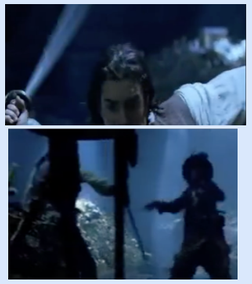
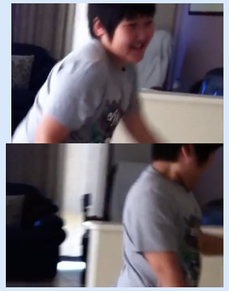
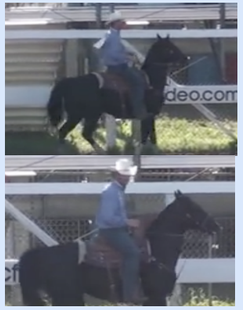
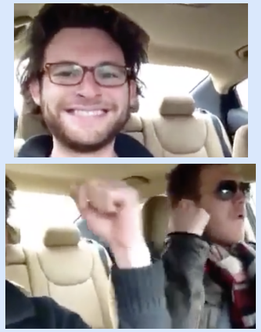
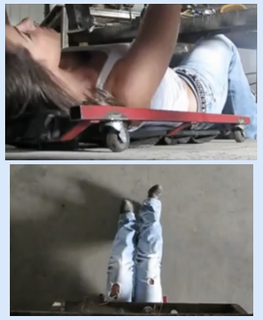
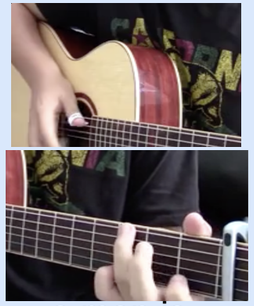
VideoCon LLM-Assisted Contrast Caption Generation

Video Frames							
Video Caption	the man has a **sword** trying to protect himself	A boy performs the dance as he **giggles**	a person riding a **brown** horse	**Two** friends traveling together	A person rolls **under** a car with a board	a man is performing on his guitar	camera passes through a gate then follows the path
Misalignment	**Object**	**Action**	**Attribute**	**Counting**	**Relation**	**Hallucination**	**Event Order Flip**
LLM Prompt	For the given **misalignment** and **video caption**, create a semantically plausible **contrast caption.** Also generate a **natural language explanation** for the difference between the video caption and contrast caption**.**						
Contrast Caption	the man has a **shield** trying to protect himself	A boy performs the dance as he **cries**	a person riding a **green** horse	**Three** friends traveling together	A person rolls **on top of** a car with a board	a man is performing on his guitar **with a band**	camera follows the path then passes through a gate
Natural Language Explanation	a man has a sword to protect himself, not a shield	A boy performs the dance as he is giggling, not crying	a person riding a brown horse, not a green horse	Two friends are traveling together, not three friends	A person rolls under a car, not on top of the car	The man is performing guitar but there is no band	Camera passes through a gate before following the path, not the other way around