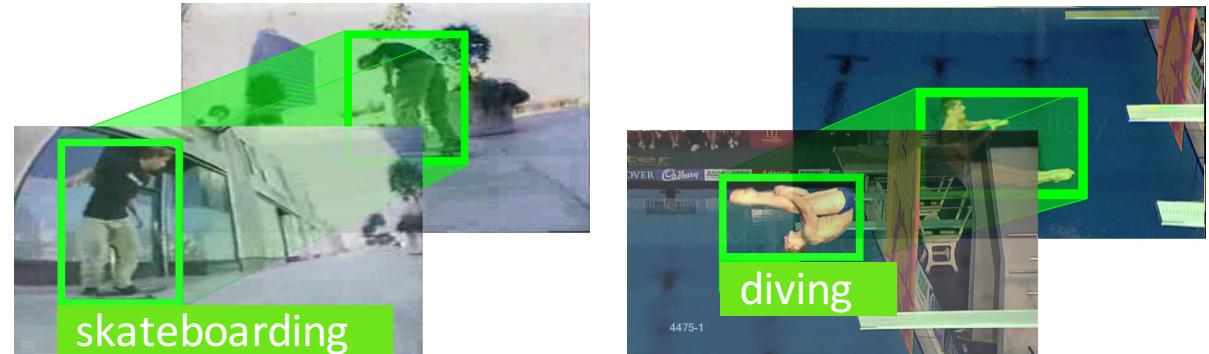


# Action tubelet detector for spatio-temporal action localization

Vicky Kalogeiton<sup>1,2</sup>, Philippe Weinzaepfel<sup>3</sup>, Vittorio Ferrari<sup>2</sup> and Cordelia Schmid<sup>1</sup>

