

Abstract

Group activities usually involve spatiotemporal dynamics among many interactive individuals, while only a few participants at several key frames essentially define the activity. Therefore, effectively modeling the group-relevant and suppressing the irrelevant actions (and interactions) are vital for group activity recognition. In this paper, we propose a novel method based on deep reinforcement learning to progressively refine the low-level features and high-level relations of group activities. Firstly, we construct a semantic relation graph (SRG) to explicitly model the relations among persons. Then, two agents adopting policy according to two Markov decision processes are applied to progressively refine the SRG. Specifically, one feature-distilling (FD) agent in the discrete action space refines the low-level spatiotemporal features by distilling the most informative frames. Another relation-gating (RG) agent in continuous action space adjusts the high-level semantic graph to pay more attention to group-relevant relations. The SRG, FD agent, and RG agent are optimized alternately to mutually boost the performance of each other. Extensive experiments on two widely used benchmarks demonstrate the effectiveness and superiority of the proposed approach.

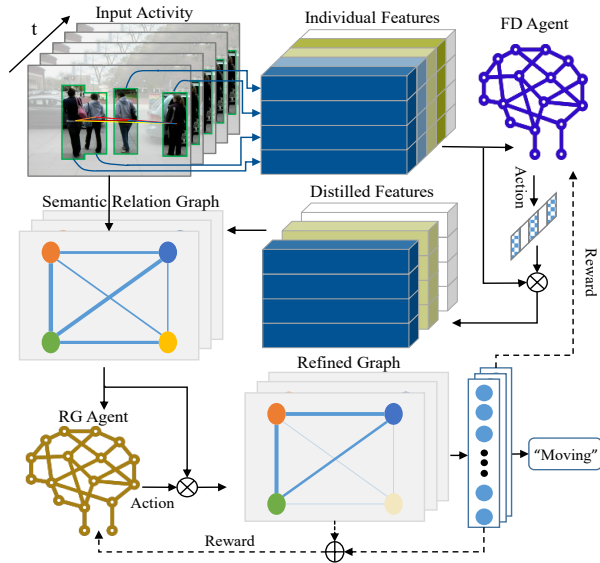


Figure 1: The overview of proposed method. A feature-distilling (FD) agent progressively selects the most informative frames of the low-level spatiotemporal individual features. A relation-gating (RG) agent further progressively refines the high-level semantic relation graph (SRG) to discover group-relevant relations.