

Segments, Residuals and Embeddings for Few-Example Video Event Detection

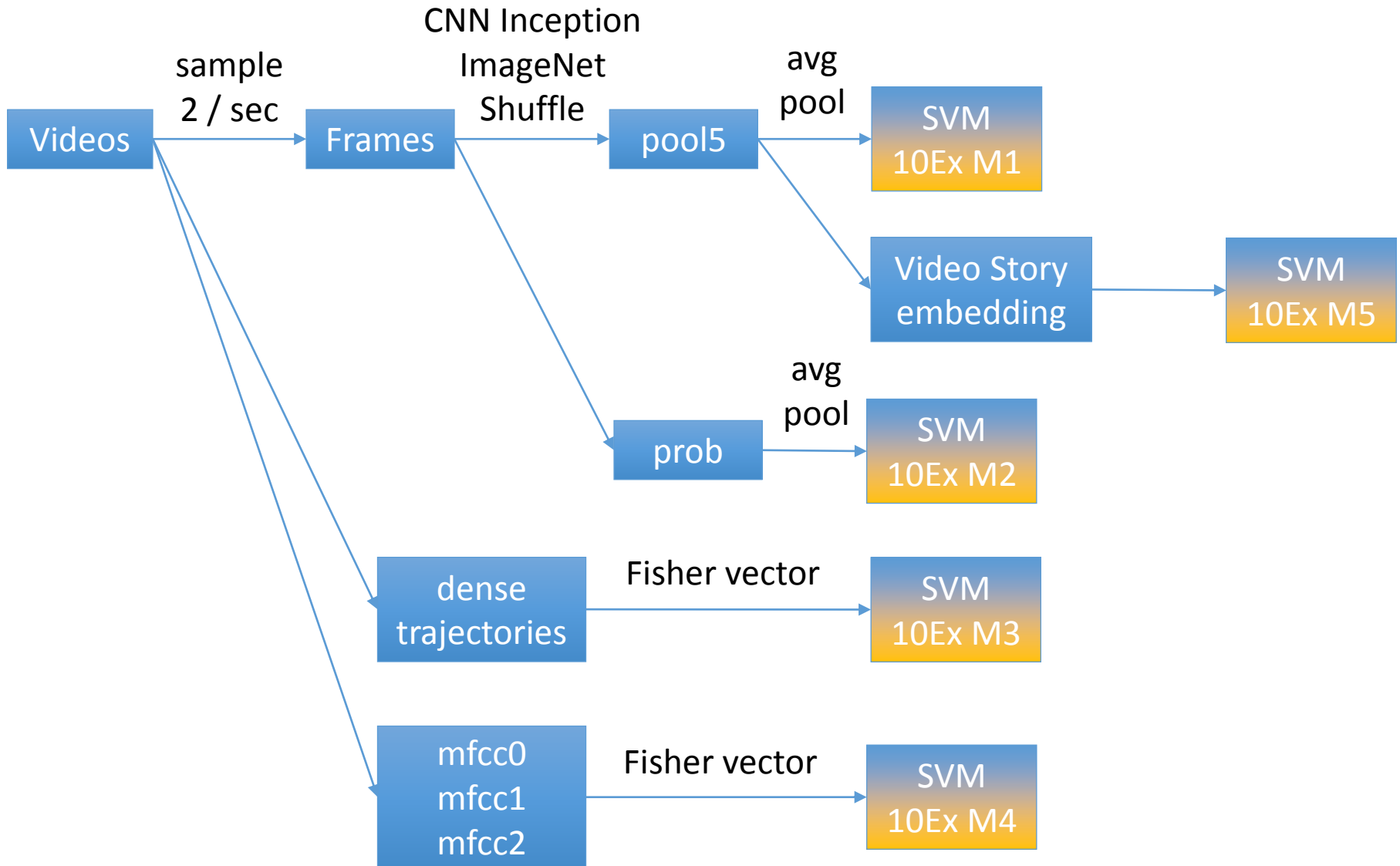
Dennis Koelma and Cees Snoek

University of Amsterdam

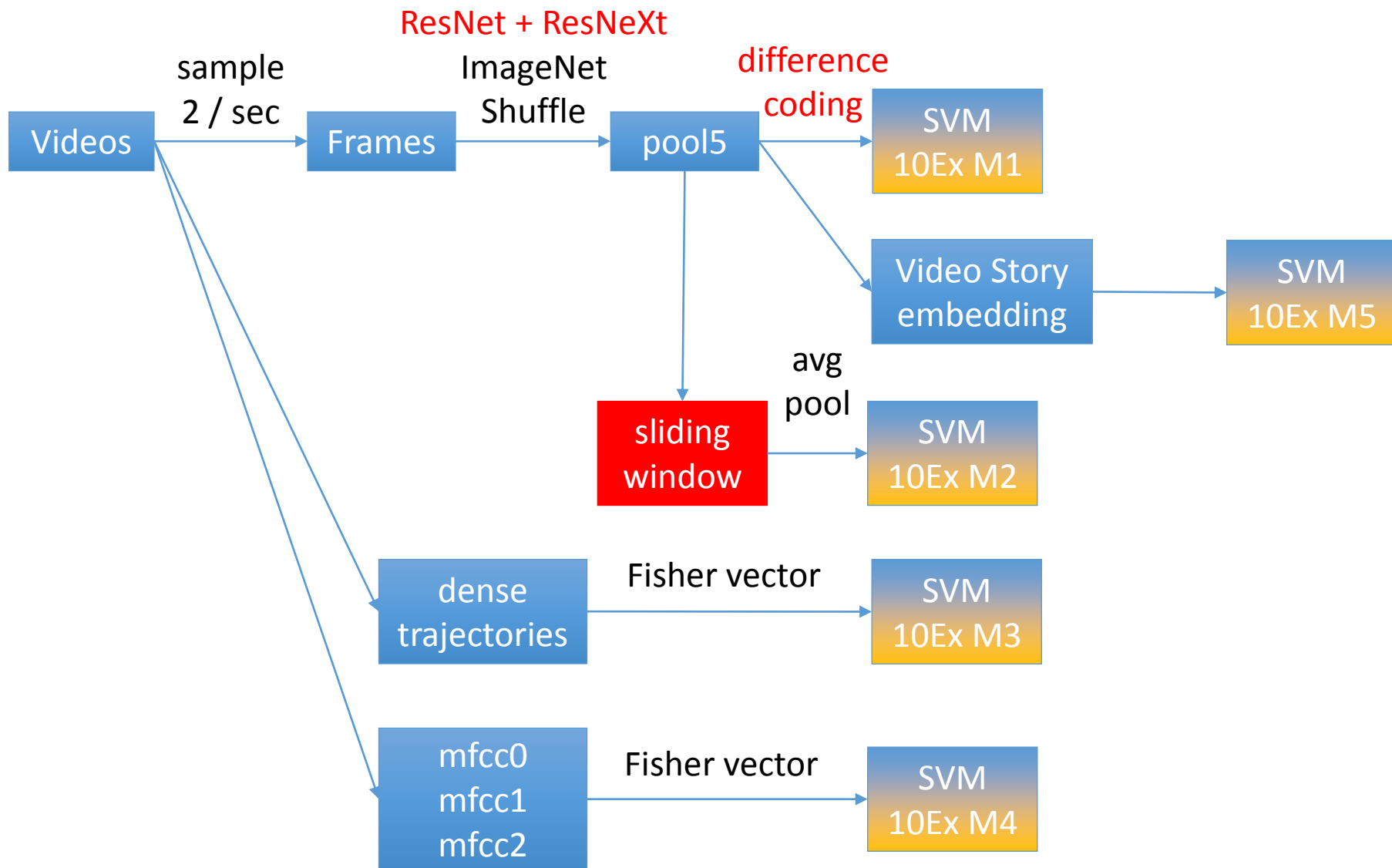
The Netherlands



Pipeline 10Ex 2016



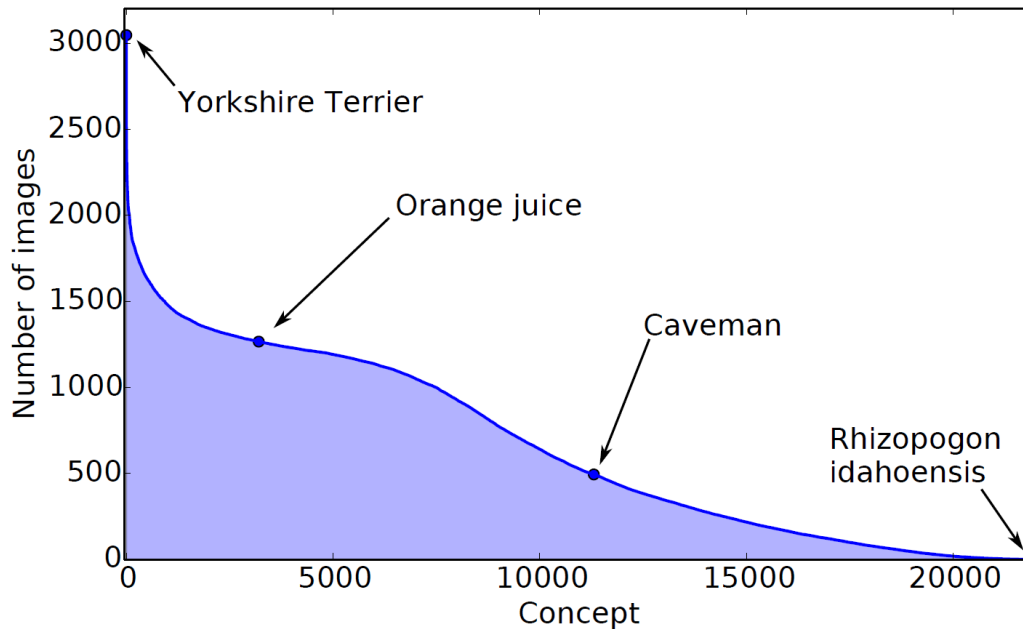
Pipeline 10Ex 2017



CNN Features from 22k ImageNet classes

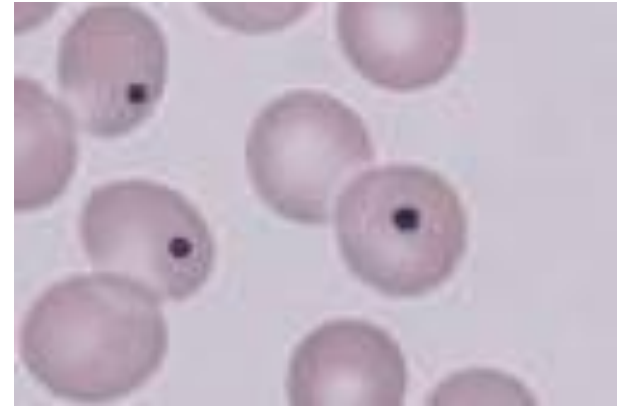
