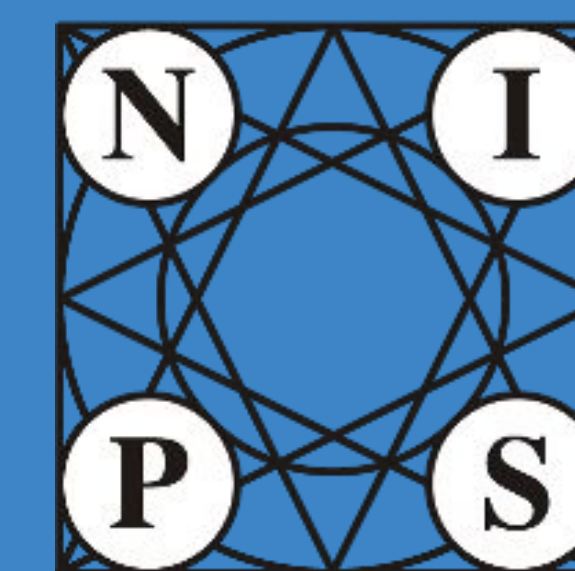


Temporal Activity Detection in Untrimmed Videos with Recurrent Neural Networks



Workshop on
Large Scale
Computer Vision
Systems 2016



Alberto
Montes



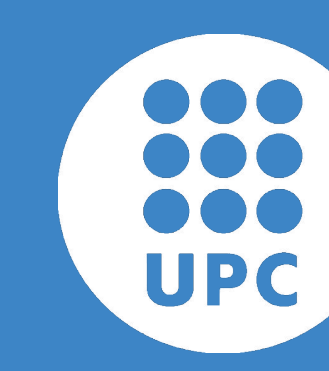
Amaia
Salvador



Santiago
Pascual



Xavier
Giró-i-Nieto



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Problem Definition: Classification + Detection

