Audio-Visual Temporal Binding

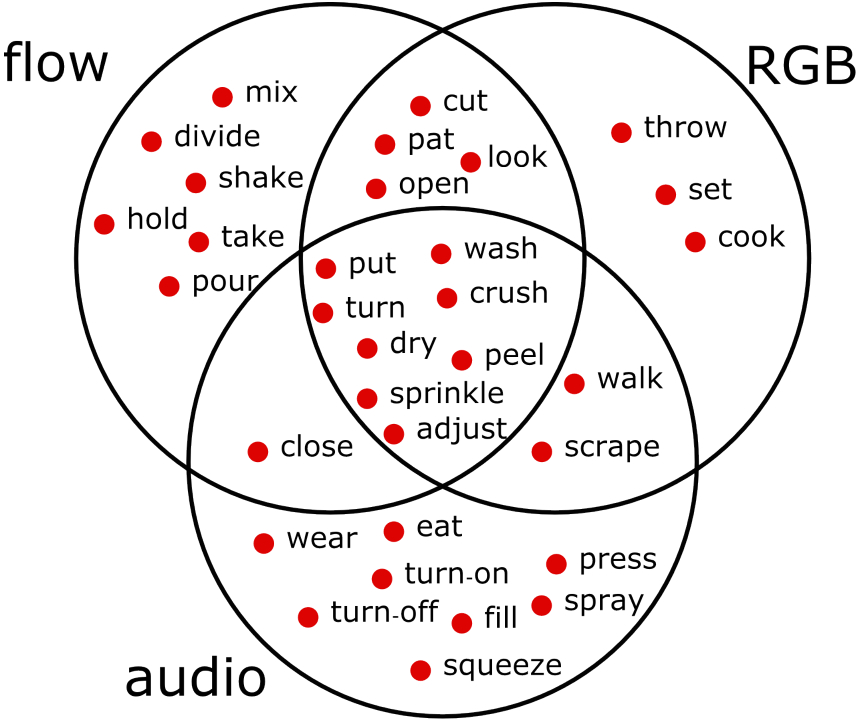
with: Vangelis Kazakos
Arsha Nagrani
Andrew Zisserman



E Kazakos, A Nagrani, A Zisserman, D Damen (2019). EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition. ICCV

Audio-Visual Temporal Binding

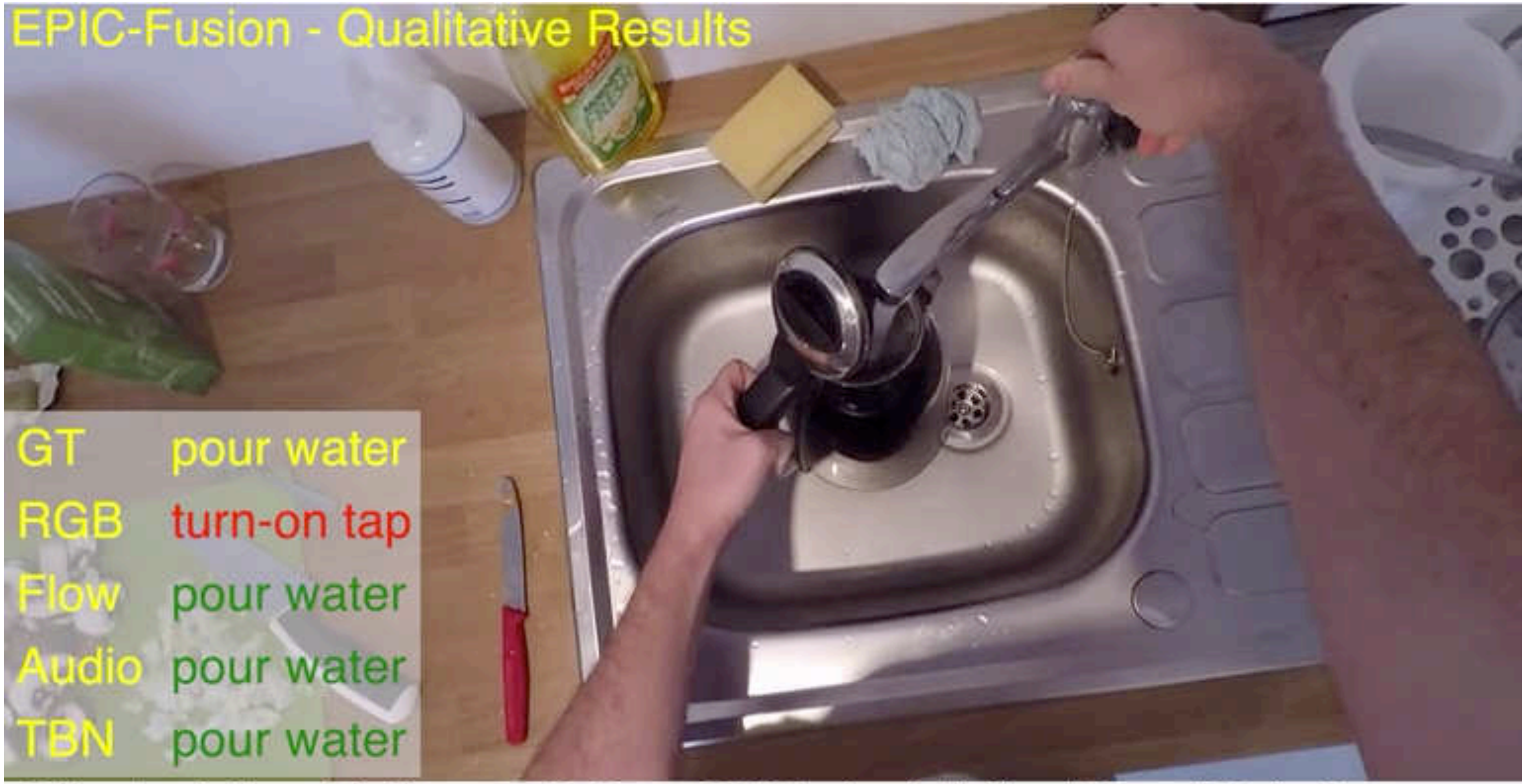
with: Vangelis Kazakos
Arsha Nagrani
Andrew Zisserman