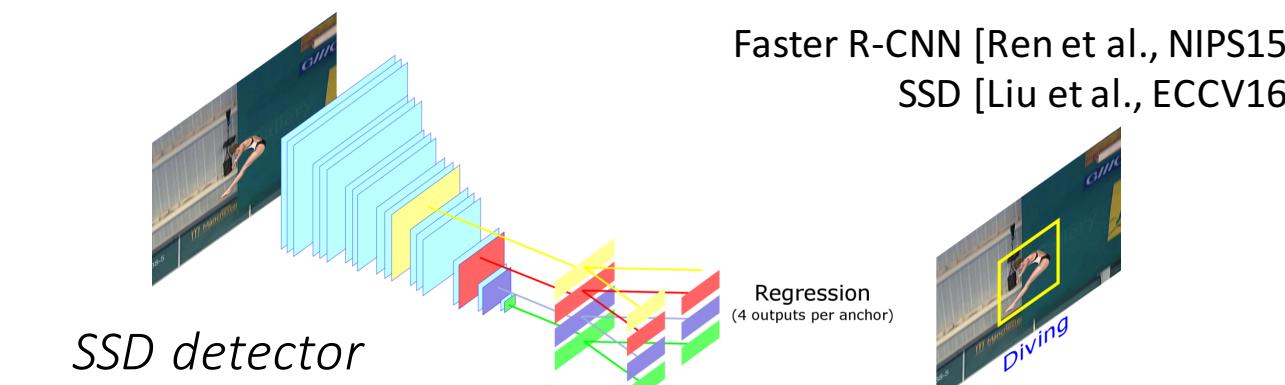
## Introduction

### Spatio-temporal action localization

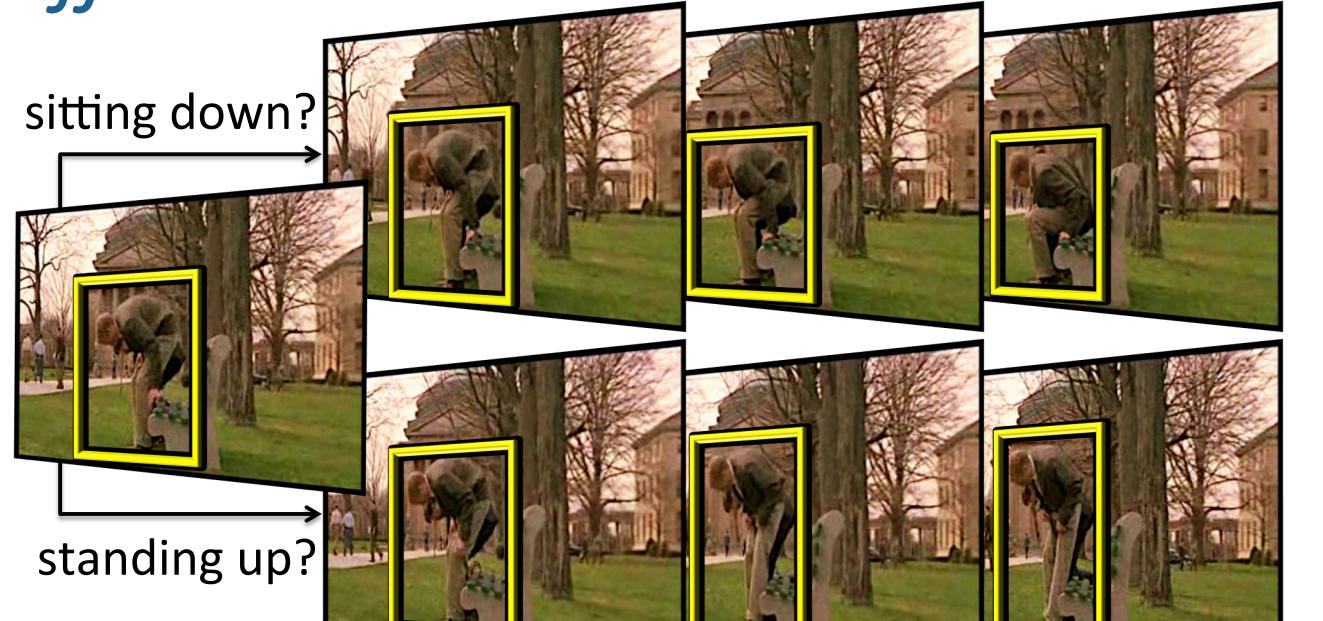


## Related work

### Per-frame detectors for actions



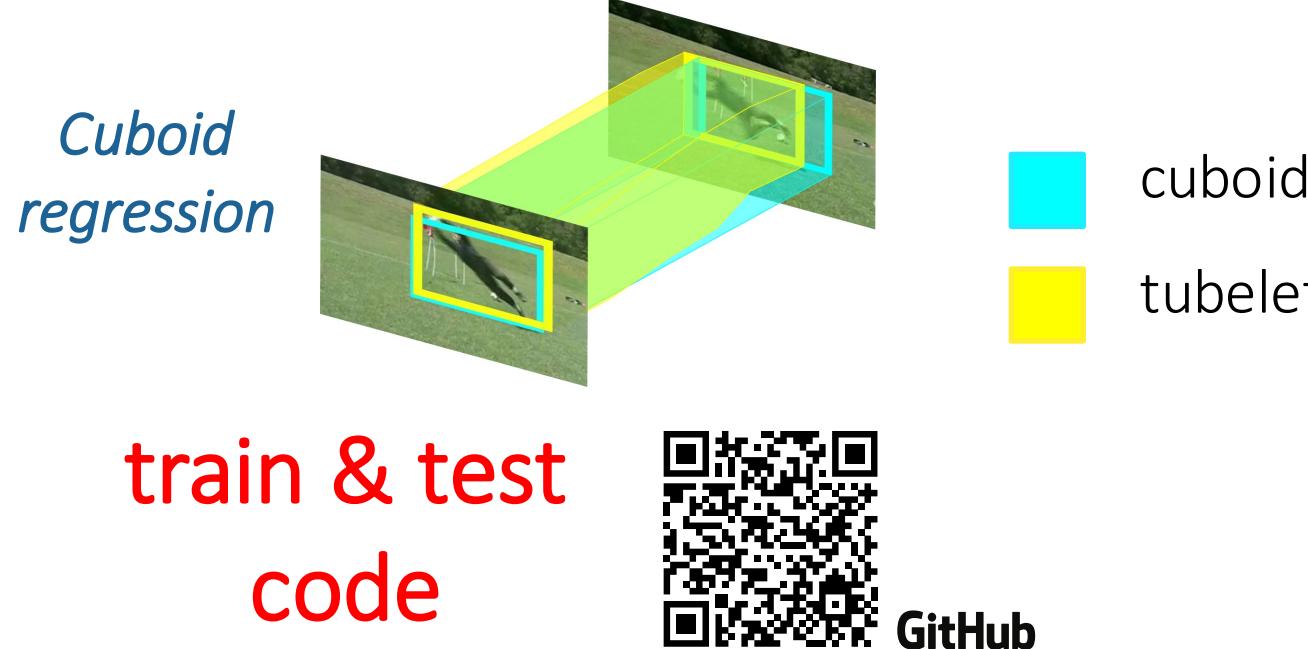
