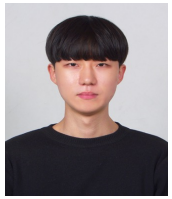
NeurIPS 2026

TraCE

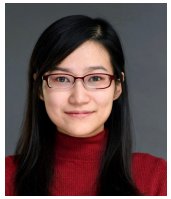
Trajectory-Aware **C**oncepts for **E**xplainable Video Understanding


Wooil Lee[1*]


Junpyo Hong[1*]


Jongseo Lee[1]


Tuomas Oikarinen[2]


Tsui-Wei Weng[2]


Jinwoo Choi[1†]

[1]Kyung Hee University

[2]UC San Diego

* Equal contribution † Corresponding author

Which Evidence Supports the Prediction?

Video models are accurate but opaque

Q What action is happening?



INPUT VIDEO

video AI model
BLACK BOX

ANSWER

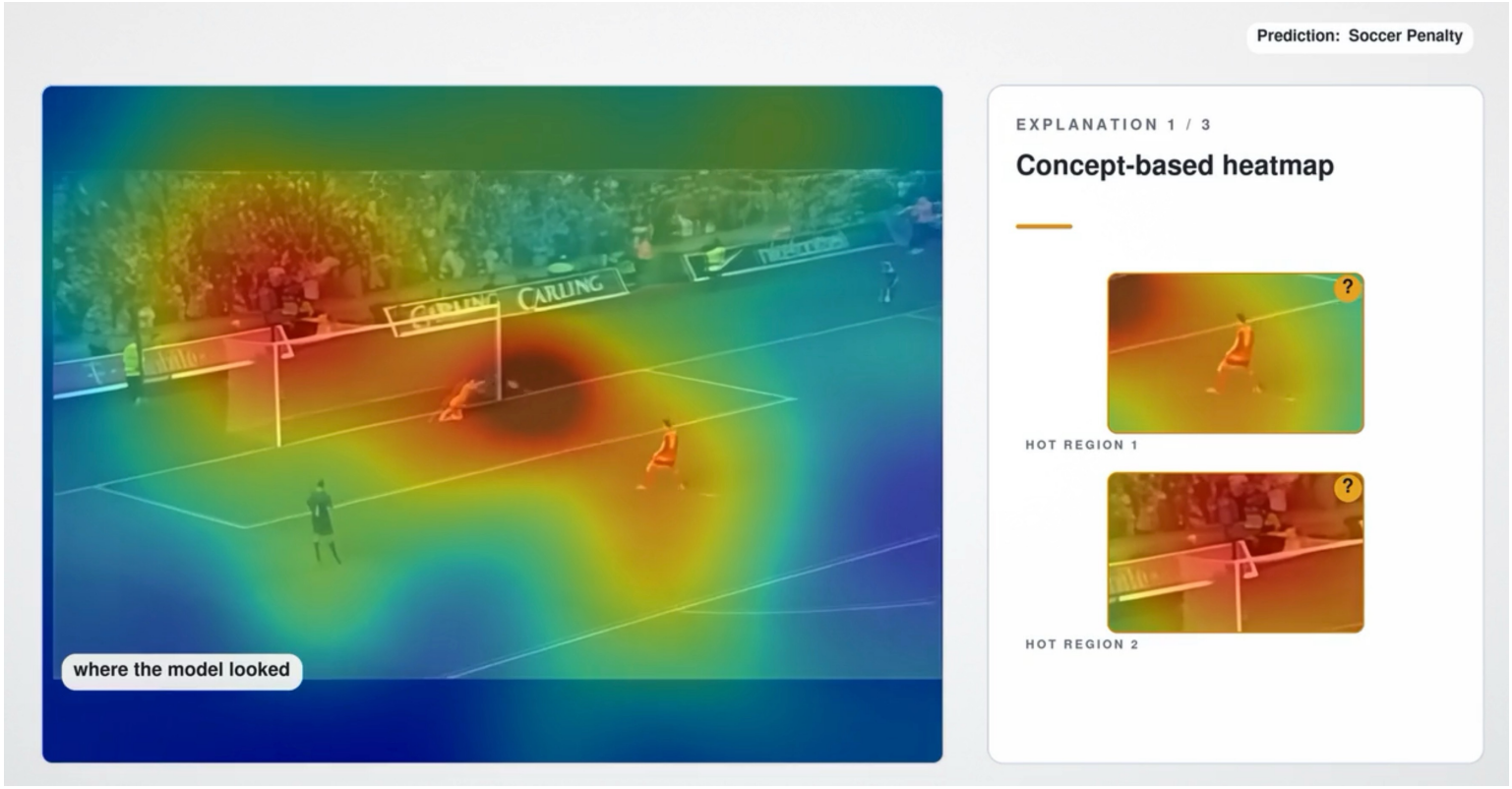
Soccer Penalty!