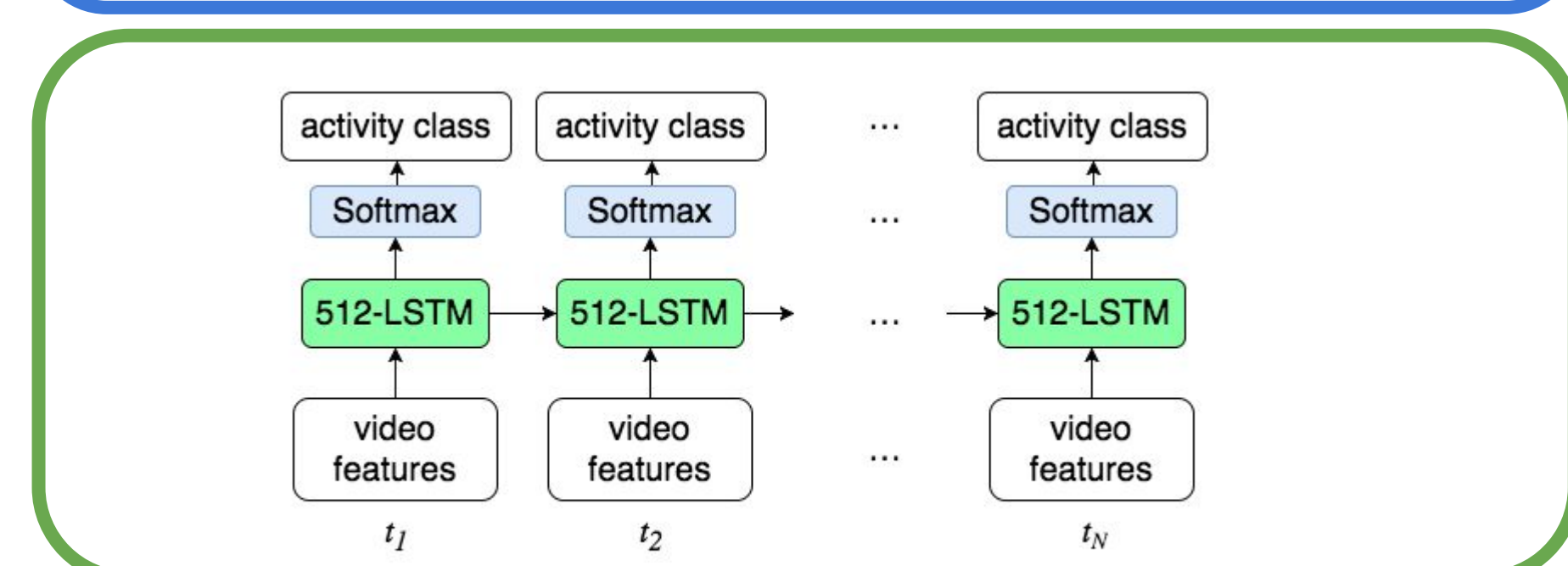
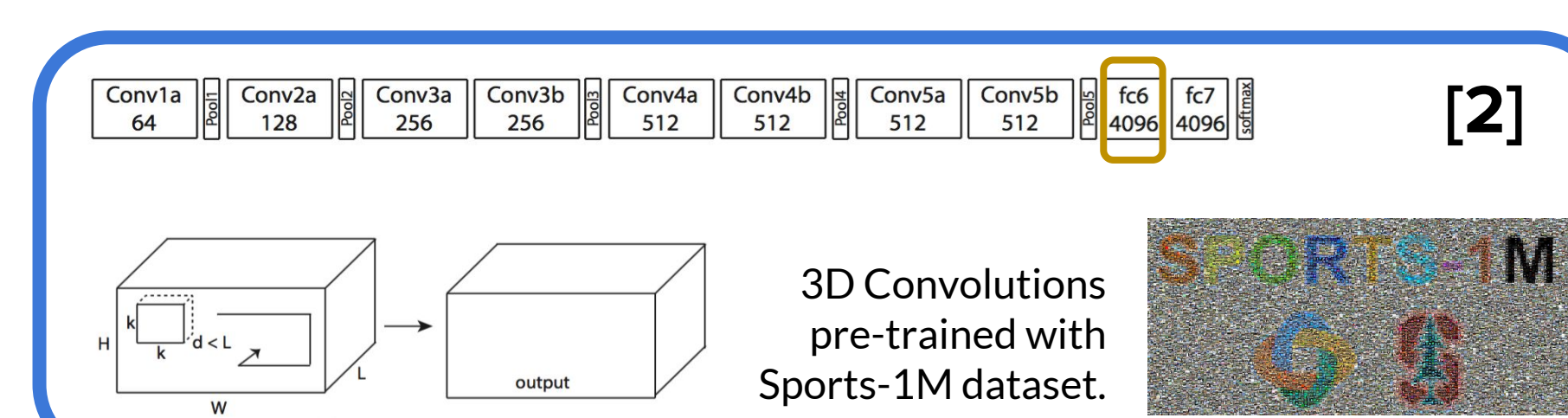
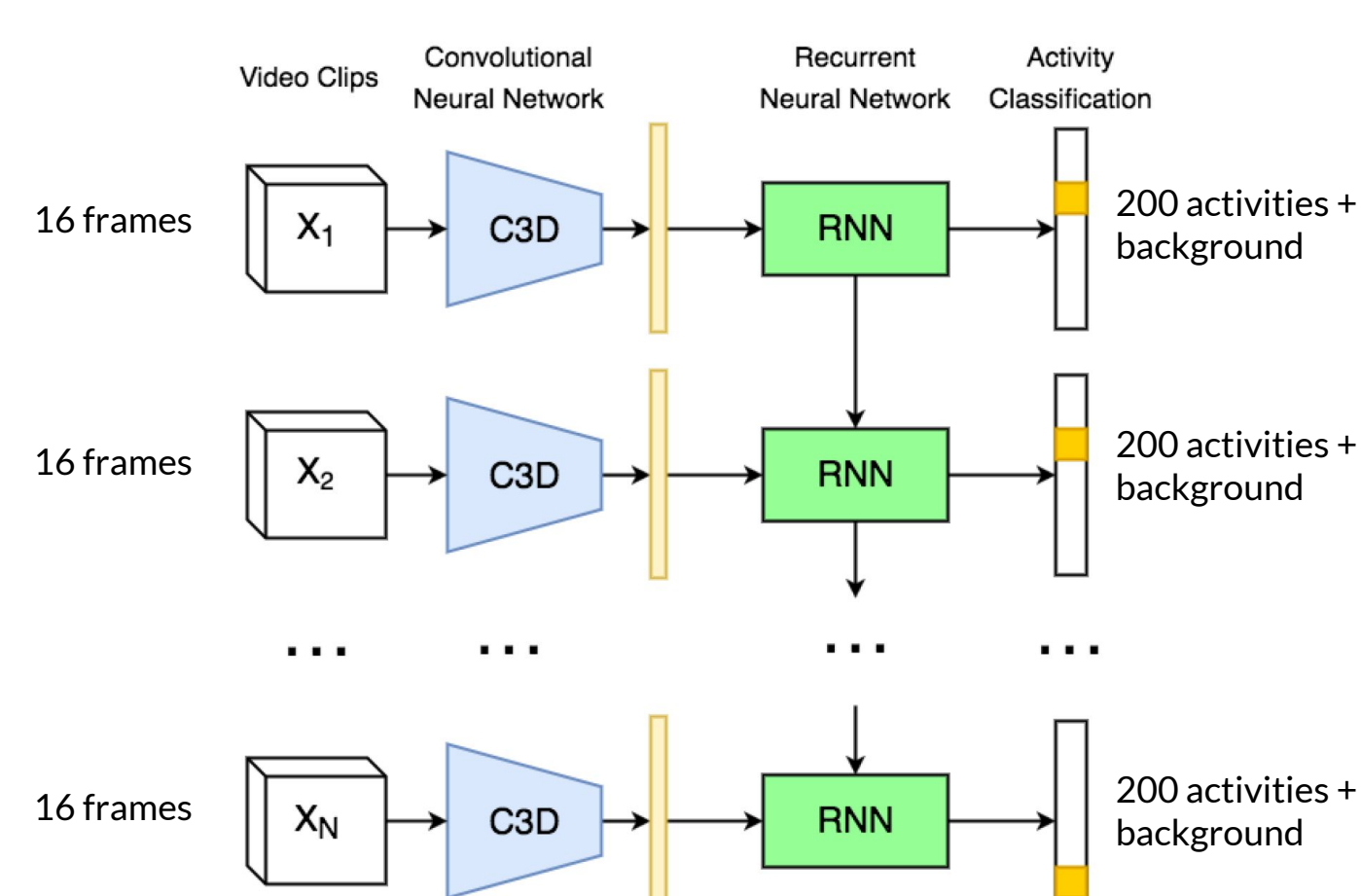


