Utilizing Driving Context to Increase the Annotation Efficiency of Imbalanced Gaze Image Data

Johannes Rehm, Odd Erik Gundersen, Irina Reshodko, Kerstin Bach

NorwAI Norwegian Research Center for AI Innovation

NTNU

Abstract

Knowing where the driver of a car is looking, whether in a mirror or through the windshield, is important for advanced driver assistance systems and driving education applications. This problem can be addressed as a supervised learning task. However, in a typical dataset of driver video recordings, some classes will dominate over others. We implemented a driving video annotation tool (DVAT) that uses automatically recognized driving situations to focus the human annotator's effort on snippets with a high likelihood of otherwise rarely occurring classes. By using DVAT, we reduced the number of frames that need human input by 94% while keeping the dataset more balanced and using human time efficiently.

Introduction

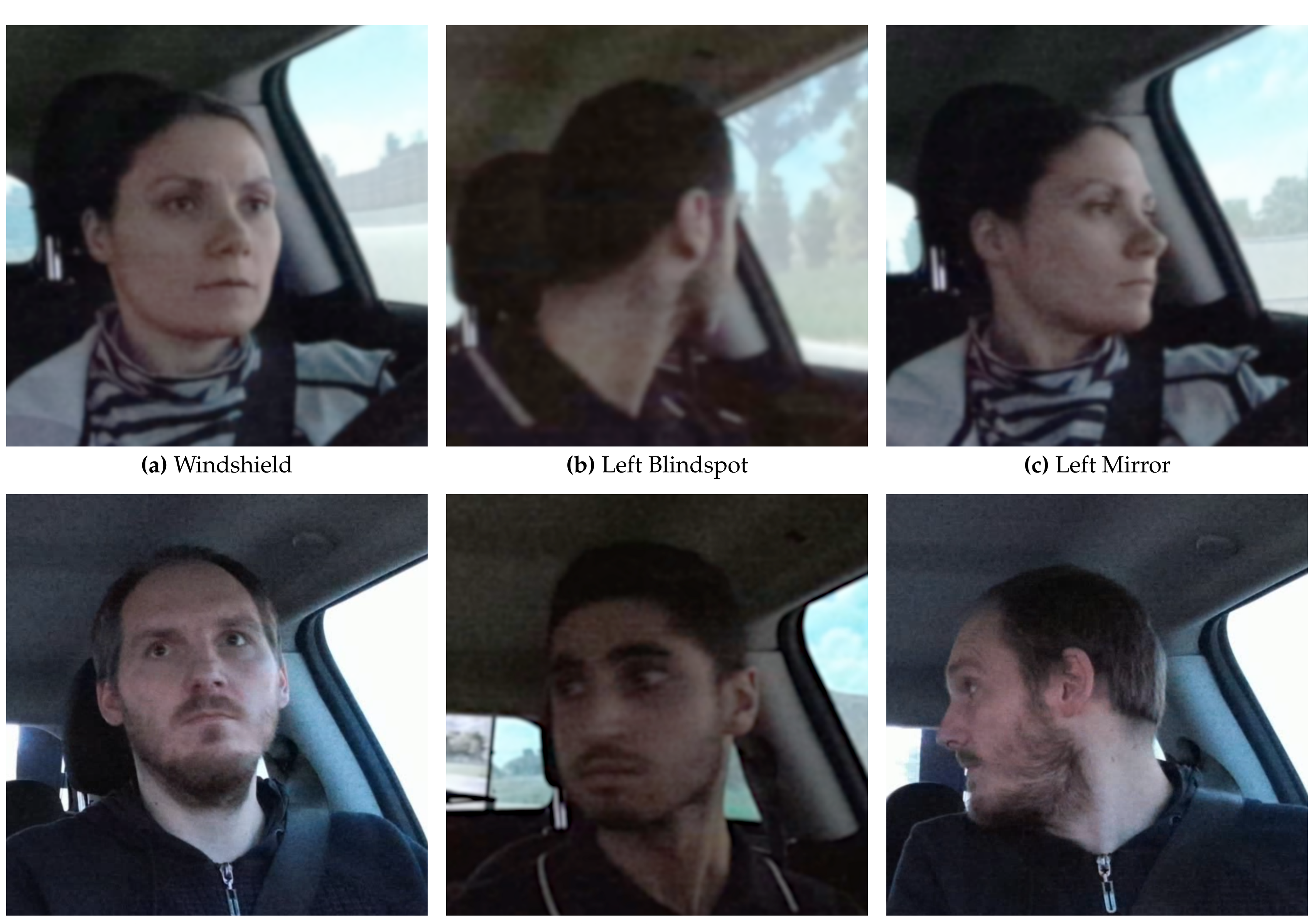
Research Objectives