correct ✓

*Correct — but **why**? Let's open the black box.*

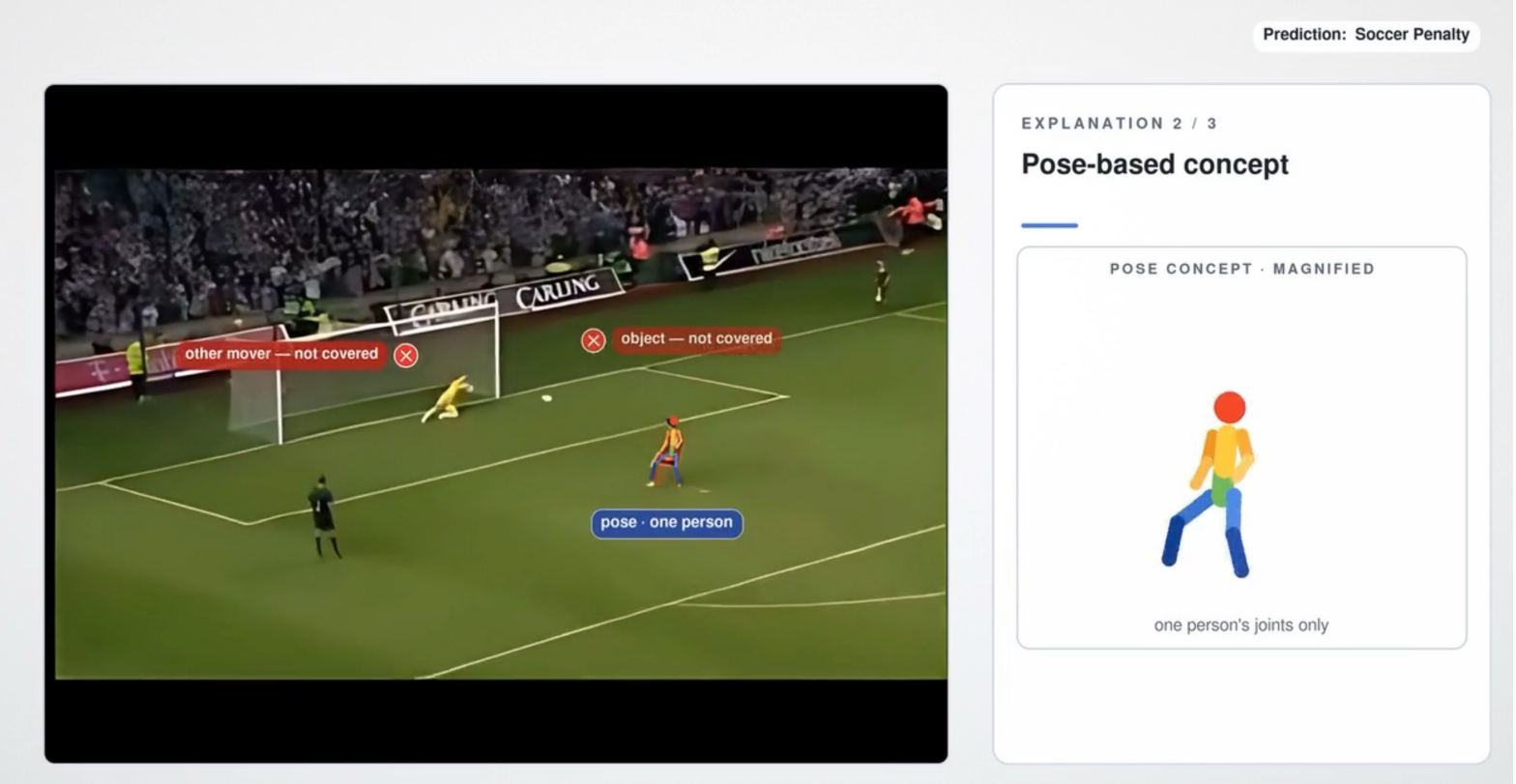
Heatmap explanations are source-ambiguous