- Use as many classes as possible
- Find a balance between level of abstraction of classes and number of images in a class

Example imbalance



296 classes with 1 image

Irrelevant classes



Siderocyte



Gametophyte

CNN training on selection out of 22k ImageNet classes

- Idea

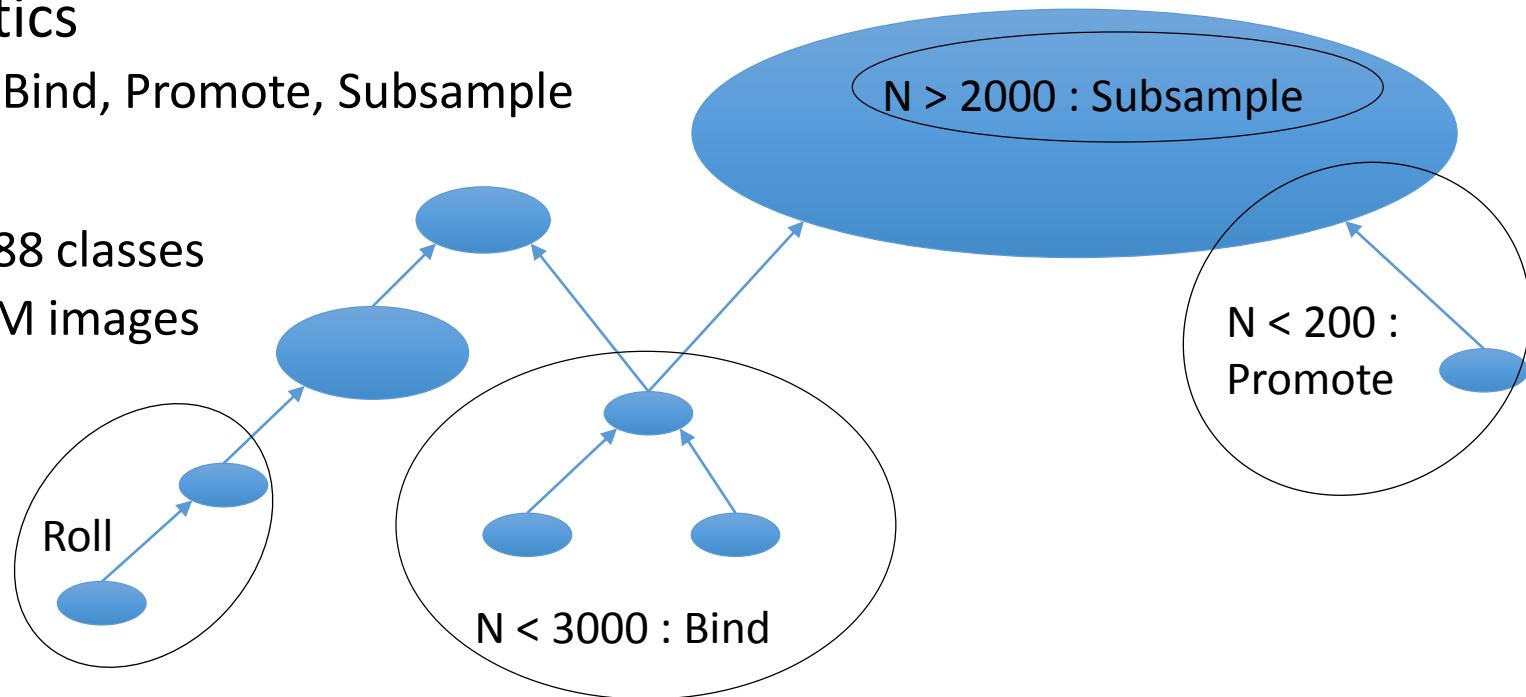
- Increase level of abstraction of classes
- Incorporate classes with less than 200 samples