- Efficient annotation of mirror and blindspot observations of drivers in a car simulator.
- Collection of a balanced dataset for training image classification algorithms.
- Minimization of the time spent on searching for sparsely distributed classes.

Proposed Solution

- Analyse videos automatically using the *Virtual Driving Instructor* (VDI) [2].
- Use context information extracted by the VDI to find situations with increased likelihood of. otherwise rare driver observation events.
- Situation-based annotation in *Driving Video Annotation Tool* (DVAT).

Example Images



(a) Windshield (b) Left Blindspot (c) Left Mirror

(d) Rearview Mirror (e) Right Mirror (f) Right Blindspot

Figure 1: Gaze Zones

Similar to [1], we call the target areas of the driver's attention *gaze zones*.

Development Process

We follow a data-centric model development process shown in Figure 2.

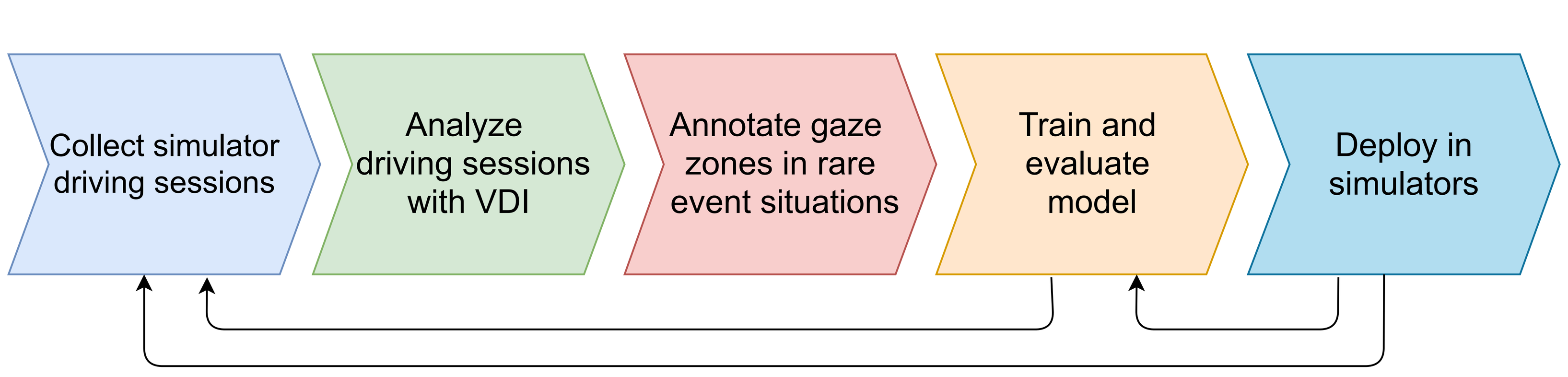


Figure 2: Annotation workflow integrated in model development process.

Driving Video Annotation Tool (DVAT)

