

Temporally localizing actions in a video is a fundamental challenge in video understanding. Most existing approaches have often drawn inspiration from image o