Why Can't I Dance in the Mall? Learning to Mitigate Scene Bias in Action Recognition



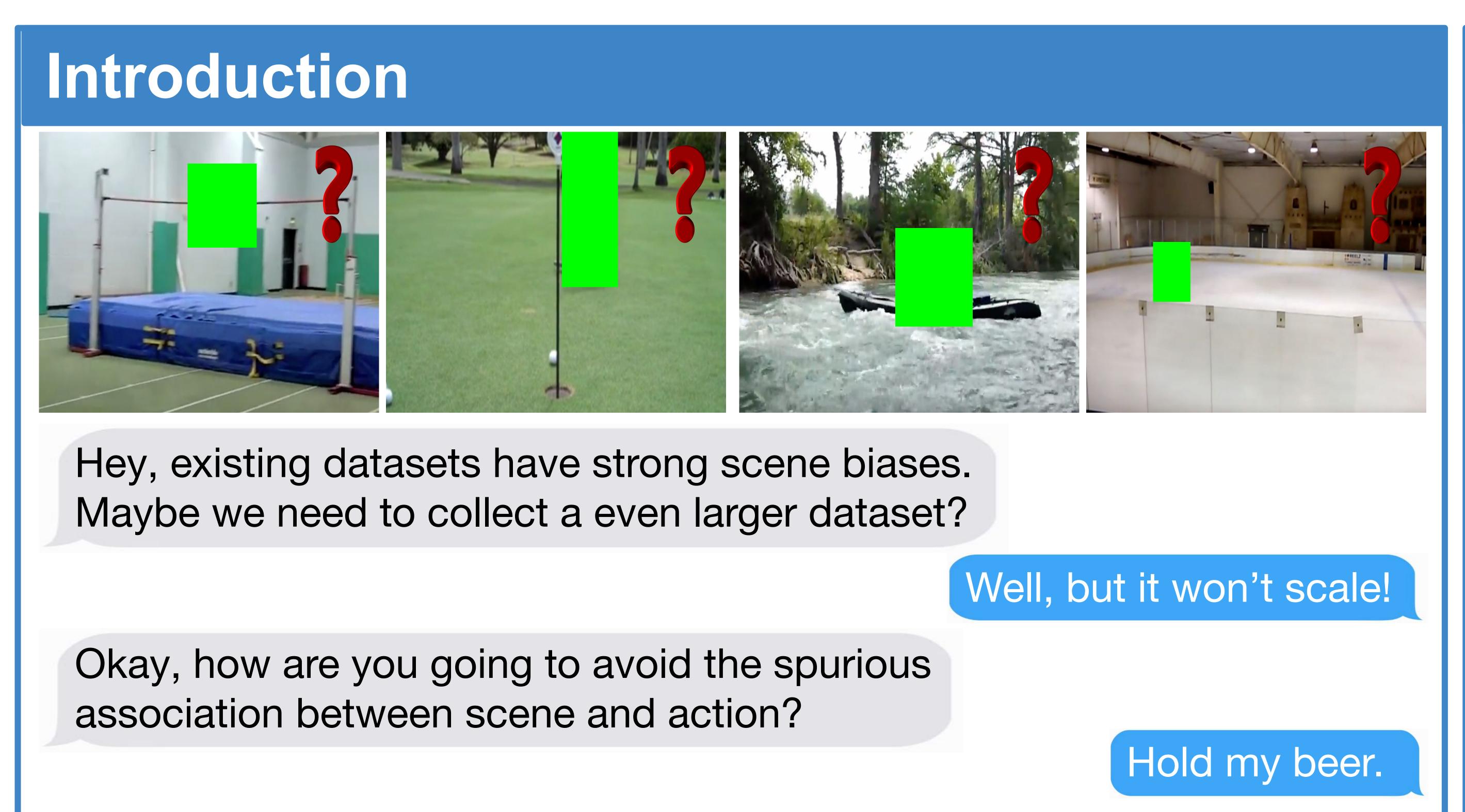
