Dual-AI: Dual-path Actor Interaction Learning for Group Activity Recognition

Mingfei Han[†*]

David Junhao Zhang[†*]

Yali Wang[*]

Rui Yan

Lina Yao

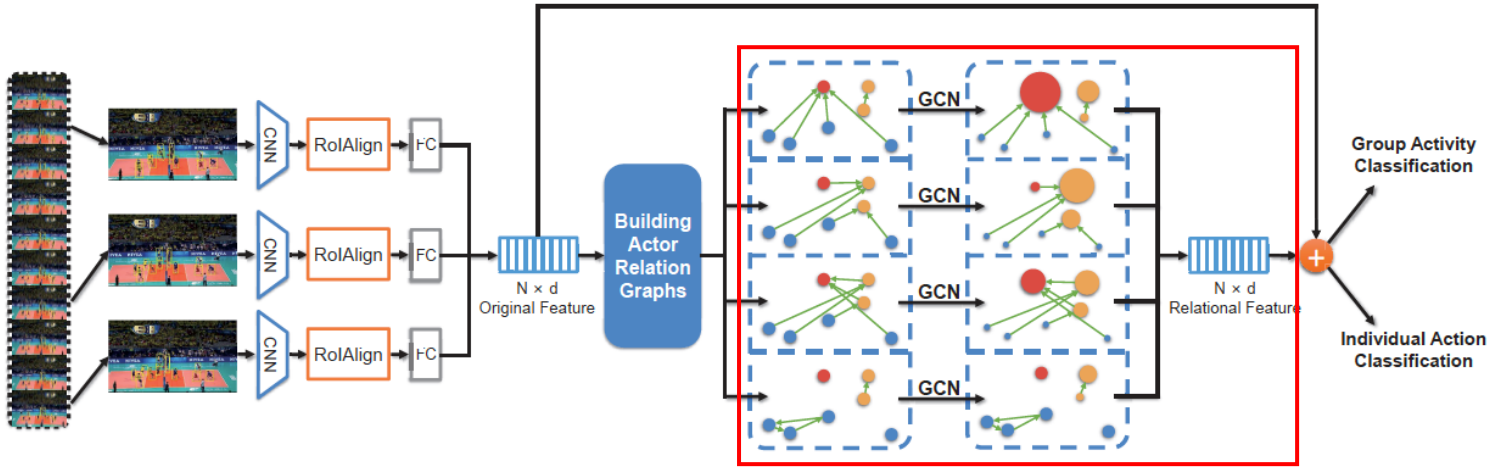
Xiaojun Chang

Yu Qiao✉

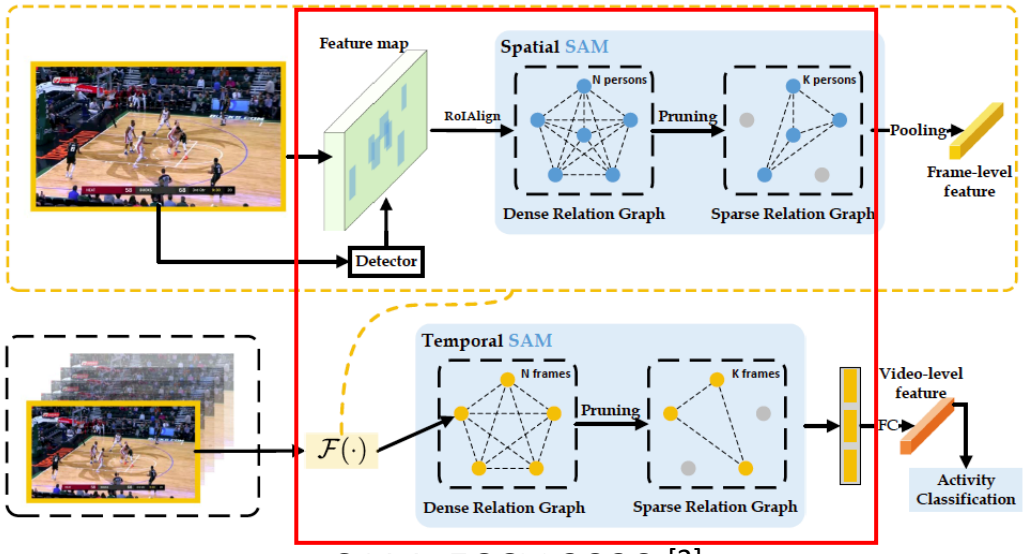
* Equal contribution; † Work was done while interning at SIAT