E Kazakos, A Nagrani, A Zisserman, D Damen (2019). EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition. ICCV

Audio-Visual Temporal Binding

with: Vangelis Kazakos
Arsha Nagrani
Andrew Zisserman



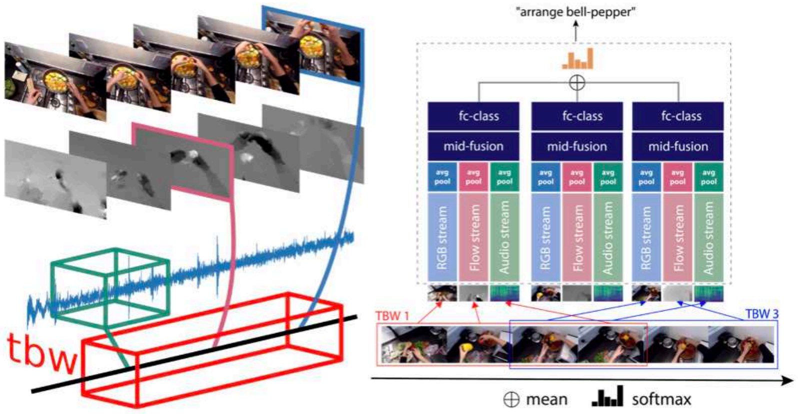
E. Kazakos, A. Nagrani, A. Zisserman, D. Damen, EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition, ICCV 2019

with: Vangelis Kazakos
Arsha Nagrani
Andrew Zisserman

EPIC-Fusion: Audio-Visual Temporal Binding for Egocentric Action Recognition

Evangelos Kazakos[1], Arsha Nagrani[2], Andrew Zisserman[2] and Dima Damen[1]

[1]University of Bristol, VIL, [2]University of Oxford, VGG



Abstract

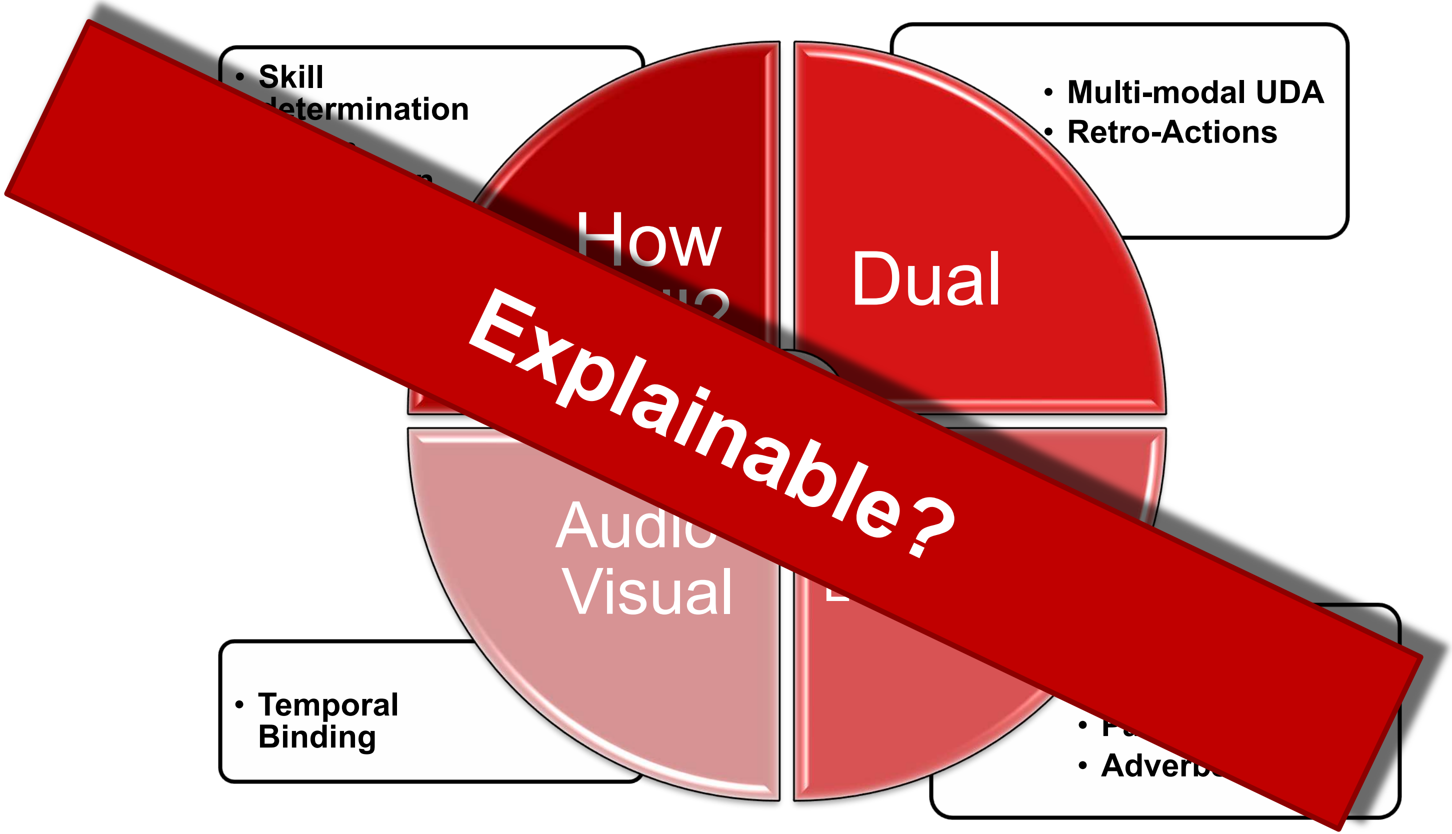
We focus on multi-modal fusion for egocentric action recognition, and propose a novel architecture for multi-modal temporal-binding, i.e. the combination of modalities within a range of temporal offsets. We train the

