

# Method Summary

Team Name: HERO\_AN

Inspired by the effectiveness of sparse sample, we combine the advantages of TRN and TSN to discriminate the action in this action challenge. Specifically, TRN has the ability of reasoning the temporal relation of video frames. However, the sparse sample strategy used in TRN would discard essential motion information if it samples in major interval or minor interval coincidentally. Besides, due to the high computation complexity, the efficiency would degrade dramatically. For TSN, although it achieves desirable performance, it just averages the predictions of three segments, without the reasoning ability. Different from TRN, TSN could capture frame information at every segments.

Based on above analyses, we propose a comprehensive sparse sample strategy. We divide video frames into four segments averagely. Then, we randomly sample four frames in each segment, which means we get 16 frames from a video. The resulted 16 frames are fed into ResNet-34 pretrained on ImageNet to extract the feature representation of each frame. These representations are concatenated into a feature vector and classified by two fully-connected layers.

That is all we used in the submission.