

Most state-of-the-art methods for action recognition consist of a two-stream architecture with 3D convolutions: an appearance stream for RGB frames and a motion stream for optical flow frames.