Dense video captioning is an extremely challenging task since accurate and coherent description of events in a video requires holistic understanding of video co