Group Activity Recognition - Case

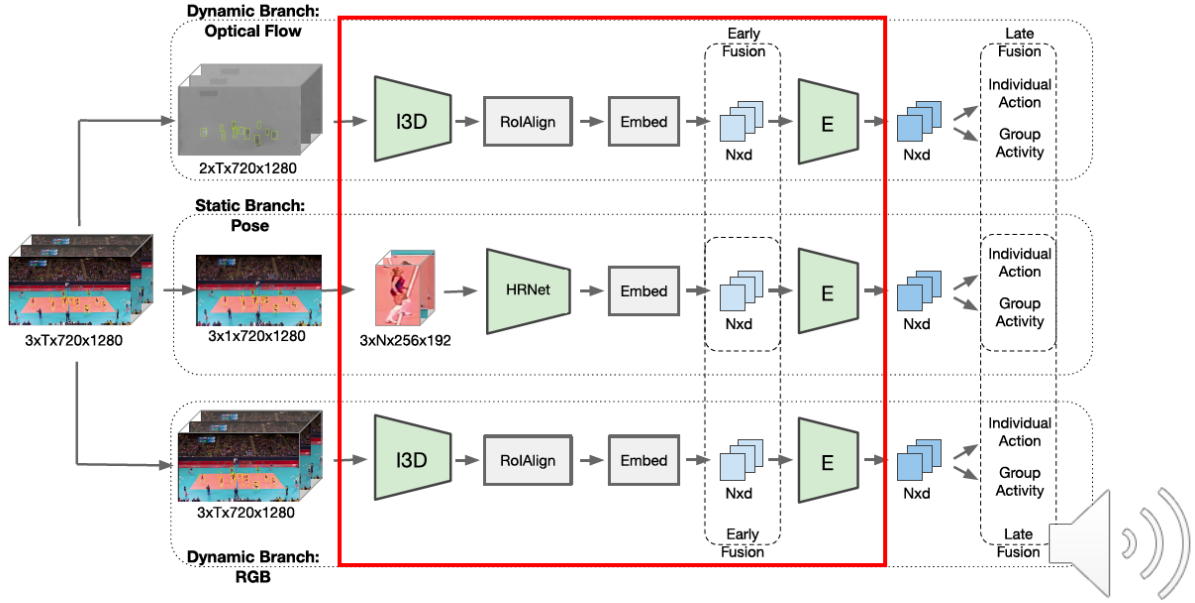
Recent Approaches – GCN, CNN, Transformer



ARG, CVPR 2019 [4]

