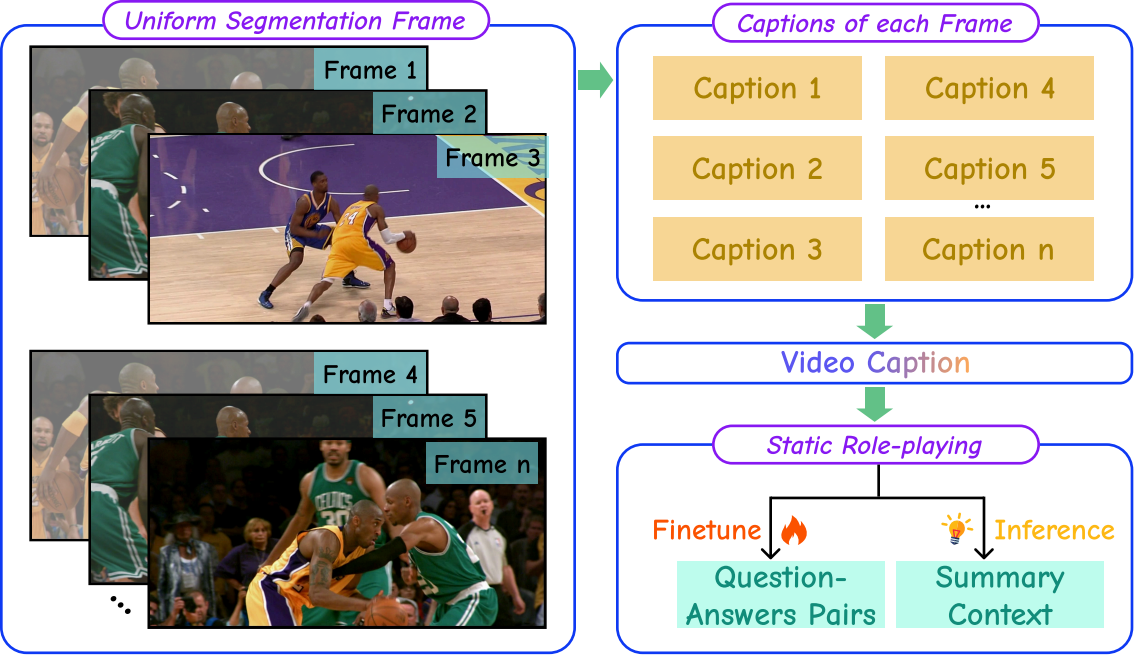
Uniform Segmentation Frame
Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame n
Captions of each Frame
Caption 1
Caption 4
Caption 2
Caption 5
...
Caption 3
Caption n
Video Caption
Static Role-playing
Finetune
Inference
Question-Answers Pairs
Summary Context