In this paper, we propose Spatio-TEmporal Progressive (STEP) action detector--a progressive learning framework for spatio-temporal action detection in videos