Downloads

- Paper [ArXiv]
- Code and models [GitHub]

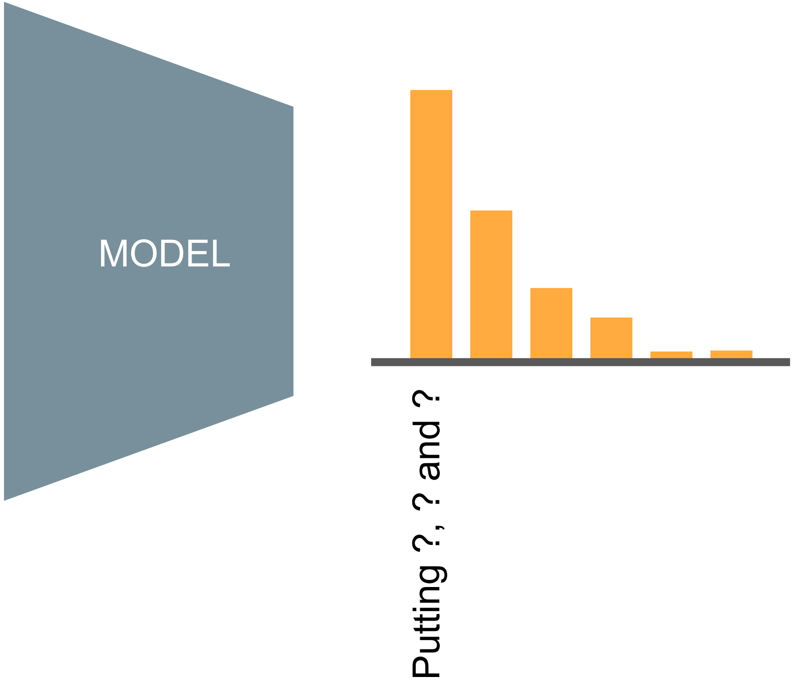
Binding for Egocentric Action Recognition. ICCV

Fine(r)-grained?