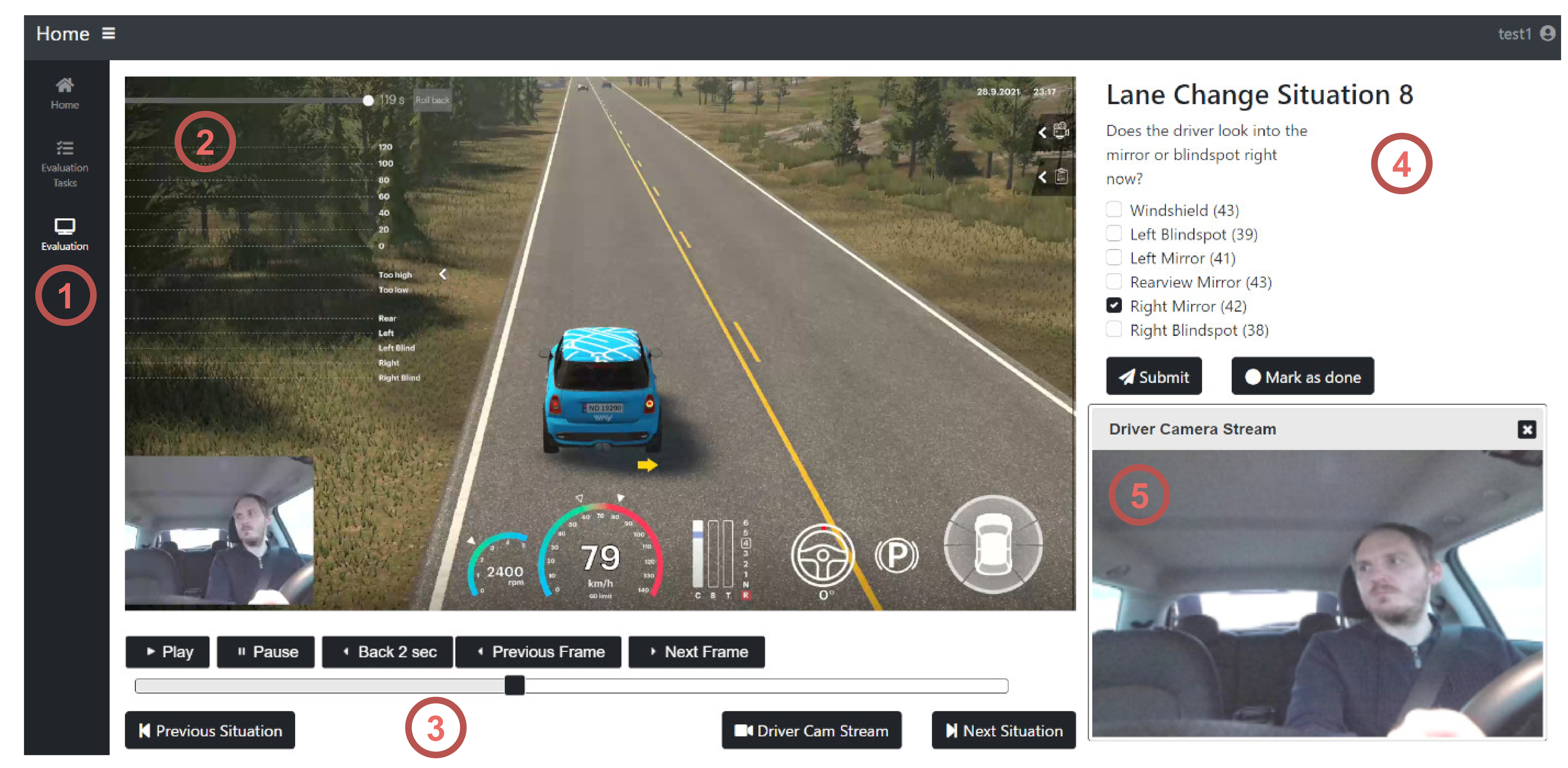


Figure 3: Annotation Framework Web Frontend.

- Web frontend showing both the video of the driver's face and the driving scene.
- Flexible setup allowing the use of DVAT for any kind of driving or traffic situation annotation.

