Reference	*CP67*
Project Title	EDR-V feasibility (EDRVf)
Coordinator	*Chalmers Tekniska Högskola*
Project Manager	*Jonas Bärgman*
Project Duration	1 year (with further delays due to parental leave)
Author(s)/contributors	Jonas Bärgman (Chalmers) Jordanka Kovaceva (Chalmers) Alberto Morando (Autoliv Development AB)
Confidentiality	Internal to the SAFER-partners

Contents

Summary

The EDRV Feasibility (EDRVf) project was launched to evaluate the potential of using data from Event Data Recorders with video (EDRV; commonly known as DashCams) for traffic safety research. These devices, installed in private and commercial vehicles, capture kinematic signals, GPS information, and video of either the roadway, the driver, or both during crash and nearcrash situations. As such, they may offer a rich, real-world source of information on crash dynamics, driver behaviour, and occupant posture.-V Feasibility (-V; commonly known as -crash situations. As such, they -world source of information on crash dynamics, driver

The project is motivated by the substantial opportunities EDR-V data could provide for both pre-crash and in-crash research. In pre-crash analyses, these data could support studies of crash causation, scenario generation, and behavioural modelling, which are increasingly important for virtual safety assessments. In the in-crash phase, EDR-V footage has the potential to supply unique observations of occupant motion immediately before impact—information valuable for qualitative assessments and for advancing human body modelling (HBM).

Despite this promise, multiple uncertainties remain regarding the practical and legal feasibility of using EDR-V data systematically in research. The EDRVf project was therefore conceived as a pre-study to clarify technical, regulatory, and methodological constraints, and to determine whether a larger-scale initiative should follow.

The project findings show that while EDRV data offers promising possibilities for both precrash and incrash research, several fundamental constraints limit its practical use. Whether GDPR becomes a barrier depends heavily on the exact project scope and requires substantial internal dialogue with data protection officers. Ethical review, on the other hand, posed no obstacle, as the application was approved without comments. Personal integrity considerations remain a central challenge; commercial providers are (correctly) highly restrictive, and access to driver facing video is essentially unattainable. From a technical standpoint, although a wide ecosystem of computer vision and analysis tools exists, no single solution can extract all relevant parameters; instead, multiple object detection, tracking, pose estimation, and depth estimation methods must be combined. Even then, high precision outputs still rely on careful calibration, high quality data, and in some cases manual annotation. -V data offers promising possibilities for both pre-crash and in-crash research, several fundamental constraints limit its practical use. Whether GDPR becomes a barrier depends heavily on the exact project scope and requires substantial internal dialogue with data-protection officers. Ethical review, on the other hand, posed no obstacle, as the application was approved without comments. Personal-integrity considerations remain a central challenge; commercial providers are (correctly) highly restrictive, and access to driver-facing video is essentially unattainable. From a technical standpoint, although a wide ecosystem of computer-vision and analysis tools exists, no single solution can extract all relevant parameters; instead, multiple object-detection, tracking, pose-estimation, and depth-estimation methods must be combined. Even then, high-precision outputs still rely on careful calibration, high-quality data, and in some cases manual annotation.

Overall, the clearest conclusion is that gaining access to EDR-V datasets—especially any in-cabin video—is extremely difficult and may be nearly impossible under current European privacy constraints.

EDR-V feasibility

1. Background

The EDR-V feasibility (EDRVf) project was initiated to assess the potential of leveraging data collected from Event Data Recorders with video (EDR-V), commonly referred to as DashCams, for traffic safety research. EDR-Vs are devices installed in consumer or commercial vehicles, designed to record information during crashes or near-crash events. These devices typically capture vehicle kinematics, GPS data, and video footage of either the forward external scene and the driver, or only the external scene, offering a dataset for understanding crash dynamics and human behavior in road traffic.