Large-Scale Video Dataset

- 19,994 Videos
- 200 Activities
- 1 activity per video (only)
- 660 hours of video
- 313 hours of activities
- **65.6 million frames**

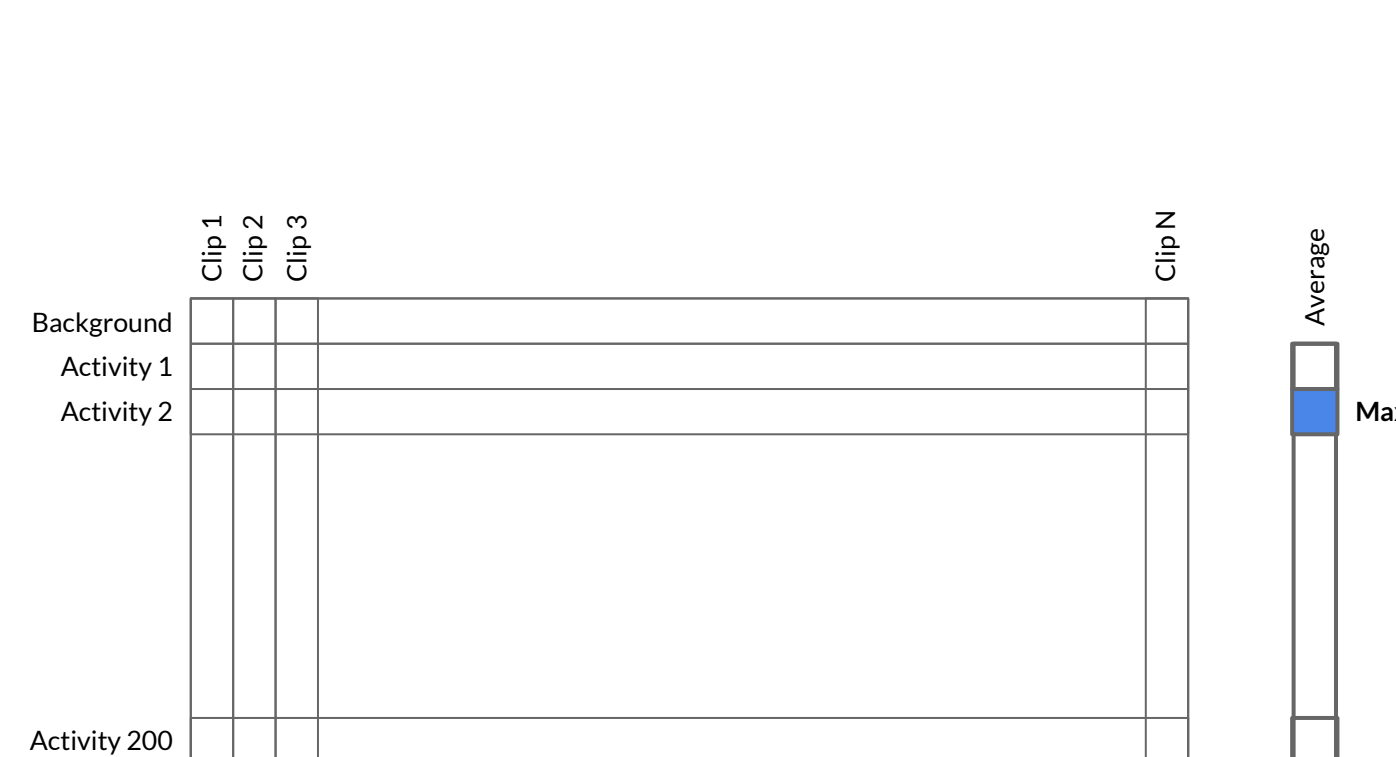


