The task of temporally detecting and segmenting actions in untrimmed videos has seen an increased attention recently. One problem in this context arises from