Jinwoo Choi

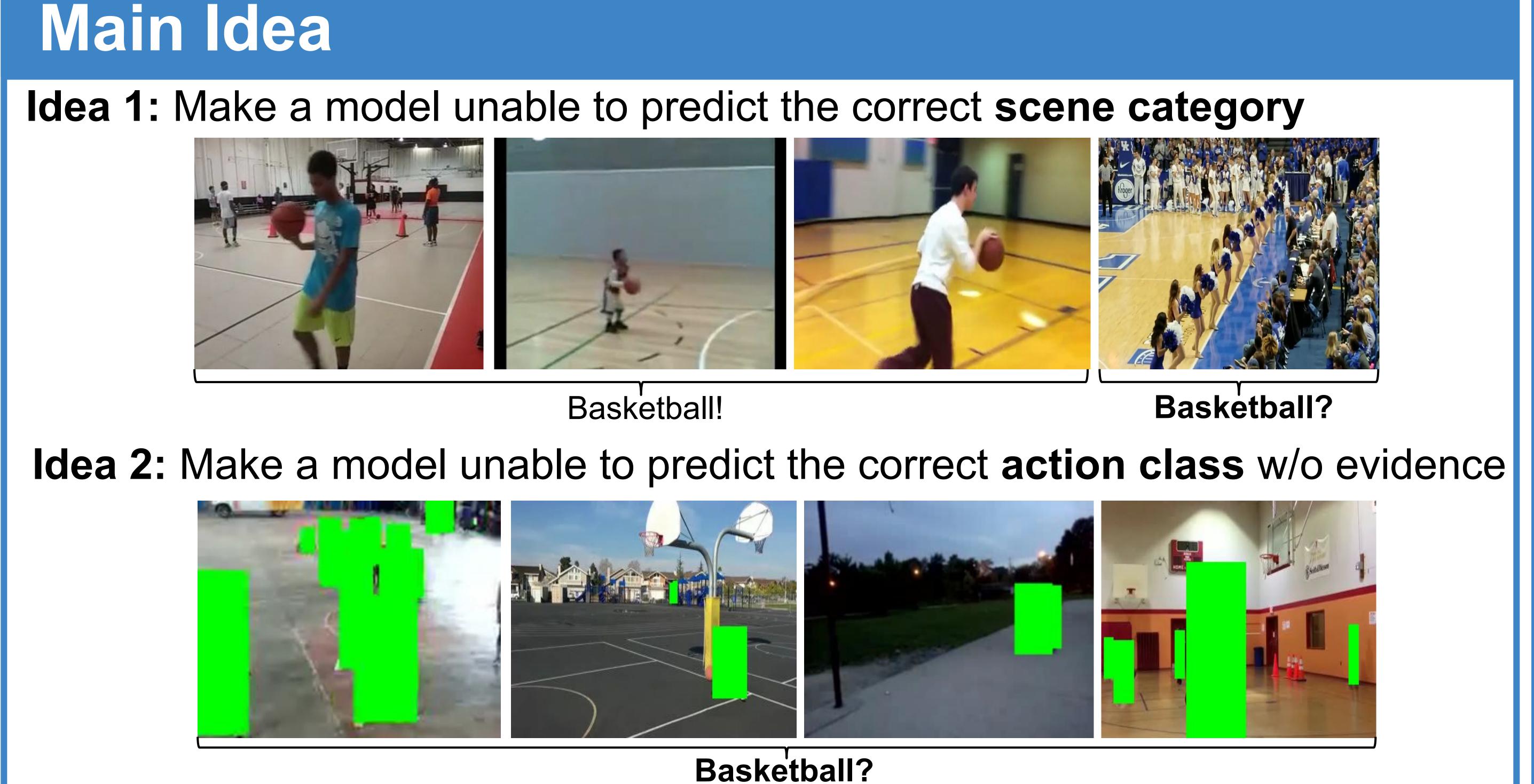
Joseph C. E. Mesou Jia-Bin Huang Chen Gao Virginia Tech