a **Concept-based heatmap**



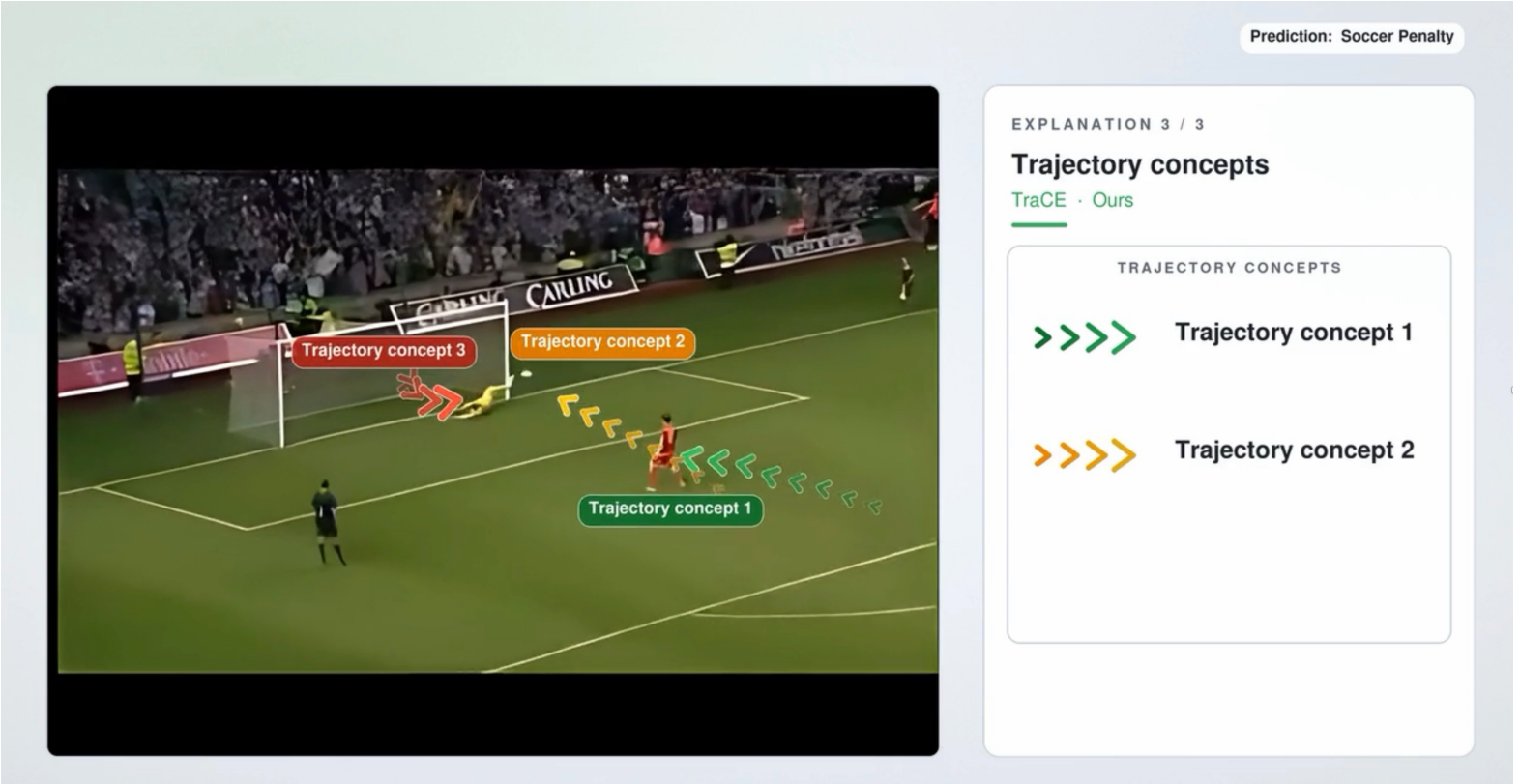
Pose-based concepts are source-incomplete