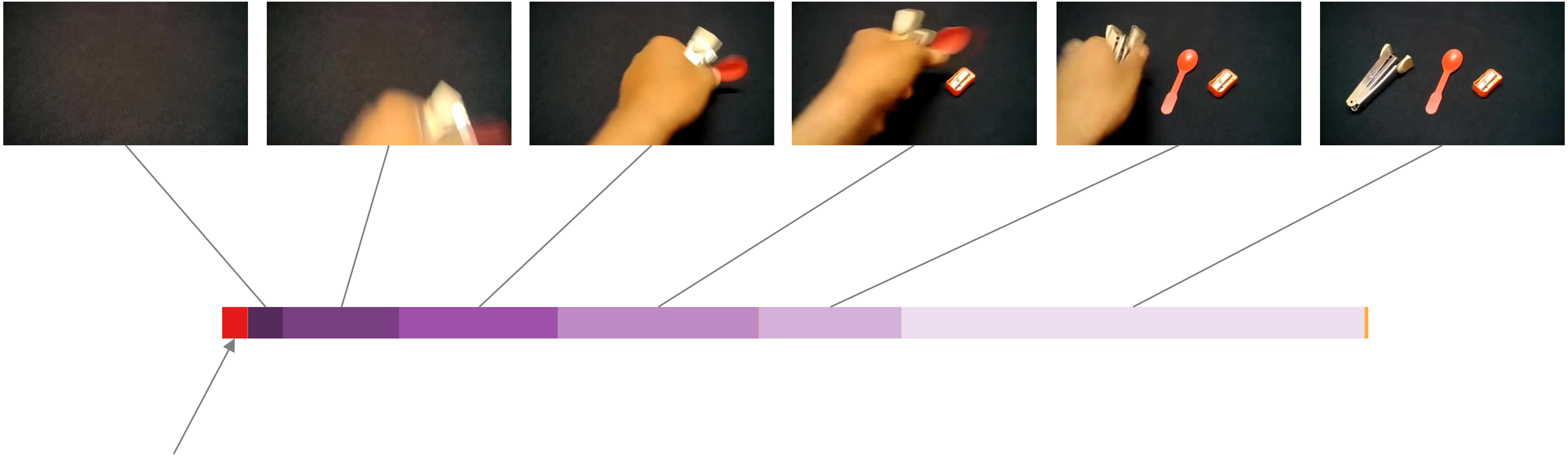
Frame Attributions in Video Models

with: Will Price



W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

Frame Attributions in Video Models

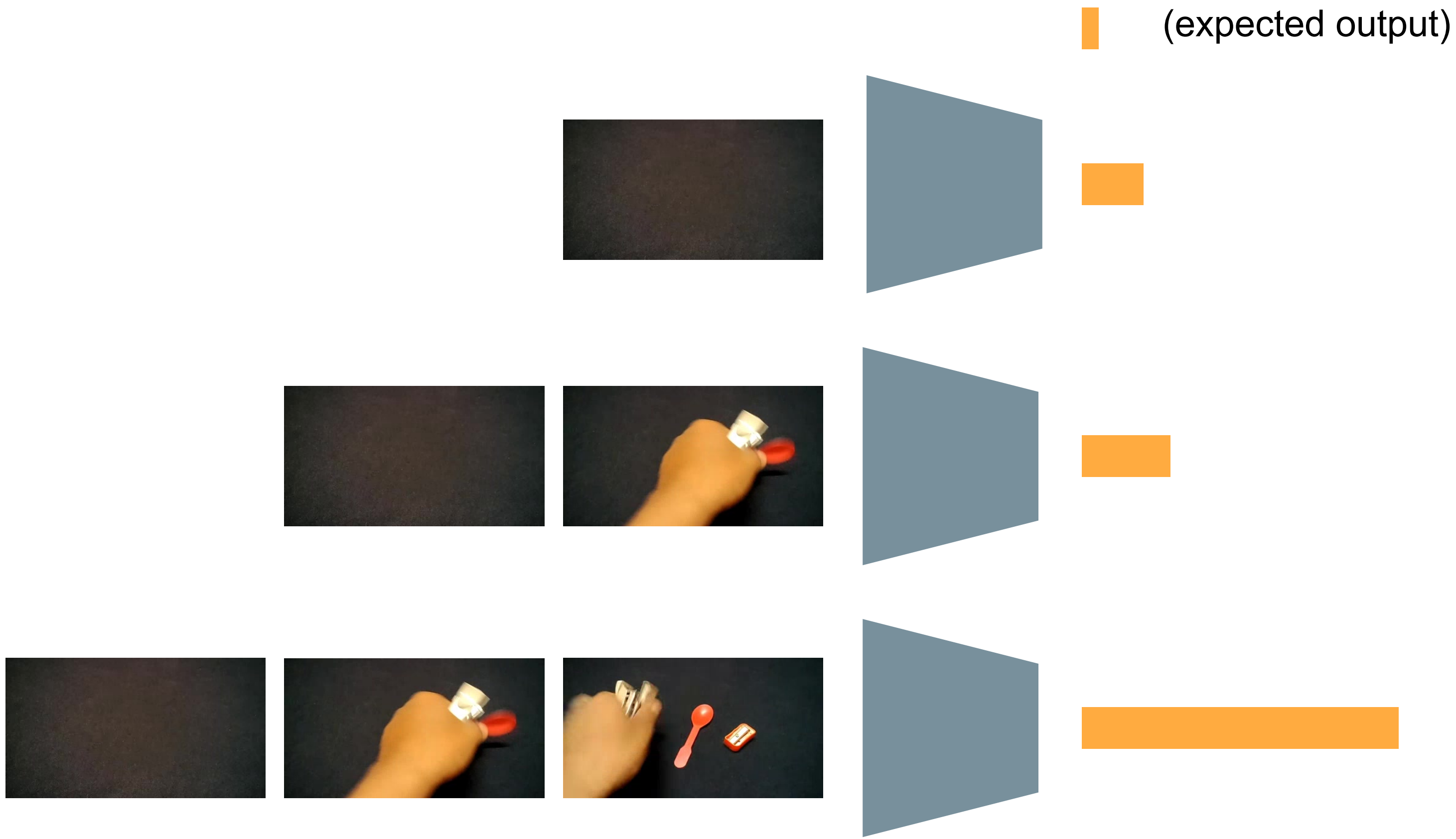


Expected output
(Prior probability for
classification model)