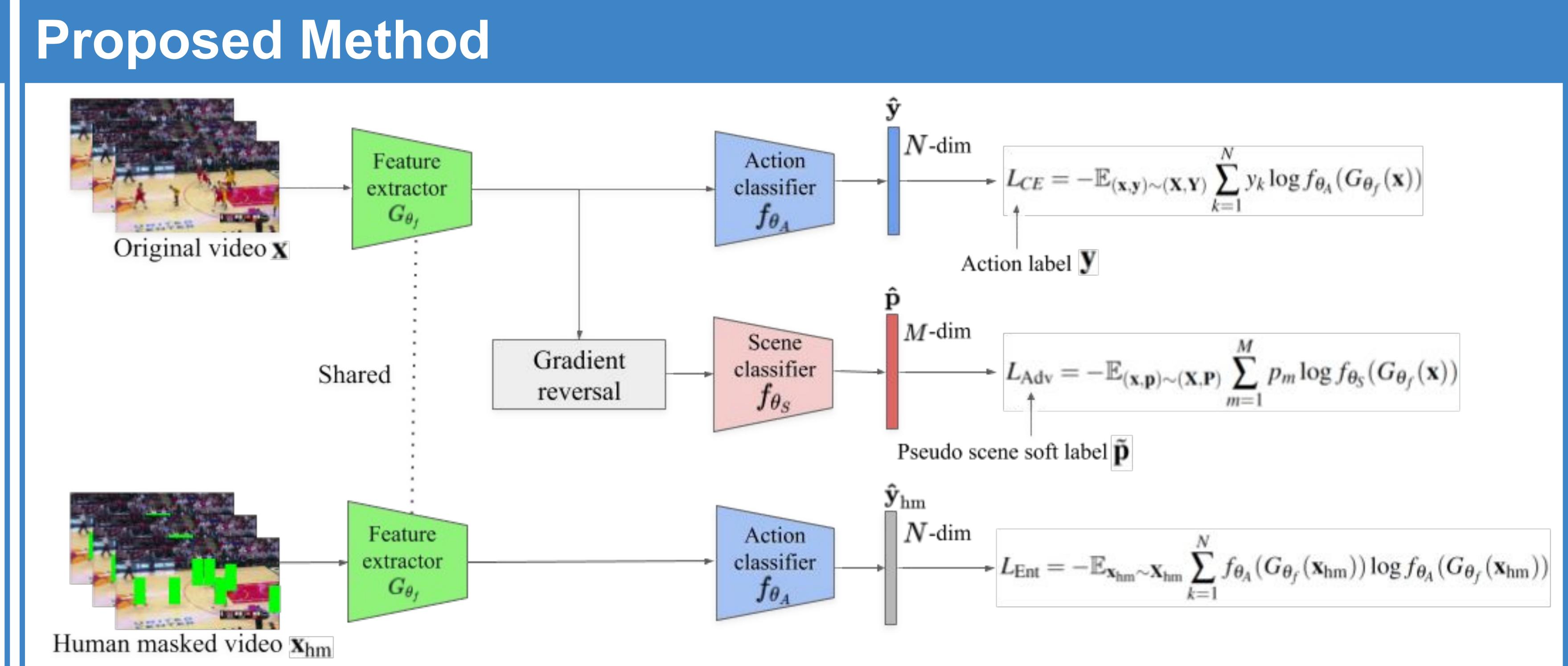
Ablation Studies

Code available at:
http://bit.ly/scene_debias

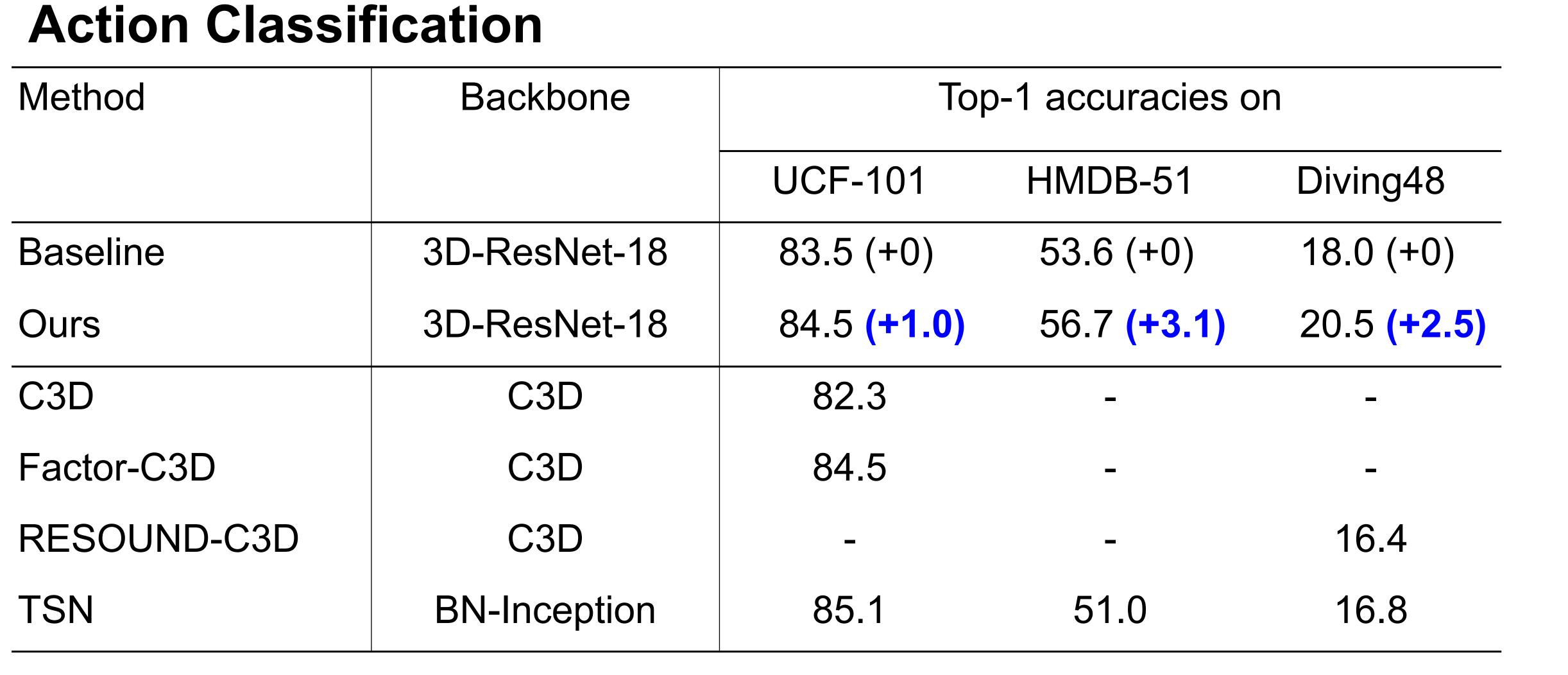


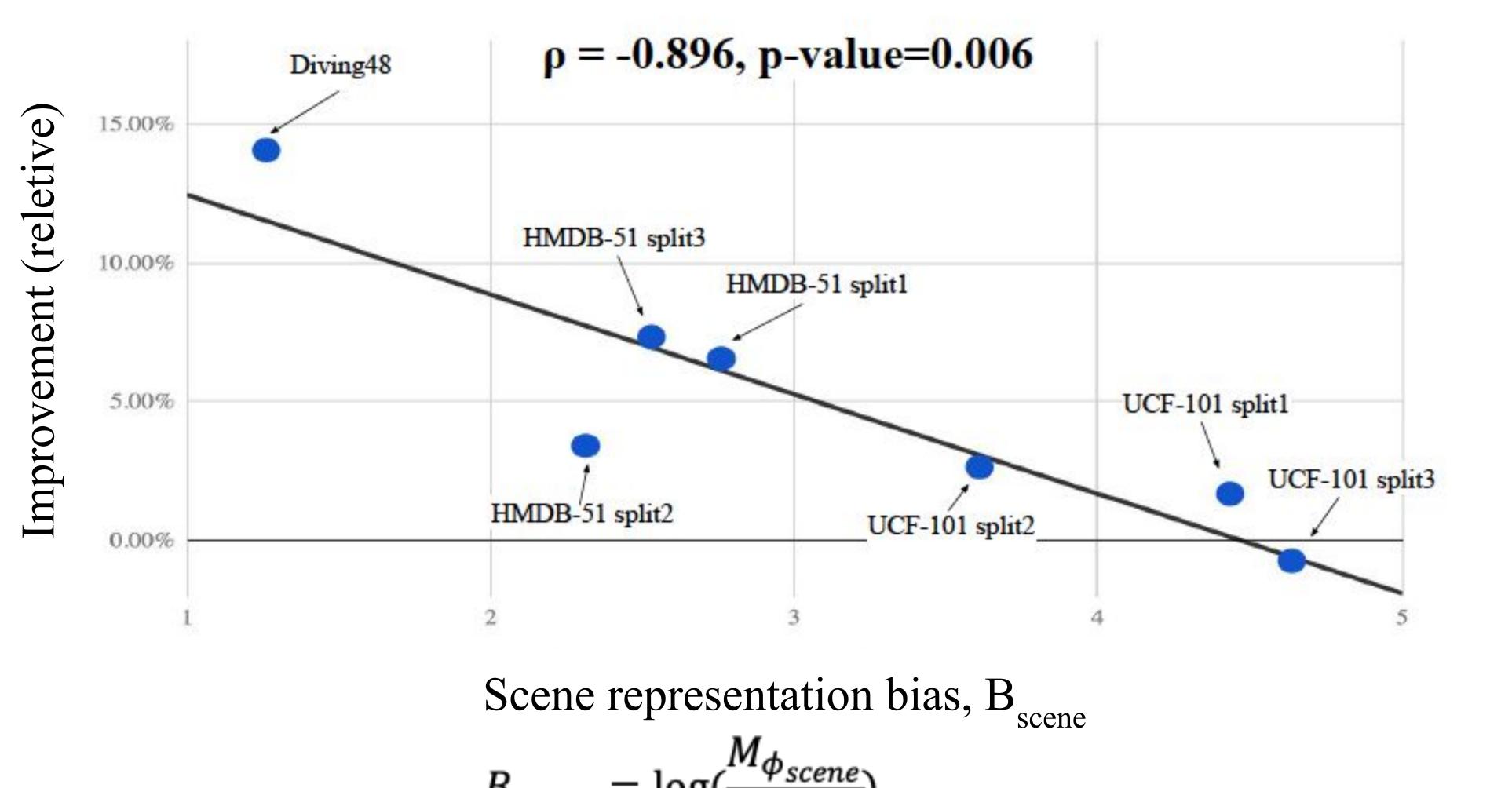












I	I .	Top-1 accuracies on HMDB-51					
LAdv	LEnt	Split-1	Split-2	Split-3	avg.		
X	X	52.9	55.4	52.6	53.6 (+0)		
X		55.0	55.3	55.1	55.1 (+1.5)		
	X	56.4	55.9	56.4	56.2 (+2.6)		
		56.4	57.3	56.5	56.7 (+3.1)		
		Top-1 accuracies on HMDB-51					