- Heuristics

- Roll, Bind, Promote, Subsample

- Result

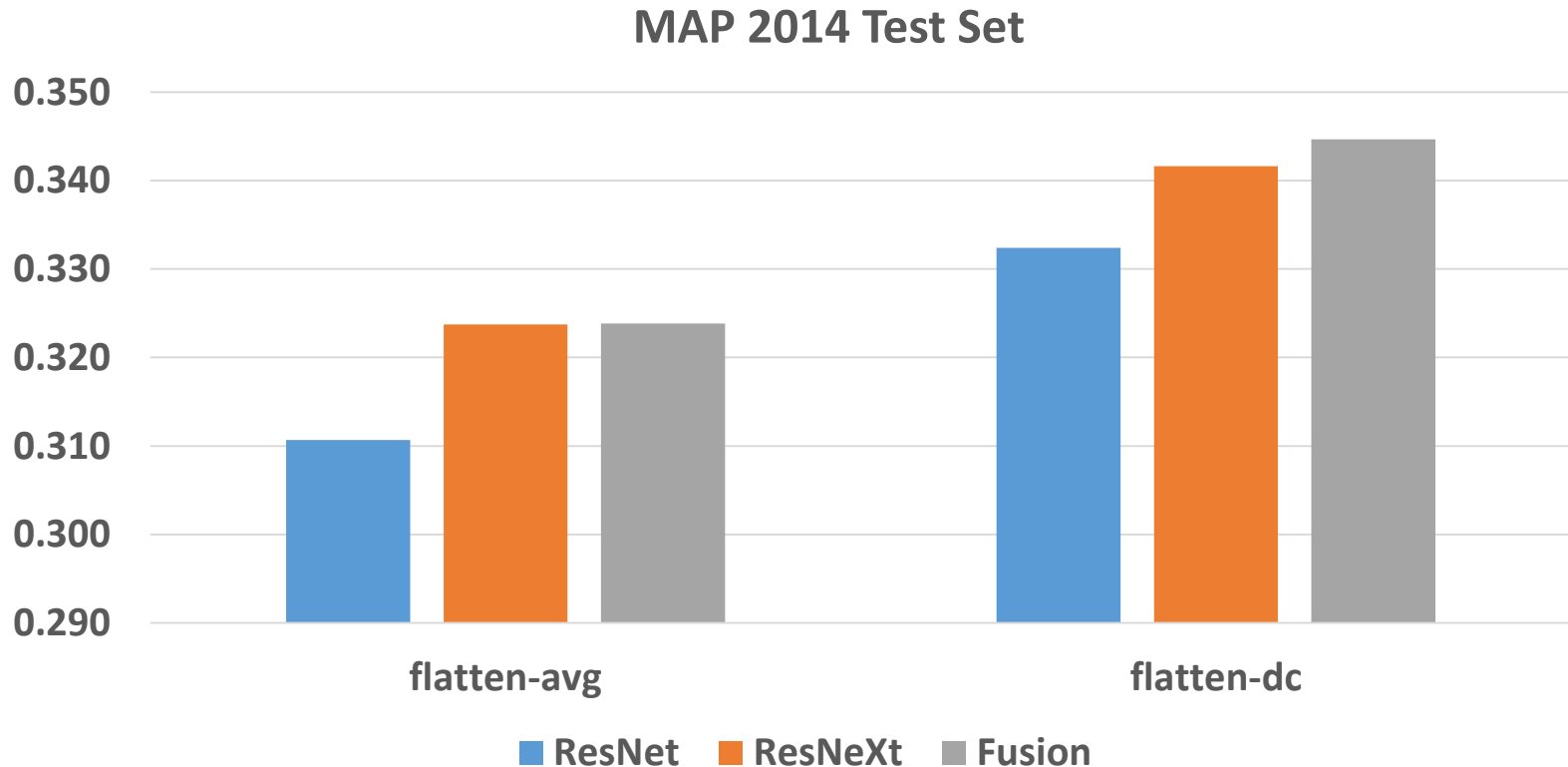
- 12,988 classes
- 13.6M images



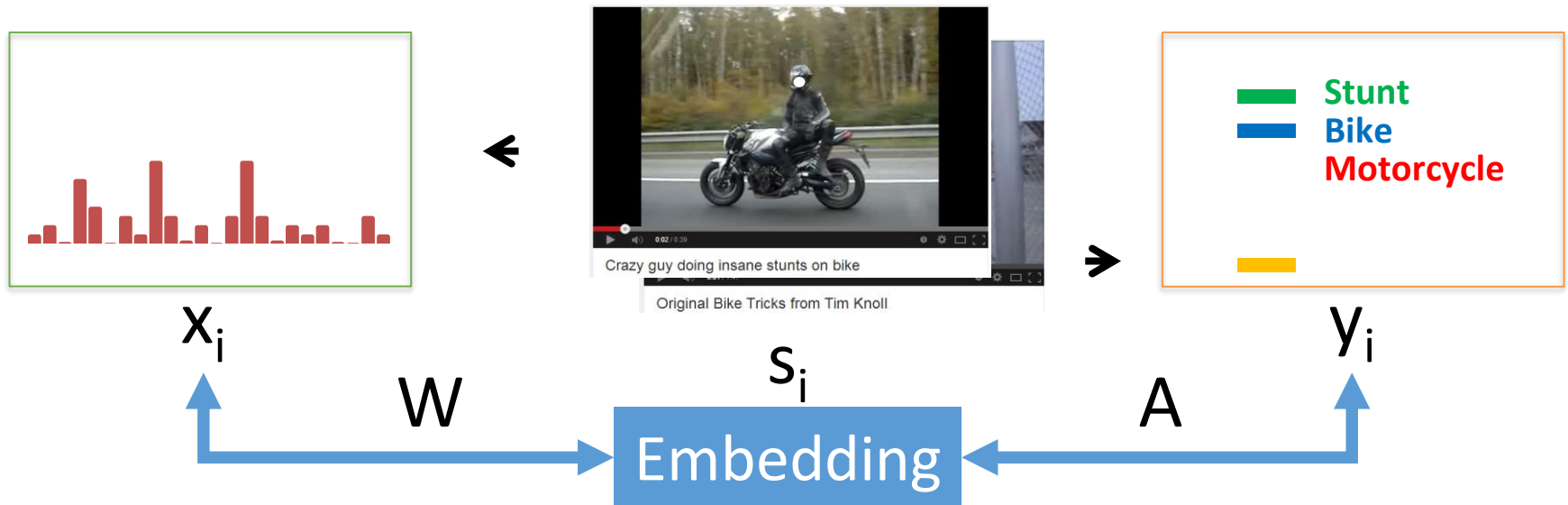
The ImageNet Shuffle: Reorganized Pre-training for Video Event Detection,
Pascal Mettes and Dennis Koelma and Cees Snoek,
International Conference on Multimedia Retrieval, 2016