W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

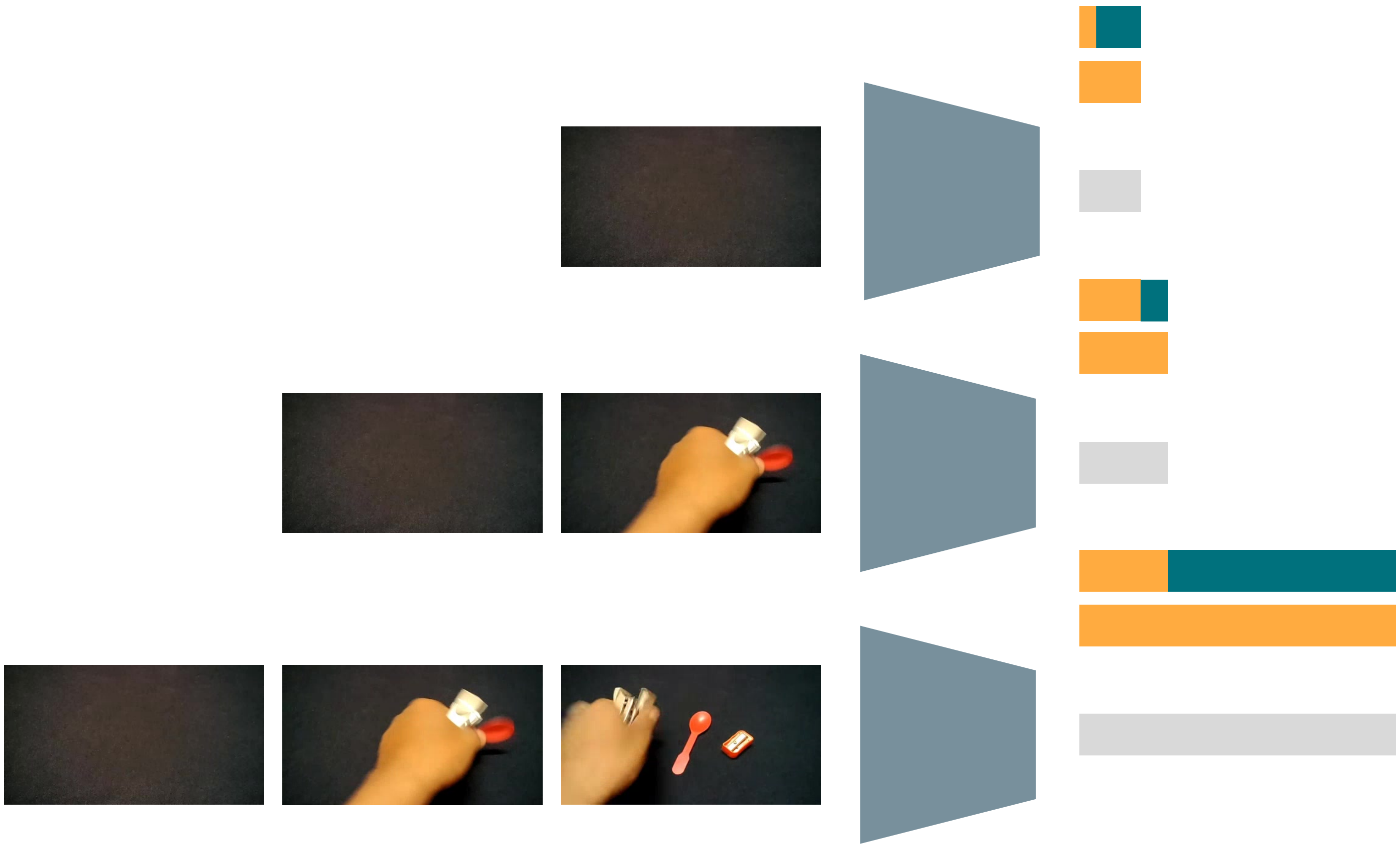
Frame Attributions in Video Models

with: Will Price



Frame Attributions in Video Models

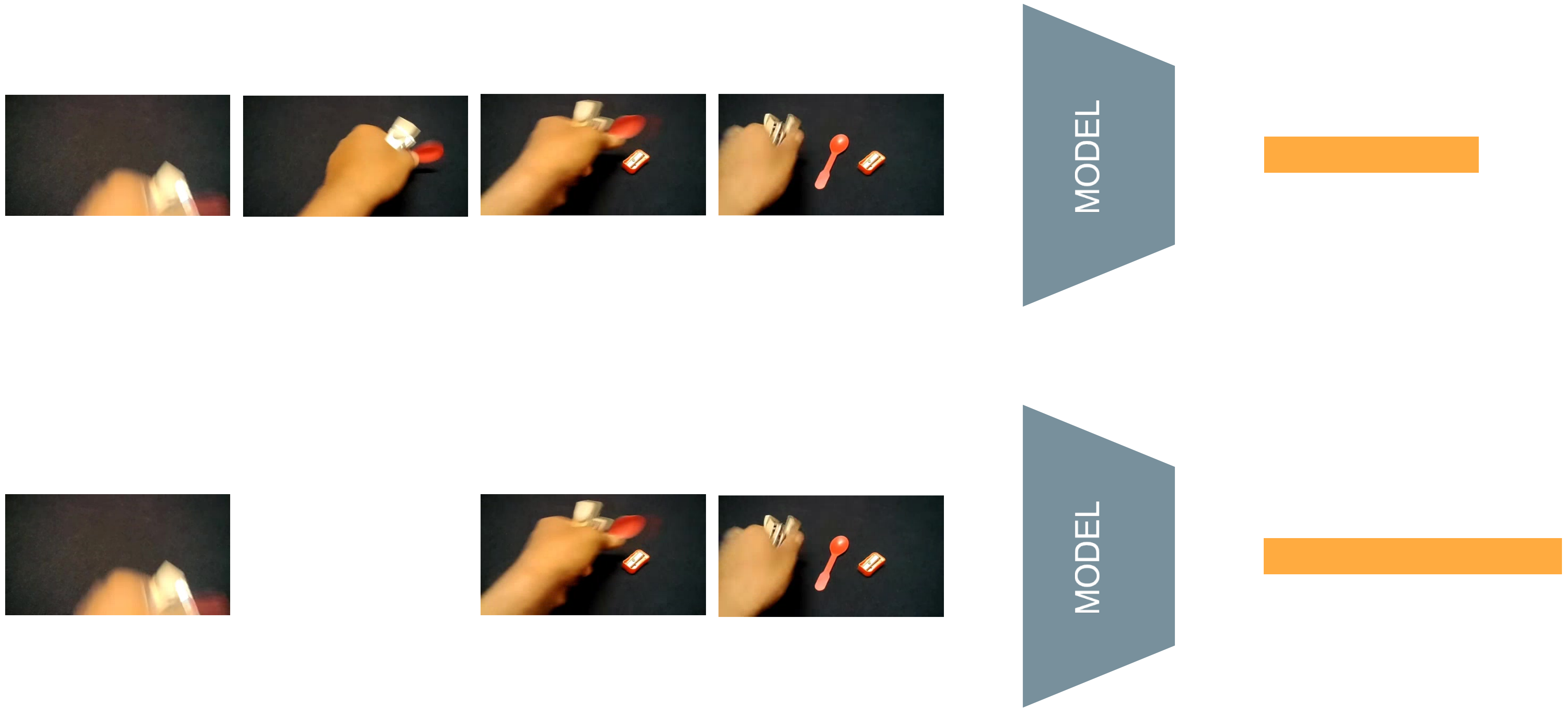
with: Will Price



W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

Frame Attributions in Video Models

with: Will Price