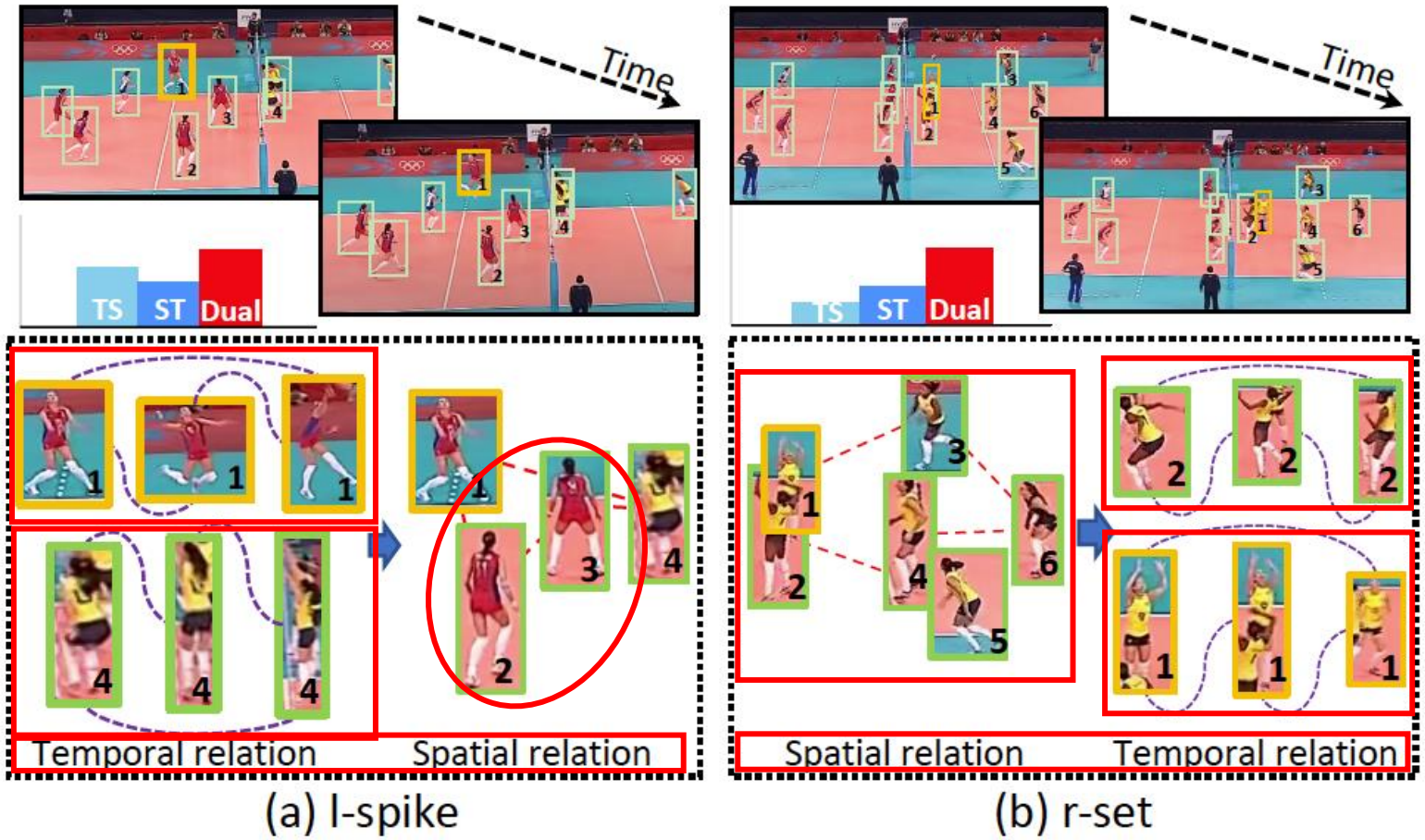


SAM, ECCV 2020 [3]



AFormer, CVPR 2020 [5]