Figure 1. Example of devices and recorded videos

The primary motivation behind this project stems from the substantial opportunities that EDR-V data presents for both pre-crash and in-crash research. In the pre-crash phase, such data can be important in crash causation studies, scenario generation, and behavioral modeling, which are essential for virtual safety assessments. For in-crash research, EDR-Vs can provide unique insights into the posture and movement of vehicle occupants immediately before a crash, supporting qualitative and quantitative analyses that are particularly valuable for human body modeling (HBM) work. Despite these promising prospects, there remain many uncertainties regarding feasibility, constraints, and limitations of using EDR-V data for research purposes. This project was therefore conceived as a pre-study to systematically address these uncertainties and determine whether a larger-scale research initiative should be pursued.

Historically, Chalmers, AB Volvo, and VCC engaged in EDR-V research prior to 2015. However, progress was halted due to regulatory challenges, notably the introduction of GDPR, and the withdrawal of key US industry partners from research collaborations (the US company Lytx that we were working closely with – they were bought by in investment company and all research activities were discontinued). With evolving data protection frameworks, a better understanding of GDPR constraints, and renewed interest in data-driven safety research, the EDRVf project aimed to revisit the feasibility of EDR-V data, considering both research/technical and regulatory perspectives.

2. Project execution

The project's approach was structured around four key activities:

1. Preparation of the ethical application to the Swedish Ethical Review Authority (Etikprövningmyndigheten, EPM). This included the formulation of a hypothetical study encompassing both pre-crash and in-crash research components, outlining the full spectrum of desired research capabilities that the project aims to support. Subsequently, the application was submitted to EPM to gauge the legal/ethics perspectives on the use of EDR-V data;
2. Discussions with organizational data protection officers to clarify the implications of GDPR and other data privacy requirements;
3. Outreach to companies that have access to and sell access to EDR-V data, with the aim of understanding data availability, costs, access procedures, contractual requirements, and the tools necessary for data extraction and analysis. The initial intent was to acquire a data sample from such companies for initial exploration (e.g., data quality assessment). Unfortunately, we did not succeed in this task (see section 2.5).
4. A comprehensive investigation into the potential for manual, semi-automatic and automatic temporal annotation of video data, focusing on the extraction of both occupant and external road user kinematics/behavior.

By clarifying the feasibility and opportunities of EDR-V data, this project has contributed to SAFER's broader mission of advancing traffic safety research. If EDR-V is identified as a viable data source (including all relevant tools), it would enable a wide range of research activities, including evaluation of traffic behavior, validation of automated systems, development of predictive models for human cognition and behavior, monitoring of occupant and road user dynamics, and improved understanding of injury mechanisms. Ultimately, the project aimed to lay the groundwork for future initiatives that can "harness" EDR-V data to inform the design, evaluation, and implementation of safety solutions in road transport.

2.3 The first step – The application to EPM

We prepared and submitted an ethics application that included almost all the different types of data and analysis types that we thought may be problematic for analysis of the pre-crash phase, to inform both pre-crash and in-crash research. The application, named "Händelseaktiverad datainsamling med video för ökad förståelse och modellering av krockar" is attached (also available at EPM as applications fall under "offentlighetsprincipen") and was approved ("Etikprövningsmyndigheten godkänner den forskning som anges i ansökan"). That is, EPM judged that the work would be ethically acceptable, but that it still falls under the Swedish Law of Ethics (i.e., any future study using EDR-V data in the way described in the application will have to apply for ethical approval).

The review process at EPM took approximately 2 months.

