Single view learning in action recognition
Gaurvi Goyal, Nicoletta Noceti, Francesca Odone
MalGa – Machine Learning Genoa center, DIBRIS, Università di Genova

Motivations and objectives

• A main challenge in human action recognition is the fact that actions may look very different depending on the point of observation.
• To attenuate view-point changes many approaches resort to additional information, such as depth or 3D poses, while others rely on videos of an action acquired simultaneously from multiple viewpoints.
• We address the challenging problem of learning high level view-tolerant representations of an action from a single view point, through a domain adaptation procedure that transfers information from a generic action recognition network.
• Our approach, with the possibility to train on small scale datasets, has the crucial benefit of requiring limited resources in terms of computation, making it cost time effective and taking a step towards sustainable methods in Computer Vision.

Our approach

We used a pre-trained model based on inception 3D [1] that provides the best accuracy on the large Kinetics [2] dataset, using optical flow as input.

• We extract mid-high level features from an intermediate point, a compromise between representation effectiveness and transferability.
• As analysis of the features in a space of low dimensionality (with TSNE) reveals that a covariate shift is present in the data.

Experimental analysis

NTU [3]

MoCA [4]

Conclusion

Pre-learned features provide a tolerance to view-point changes, with a bias towards human-human interaction viewpoints (the most frequent on large benchmark datasets). Views that are less frequent - e.g. ego as view 1 of the MoCa, or top as view 4 of IXMAS – may need further attention

REFERENCES