Motivation – Different ability of TS and ST



(a). Hitting + blocking + waiting

↓

side A Hitting

side B blocking

↓

left-spike

(b). Action scene side B

↓

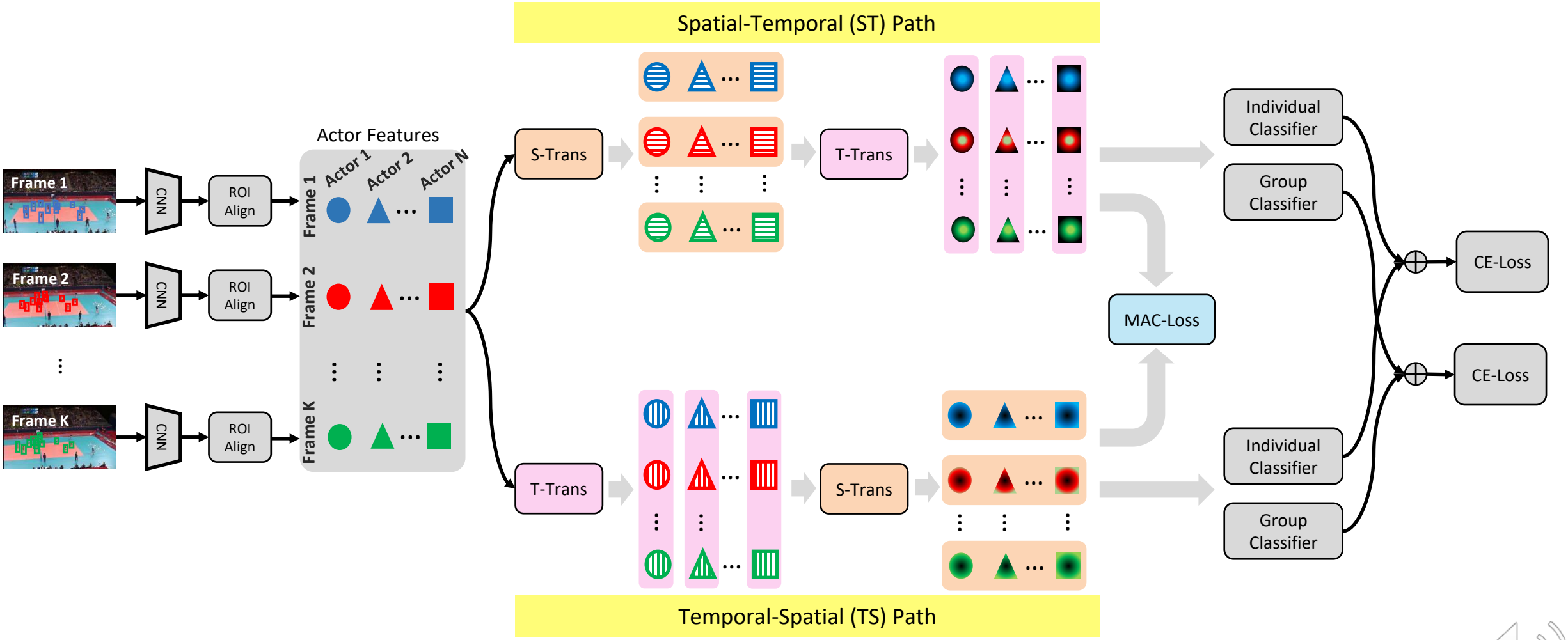
Jumping (fake spike) + Jumping (setting)

