

Recognising actions in videos relies on labelled supervision during training, typically the start and end times of each action instance. This supervision is not only