Feature Difference Coding

- K-means clustering ($k = 5$) on last fully connected layer before probability layers (called flatten)
- Fisher like encoding but sigma is based on distance of points assigned to a cluster to its center



Video Story: Embed the story of a video



Joint optimization of W and A to preserve

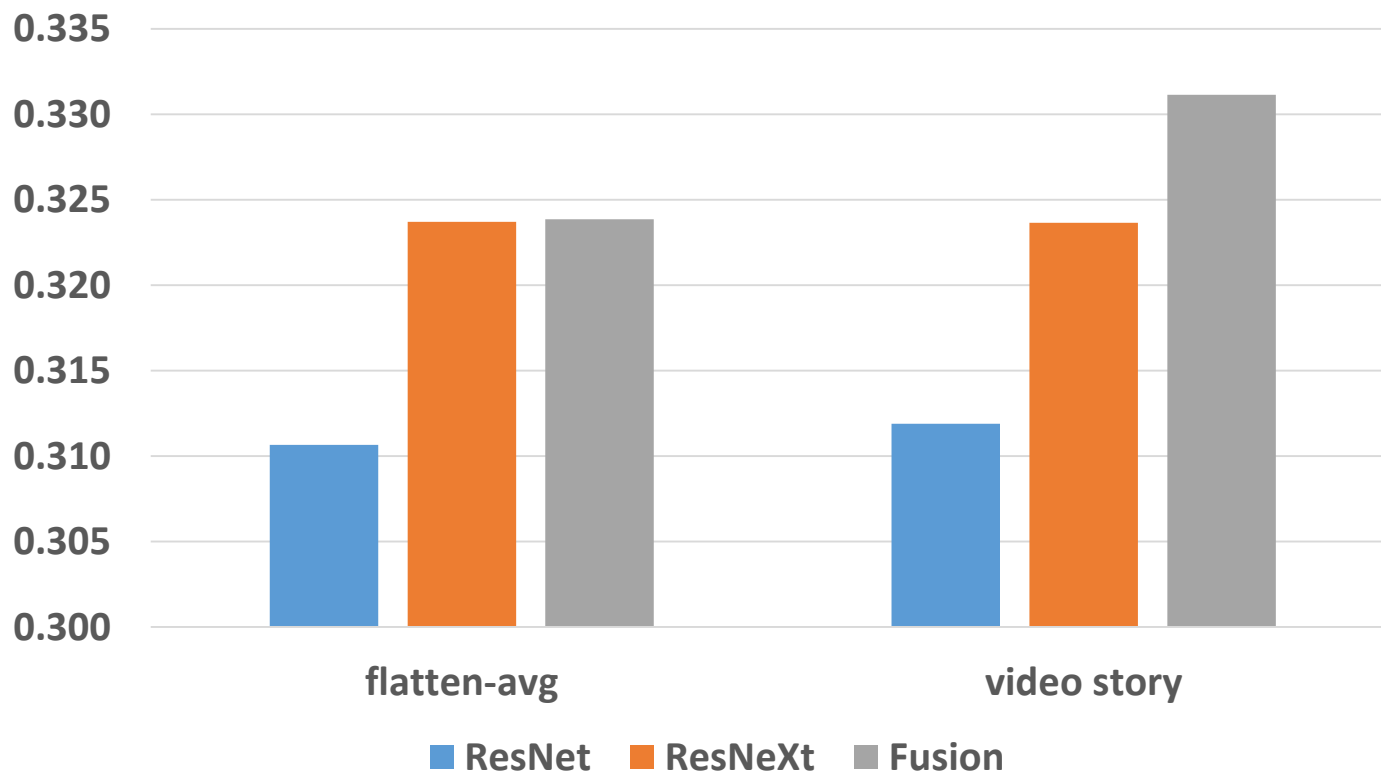
Descriptiveness: preserve video descriptions : $L(A,S)$

Predictability: recognize terms from video content : $L(S,W)$

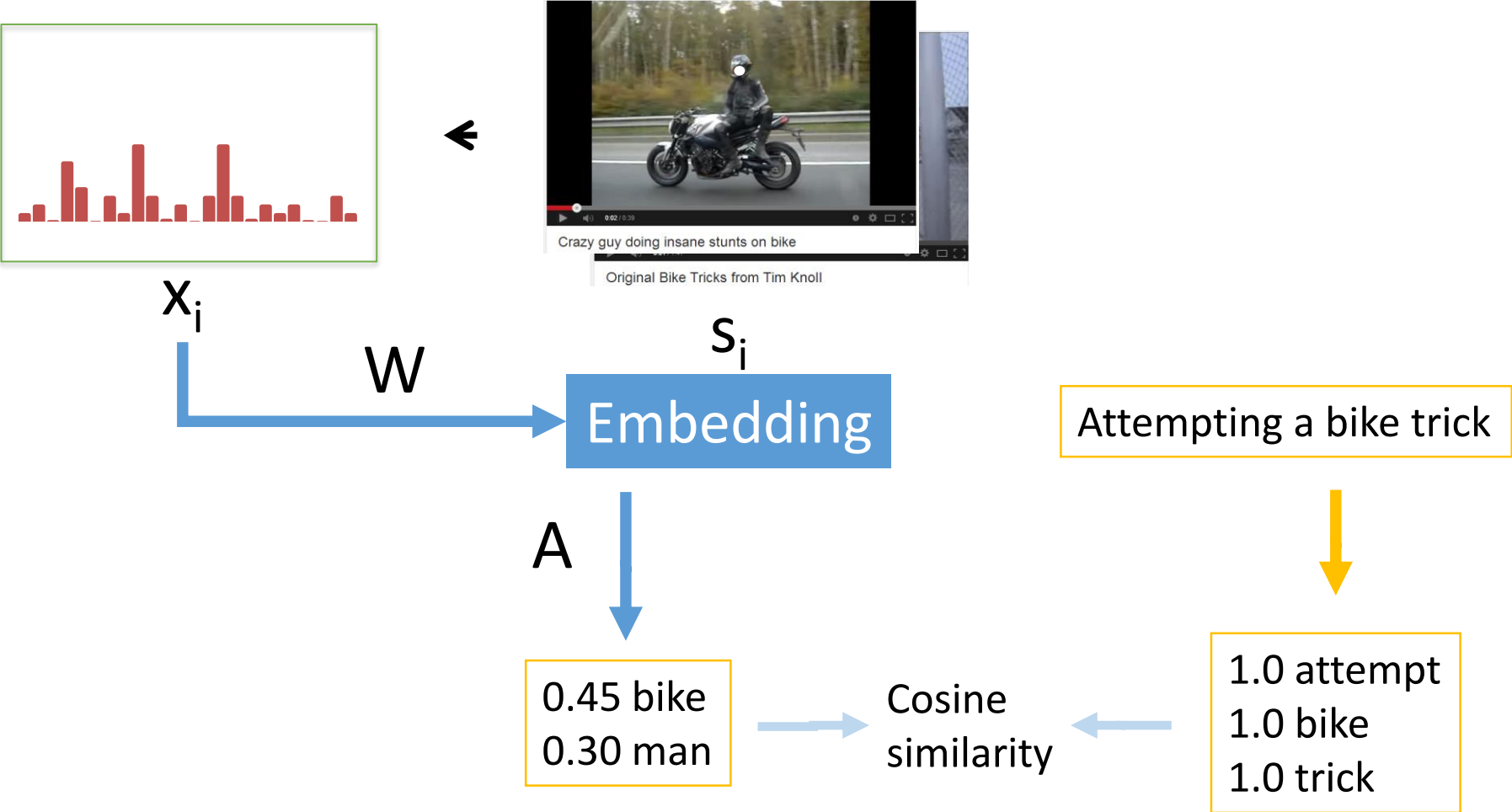
Videostory: A new multimedia embedding for few-example recognition and translation of events,
Amirhossein Habibian and Thomas Mensink and Cees Snoek,
Proceedings of the ACM International Conference on Multimedia, 2014