### Difficulties



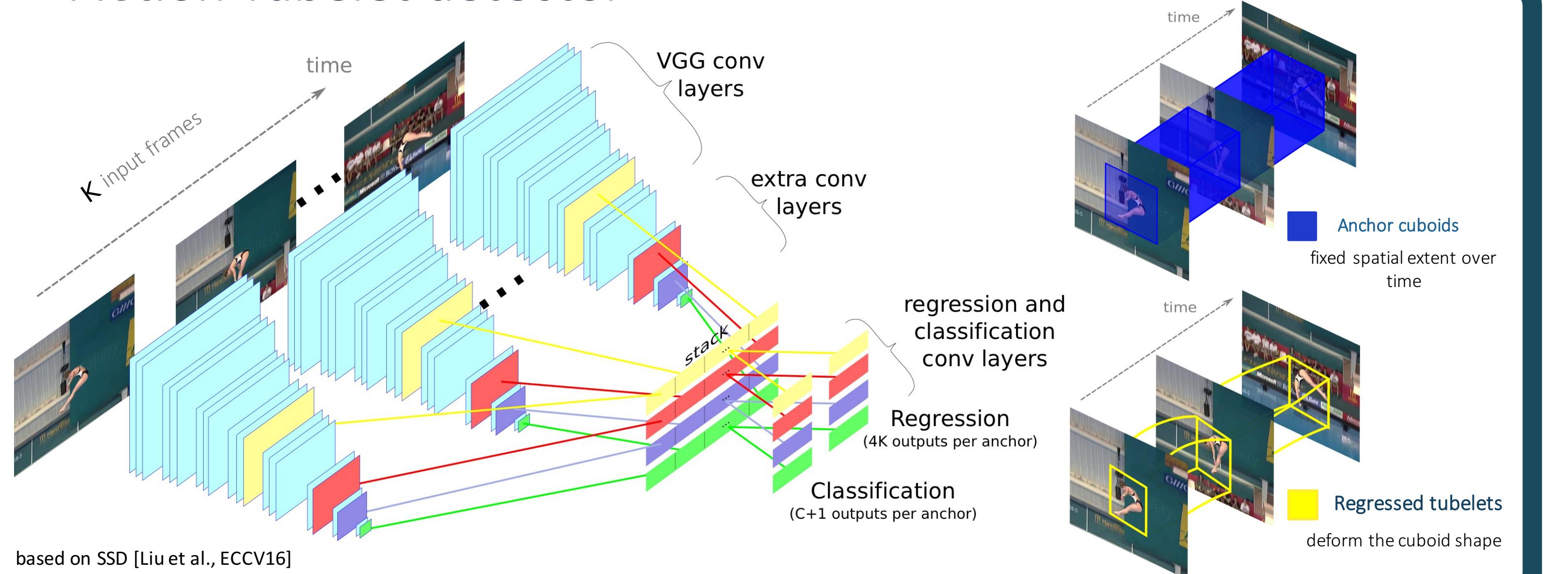
[Gkioxari et al., CVPR15; Peng et al., ECCV16; Saha et al., BMVC16; Wang et al., CVPR16; Weinzaepfel et al., ICCV15]

