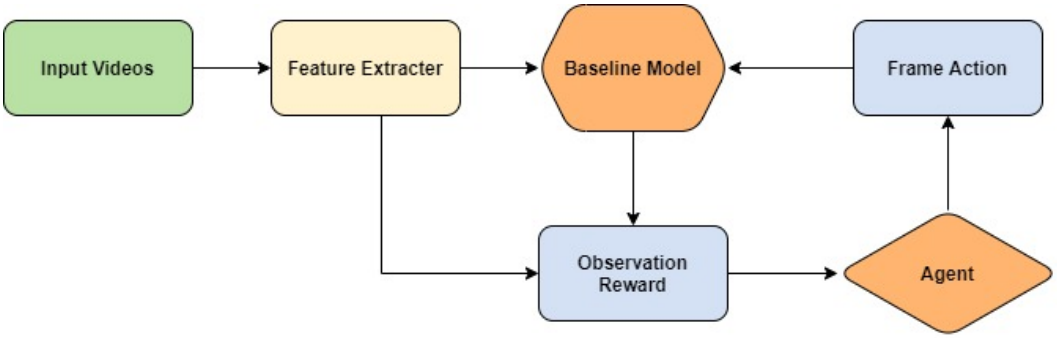
Attention-Based Sampling via Deep Reinforcement Learning for Video Emotion Recognition

Wenqian Ye
Courant Institute of Mathematical Science
New York University
NY 10003
wy2029@nyu.edu

Yunsheng Ma
Courant Institute of Mathematical Science
New York University
NY 10003
ym2382@nyu.edu

1 Proposal



Deep learning based methods have achieved remarkable progress in video emotion recognition. Existing traditional Convolutional Neural Network methods like 3D-CNN, Temporal Segment Networks (TSN), Temporal Relation Network (TRN), Temporal Pyramid Network (TPN) has achieved good performance on Video Recognition tasks in user-generated videos (UGV). However, most of these methods focus on the spatio-temporal and channel-wise property of the target video with treating all the video frames equally thus ignore this sampling factor. In this paper, combined with the state-of-the-art Video Swin-Transformer as baseline model, we propose an attention-based method to increase frame sampling efficiency using Deep Reinforcement Learning. We formulate an agent to discriminate key frames from video as a Deep-Q Learning Problem. This agent take feature and prediction as observation from baseline model without labels and generate content importance measurement of each frame.

In this project, the main focus is to figure out the agent reward and action on frames since there are no direct reward from the observation. According to the feature from dataset and the prediction from baseline model, we can design a reward mechanism as an incentive of selecting label-sensitive frames from the whole video as an expression of probability. Various agents could be trained using Q-learning or Policy Gradient based on the performance and the difficulty of implementation. Considering the task-specific problems, key frames of emotion recognition videos might be continuous, which might require mini-batchs decision on the reinforcement learning agent. The batch configuration could be discussed later. Beyond that, the action to take is also complicated since the aggressiveness of the policy may result in some kinds of dilemma.

Preprint. Under review.