b Pose-based concept

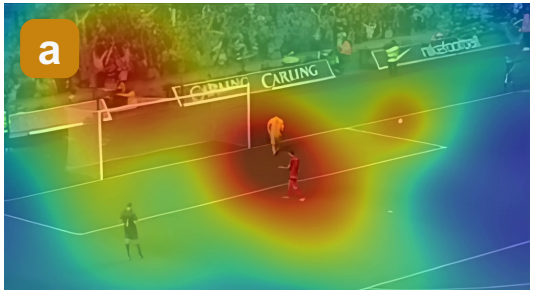


Trajectory concepts are source-flexible

C **Trajectory-based concept** **Ours**



Existing explanations under-specify the motion source



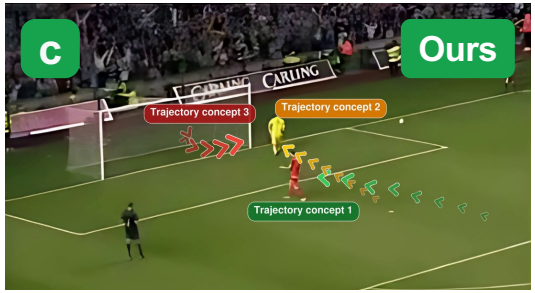
Concept-based heatmap

Source-ambiguous



Pose-based concept

Source-incomplete



Trajectory-based concept

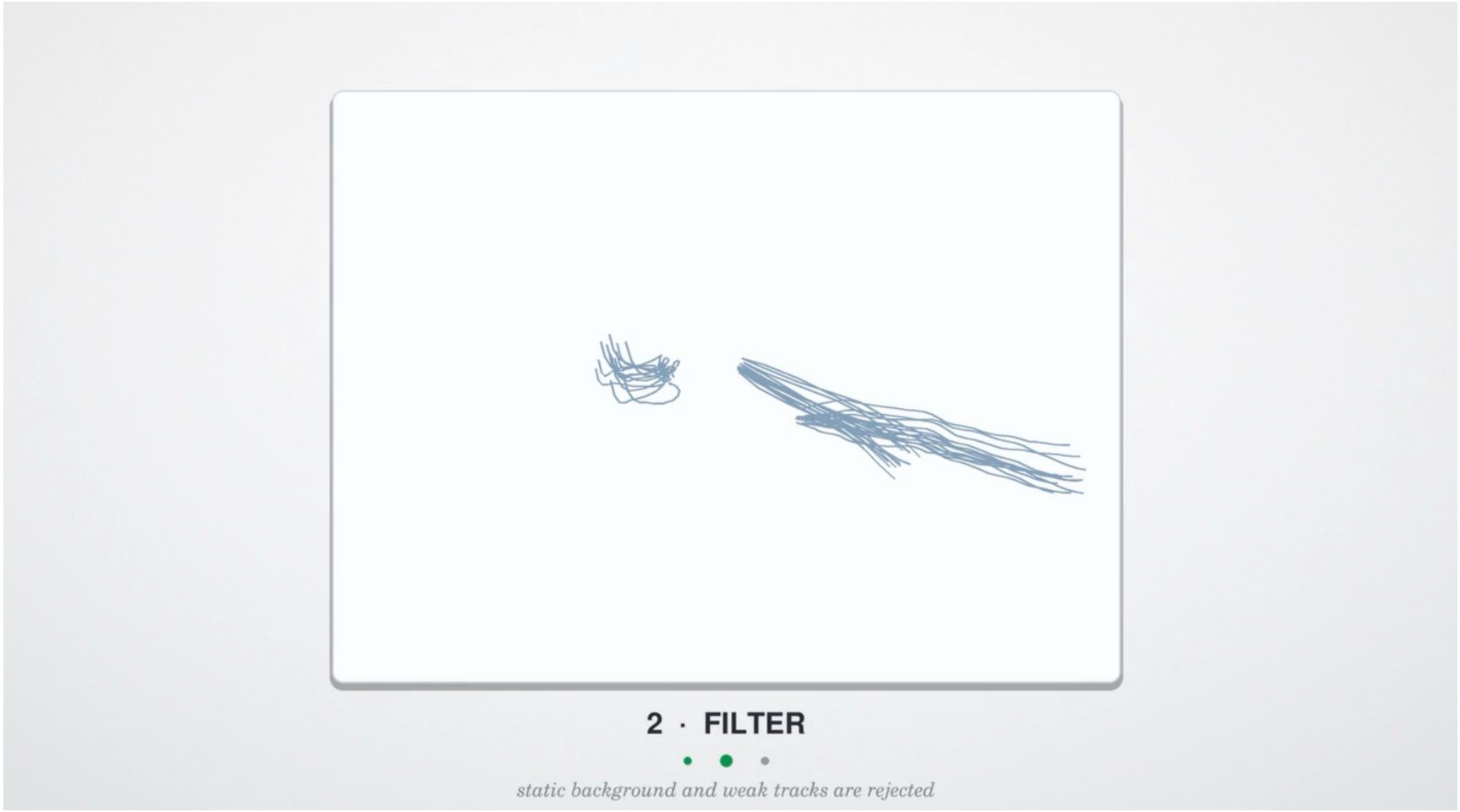
Source-flexible

We call this motion-source under-specification.
TraCE *addresses it with trajectory concepts!*

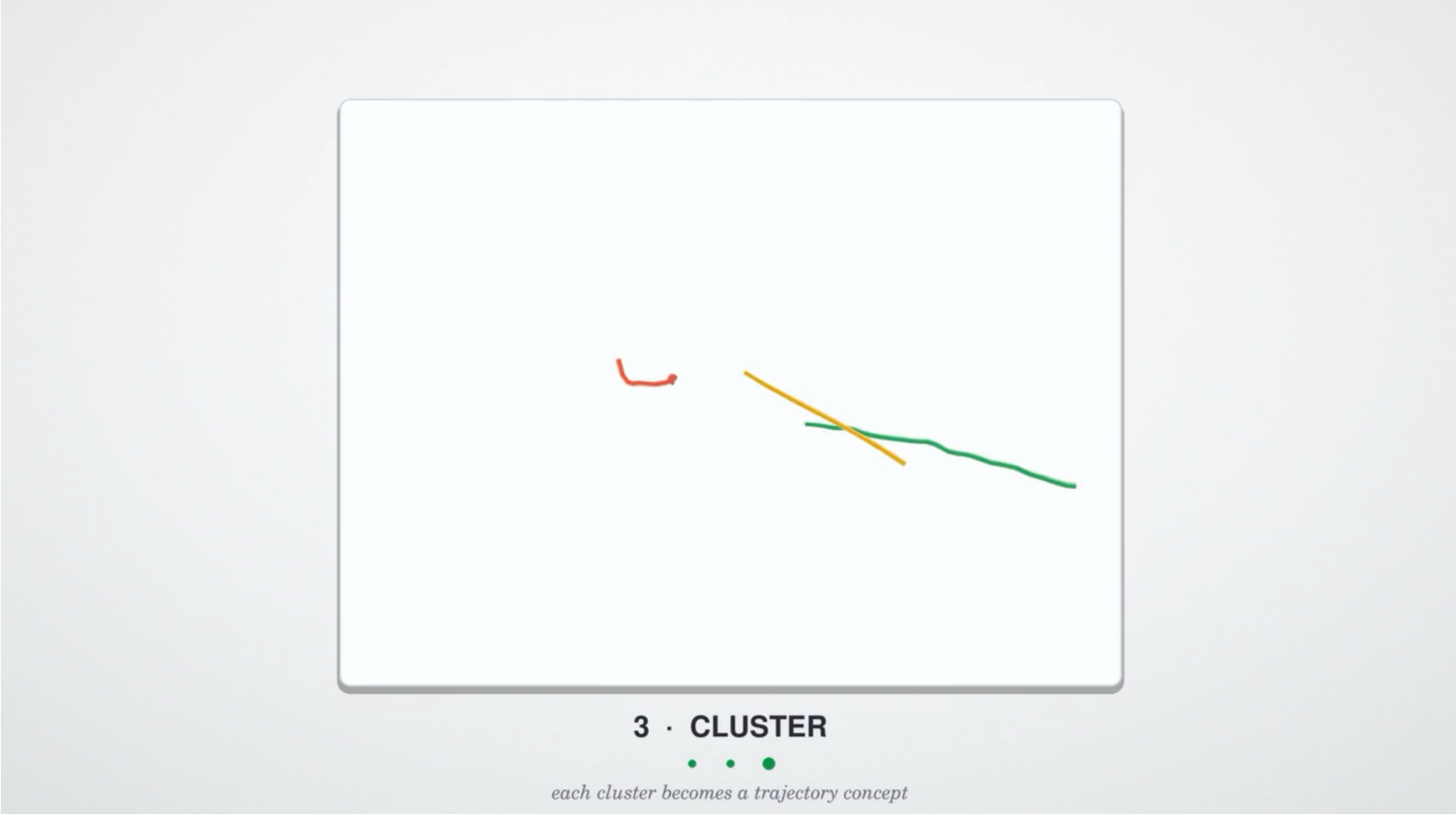
From optical flow to trajectory concepts