VideoStory Embedding as a Feature

MAP 2014 Test Set

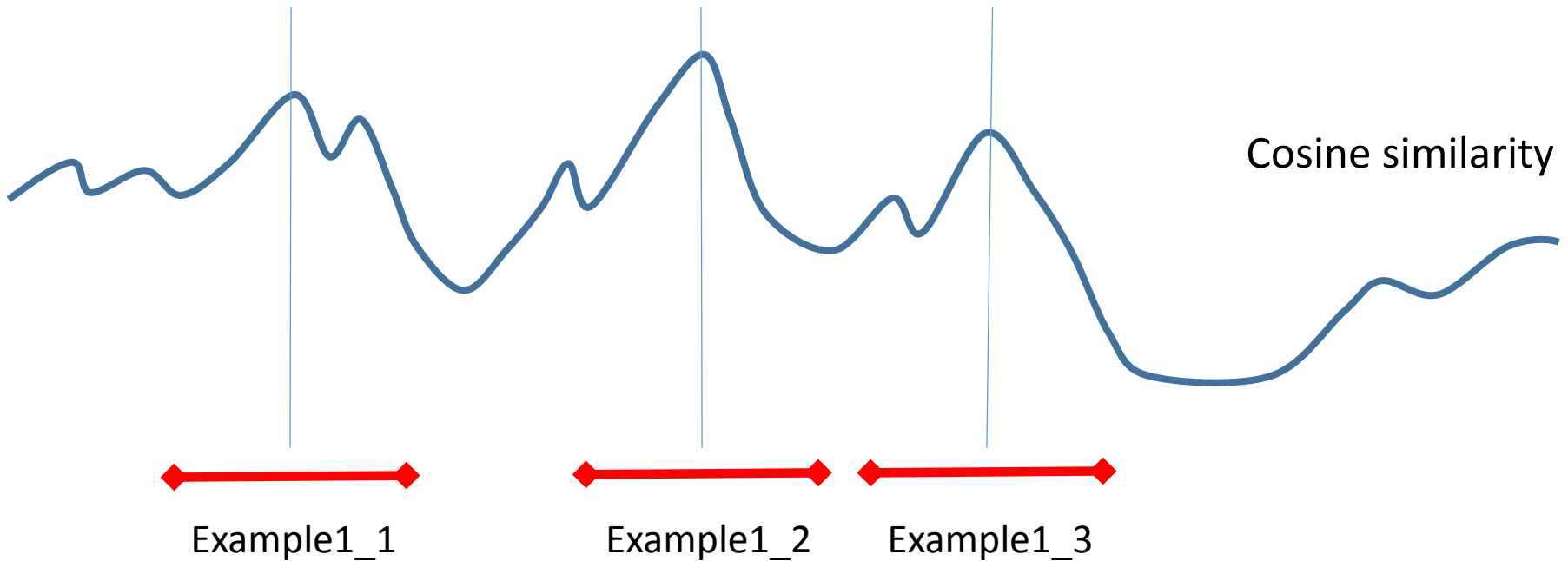


Video Story for OEx



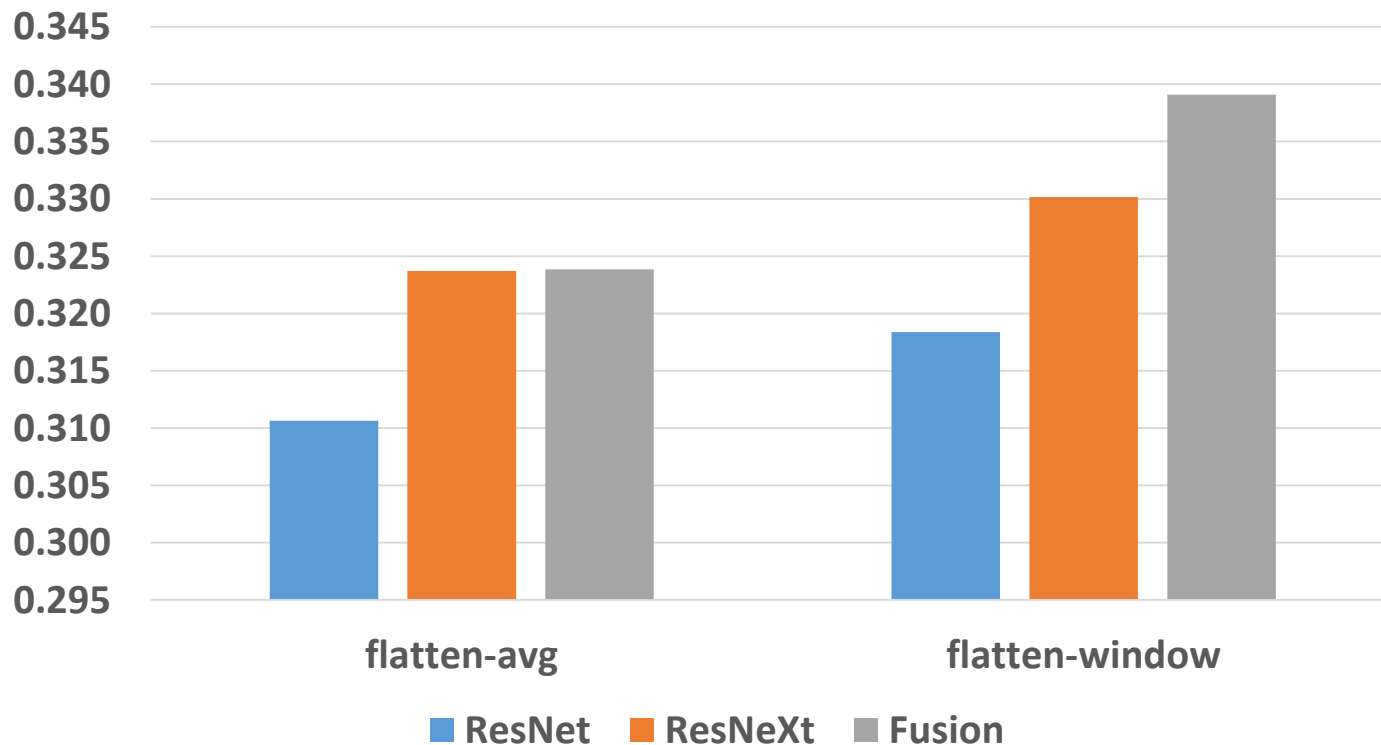
Finding Segments to Expand Training Material