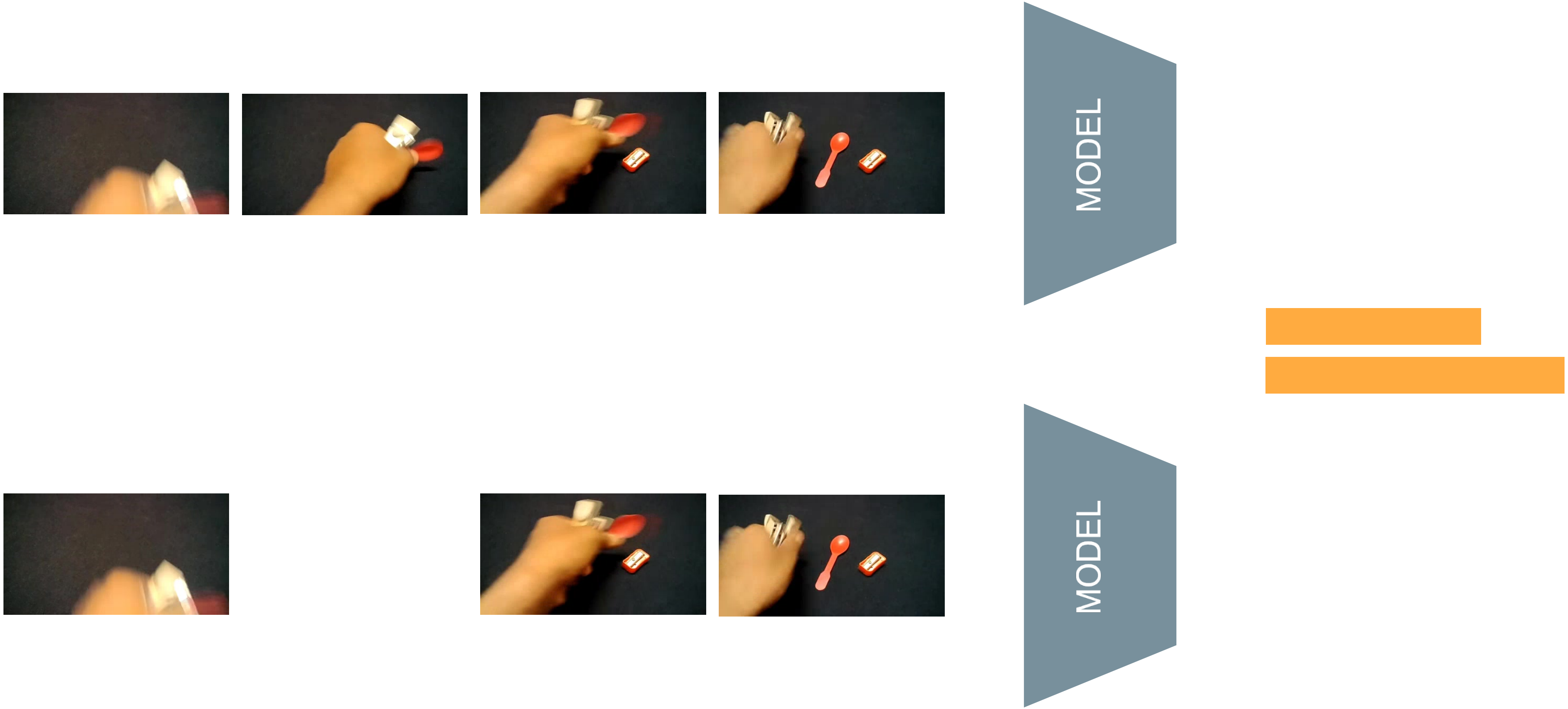


University of BRISTOL

W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

Frame Attributions in Video Models

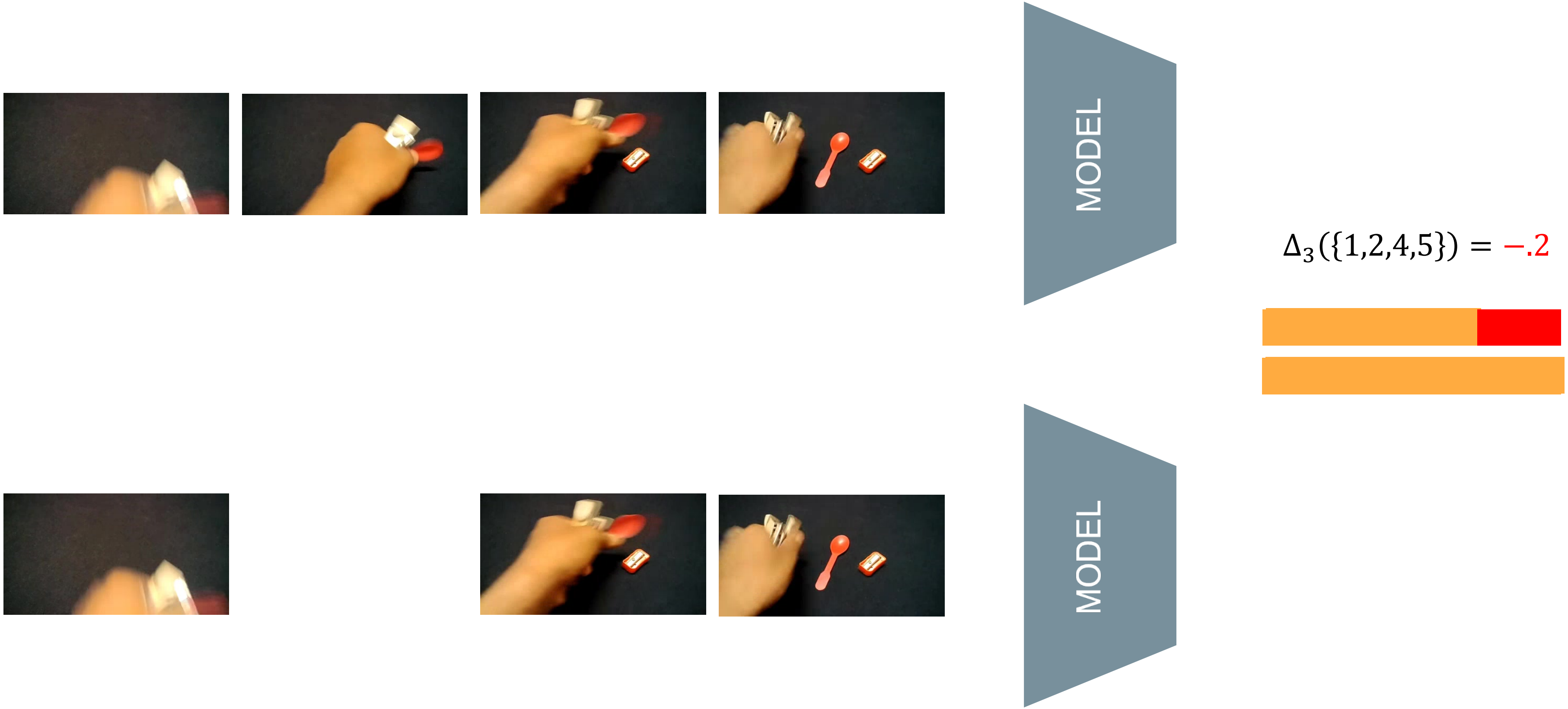
with: Will Price



W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

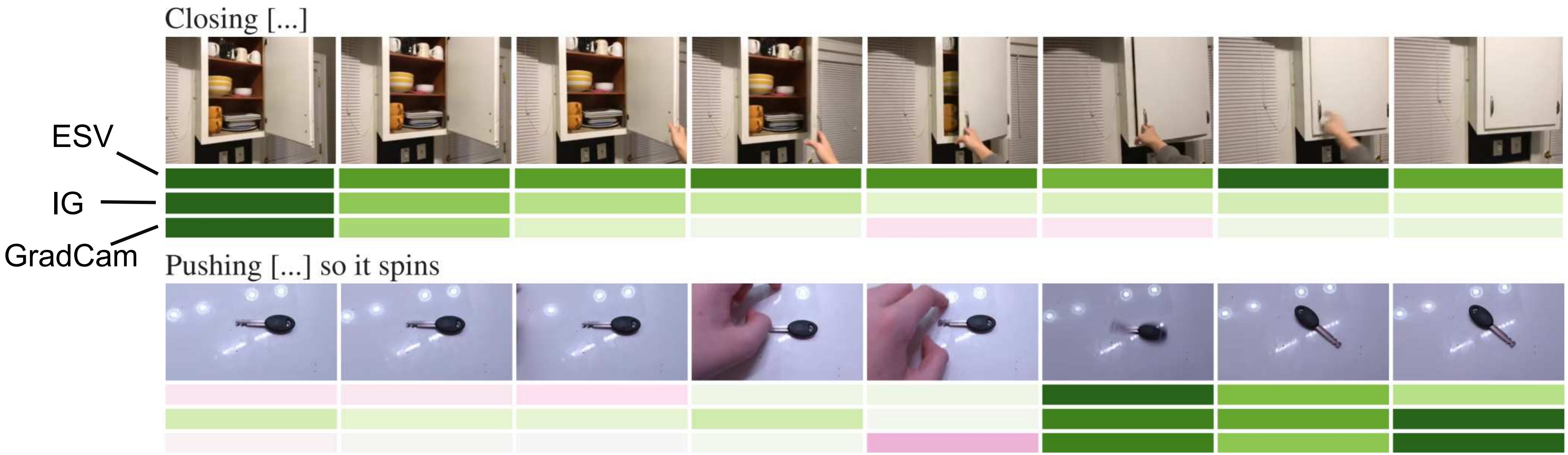
Frame Attributions in Video Models

with: Will Price



