The goal of this paper is to detect the spatio-temporal extent of an action. The two-stream detection network based on RGB and flow provides state-of-the-art a