Temporal Video Grounding

Given a '{video domain}' video, find a clip described by the sentence: '{query}'. Provide a detailed description of the clip.

Episodic Memory

Given a '{video domain}' video, find a clip that can answer the question: '{query}'. Provide a detailed description of the clip and give the answer.

Video Highlight Detection

Given a '{video domain}' video, find a highlight moment according to the sentence: '{query}'. Provide a detailed summary for the moment.

Extractive Video Summarization

Given a '{video domain}' video, summarize the video into a brief synopsis by focusing on the clips that contain the most important visual cues related to the topic.

Temporal Action Localization

Given a '{video domain}' video, find all the clips belonging to the action category: '{query}'. Provide a concise description for each clip.

Dense Video Captioning

Given a '{video domain}' video, densely summarize the video by localizing clips that depict major activity events. Describe each event by focusing on action verbs.

Step Captioning and Localization

Given a '{video domain}' video, localize a series of clips that contain the major steps in the video and briefly describe each clip.

Temporal Event Matching

Given a '{video domain}' video, identify the action in <vid> - <vid> and localize another most similar action. Provide a detailed description of the action.