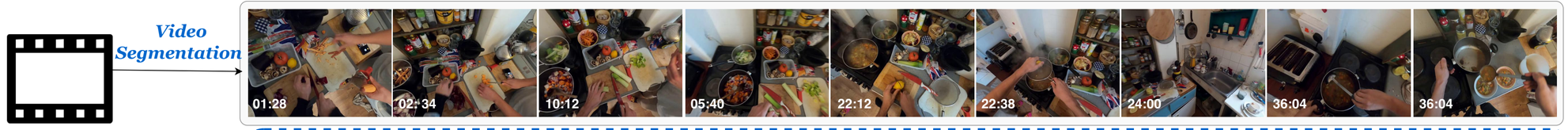


[Video QA] What type of fruit did the camera wearer cut and add to the dish?

[Summarization] Summarize key interactions and events in the video.

[Video Retrieval] #C looks around the kitchen. /#C was in the kitchen with a man Y, peeled, cut onions ...

Inference on Downstream Task

Text Retriever

[00:00-04:58] Two people are chopping up vegetables on a cutting board ...

[09:54-11:14] Two people are preparing vegetables in a kitchen. One person is chopping ...

[22:10-24:24] A person is chopping a lemon in the kitchen. They then squeeze it into a pot on the stove and add the juice to a bowl of foo

[36:20-37:36] A person is stirring soup in a pot on the stove. They are then pouring the soup into...

(i) High-Level Timeline Description

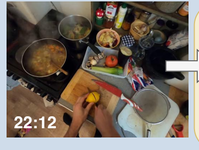
(ii) Multi-aspect Coarse-grained Description

[Action Sequence Description]
[22:10-23:48] {'action description': cutting lemons on a cutting board, 'object description': lemon and knife}, ...
[23:50-24:24] {'action description': ...

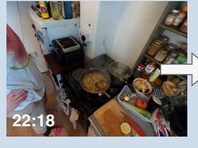
[Scene Sequence Description]
[22:10-23:48] {'scene description': A person is chopping a lemon, 'setting': kitchen, 'action': chopping}, ...
[23:50-24:24] {'scene description': ...

[Object Description]
[22:10-23:48] There are two men, a stove, pots, pans, ...
[23:50-24:24] There are a man, a knife, a chopping board, ...

(iii) Multi-modal Fine-grained Description



[Spatial Description]
{'object_name': kettle, 'number': 1, 'spatial_relationship': on the table, right of the kitchen}, ...



[Spatial Description]
{'object_name': person, 'number': 1, 'attributes': watermelon on shirt, 'spatial_relationship': left of the stove, next to the counter}, ...

...

MMViR

[Video Answer] Lemon

[Summary] People prepare and cook food in a kitchen, involving chopping vegetables...

[Retrieved Timestamp/Time Range] 00:01 / 00:10-01:00