Example1

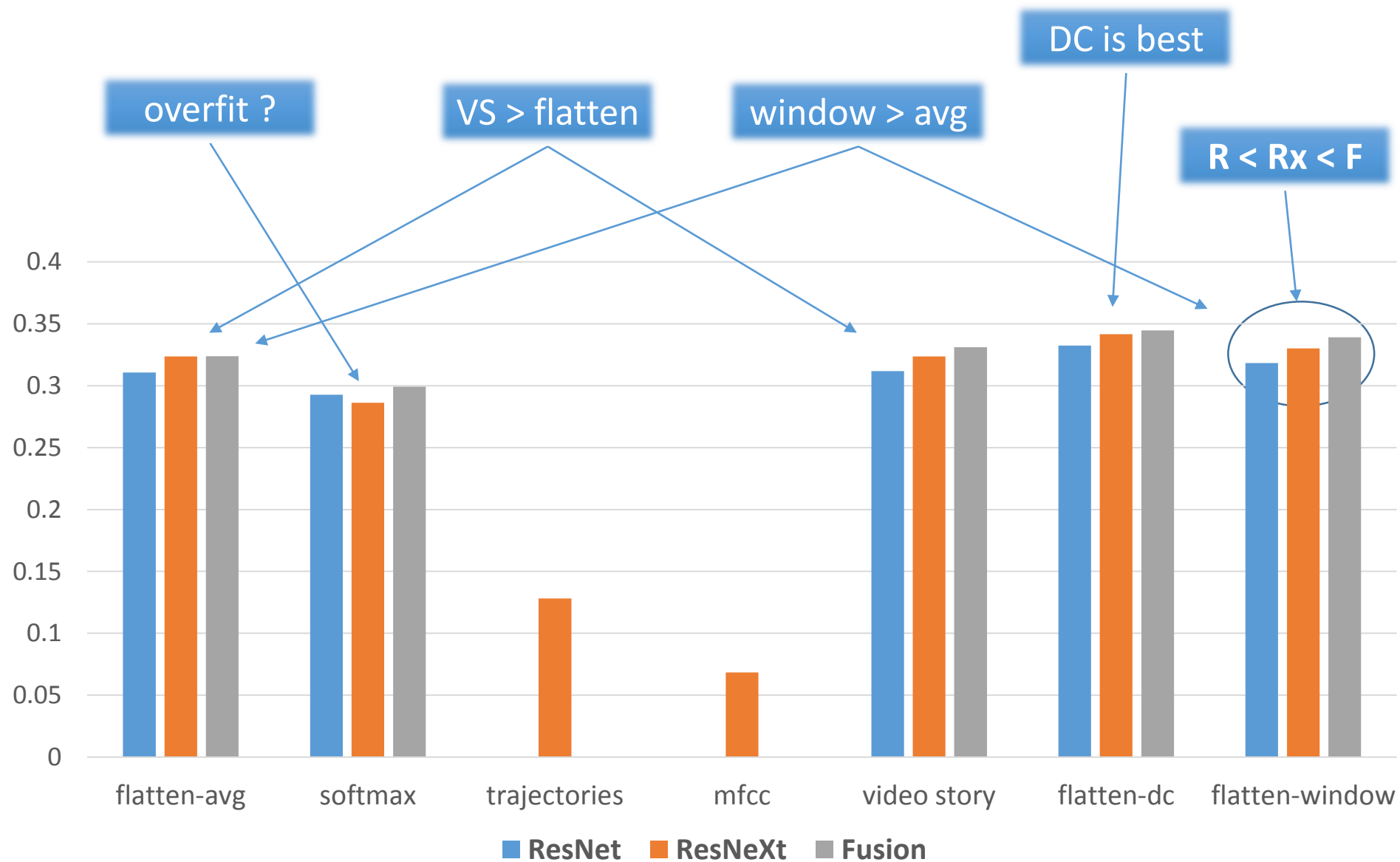


Window based Features

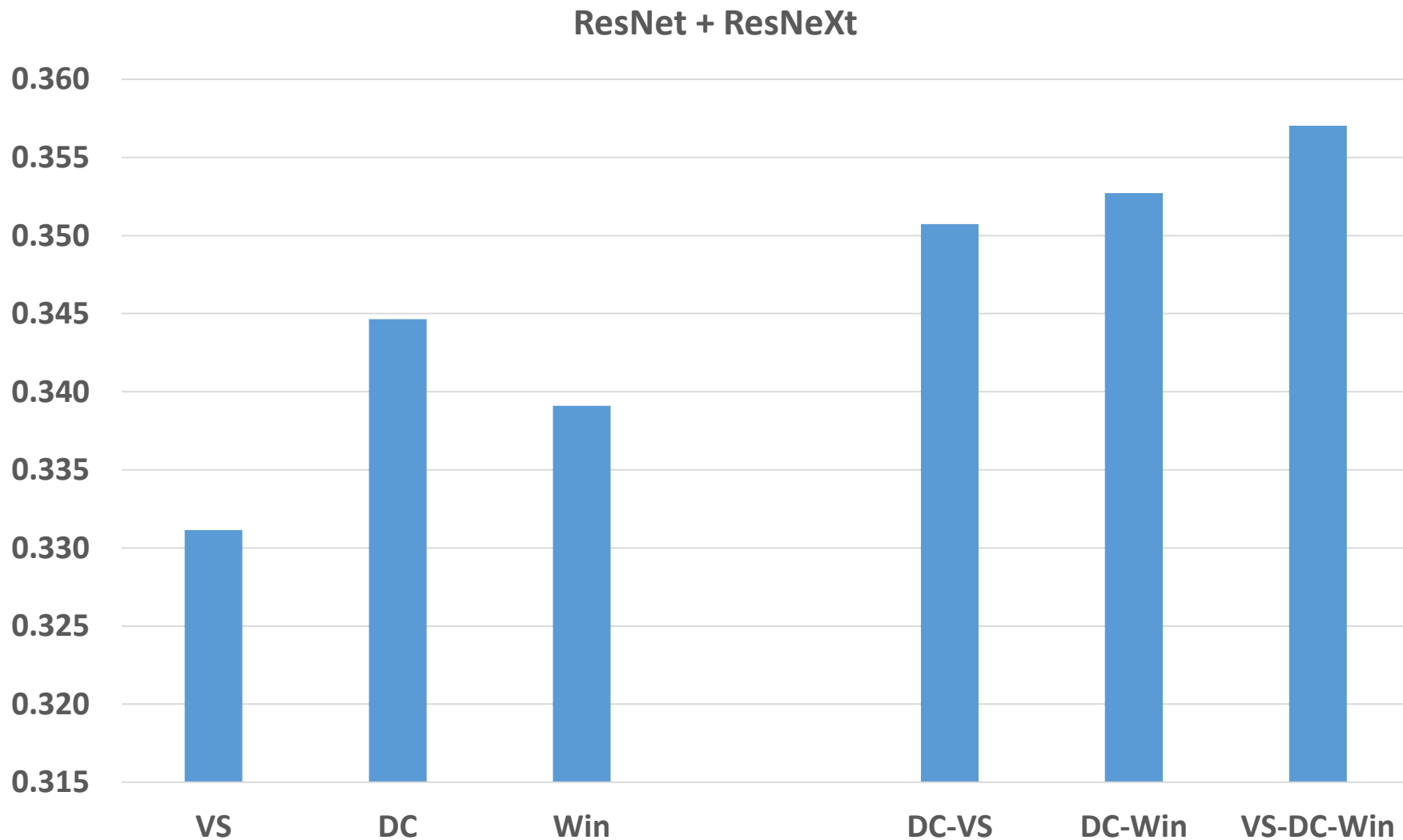
MAP 2014 Test Set



Result Individual Modalities on 2014 Test Set



Fusion Visual Modalities on 2014 Test Set

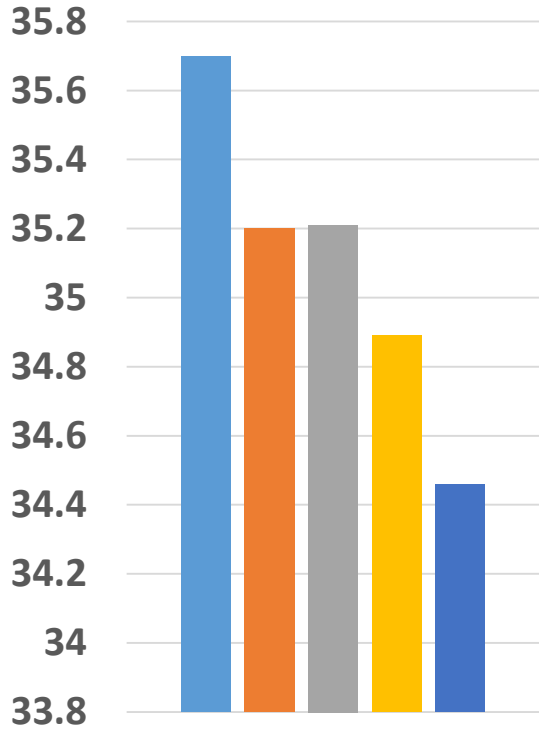


Fusion on 2014 Test Set

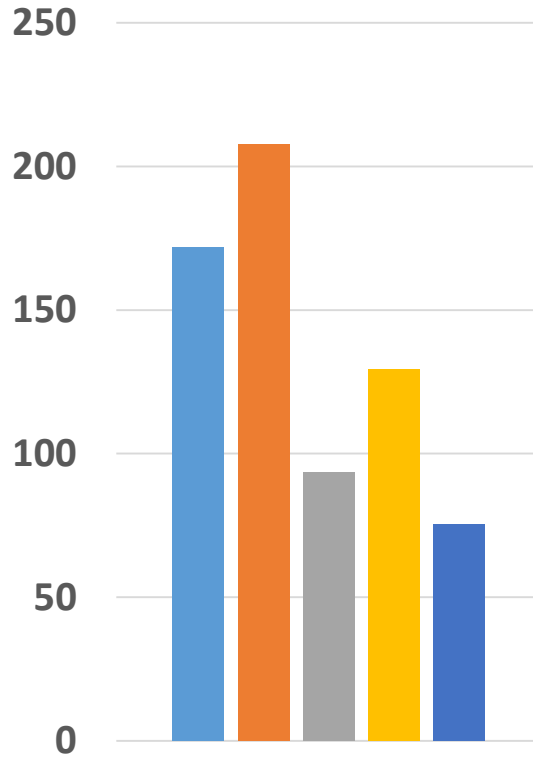


Computational Efficiency