↓

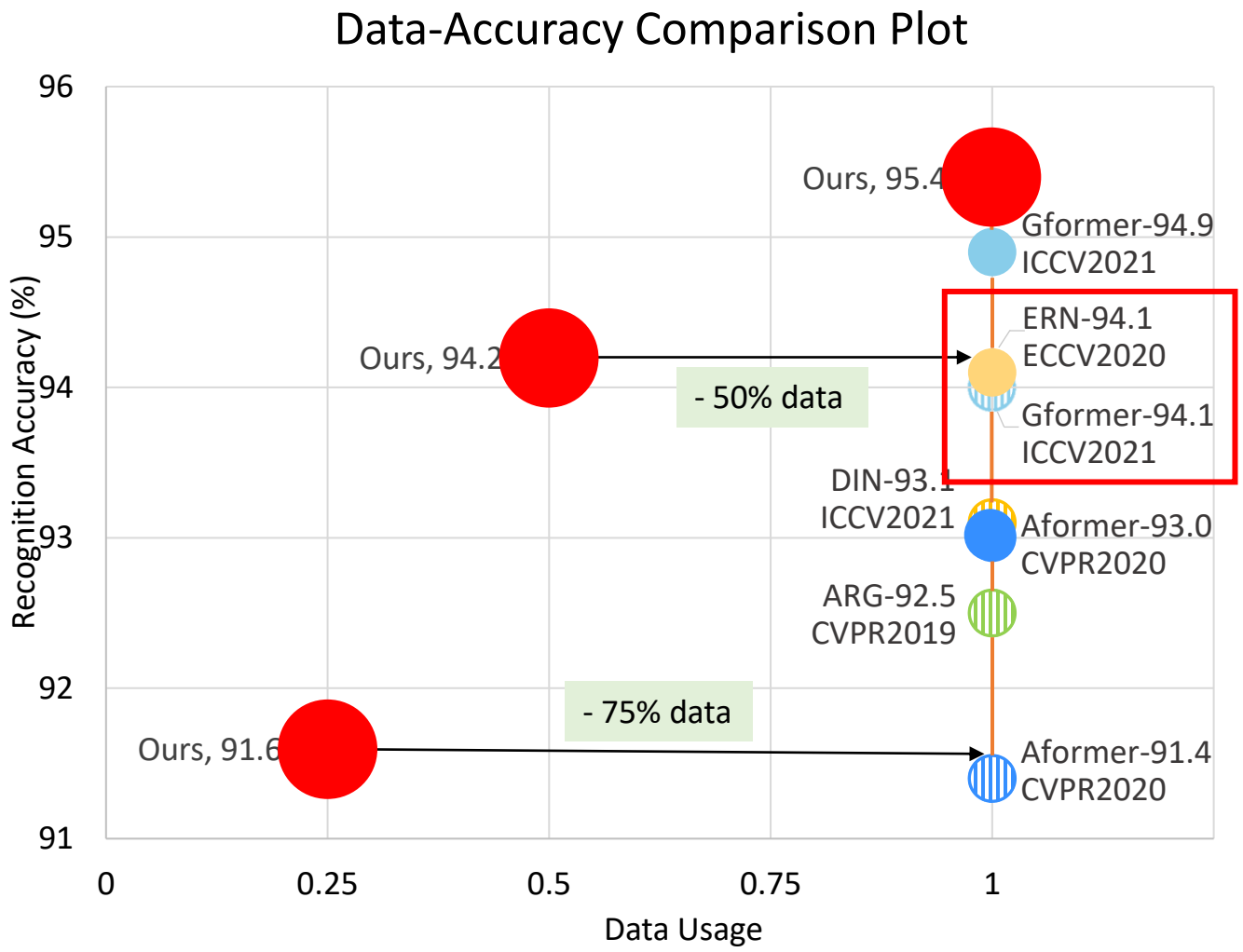
right-set

* TS: Temporal Spatial, ST: Spatial Temporal

Our Approach – Dual Spatiotemporal Paths



Experiments –SOTA comparison and Robustness



Solid point means result with additional optical flow input

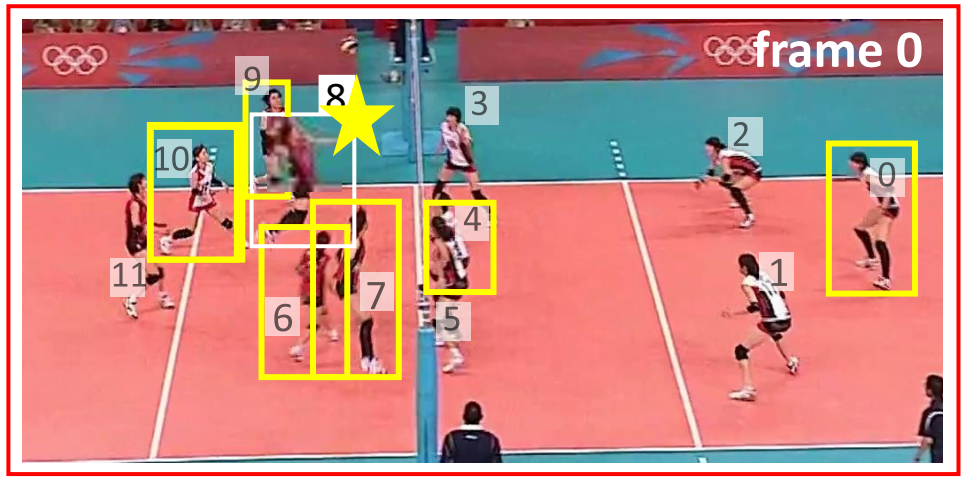
- Our method achieves current SOTA performance on Volleyball dataset.
- Our method achieves 94.2% with 50% data, which is competitive to a number of recent approaches trained with 100% data.