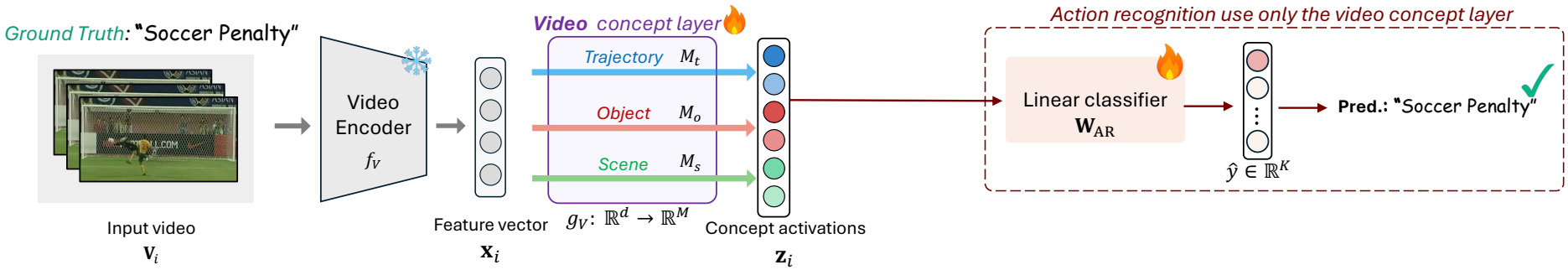
From optical flow to trajectory concepts



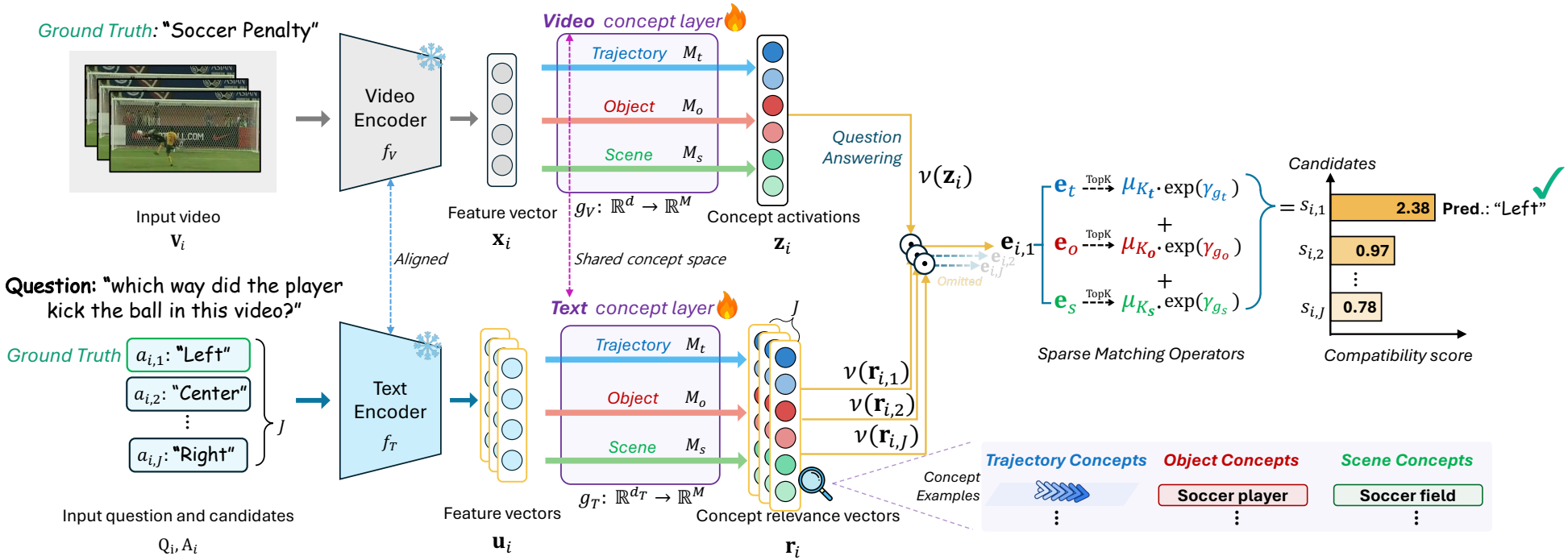
From optical flow to trajectory concepts



