STVIs for platform diving

Image and video understanding 2019 WS

C. Lenzenweger, B. Sespede

Goal

- Identify diving style from video using STVI's as a base for feature extraction

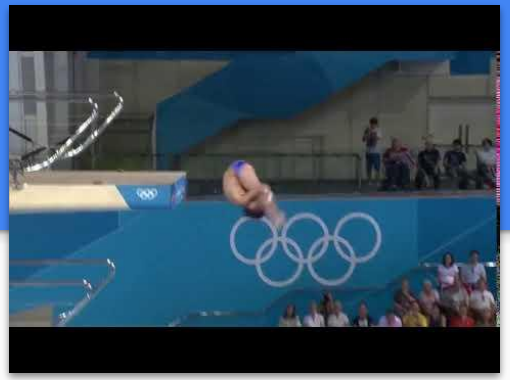
Pike

Tuck

Straight

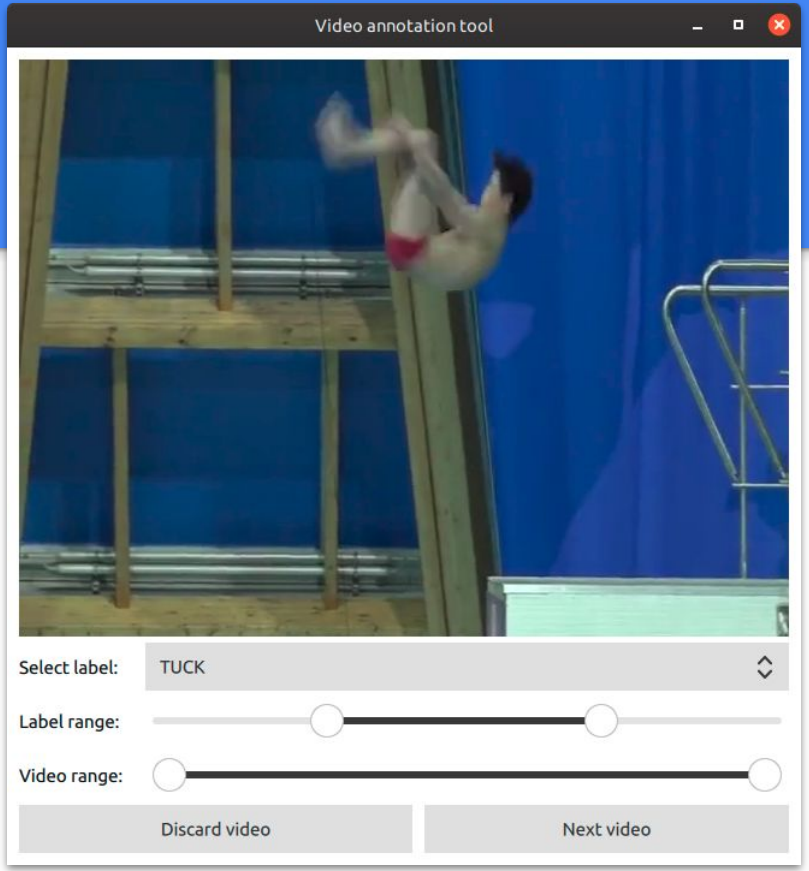
Dataset

- Diving48[1] contains 20.000 diving videos
  - Captured from a variety of angles
  - Different divers
  - Different background
  - Cover a variety of diving styles



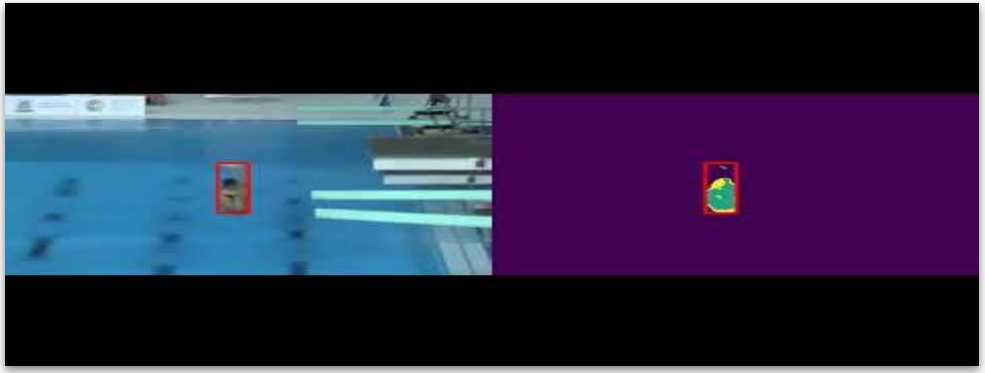
Dataset issues

- Groundtruth labels, not so much truth in them
  - Time to manually annotate!
- Implement tool and classify 300 videos (100 per diving style)
  - Always sideways
  - Low motion blur
  - Remove water splashing part
  - Remove complex twists



Space-time volume instances (STVI)

- How do they look?
  - See video
- What kind of data are we dealing with?
  - Spatio-temporal masks



Feature extraction issues

- Where has the volume gone?
  - Manually discard videos
- STVI flickering
  - Filtering in feature space

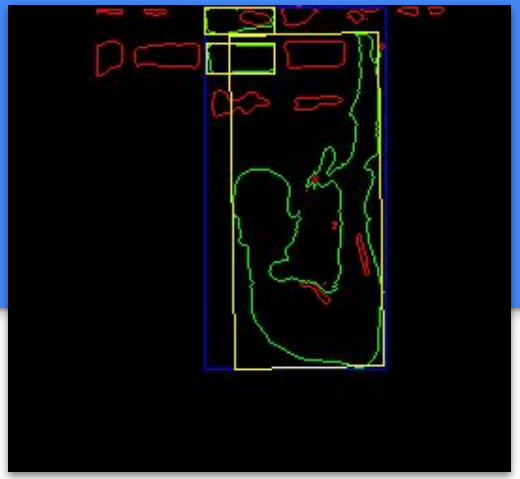
Feature extraction & token selection

- Selection of relevant STVI
  - 2-3 largest (by volume or area) ✗
    - Represent: torso / thighs / lower legs
  - Largest (by contour area) ✓
    - Represents: diver
    - Merge non-artefact contours with same STVI label

- Select features describing the diver's pose
  - Fitting primitives (MBR, ellipses, lines) to extracted contours
  - Scalar features such as angle & distance w.r.t. bounding box, elongation, compactness, fill factor
  - Scale, rotation and reflection invariant Hu moments

Feature extraction issues

- Correcting STVI artefacts
  - Minimum threshold
  - Relative area w.r.t. to biggest contour
- Example extraction (see figures)
  - Green: accepted STVI
  - Red: rejected STVI



Style classification

- Support vector machines ✓
  - Input: framewise feature vector
  - Input: stacked temporal feature vectors
- Convolutional neural networks? ✗
  - Input: feature vector over time
  - 1D convolutional layers w.r.t. Time
- Neural networks? ✗
  - Input: feature vector
  - Framewise processing

Thank you for your attention!

[1] Y. Li Y. Li and N. Vasconcelos. Resound: Towards action recognition without representation bias. ECCV, 2018