

CRCV HSAP Presentation

Megan Shah

April 15, 2021

Video Data Collection Project

- Continued with Data Collection— sorted and entered YouTube IDs, dates, and uploader names into Excel spreadsheet
- Collected Ground and Aerial Videos for Manila, Wellington, Mogadishu, Khartoum, Sanaa, Suva, Libreville, Accra, Georgetown, Reykjavik, and Tehran
- Geotags-- assigns a location to a specific photograph or video

The THUMOS Challenge on Action Recognition in the Wild

The THUMOS Challenge tasks include (1) classification and (2) temporal detection

- Participants are expected to provide temporal time and confidence intervals for 20 selected action cases-- low confidence indicates that some or none of the actions are replicated in other videos in the dataset
- Since 2014, dataset has been extended to untrimmed validation, background and test videos
- Adopting temporal detection over spatio-temporal localization in later years of the THUMOS challenge was driven by (1) computational tractability and (2) practical importance (temporal is more important in most cases)
- THUMOS dataset-- drawn from public videos on YouTube, large fraction of videos don't contain visible instances of desired action
- Manual filtering methods are used to sort relevant videos

The THUMOS Challenge on Action Recognition in the Wild

- Positive videos:
 - Youtube Data API allows video search through freebase topics
 - Each Youtube video has several freebase topics associated with it (based on annotations)
 - The dataset topic ids are manually augmented with certain keywords
 - Possible issue: viral videos and compilations may be highly rated so certain words were blacklisted
- Background videos:
 - Useful background video collecting is a bit more tricky
 - Share CONTEXT of given action without actually containing the action itself
 - One background video for one action class should not contain positive instances for other actions
 - Action classes are grouped into super classes
- Query Examples (to ensure no positive instances for other actions): venue, for sale, brands, review

The THUMOS Challenge on Action Recognition in the Wild

- Annotation and verification procedure: provides a set of potential positive and background videos for each of the 101 action classes
- Annotators were asked to go through videos of UCF101 dataset action class videos and label them positive or irrelevant
- Irrelevant videos could have contained the action but could be in the following formats
 - Slow motion
 - Sped up
 - Occlusion
 - Animation
 - Long video
- Positive videos are also annotated with second actions (associated actions dunk= basketball etc)
- Background videos should contain NO instance of the action itself being performed

The THUMOS Challenge on Action Recognition in the Wild

- Temporal annotations
 - Action boundaries are generally vague
 - Each human expert classifies things differently
 - Action classes are divided into (1) instantaneous actions (happen once) and (2) cyclical actions (like biking)
 - annotated action intervals consistently with the temporal segmentation of corresponding actions in the UCF101 dataset
 - Some action instances are ambiguous if there is any room for uncertainty
 - liberal Intersection-Over-Union threshold (small, 10%) is used to quantify the performance on this task, since actual actions are only a small fraction of the entire videos
- Actions were annotated for temporal boundaries

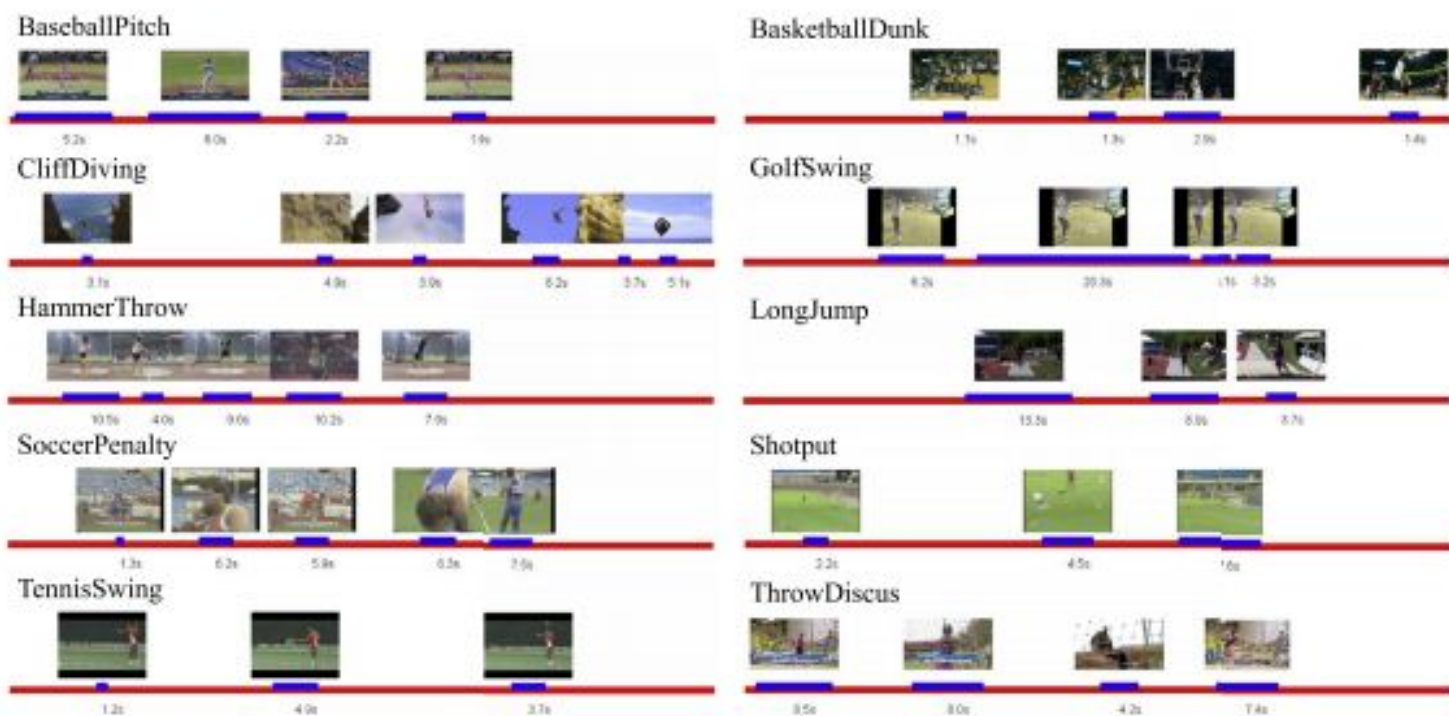


Fig. 3. Illustration of temporal annotation (shown in blue) for eight video samples from the Validation set of THUMOS'15 dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Thank you!

Any Questions?