AR: predict directly from concepts



VQA: match answers with video concepts

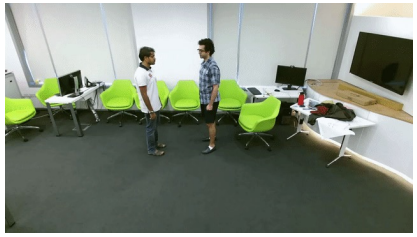


For VQA, each candidate answer is turned into the concepts it needs.
We score it by how much those concepts overlap with the video's concepts.

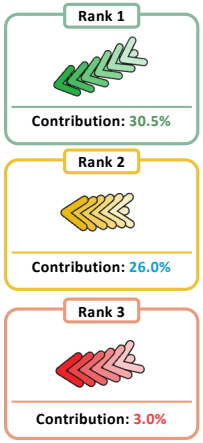
Concept contributions explain AR and VQA predictions

Sample-level AR explanation

Pred : *"Pushing other person"*



(a) Multi-agent Interaction

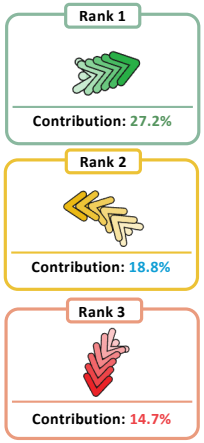


Top-3 contributing concepts

Pred : *"Moving [something] and [something] closer to each other"*

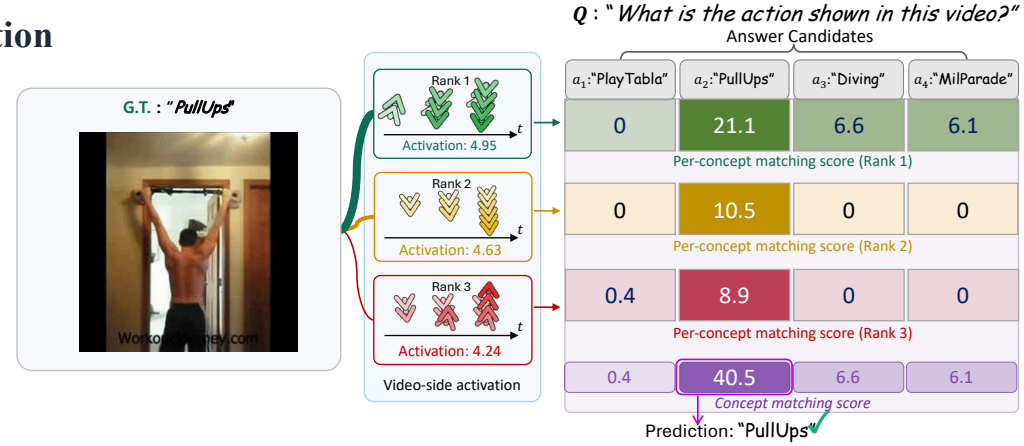


(b) Object-centric Dynamics

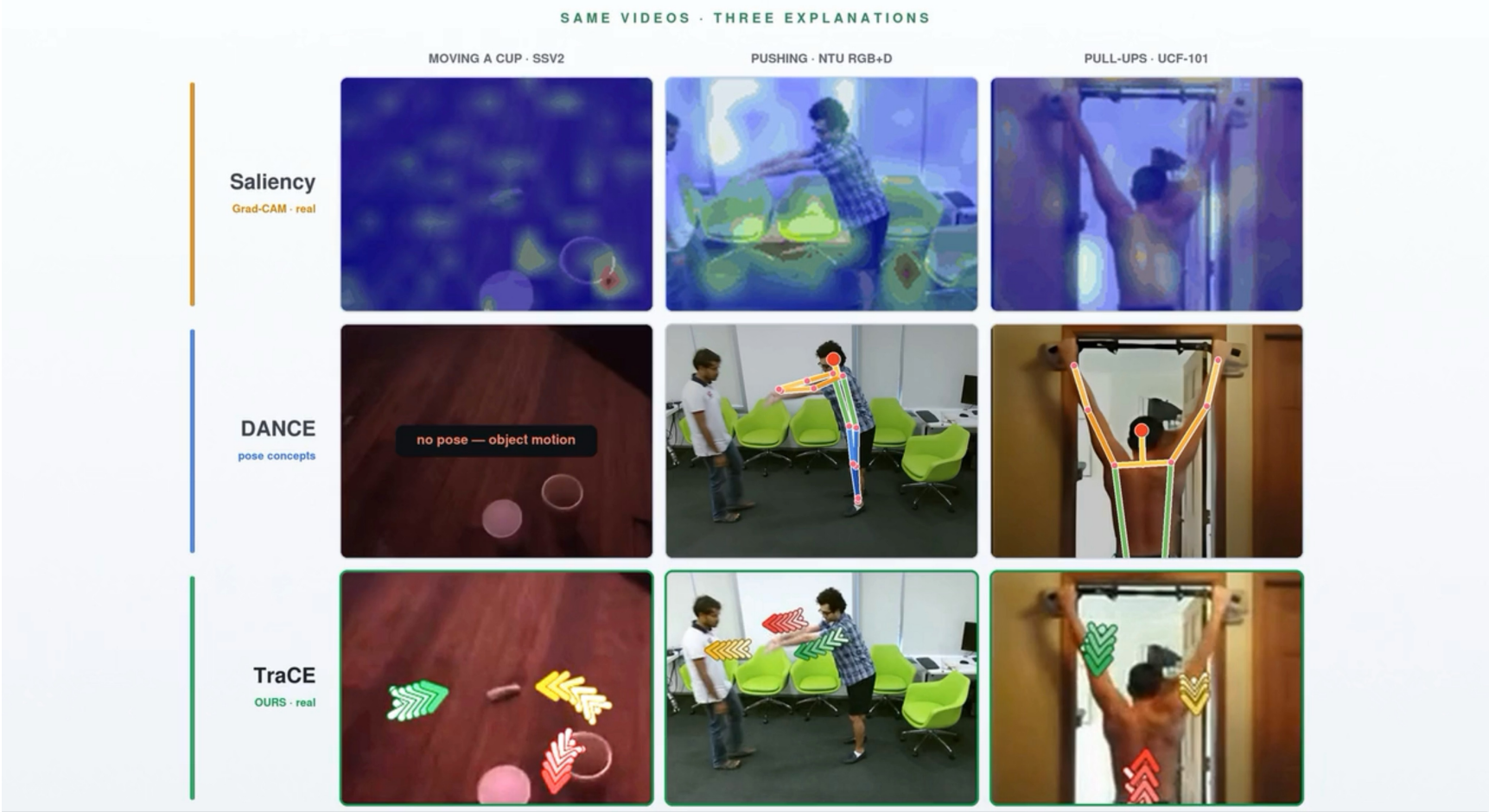


Top-3 contributing concepts

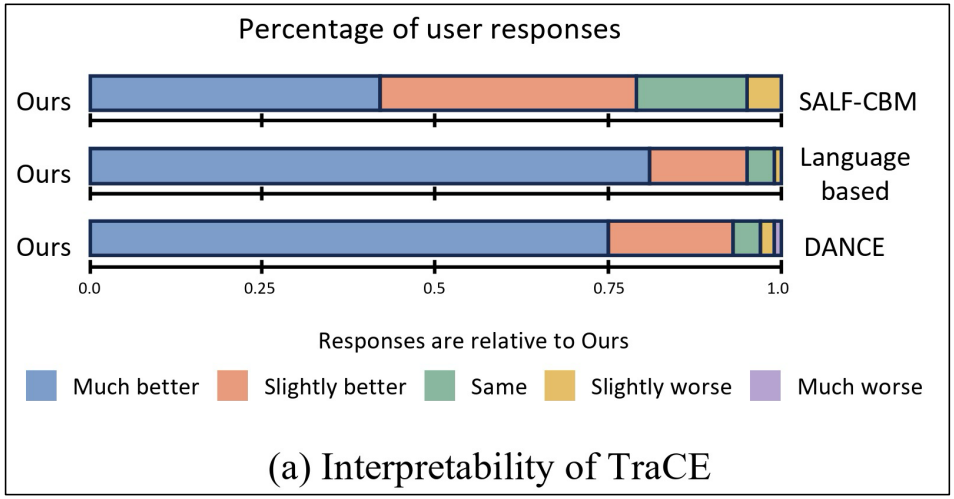
Sample-level VQA explanation



