ECCV 2018
European Conference on Computer Vision
8 – 14 September 2018 | Munich, Germany

KAUST

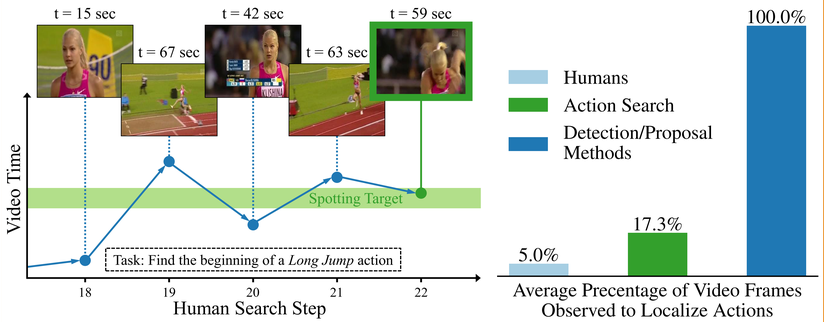
Action Search: Spotting Actions in Videos and Its Application to Temporal Action Localization

Humam Alwassel*, Fabian Caba Heilbron*, Bernard Ghanem
King Abdullah University of Science and Technology (KAUST)

goo.gl/rnpQDb

Motivation: Need for Efficiency

- Current algorithms search exhaustively for actions in videos, but Humans only observe 5%.
- **Action Spotting**: finding an action while observing as little as possible of the video.



Action Spotting: What do Humans do?

Q1: action spotting vs. localization

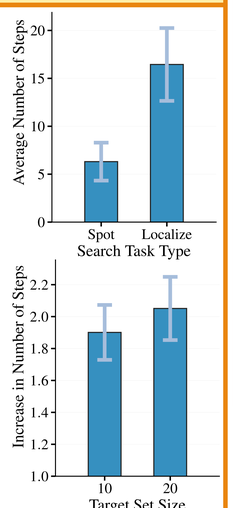
A1: 3x more search steps to localize

Q2: spotting one vs. many classes

A2: Spotting is more efficient with a smaller number of classes

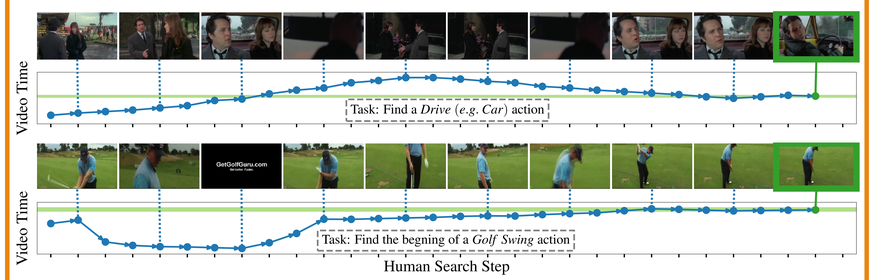
Observations: action spotting

- is a valid precursor to localization
- should be class-specific

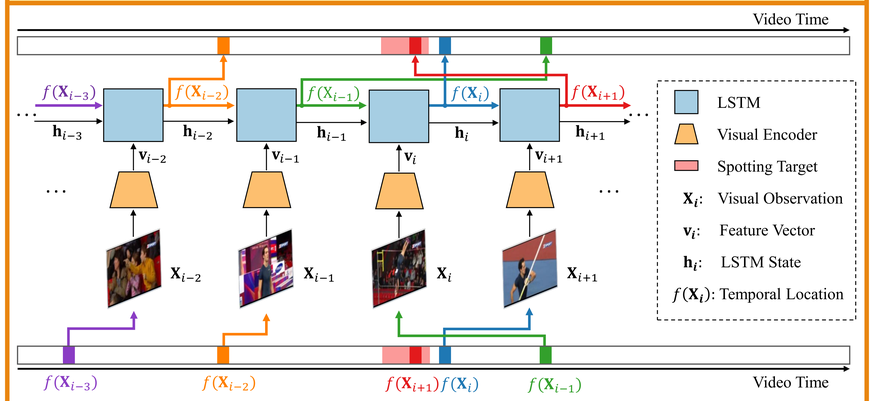


Human Searches Dataset

- Human search sequences on AVA & THUMOS14



Proposed Action Search Model



$$L = \frac{1}{n}\sum_{i=1}^{n} H_\delta(y_i, f(X_i)),$$

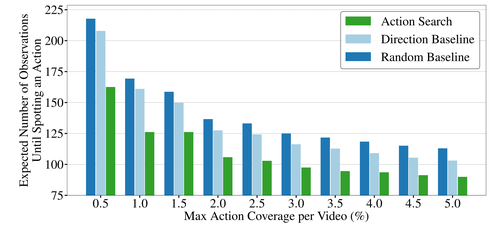
$$H_\delta(y, f(X)) = \begin{cases} \frac{1}{2}\left(y - f(X)\right)^2, & |y - f(X)| \le \delta \\ \delta|y - f(X)| - \frac{1}{2}\delta^2, & otherwise. \end{cases}$$

Acknowledgments: This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. OSR-CRG2017-3405.

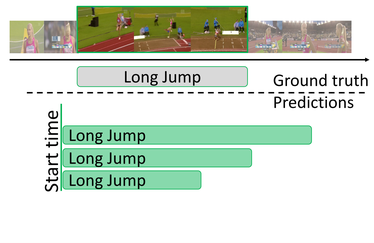
Experiments

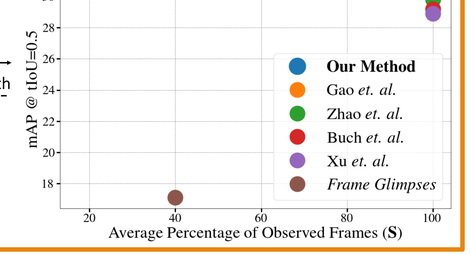
- **Action Search for Action Spotting (AVA)**

Model	Search Step	
	Direction	Length
Random Baseline	Random	Random
Direction Baseline	Learned	Random
Action Search	Learned	Learned



- **Action Search for Action Localization (THUMOS14)**





Qualitative Results