Deep Sequential Context Networks for Action Prediction
Yu Kong, Zhiqiang Tao, and Yun Fu
Northeastern University, Boston, MA, USA
(yukong, zqtao, yunfu}@ece.neu.edu

Objective: inferring the action label after the entire action execution has been observed.

Challenges: classifying actions from temporally incomplete data.

Basic idea: Learn context information from full videos, and transfer to partial videos.

Up to 300x faster! Effective and efficient!

Definitions

Uniformly segment a video into $K$ segments

Deeplearning architecture $Z = (Z_0, Z_1, \ldots, Z_K)$

Segment $x$ Partial video $x^{(k)}$ Observation ratio $r = k/K = 0.3$

Visual features: C3D, dense trajectory

Table 3. Top 220 action categories having over 50% occurrence ratios are summarized in testing.

4.6 Effectiveness of Components and Parameters

In addition, training ONE method is significantly faster than MTSSVM and C3D+Chi-square SVM.

Results indicate that our method is significantly faster than both ONE and the no-CS method. The constraints implications in Eq. (7) lower the performance. The strength of the constraints in Eq. (7) is unknown without too much performance decrease.

Practical

Use single one SVM

Impractical

One SVM needs to be picked from K SVMs

UCF-101 dataset

Sports-1M dataset

Performance in practical and impractical scenarios.

Impractical

References

3. Y. Cao. etc. Recognizing human activities from partially observed videos. CVPR, 2013.