## Contributions

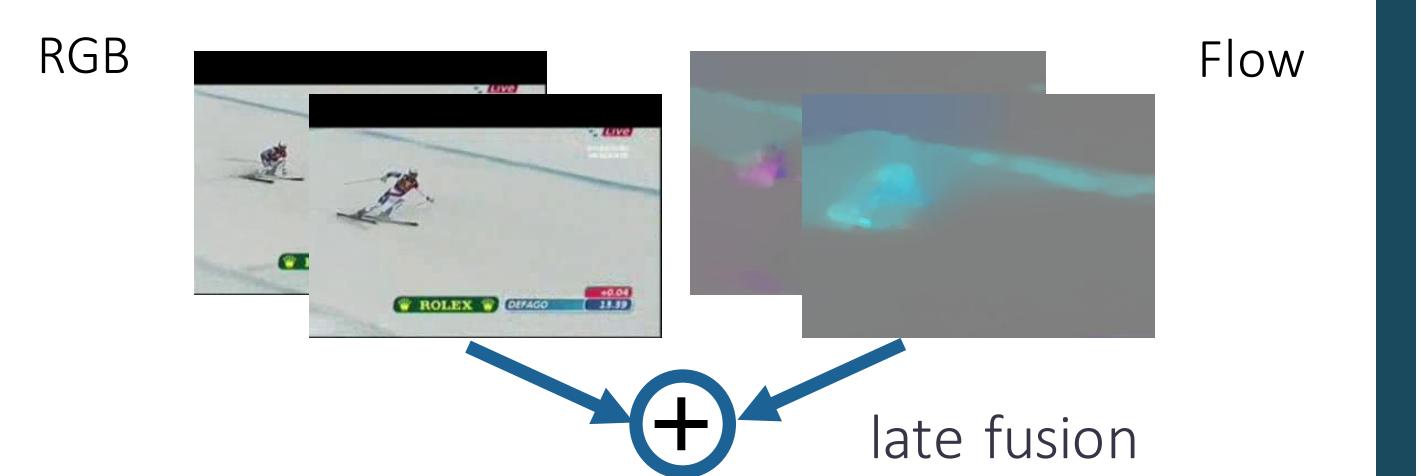
- ACtion Tubelet detector:  
input: sequences of frames  
output: tubelets
- Scoring and regression of anchor cuboids
- Anchor cuboids handle moving actors