2.4 The second step – Discussions with data protection officers

We will not go into the specific feedback provided by each data protection officer at each organization as the discussions were specific for the individual organizations. Instead, we summarize a few overall conclusions. None of these statements are the official position of any organization – they are the interpretation based on this discussion and available documentation (e.g., Integritetsskyddsmyndigheten | IMY):

- Purchasing and using EDR-V data is possible. But as expected, data protection is critically important, both to safeguard the individuals appearing in the videos and to prevent companies from facing fines for mishandling this sensitive data. Each party, whether universities and companies, needs to go through internal processes first for data acquisition, storage, and processing. Lawyers would also need to be involved in drafting contracts between parties. These are additional steps that are lengthy and expensive and need to be taken into consideration before starting a research collaboration.
- when we have video data we have no way of (legally) contacting the individuals, which is problematic as both for legal basis "public interest" and (obviously) "consent" there is a requirement to contact individuals. This makes it impossible to conduct research under the legal basis of consent. For the legal basis of public interest it may be sufficient to try to inform them by providing some written public information on a webpage; although it is unlikely that they will see it, we have done what we can to reach out. If this is a reasonable approach or not will depend on the study and will be a decision for each individual data protection officer.
- "data protection impact analysis" (konsekvensanalys; Konsekvensbedömning enligt GDPR | IMY) will always have to be prepared for such research. This is a tool for the data protection officer where they need to make sure that everything (e.g., all risks) is/are considered. A GDPR data protection impact assessment should provide a systematic description of the planned processing and its purpose, including which personal data will be processed, how they are collected, who has access, and which technologies, systems, data sources, and transfers are involved. It should also evaluate necessity and proportionality by assessing whether the processing is required to achieve its purpose, whether less privacy intrusive alternatives exist, and whether the amount of data used is proportionate. Finally, it must include an assessment of the risks to individuals' rights and freedoms along with the technical and organizational measures planned to mitigate these risks and demonstrate GDPR compliance
- Even data collections performed internally (e.g., to test individual data collection systems for ground truth calibration/modelling) would have to be considered independently as a separate GDPR issue, as it may include people both inside and outside of the vehicle (including people that are captured randomly on the external cameras as they go about their daily lives.

2.5 The third step – Outreach to EDR-V data providers.

We took advantage of our extended professional network to scout for as many providers as we could. Since the beginning, the most promising companies were *Nexar* (https://www.getnexar.com/) and *SmartDrive* (https://www.smartdrive.net/). Both are based in the US, and both provide devices that record internal and external views. Moreover, these two companies are based in the US, where data protection law is, at the moment, more lenient than in Europe, resulting in EDR-V services having been deployed in the US for decades. We wanted to evaluate the GDPR, ethics and practical accessibility of data we know is voluminous.

We had three meetings with Nexar. While they were keen to collaborate, they would not distribute the driver-facing video (they would not provide raw data and did not have processes to anonymize or in other ways postprocess the video), making the use of their EDR-V data for in-crash research unfeasible. Nexar could, however, provide data for pre-crash research, based on external video and ego-vehicle-kinematics sensors. They could provide post-processed information at different levels of details (at different cost tier). They would then process the forward facing video and provide information (e.g., reconstruction of the kinematics of the own vehicle and other road-users), as well as lane markings and other infrastructure aspects. We attempted to get a data sample for initial exploration, but we stopped after some of the NDA requirements was prohibiting – we may pursue this further in the future, but within the scope of this project we decided to close the data extraction attempts with this. We did not manage to contact SmartDrive, despite multiple attempts and despite having collaborated in the past. The reason may be their recent acquisition by Solera Fleet solutions.

We have not identified any European companies that could provide the in-cabin videos in addition to the forward-facing one at scale. GDPR and other privacy/integrity laws in Europe makes this business case for EDRV data collection for commercialization highly problematic. Although there are sources of EDRV data in Asia (e.g., Japan, South Korea, and China), in the project we decided to primarily pursue US or European data. Over the duration of the project we have had some discussions with individuals at various Asian universities working with EDRV data, but although they seemed willing to collaborate, it was far from obvious if or how such data may be made available to us. This will have to be investigated further in future work. In China, the China In-Depth Accident Study (CIDAS) have internal and external video data for a small subset of their data, as companion to in-depth crash investigation data, but at too low frame rate and resolution.