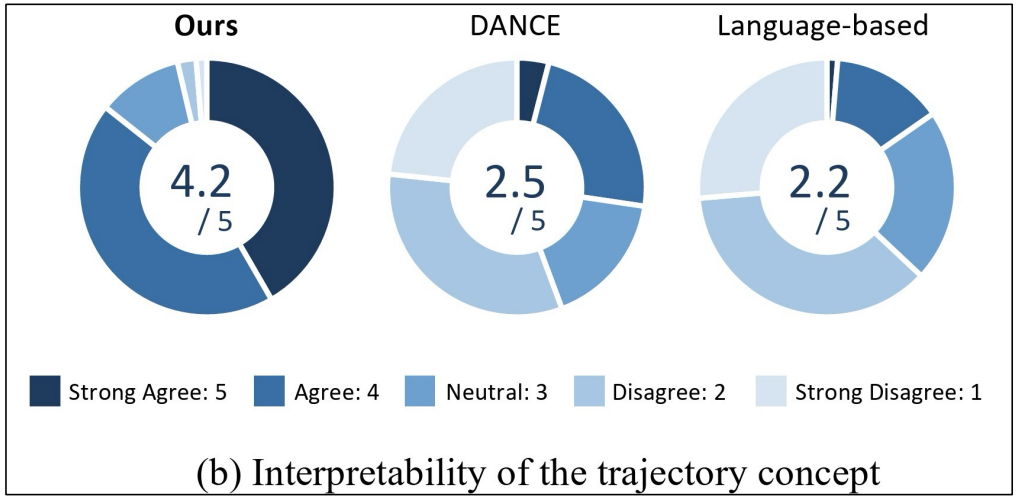
Qualitative evaluation



Users prefer TraCE's trajectory-based explanations



(a) Interpretability of TraCE



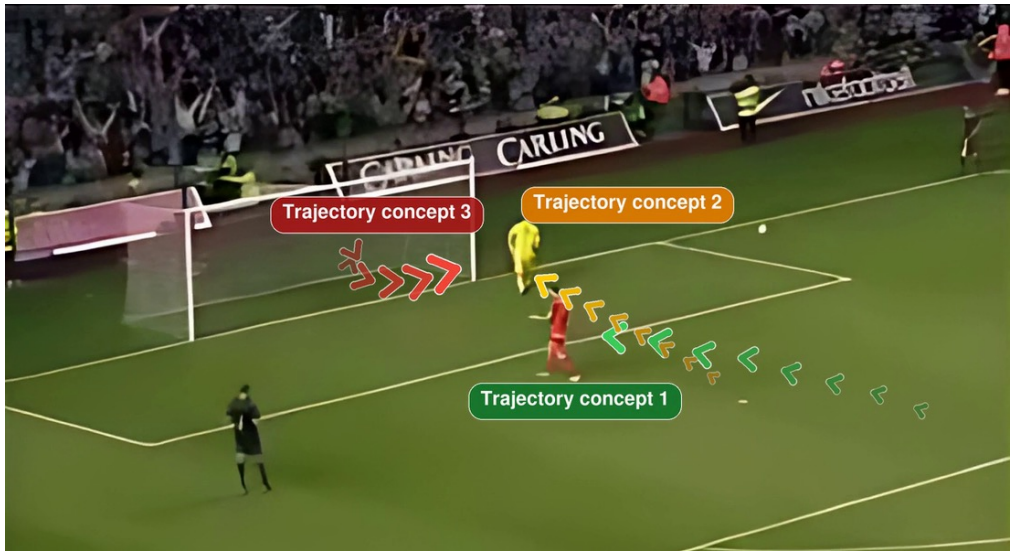
(b) Interpretability of the trajectory concept

TraCE preserves accuracy while adding interpretability

Method	NTU RGB+D Mutual		Chiral SSv2		UCF-101		KTH	TOMATO	STAR	Avg.	
	AR	VQA	AR	VQA	AR	VQA	AR	VQA	VQA	AR	VQA
Baseline w/o interpretability	99.7±0.2	97.4±0.1	86.6±0.2	82.6±0.2	89.8±0.1	99.8±0.0	91.1±0.3	35.9±2.2	54.5±0.4	91.8	74.0
LF-CBM [40] w/ language concepts	96.9±0.3	–	77.6±0.5	–	88.3±0.2	–	89.1±0.7	–	–	88.0	–
DANCE [26]	97.1±0.3	–	81.7±0.3	–	89.0±0.2	–	91.3±0.4	–	–	89.8	–
SALF-CBM [7]	99.7±0.0	–	80.4±0.4	–	87.5±0.2	–	87.9±0.4	–	–	88.9	–
SpLiCE [9] adaptation	–	53.0	–	57.3	–	86.7	–	22.4	49.2	–	53.7