## ACtion Tubelet detector



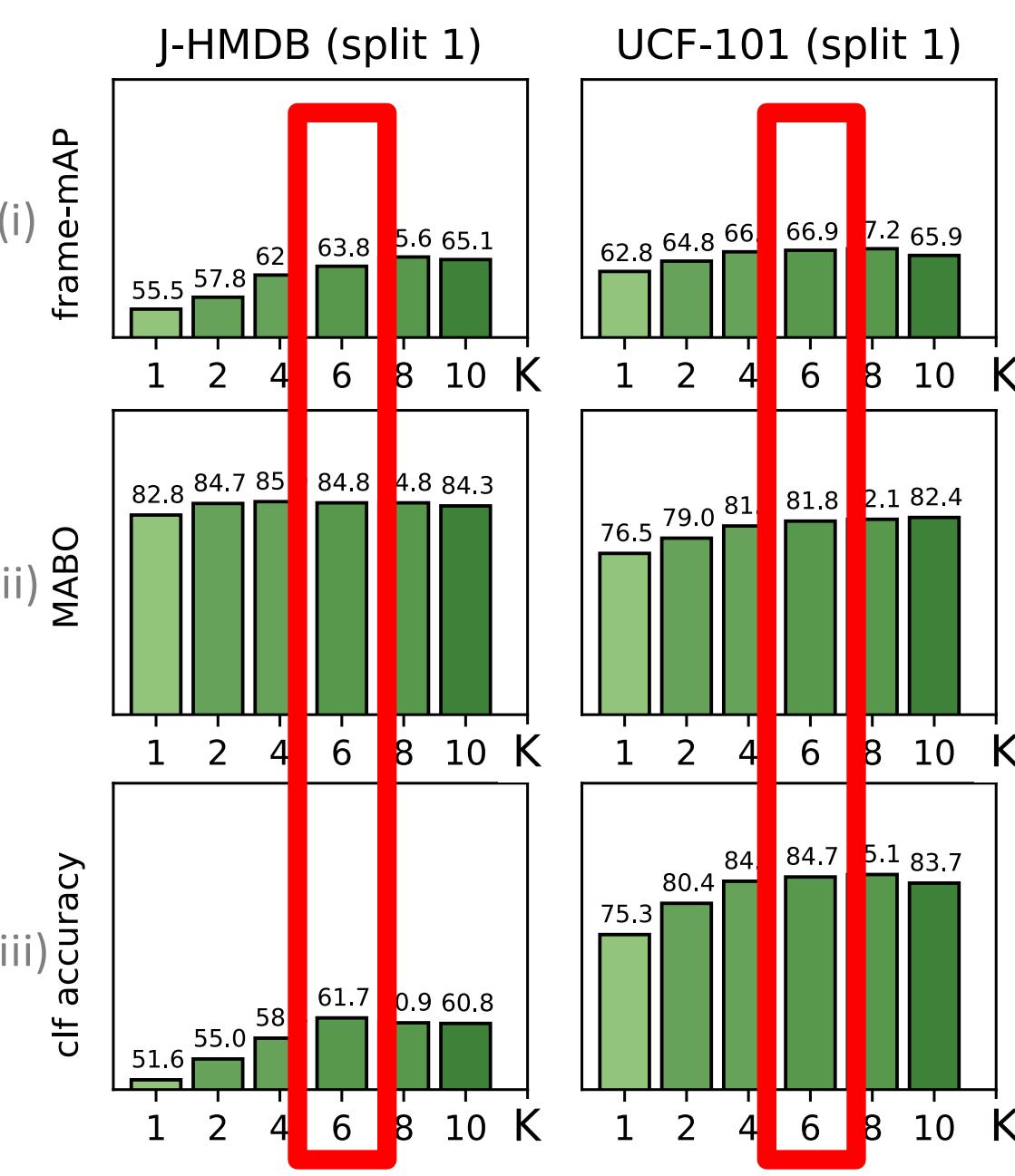
## Two streams



## Spatio-temporal tubes

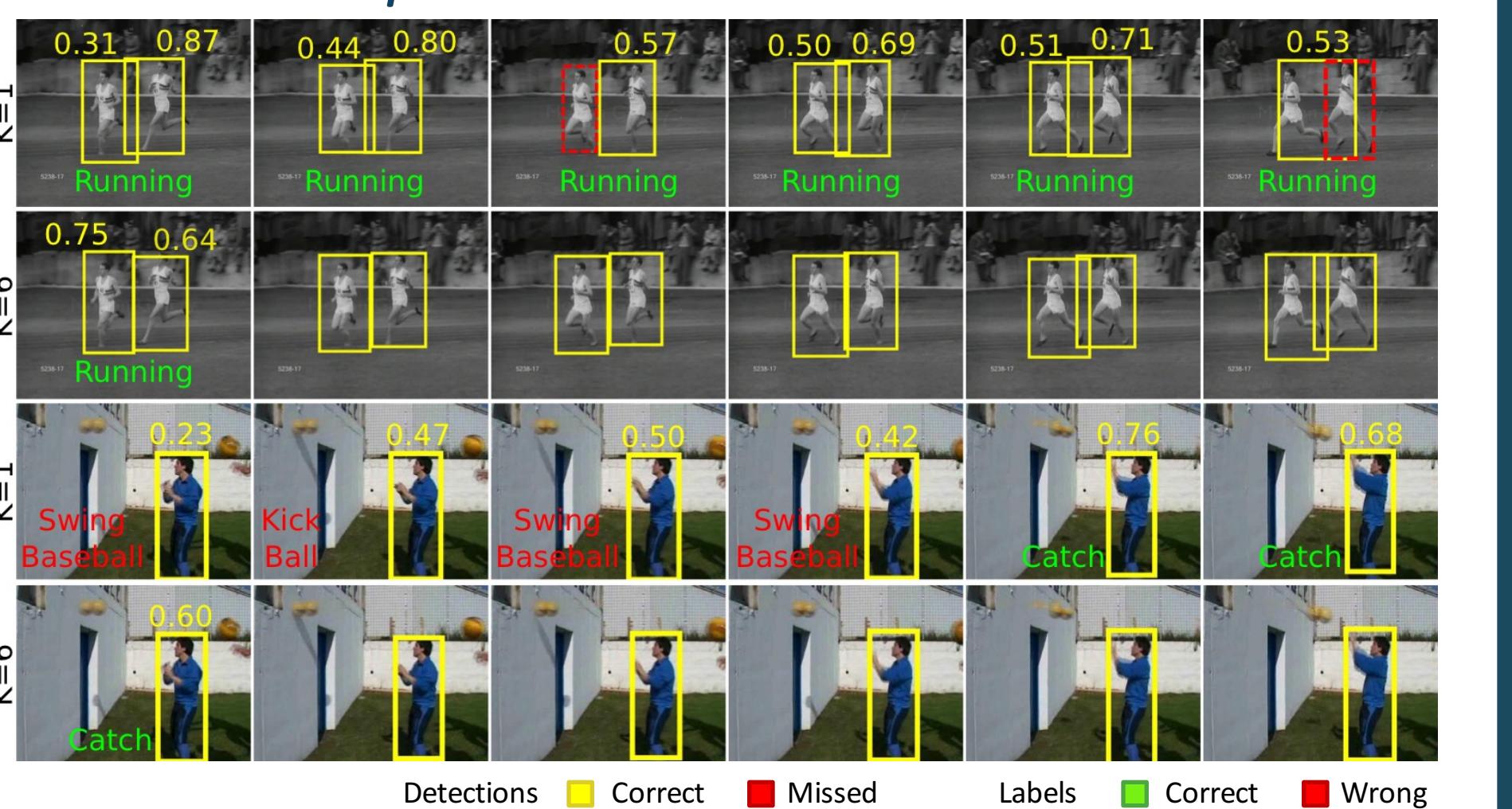
- Online algorithm, robust to missed detections
- Linking tubelets with consistency of scores and overlap
- Temporal score and box smoothing:  
average score of tubelets  
average bounding box coordinates

## Impact of tubelets



Better detection *performance* (i):  
 ✓ improved *localization* (ii)  
 ✓ improved *scoring* (iii)

## Tubelet examples



## Frame-mAP of actors with different motion speed

	J-HMDB (split 1)			UCF-101 (split 1)		
	slow	medium	fast	slow	medium	fast
K=1	61.2	55.5	49.0	69.6	73.5	67.3
K=6	<b>69.8</b>	<b>66.9</b>	<b>58.0</b>	<b>75.4</b>	<b>78.5</b>	<b>70.7</b>