Frame Attributions in Video Models

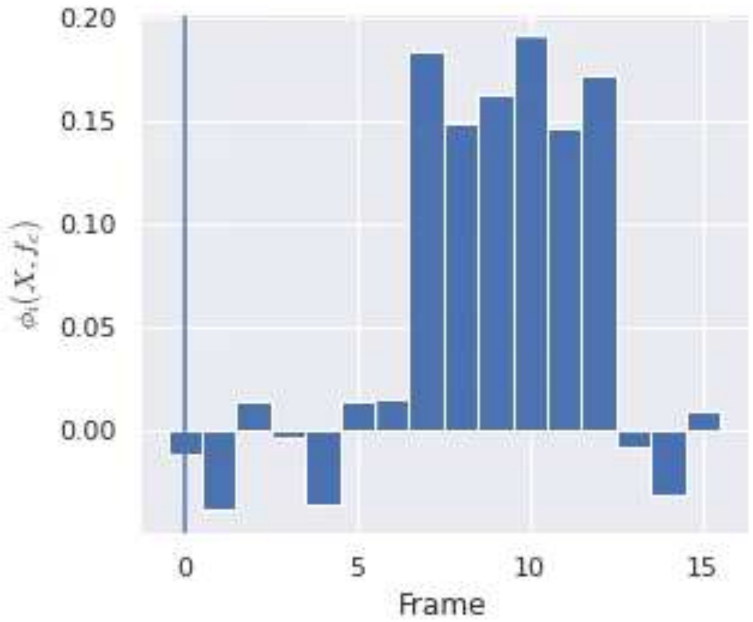
with: Will Price



W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

Frame Attributions in Video Models

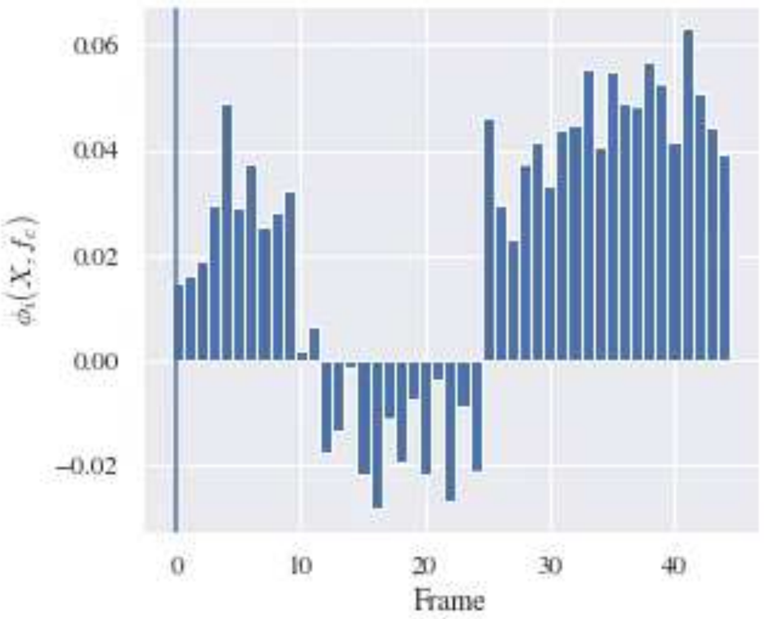
with: Will Price



Twisting (wringing) something wet until water comes out

Frame Attributions in Video Models

with: Will Price

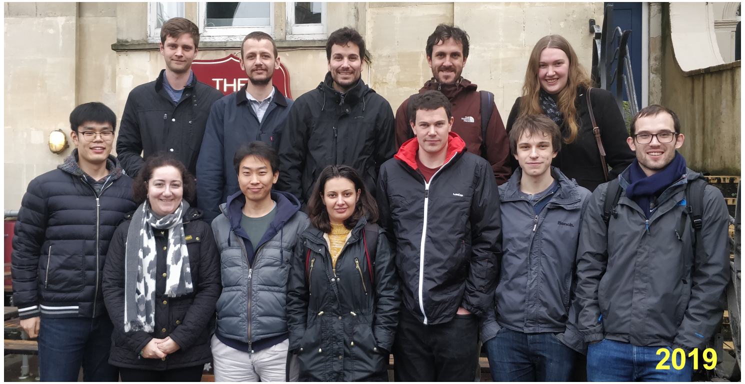


Showing that something is empty

W Price, D Damen (2020). Play Fair: Frame Attribution in Video Models. Asian Conference on Computer Vision (ACCV)

Dashboard

The Team



University of BRISTOL

Thank you

For further info, datasets, code, publications…

http://dimadamen.github.io

@dimadamen

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

Q&A