Qeios

Peer Review

Review of: "DAVE: Diverse Atomic Visual Elements Dataset with High Representation of Vulnerable Road Users in Complex and Unpredictable Environments"

Sumit Mishra[1]

1. Robotics, Korea Advanced Institute of Science and Technology, Korea, Republic of

The **DAVE dataset** presents a valuable contribution to video recognition for autonomous driving, particularly in unstructured traffic environments. However, there are some limitations that may impact its generalizability, fairness, and effectiveness across different scenarios. Below are some key shortcomings that should be considered:

Limited Geographical Scope and Driving Orientation Bias

- The dataset is collected exclusively in India, which may limit its generalizability to other unstructured traffic environments in Asia, Africa, and Latin America. Even within Asia, there are significant differences in driving orientation—countries like South Korea follow right-hand driving, whereas India follows left-hand driving. Training models only on Indian data may introduce biases that hinder adaptation to other regions with different traffic rules, road layouts, and driver behaviors.

Potential Bias in Data Collection Due to Limited Vehicle Perspectives

- The dataset relies on dashcams installed in only two specific vehicle models (MG Hector and Maruti Ciaz), which could introduce biases in viewpoint, height, and motion patterns. This is particularly problematic in developing countries like India, where roads are shared by diverse vehicle types such as motorcycles, auto-rickshaws, bicycles, and buses. Capturing data from a broader set of vehicles would ensure that the dataset better represents real-world traffic conditions.

 doi.org/10.32388/7CNUVR

Annotation Process and Potential Errors

- While the dataset boasts rich annotations, it relies on manual annotation using CVAT. However, the paper does not discuss inter-annotator agreement or validation processes in detail. Datasets are often annotated by multiple human annotators (e.g., five individuals) with the final label determined by majority voting or mode selection to improve consistency. Without such measures, annotation inconsistencies—especially in complex traffic scenarios—could affect model training and benchmarking reliability.

Comparative Evaluation with Existing Datasets Lacks Failure Analysis

- The paper frequently states that existing models perform worse on DAVE compared to other datasets. However, it does not provide an in-depth analysis of **why** the performance drops—whether due to dataset complexity, model limitations, or annotation differences. A detailed ablation study, including failure case analysis, would provide more insight into what aspects of DAVE make it particularly challenging and how models could be improved.

Need for a Larger Dataset for Diverse Applications

- While DAVE introduces valuable complexity, certain applications—such as accident scenario detection, risk assessment, and anomaly detection—require a significantly larger dataset to train robust models. Real-world accident scenarios are rare events, requiring extensive video footage to capture diverse cases and edge conditions. Expanding the dataset with more videos across different environments would enhance its usability for broader autonomous driving and safety applications.

Declarations

Potential competing interests: No potential competing interests to declare.