

Spatiotemporal feature learning is of central importance for action recognition in videos. Existing deep neural network models either learn spatial and temporal f