Reviewer #2

Questions

- 1. Paper summary
    - DeepLocalization introduces an approach for real-time localization of distracted driving actions, leveraging Graph-Based Change-Point Detection and a Video Large Language Model (Video-LLM). This dual approach aims to pinpoint actions in time and categorize activities accurately, optimizing for consumer-grade GPUs. Tested on the SynDD2 dataset, DeepLocalization demonstrated promising performance, achieving 57.5% accuracy in event classification and 51% in event detection. The framework's lightweight design and efficacy in identifying diverse driver behaviors mark a significant stride toward enhancing road safety by monitoring driver distraction.
- 2. Paper strength
    - DeepLocalization introduces a methodology by synergizing Graph-Based Change-Point Detection with a Video Large Language Model (Video-LLM) for real-time localization and precise classification of distracted driving behaviors. This innovative blend not only addresses the critical challenge of distracted driving—a significant factor in road accidents—but also showcases the framework's optimization for consumer-grade GPUs, highlighting its feasibility for broad implementation. Tested on the complex SynDD2 dataset, DeepLocalization's commendable performance, achieving significant accuracy in both event detection and classification, vividly demonstrates its robust potential in enhancing road safety by effectively identifying and analyzing diverse driver behaviors. This synergy of novel approach and practical application positions DeepLocalization as a significant advancement in leveraging deep learning for real-time driver monitoring and distraction assessment.
- 3. Paper weakness
    - DeepLocalization's approach to distracted driving detection exhibits certain limitations, particularly in its handling of specific classes where the change point algorithm's performance is inconsistent. Activities characterized by minimal movement or susceptibility to noise in key-point data challenge the algorithm's efficacy, indicating a potential area for refinement. Moreover, the framework's heavy reliance on precise key-point extraction underscores a vulnerability in environments with noisy or incomplete data, necessitating advancements in key-point identification techniques for robust application. Additionally, while the framework achieves notable successes in event detection and classification, there remains a significant opportunity to improve the accuracy of event classification. This is especially true for distinguishing

between activities that are closely related or differ subtly, pointing to a need for further sophistication in the classification algorithm. Addressing these areas could substantially enhance DeepLocalization's utility in accurately monitoring and assessing distracted driving behaviors across various scenarios.
Shaiqur:- Thank you for the suggestions, and we certainly believe that the accuracy of change point algorithm could be improved for cases where it fails. We could research more by refining the algorithm to locate such scenarios. Also, other approaches may be tried to find the key points which would address noisy or inconsistent data. And finally, the classification model could be fine tuned to classify similar classes.

- **4. Additional comments to authors**
  - DeepLocalization's approach to distracted driving action localization and classification is commendable. However, addressing the challenges related to the change point algorithm's sensitivity to noise and the dependency on precise key-point extraction could further bolster the framework's robustness and applicability. Exploring advanced techniques for key-point extraction and refining the model's capability to differentiate closely related activities may enhance performance and generalization across diverse driving scenarios.
- **6. Paper decision**
  - Borderline

Reviewer #4

Questions

- **1. Paper summary**
  - This paper proposes to classify the act of driver, leveraging graph-based change-point detection with video large language model. They tested on SynDD2 dataset from 7th AI City Challenge(2023) Track 3 with 57.5% accuracy.
- **2. Paper strength**
  - 1. Utilizing the change detection algorithm with graph-based representation, the method can classify the driving scenario in real time and low computation.
    2. Based on temporal activity localization with GNN, it can extract the duration of driver's action.
- **3. Paper weakness**
  - 1. Just usage of pretrained VideoChatGPT
    Shaiqur:- The idea is to use a lightweight framework, and our approach uses one 24GB RAM, which is much less compared to recent approaches.
  - 2. Tested on last year challenge dataset, and insufficient experimental results (accuaracy comparison with other methods, fps) for proving efficiency.
    Shaiqur:- I compared the action proposal with last year's top-performing team.

- **4. Additional comments to authors**
  - None
- **6. Paper decision**
  - Borderline

Reviewer #5

Questions

- **1. Paper summary**
  - This paper presents DeepLocalization, a framework for temporal action localization specifically designed for monitoring driver behavior using change point detection. It is evaluated on the SynDD2 dataset, which focuses on distracted driving behaviors.
- **2. Paper strength**
  - The language is clear throughout the paper, and the introduction and related works sections are well-articulated.
- **3. Paper weakness**
  - 1. The rationale behind the choice of the change point algorithm is not sufficiently justified. Merely stating that "using a change point algorithm has not been explored" lacks a solid foundation.
    Shaiqur:- The idea behind our approach is to introduce the framework for activity recognition. Our approach is very lightweight compared to other approaches used last year; for example, the top-performing team last year used eight V100 GPU cards with 32GB of memory, while we used only one GPU with 24 GB RAM.

    2. The approach lacks novelty as it heavily relies on existing algorithms such as gseg2 and YOLO, with the main technical contribution being the refinement of the Video-LLM through prompt engineering.
    Shaiqur:- Same argument as above. The novelty lies in the overall framework.

    3. The paper lacks detailed descriptions of the models and methods, hindering reproducibility for readers….Shaiqur
    Shaiqur- explained the methods in more detail.
    Farabi- VideoChatGPT is designed to detect distracted behaviors in video clips and followed the exact same way it is in the official repository and the paper . So following the main paper would be sufficient enough
    4. There is a notable absence of comparison between the results obtained in this study and other state-of-the-art results, which limits the assessment of the proposed method's effectiveness
    Shaiqur: I compared the action proposal accuracy with last year's top team.

5. The presentation of figures and tables needs improvement, for instance, Figure 2 would be better as a table rather than a screenshot of table.
Shaiqur:- I improved the tables and worked on the figures.

- **6. Paper decision**
  - Strong Reject

Reviewer #6

Questions

- **1. Paper summary**
  - The paper proposes a framework designed for the real-time localization of actions to monitor driver behavior. It leverages existing methods to suggest a framework.
- **2. Paper strength**
  - 1. The paper effectively combines existing algorithms to enhance performance.
    2. It includes an ablation study that identifies which combinations of algorithms are most effective.
    3. The paper is well written and easy to understand.
- **3. Paper weakness**
  - 1.The methods are described briefly…Shaiqur
    Shaiqur- explained the methods in more detail.
    2.There is a limited number of figures…..Shaiqur
    Shaiqur- Added three figures related to camera positions, respective camera view images, and key point generation.
    3.There is a lack of novelty…..Shaiqur
    Shaiqur- we introduced a novel approach that used key points to find the action proposals and classify those proposals using VideoChatGPT. Our approach is lightweight, using only one GPU with 24 GB RAM.
- **4. Additional comments to authors**
  - The paper uses popular methods in computer vision to produce a novel method. I consider it interesting.
- **6. Paper decision**
  - Borderline