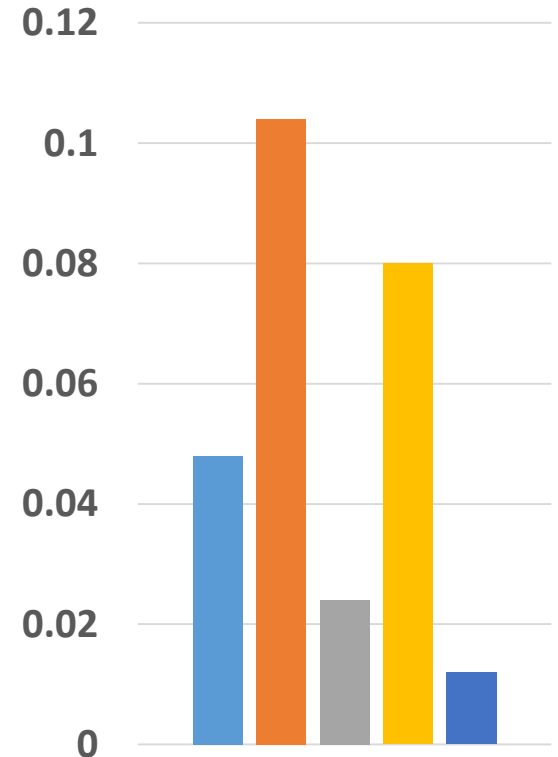
MAP



Feature Extraction



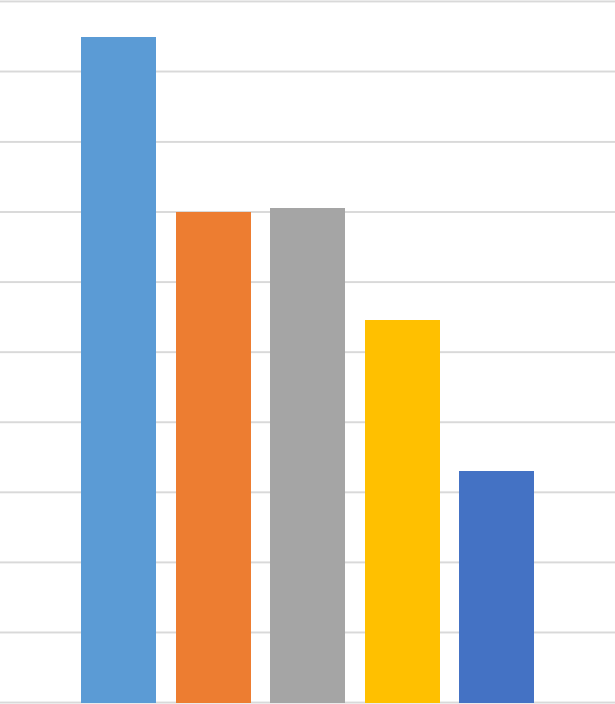
Classification



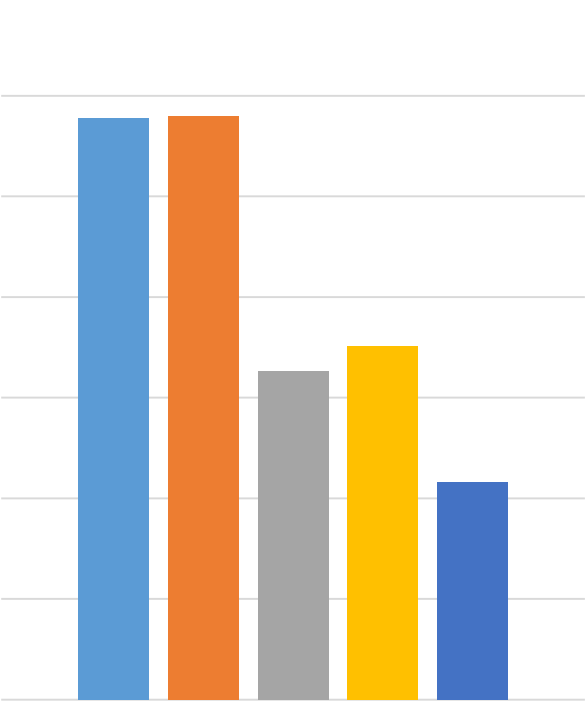
- p-visualFusionTwoCNN
- c-mmFusionTwoCNN
- c-visualFusionOneCNN
- c-mmFusionOneCNN
- c-visualSingle

Our MED Submission

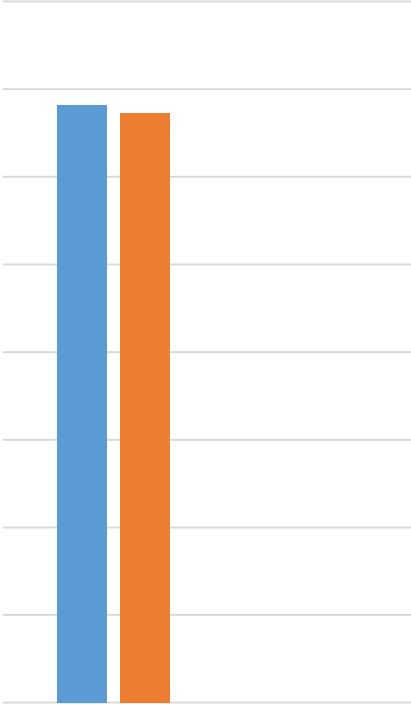
Test 2014



PS



