Current studies on action detection in untrimmed videos are mostly designed for action classes, where an action is described at word level such as jumping, tur