Architecture: C3D + LSTM

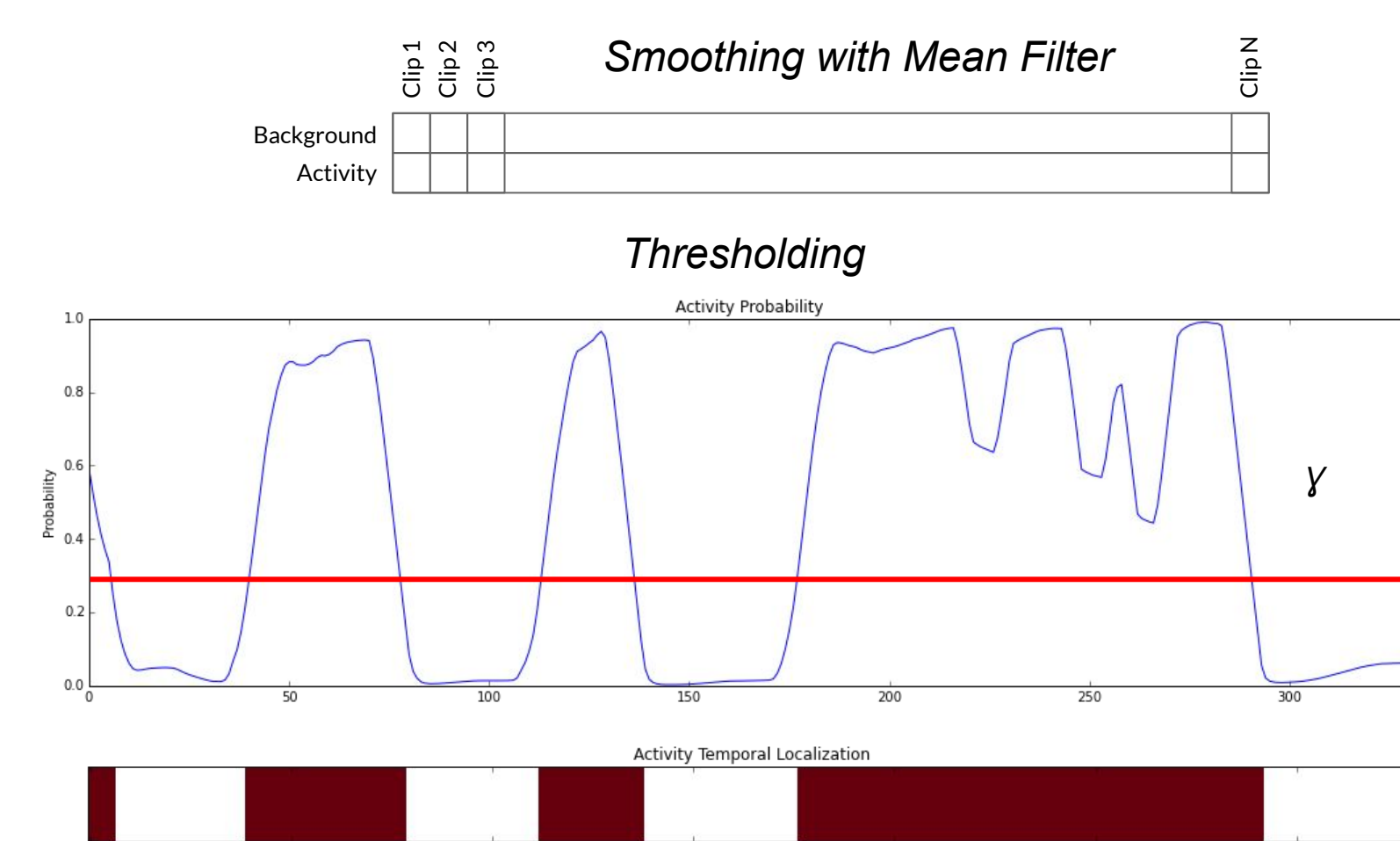


Post-Processing

From Clip to Video Classification

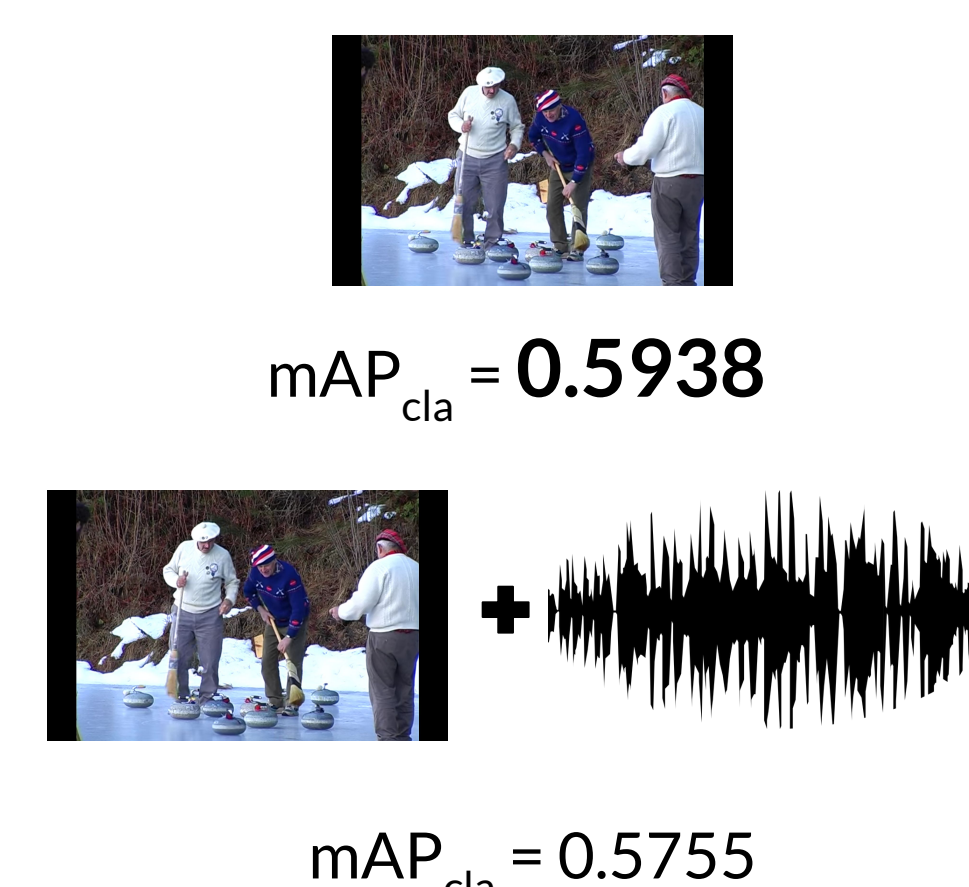


From Clip to Activity Temporal Localization

