

Automatic generation of video captions is a fundamental challenge in computer vision. Recent techniques typically employ a combination of Convolutional Neural