In summary, although there may be opportunities to access EDRV data from sources in Asia, the US company we contacted (Nexar) would not share data on vehicle occupants, and the European market for commercially collected EDRV data is severely constrained by GDPR. Future work will need to explore the potential use of Asian data, while also monitoring developments in both Europe and the United States.

2.6 The fourth step – A review of data extraction tools

The objective of this review was to identify approaches and tools capable of extracting a set of parameters from EDR-V to support pre-crash and in-crash analysis. The parameters of interest include **lane positions**, number of road users or other **objects, their positions, velocities, accelerations, headings, and relative speeds**. In addition, the review considers higher-level occupant-related outputs, such as **head orientation** and **gaze direction**, which can provide insight into driver attention and behavior. In addition to extracting the kinematics of surrounding road users, the internal camera of the EDR-V can be leveraged to derive time-series measures of driver behavior, **driver glance** behavior, indicating where the driver is looking over time.

Further, EDR-V data can be used to model how people sit while driving and how they respond to impending traffic conflicts. Such data enable the derivation of realistic initial conditions for **occupant posture** and **behavior** in pre-crash scenarios, for example bracing responses. Crash test configuration also requires specification of anthropometric characteristics and vehicle interior geometry. While some of this information may be inferred from video data, such as approximate **occupant size** or **distance to the steering wheel**, the achievable accuracy may be insufficient for certain high-fidelity safety analyses.

To address these requirements, several tools and platforms are reviewed and compared in terms of their ability to support accurate parameter extraction in an **automatic** or **semi-automatic** fashion, as well as their limitations. We list available annotation tools and computer vision techniques that enable the extraction of information from video data in a highly automated manner, with minimal reliance on human supervision, by leveraging recent advances in computer vision. This Section presents an inventory of the reviewed tools, including both open-source and commercial software, along with their capabilities and limitations.

Label studio (https://labelstud.io/) supports data annotation for machine learning, with a focus on video and image labeling to enable supervised model development (Tkachenko et al. 2025). It provides functionality to annotate objects using bounding boxes, segmentation masks, and associated attributes across individual frames or video sequences, allowing the creation of high-quality labeled datasets for training custom models such as YOLO-based detectors or semantic and instance segmentation networks. Within a typical workflow, this tool is used to manually label roadway infrastructure, for example lanes, traffic signs, or barriers, as well as dynamic road users such as vehicles, cyclists, and pedestrians, and then export the annotations in standard formats suitable for downstream model training such as for training object detection or segmentation models. Its limitation is that it does not include pre-trained perception models or automated computation of kinematic parameters, requiring additional tools or models to perform detection, tracking, or motion analysis.

Ultralytics (https://www.ultralytics.com/) provides state-of-the-art YOLO models for object detection, widely used in computer vision pipelines for traffic and road-user analysis. Its capabilities include real-time object detection and tracking, with pretrained YOLOv8 (and its subsequent versions) models that can reliably detect vehicles, pedestrians, cyclists, and other common roadway objects, and straightforward integration into Python-based workflows for downstream processing. In a typical workflow, Ultralytics YOLO models are used to perform object detection and extract bounding boxes from video frames, which can then be combined with multi-object tracking algorithms such as DeepSORT to estimate object trajectories and derive proxy measures of velocity and acceleration over time. However, Ultralytics does not provide depth estimation or direct kinematic analysis out of the box, and therefore must be integrated with additional tools, such as OpenCV for motion analysis or monocular depth models like MiDaS, to compute distances and more complete motion parameters.

YOLOE (https://docs.ultralytics.com/models/yoloe/; "Real-Time Seeing Anything") is an object detection approach, in the Ultralytics workflow, designed to support flexible and efficient recognition of user-defined object categories. It can be useful in scenarios where specific objects must be annotated from scratch, as it can accelerate the annotation process by providing rapid, model-assisted predictions that reduce manual labeling effort for new datasets. Within a data preparation workflow, YOLOE can be used to bootstrap annotations for custom object classes, allowing human annotators to focus on correction and refinement rather than drawing labels from the ground up. While its primary role is object detection rather than kinematics analysis, it can serve as a front-end tool to speed up annotated dataset creation for downstream perception and motion-estimation pipelines.

