Adversarial Cross-Domain Action Recognition with Co-Attention
Boxiao Pan*, Zhangjie Cao*, Ehsan Adeli, Juan Carlos Niebles

Motivation and Idea

**Motivation:**
1. Transfer learning for video action recognition;
2. Temporal misalignment problem leads to suboptimal performance.

**Idea:**
Find the semantically similar video segments across domain and align them (co-attention).

Model

**Co-attention Module:**
1. Co-attention matrix: assigns high scores to segments that are both important action indicators and common across domains;
2. Co-attention vectors: row / column summation of co-attention matrix, which are used during classification;
3. Attention network: predicts target co-attention vector at test time (no access to source videos during then).

**Discriminator:**
1. Video-level discriminator: traditional distribution matching;
2. Segment-level discriminator: ensures that the aligned features follow the source feature distribution.

Experiments: Quantitative

**Datasets:**
1. Olympic Sports - UCF50;
2. HMDB51 - UCF101;

**Observations:**
1. TCoN outperforms all previous methods on all datasets, with the largest margin on Jester;
2. Optical flow model consistently outperforms RGB model.

Comparison with prior works

Experiments: Qualitative

Co-attention Matrix Visualization:
We visualize the co-attention matrix for a pair of source-target videos. Only the segments that are both important and common across domains will be paid high attention to.