Towards Universal Representation for Unseen Action Recognition

Unseen Action Recognition (UAR) aims to recognized novel action categories without example. Previous methods focus on inner-dataset seen/unseen splits, but it can’t recognize the unseen action outside training examples. UAR uses a large-scale training source to achieve a Universal Representation (UR). UAR first leverages the power of deep neural networks to extract visual features in training. It extracts deep features for each frame and summarise the video by essential components thatare kernelised by Generative Multiple Instance Learning (GMIL) that all the visual features (instances) in a video share the label while only a small portion is determinative. UAR preserves shared components with the label embedding to achieve UR using NMF with JSD and encodes unseen actions with GMIL using the same essential components in ActivityNet to achieve a matching using UR. UAR is extended to more realistic Cross-Dataset UAR (CD-UAR) scenarios so we can directly recognise unseen categories from new datasets without further training or fine-tuning.. Unseen actions in new datasets can be directly recognised via the UR without further training or fine-tuning on the target dataset.