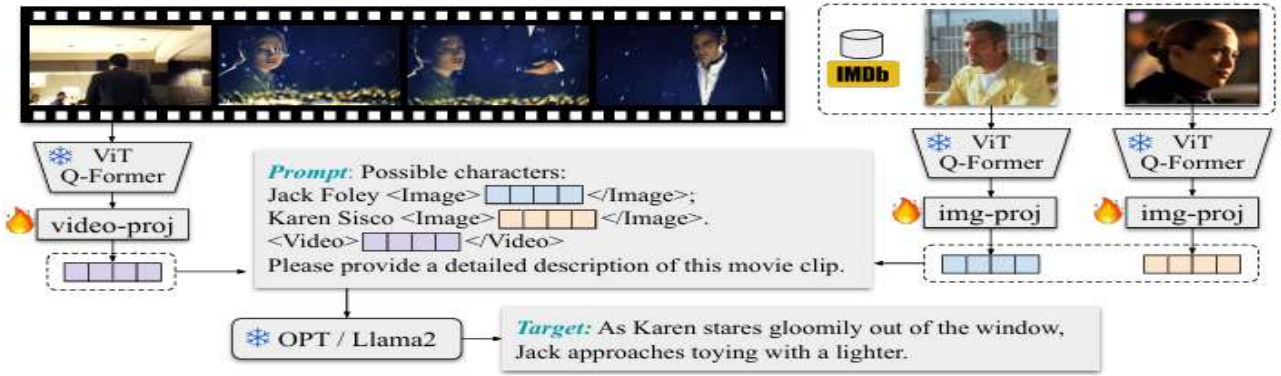
AutoAD: Movie Description in Context

What is the problem?

•**Automatic Audio Description (AD) Generation**: The primary challenge is to generate coherent and contextually relevant audio descriptions for movies to enhance accessibility for visually impaired viewers.
•**Narrative Coherence**: Ensuring that the AD maintains a consistent narrative flow without repetition or irrelevant details is crucial for a seamless viewing experience.

What has been done earlier?

•**Manual and Semi-Automated Approaches**: Previous methods relied heavily on manual processes or basic automation, which often led to:
•Incoherent descriptions that did not align with the visual content.
•Lack of contextual relevance, resulting in a poor user experience.
•**Limited Use of Data**: Earlier models did not effectively utilize comprehensive datasets, leading to suboptimal performance in AD generation.
•**Basic Evaluation Metrics**: Existing metrics for assessing AD quality were often inadequate, lacking sensitivity to nuances in character naming and context.



Amit Kumar Mohapatra,B421005

What are the remaining challenges? What novel solution proposed by the authors to solve the problem?

Challenges:

•**Coherence Across Narration**: Maintaining narrative coherence throughout the movie remains a significant challenge, as AD should not repeat the same information or provide irrelevant details.

•**Contextual Integration**: The integration of contextual elements, such as character actions and emotional tone, into the AD is often insufficient.

•**Evaluation of Generated Descriptions**: Current evaluation methods may not accurately reflect the quality and relevance of generated descriptions, necessitating the development of more robust metrics.

Solution Proposed:

New Architectures:

•**Movie-BLIP2:**

- •Integrates advanced video-language models for better AD generation.

•**Movie-Llama2:**

- •Utilizes a stronger language model to enhance description quality.

Data Collection Methodology:

•**Audio-Audio Alignment:**

- •Collects AD data by aligning audio descriptions with pixel data from public movie snippets.

•**Pseudo-Labeling of Instruction Videos:**

- •Augments training datasets by generating labels from instruction videos.

Enhanced Evaluation Metrics:

•**CRITIC Metric:**

- •Assesses accuracy of character references in generated AD versus human references.

•**LLM-AD-eval:**

- •Uses large language models to evaluate matching quality between predicted and ground-truth ADs.

Performance Improvement:

•Proposed methods show significant performance boosts across metrics, nearing practical applications.

Future Directions:

•Focus on enhancing coherence in AD narration.

•Utilize external knowledge like plot summaries.

•Explore alignment between narration tone and movie content.

Amit Kumar Mohapatra,B421005