## Comparison to the state of the art

### Frame-mAP

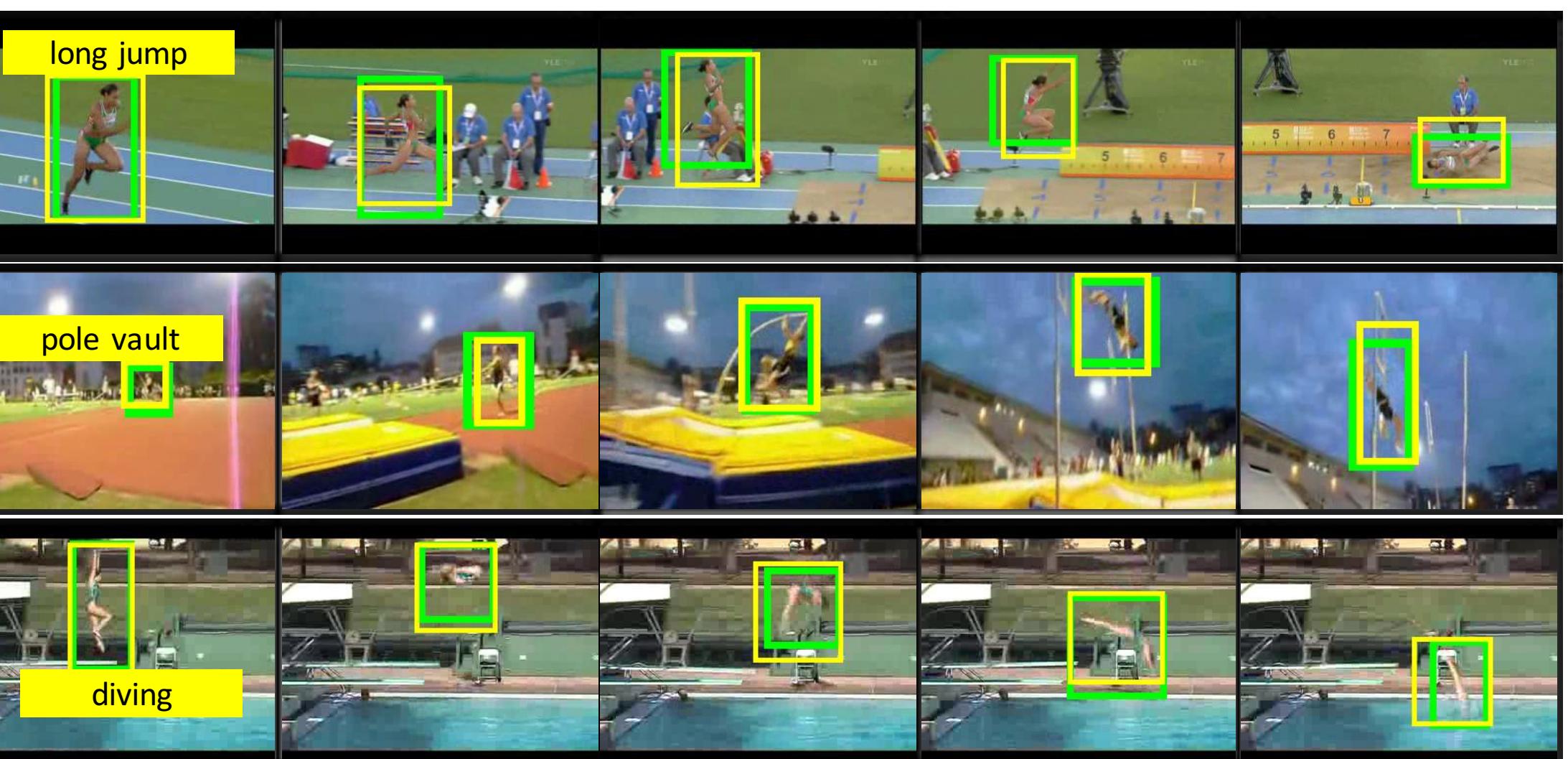
detector	method	J-HMDB			UCF-101		
actionness	Wang et al., CVPR 16			39.9			-
R-CNN	Gkioxari et al., CVPR 15			36.2			-
Weinzaepfel et al., ICCV 15				45.8			35.8
Faster R-CNN	Peng et al., ECCV 16, -MR			56.9			64.8
	Peng et al., ECCV 16, +MR			58.5			65.7
SSD	ACT-detector			<b>65.7</b>			<b>67.1</b>

### Video-mAP

detector	method	J-HMDB			UCF-101		
		0.5	0.75	0.5:0.95	0.5	0.75	0.5:0.95
actionness	Wang et al., CVPR 16	56.4	-	-	-	-	-
R-CNN	Gkioxari et al., CVPR 15	53.3	-	-	-	-	-
Weinzaepfel et al., ICCV 15		60.7	-	-	-	-	-
Faster R-CNN	Peng et al., ECCV 16, -MR	70.6	48.2	42.2	35.9	1.6	8.8
	Peng et al., ECCV 16, +MR	73.1	-	-	-	-	-
Saha et al., BMVC 16		71.5	43.3	40.0	35.9	7.9	14.4
Singh et al., ICCV 17		72.0	44.5	41.6	46.3	15.0	20.4
SSD	ACT-detector	<b>73.7</b>	<b>52.1</b>	<b>44.8</b>	<b>51.4</b>	<b>22.7</b>	<b>25.0</b>

## Video tubes on UCF-101

### Successful detections



### Failure case

