

NLMR (Text Query): " the man then rides the bike and it moves across the snow."

SLMR (Audio Query):

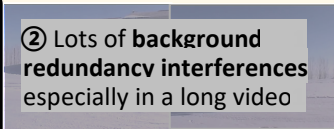


Backward

Current Clip

① Fading temporal association

Forward



② Lots of **background redundancy interferences** especially in a long video

③ **Specific modal constraints** (①② for video) in cross-modal guidance

Ground Truth (GT): 80.75s |----->| 147.34s