`TraCE`	97.4±0.1	97.1±0.3	84.9±0.4	82.6±0.6	89.0±0.3	99.7±0.1	91.8±0.8	37.0±1.7	55.4±1.2	90.8	74.4

Best average accuracy among interpretable models

Method	NTU RGB+D Mutual		Chiral SSv2		UCF-101		KTH	TOMATO	STAR	Avg.	
	AR	VQA	AR	VQA	AR	VQA	AR	VQA	VQA	AR	VQA
Baseline w/o interpretability	99.7±0.2	97.4±0.1	86.6±0.2	82.6±0.2	89.8±0.1	99.8±0.0	91.1±0.3	35.9±2.2	54.5±0.4	91.8	74.0
LF-CBM [40] w/ language concepts	96.9±0.3	–	77.6±0.5	–	88.3±0.2	–	89.1±0.7	–	–	88.0	–
DANCE [26]	97.1±0.3	–	81.7±0.3	–	89.0±0.2	–	91.3±0.4	–	–	89.8	–
SALF-CBM [7]	99.7±0.0	–	80.4±0.4	–	87.5±0.2	–	87.9±0.4	–	–	88.9	–
SpLiCE [9] adaptation	–	53.0	–	57.3	–	86.7	–	22.4	49.2	–	53.7
`TraCE`	97.4±0.1	97.1±0.3	84.9±0.4	82.6±0.6	89.0±0.3	99.7±0.1	91.8±0.8	37.0±1.7	55.4±1.2	90.8	74.4



TraCE

Trajectory concepts make motion sources flexible.

Project page

leewooil.github.io/TraCE_project_page