Kognic (https://www.kognic.com/) focuses on annotation and validation workflows tailored to autonomous driving systems, with an emphasis on producing high-quality ground truth data at scale. Its capabilities include advanced video annotation tools for object labeling and lane marking detection, structured validation pipelines for training and evaluating perception models, and robust infrastructure for managing large multimodal datasets, including video and sensor data such as LiDAR. Within a typical workflow, Kognic is used to generate and curate ground truth annotations for vehicles, road users, and roadway infrastructure, as well as to validate object detection and perception models whose outputs may later be used for kinematics estimation. However, Kognic is primarily an annotation and dataset management platform and does not perform kinematic analysis or depth estimation itself, requiring integration with additional perception and motion-analysis tools to derive such parameters.

Roboflow (https://roboflow.com/) is a computer vision platform oriented toward dataset management and model training, streamlining the development of object detection systems. Its capabilities include tools for annotating, preprocessing, and augmenting image datasets, integrated pipelines for training object detection models using frameworks such as **YOLO** or **TensorFlow**, and deployment options for running trained models in inference environments. In a typical workflow, Roboflow is used to manage and curate datasets for training models that detect vehicles, VRUs, and roadway infrastructure, and to export the resulting trained models for integration into a broader kinematics or analytics pipeline. Its main limitation is that it does not support depth estimation or motion tracking, so additional tools are required to compute distances, trajectories, or kinematic parameters.

RF-DETR (https://github.com/roboflow/rf-detr) is a transformer-based detector provided by Roboflow. Unlike traditional convolution-based detectors, RF-DETR (Robinson et al. 2025) uses a transformer architecture to model global context within an image, which can improve detection performance in complex scenes with multiple interacting objects. Within a computer vision workflow, RF-DETR can be used to detect vehicles, road users, and infrastructure elements in images or video frames, and its integration within the Roboflow ecosystem allows deployment alongside dataset management and annotation tools. As with other object detectors, RF-DETR focuses on object localization and classification and must be combined with tracking, depth estimation, and motion analysis modules to derive kinematic parameters. Note, this has been used for object detection in the MicroVision (https://www.drivesweden.net/sites/default/files/2026-01/DriveSweden - MicroVision project report.pdf) project (at Chalmers).

V7 Labs (https://www.v7labs.com/video-annotation) provides an advanced annotation platform for video and image datasets, designed to support efficient creation of high-quality training data for computer vision models. Its capabilities include annotating video sequences with bounding boxes, segmentation masks, and keypoints, as well as AI-assisted labeling features that help automate annotation tasks and improve consistency at scale. The platform also supports integration with custom models to iteratively refine annotations. Within a typical workflow, V7 Labs is used to annotate road users, lanes, and other infrastructure elements in video data for training object detection or segmentation models, with AI-assisted tools accelerating the labeling of vehicles and pedestrians. Its primary limitation is that it focuses on data annotation and curation rather than providing automated depth estimation or kinematic calculations, which must be handled by complementary tools.

VideoMAEv2-Base (https://huggingface.co/OpenGVLab/VideoMAEv2-Base) is a self-supervised video representation model designed to *extract meaningful spatiotemporal features* from raw video, not a ready-made tool for object detection or metric measurements like distances or kinematics. It is a variant of Video Masked Autoencoders (VideoMAE) pretrained on millions of unlabeled videos to learn general spatiotemporal embeddings (features) from video clips. It produces feature vectors (representations) for video segments that encapsulate appearance and motion patterns, not direct labels or measurements. The model is intended primarily for feature extraction and downstream tasks such as action recognition or temporal action detection after fine-tuning. VideoMAEv2-Base does not directly provide object detection (cars, pedestrians, belts, steering wheel), pose estimation (body joints or angles), depth or distance measurements, velocity, acceleration, heading outputs. However, the model produces latent features that describe the visual content of a clip, which is useful as inputs to other models that can provide those outputs.

