



Prompt) A video and multiple pieces of information will be provided to you. One of these pieces of information matches the content of the video, while the remaining ones do not. Your objective is to pinpoint the information that is in harmony with the video and craft a suitable video caption.\nInformation A: {'subject': 'entire video', 'order': 'a white car is driving while snow particles are melting simultaneously'}\nInformation B: {'subject': 'entire video', 'order': 'snow particles are melting, then a white car is driving'}\nInformation C: {'subject': 'entire video', 'order': 'a white car is driving, then snow particles are melting'}\n

Ground Truth) B. snow particles are melting, then a white car is driving

LLaVA-OV-7B) A white car is driving **while** snow particles are melting **simultaneously**.

+TCD) A white car is driving **while** snow particles are melting **simultaneously**.

+DINO-HEAL) A white car is driving **while** snow particles are melting **simultaneously**.

+SEASON) As the video progresses, the **snow particles continue to melt**, revealing more green grass and twigs underneath. **The scene then transitions** to a highway at sunset, where **a white car drives towards** a row of wind turbines on the left side of the road. **The car moves forward**, passing by the wind turbines under a sky filled with scattered clouds, with the sun setting in the distance."