



CRCV HSAP PRESENTATION

EMILY PARK
MARCH 18, 2020



DATA COLLECTION FOR KRISHNA

Ground Videos: 115 new (~570 total)

Aerial Videos: 108 new (~578 total)

Total: 223 new videos (~1148 total videos collected)

+ 12 total cities



RESEARCH PAPER READING (1)

Computer Vision and Image Understanding: The THUMOS challenge on action recognition for videos “in the wild”

- Video Action Recognition (including THUMOS challenge) focused primarily on the classification of pre-segmented (i.e. trimmed) videos
- **Introduction**
 - Large Datasets: UCF101 & HMDB
 - Performances on these datasets has made great progress in the past few years
 - **THUMOS challenge**
 - Based on the UCF101 dataset
 - UCF101’s cleanly trimmed videos could be reasonable during training time since it provides methods with **strongly supervised data**, but the same restriction during testing isn’t realistic:
 - 1.) it makes an unrealistic process to trim videos that surround the desired action
 - 2.) it creates a test set distribution that does not match the real world distribution
 - 3.) it can allow methods to take in false side information, such as the length of the test video clip

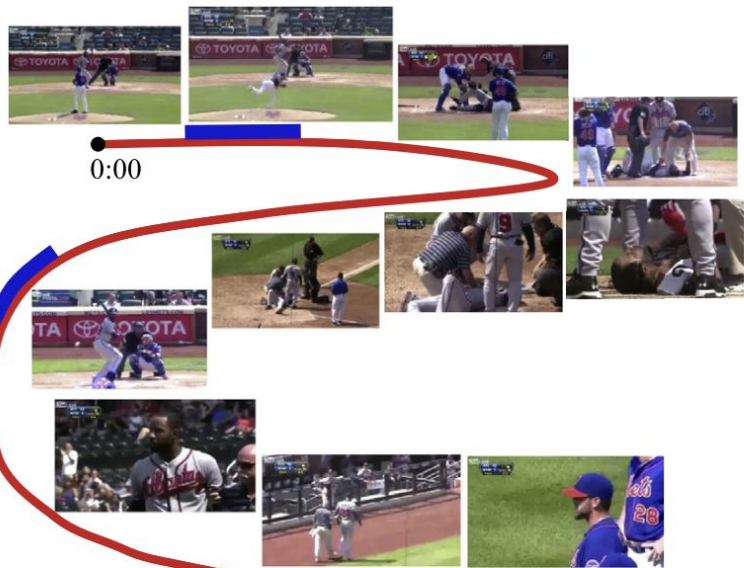
UCF 101
(trimmed)



0:00

0:03 Time

THUMOS
(untrimmed)



0:00

2:49 Time



RESEARCH PAPER READING (2)

Computer Vision and Image Understanding: The THUMOS challenge on action recognition for videos “in the wild”

- Because of this, it becomes difficult to predict the performance of different methods in real applications
- THUMOS consists of 2 principal challenges:
 - 1.) *classification* — goal = to detect whether a video contains a particular action or not
 - 2.) *temporal detection* — goal = to classify an action and find its temporal locations in each video
- THUMOS divides into 5 main categories
 - 1.) Human-Object Interaction
 - 2.) Body-Motion Only
 - 3.) Human-Human Interaction
 - 4.) Playing Musical Instruments
 - 5.) Sports
- THUMOS has 2 objectives:
 - a.) to serve as a benchmark & enable a comparison of different approaches in large-scale realistic video analyses and settings
 - b.) to advance the current techniques (AKA state of the art)



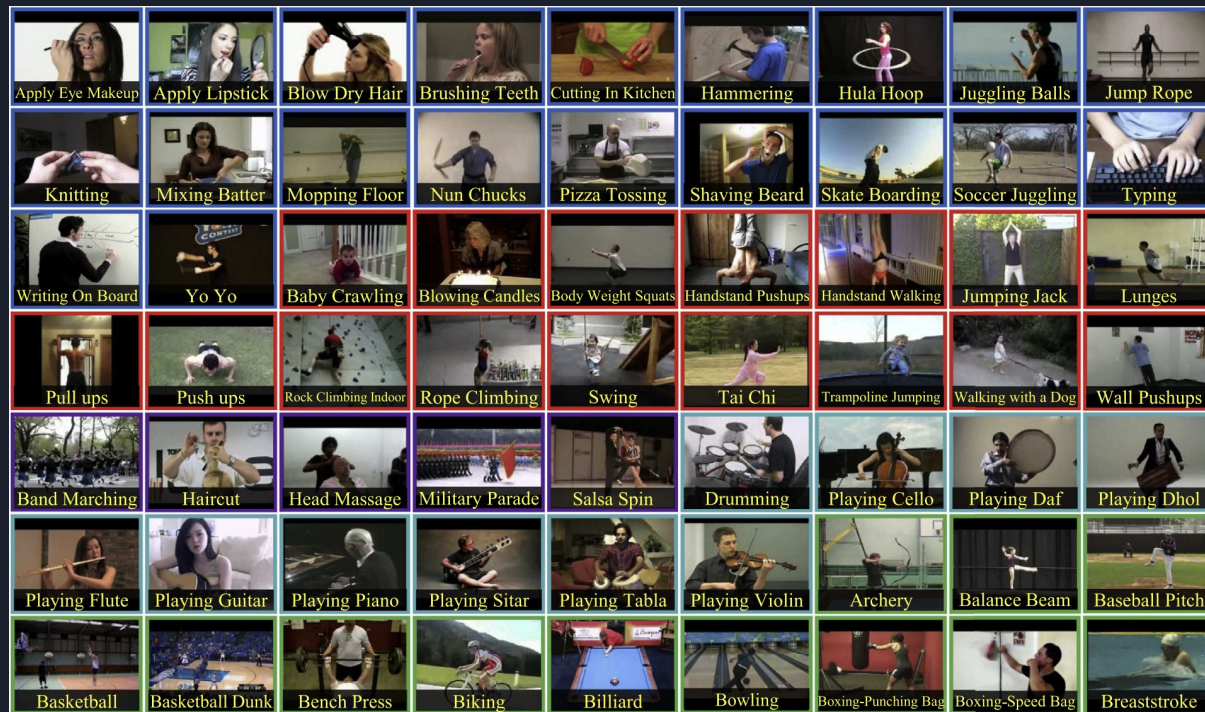
RESEARCH PAPER READING (3)

Computer Vision and Image Understanding: The THUMOS challenge on action recognition for videos “in the wild”

- One of the main ideas – the THUMOS dataset extends the prior work and show show that it differs from other trimmed videos
- For ex.)
 - Sports-1M dataset (released in 2014) – contains over 1 million untrimmed videos from almost 487 classes (1000-3000 videos per action class) with different categories (such as aquatic sports, winter sports, ball sports, etc.)
 - By contrast, THUMOS has videos that are annotated and it includes negative background videos (false alarm) for each action class in both the validation and test sets
 - Makes the action recognition task more difficult
- ActivityNet Dataset
 - Recent dataset for recognition of human activities
 - THUMOS is different from ActivityNet, since it contains a large number of background videos making the problem of action recognition more realistic.

RESEARCH PAPER READING (4)

Computer Vision and Image Understanding: The THUMOS challenge on action recognition for videos “in the wild”





Thank you!

Any questions?