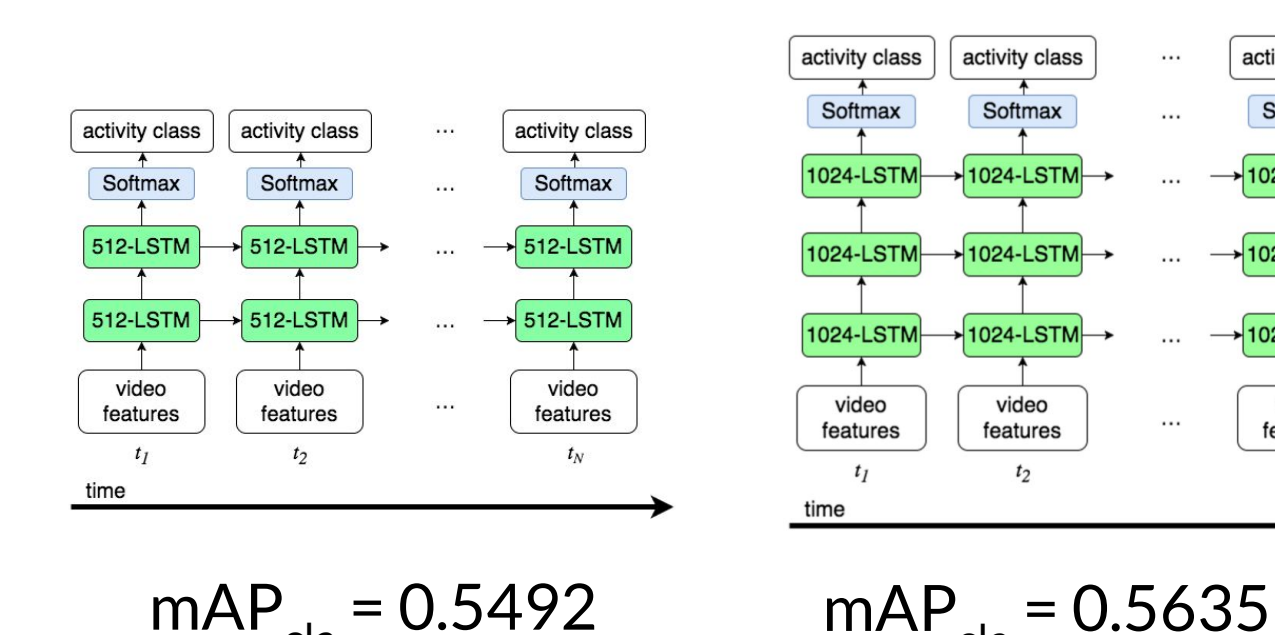


Other Configurations Tested*

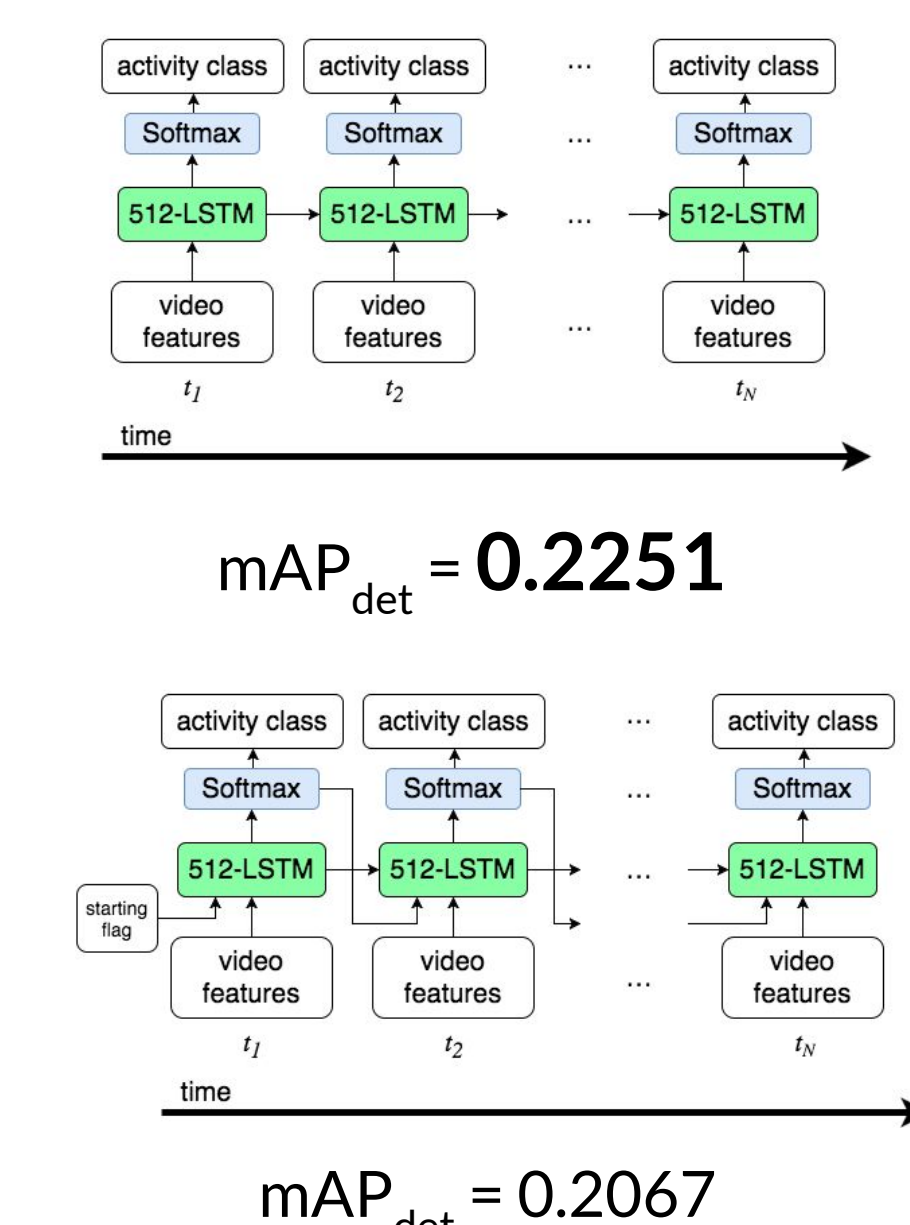
MFCC Audio Features



Stacked LSTMs



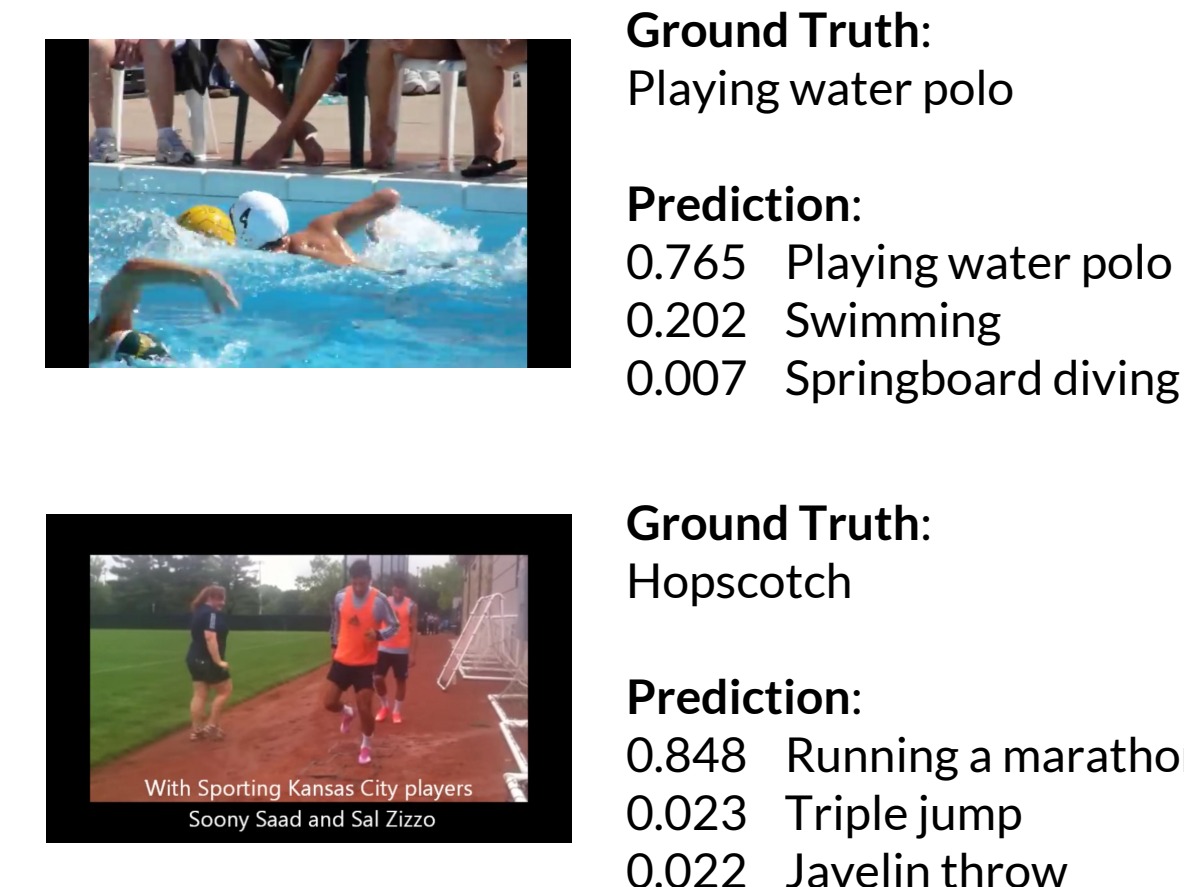
Feedback