VideoMAEv2 (Wang et al. 2023) can be integrated into a larger workflow as a feature extractor for downstream models that do actual perception or measurement tasks. The workflow can include **feature extraction, pretraining, parameter extraction.** First, VideoMAEv2 can embed segments of a dashcam video into feature vectors. Those feature vectors can be used as input to a supervised model trained to recognize actions or states relevant to kinematics (e.g., vehicle braking, driver leans forward). This approach can improve classifier performance compared to raw pixel input because the features capture temporal structure. Second, is pretraining for custom perception tasks, VideoMAEv2 can be fine-tuned on EDR-V specific datasets to make its features more relevant for automotive scenes (e.g., crash vs non-crash, occupant posture categories). The fine-tuning requires annotated data. Third, to extract concrete in-crash/in-vehicle parameters from EDR-V (like driver posture, seat-belt placement, distances), we would need specialized models or modules such as: a) **object detection** for vehicle occupants, belts, steering wheel (e.g., YOLO, Detectron2); b) **pose estimation** to locate human joints (e.g., OpenPose, ViTPose, available on Hugging Face); c) **depth estimation** to infer distances from camera view (e.g., MiDaS or depth-aware networks); d) **tracking and motion** modules (e.g., DeepSORT, optical flow methods). VideoMAEv2's features **can be used as an intermediate representation** to feed into these modules (for classification or temporal context), but the model itself does not output the physical parameters directly.

DeepScenario (https://www.deepscenario.com/automotive) is an automotive-focused platform that provides autolabeling capabilities for annotating dashcam video data at scale. It leverages computer vision algorithms to automatically detect and label objects present in driving scenes, including vehicles, road users, and infrastructure elements, and to reconstruct these objects in three-dimensional space. By generating 3D object representations and scene context directly from dashcam footage it supports the creation of annotated datasets suitable for training and validating perception models. This platform reduces the manual effort required for video annotation and enables downstream analysis that benefits from spatially consistent, scene-level information, although integration with additional tools is still to be required for detailed kinematic and behavior modeling.

BORIS (Behavioral Observation Research Interactive Software) **and TEMA** (Tracking and Motion Analysis Software) are both established tools used for motion analysis, and they *can* be part of a workflow for extracting in-vehicle kinematics, but they serve very different purposes and have limitations relative to the full set of parameters (driver posture, belt placement, steering wheel distance, etc.). Below is a breakdown of each tool, what they do well, and how they would fit into an in-crash/in-vehicle analysis pipeline.

BORIS (https://www.boris.unito.it/) is a manual and semi-automated behavioral annotation tool originally developed for ethological research. It can be used to annotate videos with timestamps, label behaviors, and extract event durations. Capabilities relevant to in-vehicle analysis are a) Frame-by-frame coding of events or postures e.g., driver leaning forward, hands off steering wheel, belt slack; b) Timed annotations (onset, offset), which are useful for labeling when a driver's posture changes; c) Exportable CSV data for downstream quantitative analysis. BORIS does not perform automatic detection, an annotator must label everything manually. BORIS does not have built-in distance measurement, it is purely annotation tool and it doesn't generate object coordinates or depth. It does not do kinematic computation for e.g., velocity, acceleration, joint angles, etc. BORIS is useful as a ground truth annotation tool when creating labeled data for training detection models, manually validating posture events in video segments, annotating sequence onset/offset of behaviors. It does not automatically extract posture or distances.