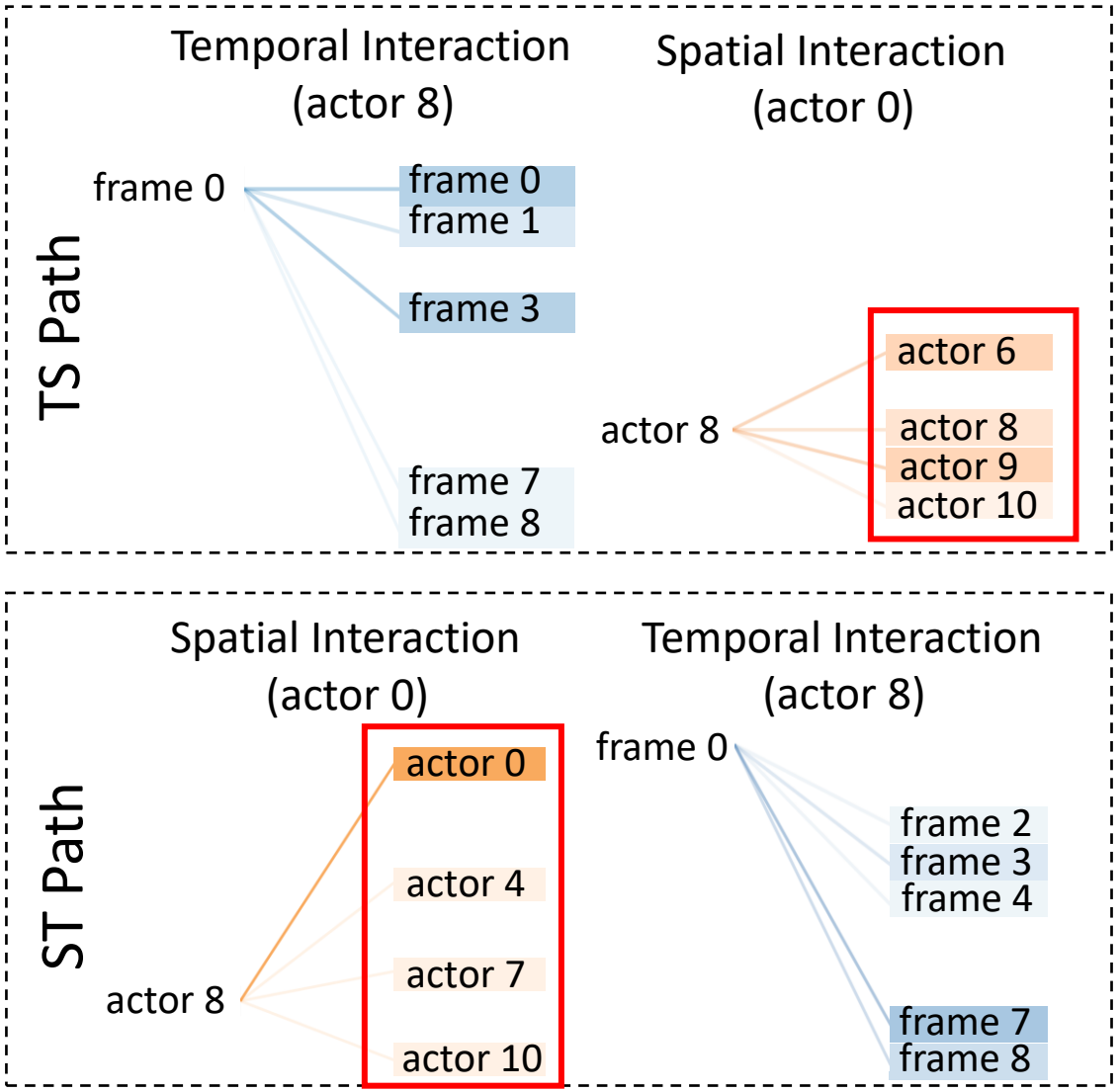
Experiments – Robustness (SOTA comparison with Weak Supervision)

- Without bells and whistles, our method surpasses all the existing methods by a good margin, establishing new state-of-the-art results.

Method	Backbone	Mod-ality	NBA Acc./Mean Acc.	Weak Vlb. -M Acc.
TSN* [40]	Incep-v1	RGB	– / 37.8	–
I3D* [10]	I3D	RGB	– / 32.7	–
Nlocal* [44]	I3D-NLN	RGB	– / 32.3	–
ARG* [46]	Incep-v3	RGB	– / –	90.7
SAM [49]	Res-18	RGB	– / –	93.1
SAM [49]	Incep-v3	RGB	49.1 / 47.5	94.0
Ours	Incep-v3	RGB	51.5 / 44.8	95.8
	Incep-v3	Flow	56.8 / 49.1	96.1
	Incep-v3	Fusion	**58.1 / 50.2**	**96.5**

Experiments – Visualization





Thank you

Project page: mingfei.info/Dual-AI