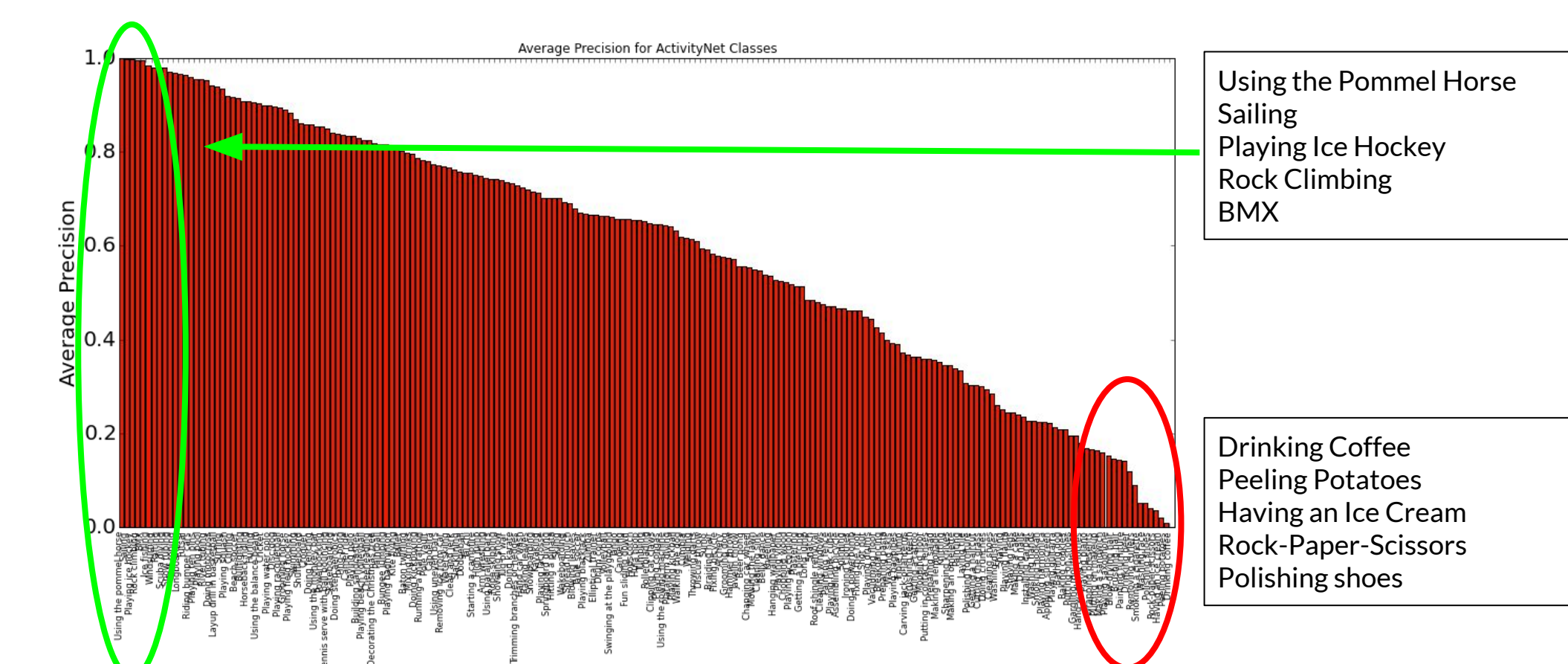
(*) results on validation dataset

Results

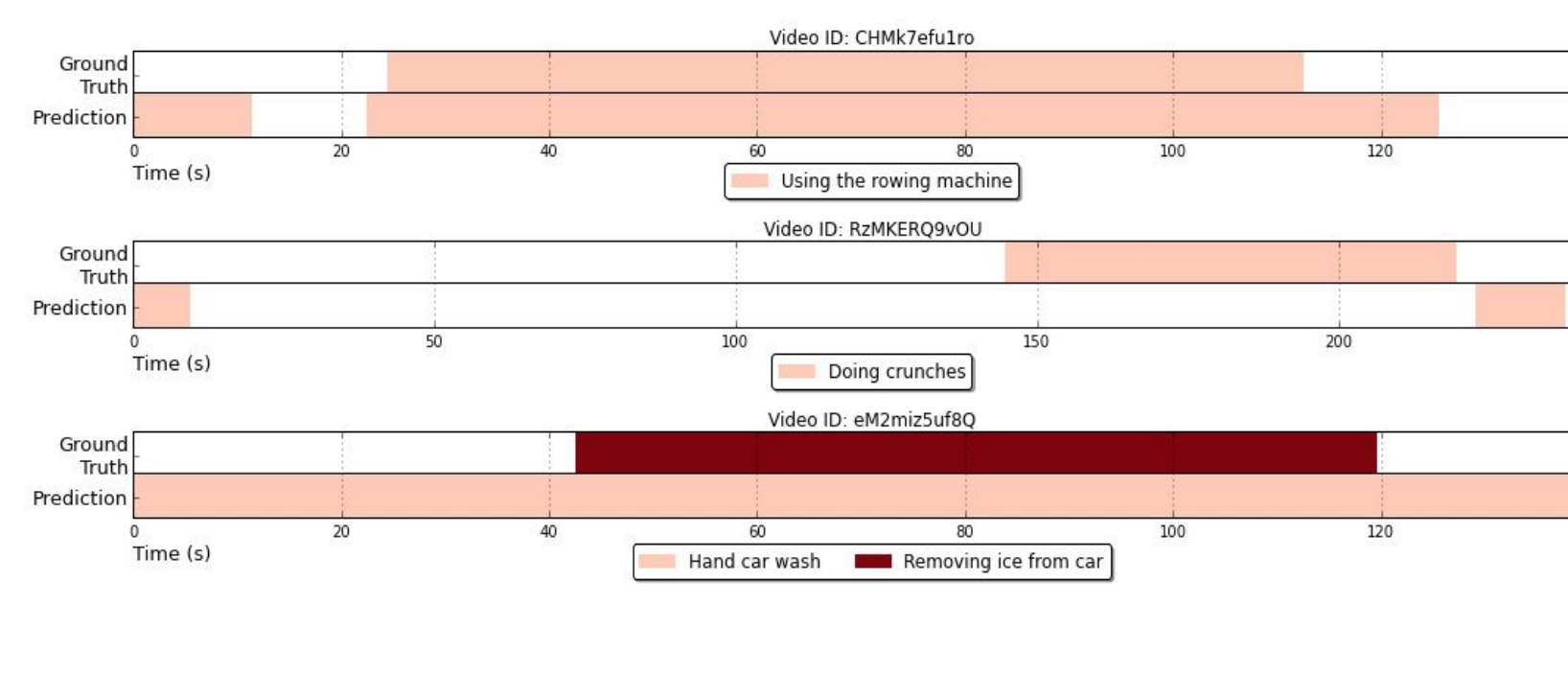
Classification



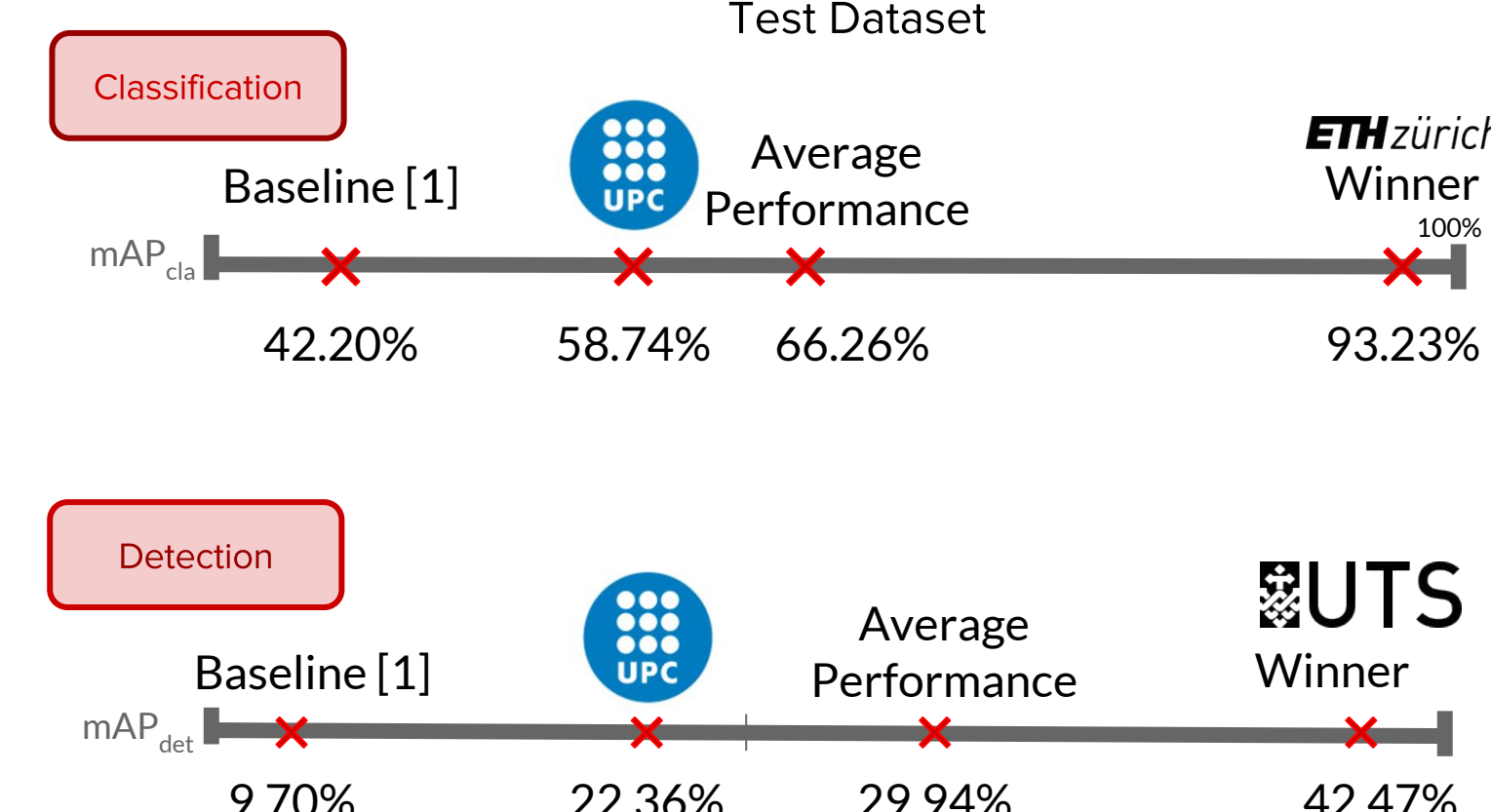
C3D Features Bias



Temporal detection



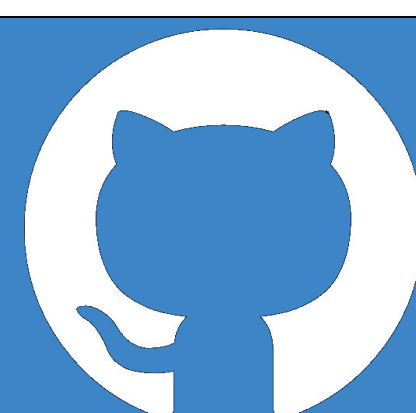
CVPR 2016 Challenge



[1] Caba Heilbron, Fabian, et al. "Activitynet: A large-scale video benchmark for human activity understanding." CVPR 2015.
[2] Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." ICCV 2015.



malberto@student.ethz.ch
amaia.salvador@upc.edu
santiago.pascual@tsc.upc.edu
xavier.giro@upc.edu



<https://imatge-upc.github.io/activitynet-2016-cvprw/>