Pseudo label		Split-1	Split-2	Split-3	avg.		
None (w/o deb	piasing)	52.9	55.4	52.6	53.6 (+0)		
Hard		54.8	54.2	54.6	54.5 (+0.9)		
Soft (ours)		56.4	57.3	56.4	56.2 (+2.6)		

									rar	nd
emporal Action Localization										
Method	Inputs	uts Backbone -	mean AP @ IoU threshold							
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	avg.
Baseline	RGB	3D-ResNet-18	48.6	48.6	45.6	40.8	32.5	25.5	15.5	36.7 (+0)
Ours	RGB	3D-ResNet-18	50.2	50.5	47.9	42.3	33.4	26.3	16.8	38.2 (+1.5)
CDC	RGB	C3D	_	_	40.1	29.4	23.3	13.1	7.9	_
TAL-Net	RGB+Flow	I3D	59.8	57.1	53.2	48.5	42.8	33.8	20.8	45.1
SSN	RGB+Flow	InceptionV3	66.0	59.4	51.9	41.0	29.8	19.6	10.7	39.8

Spatio-Temporal Action Detection						
Method	Inputs	Backbone	Pre-train on	mean AP		
Baseline	RGB	VGG	ImageNet+MiniKinetics	32.5 (+0)		
Ours	RGB	VGG	ImageNet+MiniKinetics	34.5 (+2.0)		
ACT	RGB+Flow	VGG	ImageNet	65.7		
S3D-G	RGB+Flow	Inception (2+1)D	ImageNet+FullKinetics	75.2		