TEMA (https://temaplatform.com/tema-classic/) is used for frame-by-frame annotation for object tracking. TEMA (by Image Systems) is a professional motion tracking tool used in biomechanics and vehicle crash testing. It tracks markers or features in high-speed video to compute: 2D or 3D coordinates over time, velocities and accelerations, angles between segments, and distance measurements. It has been used for precise motion capture from video when there is a controlled video (e.g., multiple calibrated cameras or pre-marked points), there is need of kinematics (displacements, angles, distance to wheel), and need to validate biomechanical simulations. From the perspective of in-crash analysis, this class of tools offers capabilities for precise motion measurement when marker-based tracking is feasible. By placing physical markers on body segments, seats, seat belts, or the steering wheel, high-precision tracking of occupant motion can be achieved, particularly when combined with calibrated multi-camera setups that enable true 3D spatial measurements. Automatic tracking can be performed across long video sequences with user supervision, allowing consistent tracking of specific anatomical points such as the elbow, shoulder, or pelvis. Based on these tracked points, the tools can compute detailed kinematic parameters, including velocities, accelerations, and segment or joint angles. However, these tools do not provide built-in semantic understanding of occupant states, they track points rather than concepts such as driver posture. Their accuracy is highly dependent on careful marker placement and visibility, and they do not support seat belt detection or belt geometry estimation without explicit markers or additional custom workflows.

SAM 3D Pose (https://ai.meta.com/sam3d/), developed by Meta, extends the Segment Anything framework to support 3D human body pose estimation from videos and images. The tool enables conversion of segmented human bodies into 3D pose representations, providing estimates of body orientation and joint configuration that can be useful for analyzing occupant posture and movement. In addition to pose-related capabilities, the Meta ecosystem includes functionality for tracking and annotating moving objects across frames, as well as built-in support for privacy-preserving operations such as blurring faces and license plates. The SAM 3D Body approach is therefore relevant for workflows that require body pose estimation, object tracking, and compliant handling of sensitive visual information, although it remains primarily a perception and representation tool rather than a full kinematics or biomechanics analysis solution.

Depth estimation models such as **Depth Anything V2** (https://github.com/DepthAnything/Depth-Anything-V2)provide an approach for inferring scene geometry and relative or metric distances from monocular images and video. Depth Anything V2 builds on large-scale training to deliver robust depth predictions across diverse environments, and its metric depth variant extends this capability by estimating depth values in real-world units when appropriate calibration or assumptions are available. Within a video analysis workflow, these models can be used to approximate distances between the camera and detected objects, such as vehicles, road users, or in-vehicle elements, and to support downstream calculations of relative position and motion. While monocular depth estimation cannot match the accuracy of stereo or LiDAR-based systems, Depth Anything V2 offers a scalable, camera-only solution that is well suited for integration with object detection, tracking, and kinematics estimation pipelines.

OpenPose (https://github.com/CMU-Perceptual-Computing-Lab/openpose) is an open-source pose estimation framework developed by the CMU Perceptual Computing Lab that enables the detection of human body keypoints directly from images or video without the need for physical markers. It estimates 2D skeletal joint locations by analyzing visual cues, making it suitable for non-intrusive analysis of human posture and movement in naturalistic settings, such as in-vehicle recordings. Within a broader workflow, OpenPose can be used to infer occupant pose and motion over time, supporting analyses of posture changes, bracing behavior, or head and upper-body movement. However, as a monocular, markerless approach, its outputs are limited to image-space coordinates unless combined with depth estimation or multi-camera setups to derive accurate 3D positions and kinematic measures.

OpenFace (https://github.com/TadasBaltrusaitis/OpenFace) is an open-source toolkit for facial behavior analysis that provides robust face detection, landmark localization, head pose estimation, and eye gaze tracking from video or images. Developed to enable research in human affective and social behavior, it uses a combination of classical computer vision and machine learning techniques to output expressive facial features such as 2D/3D facial landmarks, head orientation, eye gaze vectors, and facial action units. In an in-vehicle analysis context, OpenFace can be applied to driver-facing video to estimate where the driver is looking, how the head is oriented, and subtle facial movements that may relate to attention, distraction, or fatigue. While it does not directly compute full body kinematics, it provides facial and head pose signals that can complement pose estimation and object detection modules when assessing driver behavior or glance dynamics.