Results

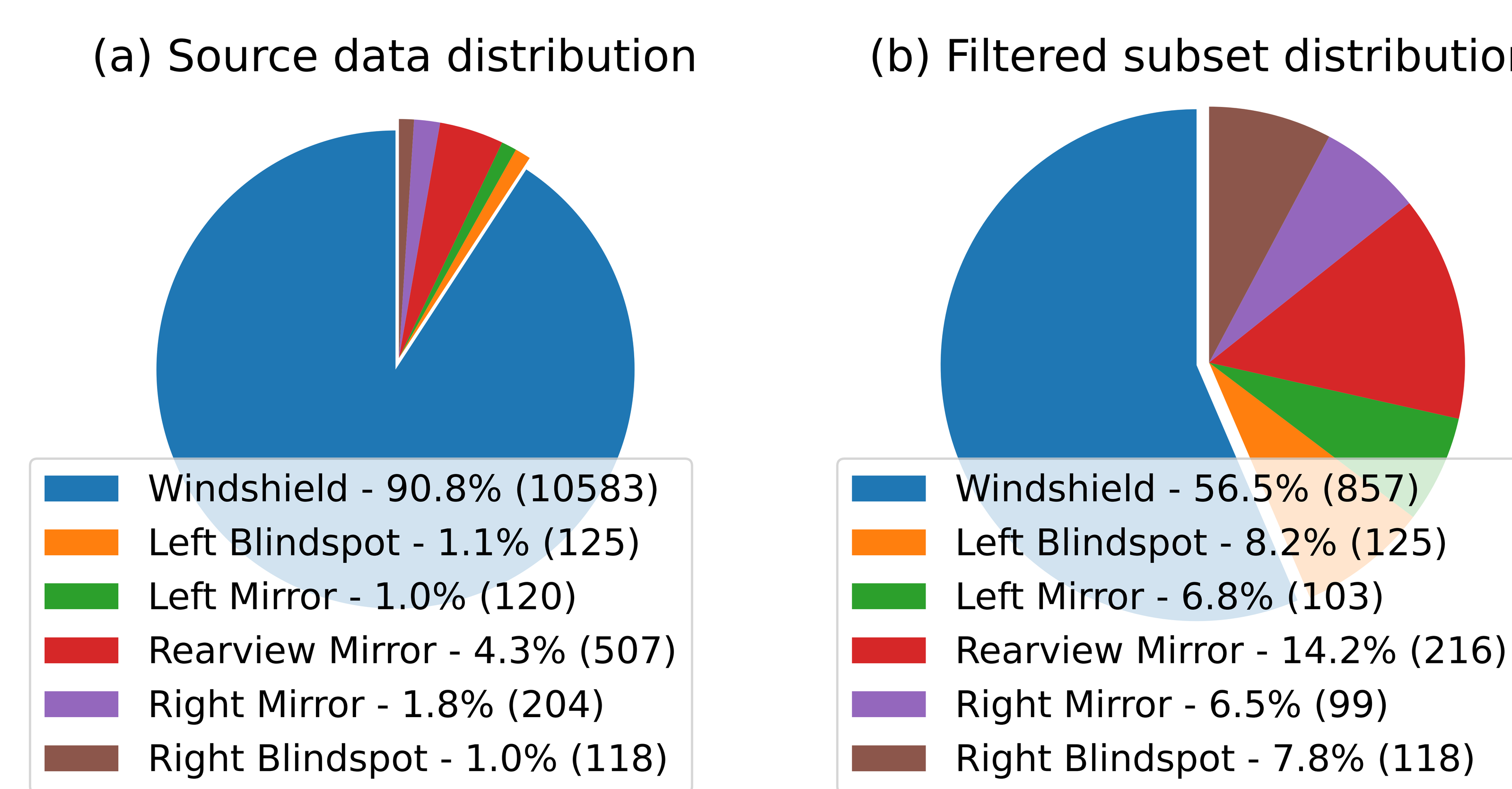


Figure 4: Source data class distribution vs. class distribution after filtering by driving context-based situations

- We annotated the gaze zones of a simulator driving session for every frame.
- This gives us an approximation of how imbalanced such a dataset would be and how more balanced the dataset is if we apply our context filtering approach.
- The used context information are lane changes in an overtake exercise.

Conclusion

We presented an efficient method to label gaze image data. By utilizing context information about driving situations, we automatically get a subset of the complete data which includes a higher proportion of otherwise rarely occurring classes. This considerably reduces the effort for human annotators to label a fairly balanced dataset from very unbalanced source data.

References

[1] R. F. Ribeiro and P. D. Costa. Driver gaze zone dataset with depth data. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–5. IEEE, 2019.

[2] M. K. Sandberg, J. Rehm, M. Mnoucek, I. Reshodko, and O. E. Gundersen. Explaining traffic situations–architecture of a virtual driving instructor. In *International Conference on Intelligent Tutoring Systems*, pages 115–124. Springer, 2020.