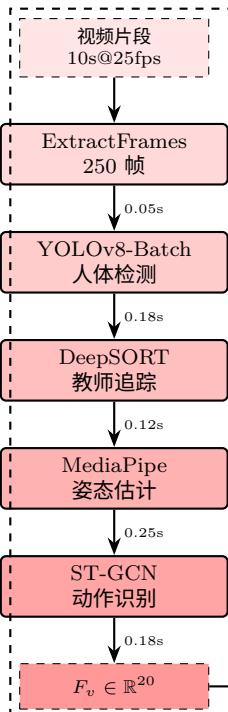


多模态特征提取流程图 (Algorithm 1)

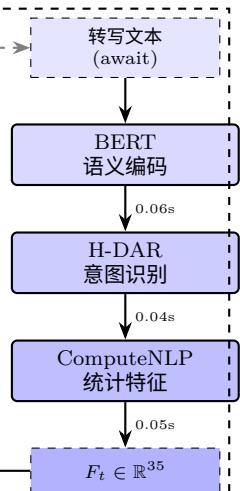
视频流水线 (0.82s)



音频流水线 (0.37s)



文本流水线 (0.15s)



特征合并

$$F = \{F_v, F_a, F_t\} \in \mathbb{R}^{70}$$