3. Results and deliverables

This report constitutes the main deliverable of the project. The results are presented in Section 2 and address the questions outlined in the project proposal, with the overarching aim of improving our understanding of the feasibility of using EDR data with video (DashCams) for pre-crash and in-crash research.

4. Main takeaways and next Steps

The main takeaways from this work are:

- Depending on the scope of the research, GDPR may apply, but to be able to understand if the research plan is compatible with GDPR is a problem the partners involved will have to have substantial internal interactions with the data protection officer, and the project scope needs to be well defined.
- Etikprövningsmyndigheten does not have a concern with this type of study. We got the application accepted without any comments. For the application, see the appendix.
- The personal integrity constraint is (and should be) a main issue for research related to the driver. The company we talked to was clear that we would not be able to get driver-facing video for analysis, and that they mostly did not have that either (our interpretation).
- There may be opportunities to get access to Asian EDR-V data, but will have to be investigated in future projects.
- A wide range of tools exists to extract parameters from EDR-V data, but no single solution can automatically deliver all required pre-crash and in-crash variables. Object detection frameworks, annotation platforms, pose-estimation models, depth-estimation tools, and motion-tracking software each provide partial capabilities, but they must be combined to obtain lane position, road-user kinematics, occupant posture, gaze behavior, and other parameters of interest.
- While modern computer-vision methods can significantly automate extraction of scene and occupant information, their accuracy and applicability depend on workflow integration, data quality, and task-specific fine-tuning. High-fidelity analyses—such as detailed occupant posture or metric distance estimation—often still require calibrated setups, additional models, or manual annotation, highlighting both the potential and the current limitations of EDR-V-based parameter extraction.

EDR-V data remains an unique source of information for traffic safety research. Gaining access to this data is indeed challenging—among others due to the limited number of providers and strict internal and European data policies—but obtaining in-cabin videos from commercial data suppliers (even when not directly affected by European laws) is nearly impossible because of their stringent privacy concerns. These challenges currently limit the use of EDR-V for, especially, granular in-crash research based on EDR-V.

5. Acknowledgement

We want to thank SAFER, and especially its partners, for contributing to the Idea Exploration Program funding.

6. References

- Tkachenko M, Malyuk M, Holmanyuk A, et al (2025) Label Studio: Data labeling software. URL https://github.com/HumanSignal/label-studio
- Label Studio https://github.com/HumanSignal/label-studio
- RF-DETR https://github.com/roboflow/rf-detr
- Robinson I, Robicheaux P, Popov M, et al (2025) RF-DETR: Neural Architecture Search for Real-Time Detection Transformers. URL https://arxiv.org/abs/2511. 09554, arXiv:2511.09554
- Wang L, Huang B, Zhao Z et al. (2023) VideoMAE V2: Scaling Video Masked Autoencoders with Dual Masking. URL https://arxiv.org/pdf/2303.16727
- Image Systems Motion Analysis (TEMA) https://ssba.org.se/image-systems-motion-analysis/
- TEMA https://temaplatform.com/tema-classic/
- DeepScenario https://www.deepscenario.com/automotive

- OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields https://arxiv.org/abs/1812.08008
- OpenPose https://github.com/CMU-Perceptual-Computing-Lab/openpose
- SAM3D https://ai.meta.com/sam3d/
- SAM3D Body https://aidemos.meta.com/segment-anything/editor/convert-body-to-3d
- Depth Anything V2 https://github.com/DepthAnything/Depth-Anything-V2
- Depth Anything V2 for Metric Depth Estimation https://github.com/DepthAnything/Depth-Anything-V2/tree/main/metric_depth
- OpenPose https://github.com/CMU-Perceptual-Computing-Lab/openpose
- OpenFace https://github.com/TadasBaltrusaitis/OpenFace
- DriveSweden - MicroVision project report.pdf