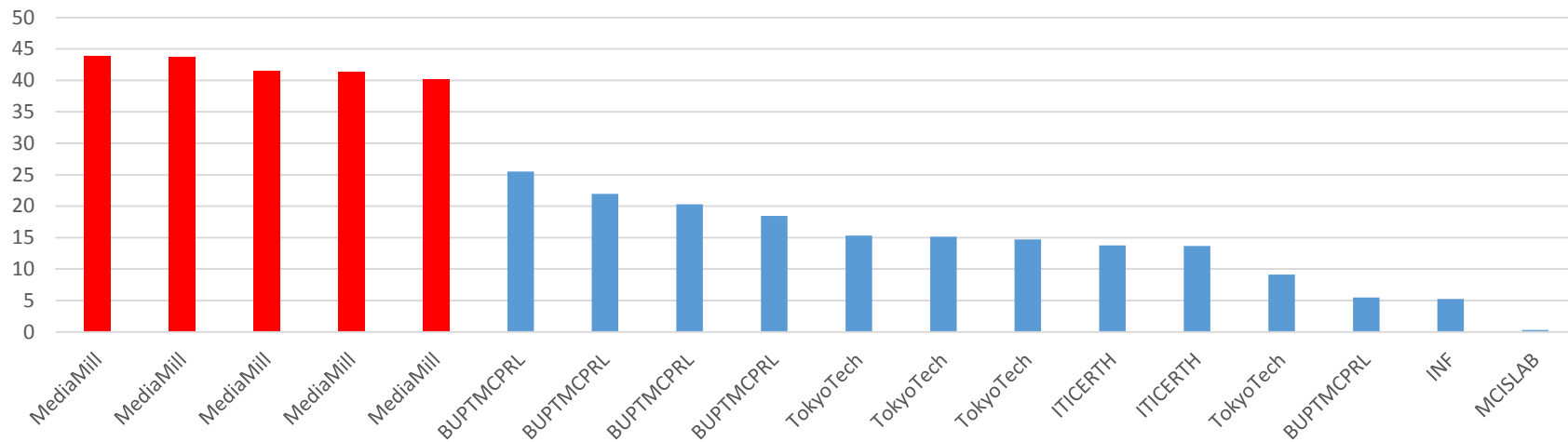
AH



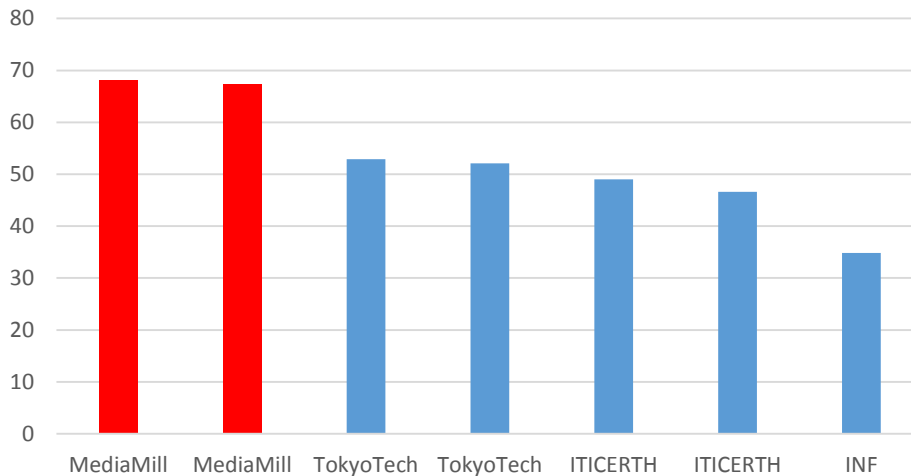
- p-visualFusionTwoCNN
- c-mmFusionTwoCNN
- c-visualFusionOneCNN
- c-mmFusionOneCNN
- c-visualSingle

All MED Submissions

PS



AH



Conclusions

- Visual features are still improving
- Fusion still works but other modalities need work
- Oex helps to get more out of your examples

Thank You