

Effect of Data or Knowledge Gaps on AI Model Performance

Dr. Andreas Albrecht, Istvan Bognar (Robert Bosch GmbH)
Dinish K Karunakaran, Derya Sahin, Himanshu Walia (AVL Deutschland GmbH)



Knowledge Gap Demonstrator Setup

AI models such as DNNs learn their I/O behavior from data in a numerical optimization. Missing or underrepresented data in the training set leads to systematic (epistemic) knowledge gaps and AI model insufficiencies. This demonstrator shows the effect of such gaps. It uses AVL's dynamic ground truth (DGT) system equipped with 6 cameras, 3 LiDARs and dGPS (Fig. 1). A baseline model, two test models with designed data gaps, and a monitor run in real-time on an NVIDIA AGX Orin unit. The data gap effects on the AI models are visualized using the Lichtblick suite [1].



Figure 1: Knowledge gap demonstrator setup (left: DGT system mounted on top of a test vehicle, right: close-up of the DGT system) (© AVL Deutschland GmbH)

Training Experiments with Knowledge Gaps

Using ignore regions and data feature filters, we design datasets with gaps in well-defined data space regions to analyze AI model sensitivity towards the selected data features (Fig. 2).

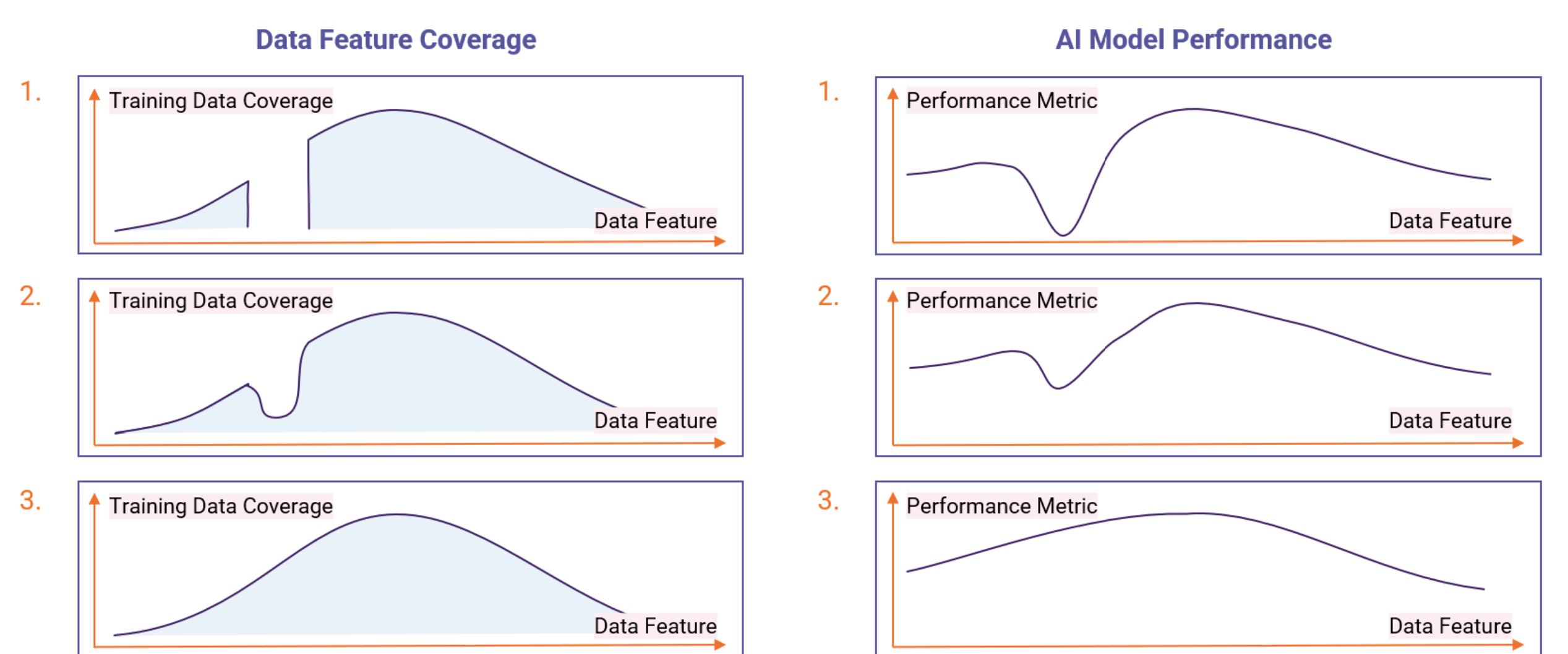


Figure 2: Dataset design to create data gaps using filters to analyze effects on AI performance (© Robert Bosch GmbH)

Identifying Data-related AI Model Insufficiencies

To identify data-related AI model insufficiencies, we have to stimulate the model in the gap region and analyze AI model performance and coverage versus relevant data space dimensions on both test and training datasets (Fig. 3 and Fig. 4).

Online Monitoring

We monitor consistency of the predicted model outputs during runtime to identify False Positive (FP) and False Negative (FN) detections (Fig. 5).



Gaps in training data cause model insufficiencies. Due to gaps in test data, AI insufficiencies remain undiscovered.

Figure 5: Pedestrian detection using AI models [4] with knowledge gaps (left: baseline trained on nuScenes [2] default train set, center: with close-by gap, right: with left-side gap, green boxes: model detections) on ZOD data [3] (© Robert Bosch GmbH) and monitoring results of the AI models (red boxes: identified FN detections) (© AVL Deutschland GmbH)

*) Disclaimer on [3]: For this dataset, Zenseact AB has taken all reasonable measures to remove all personally identifiable information, including faces and license plates. To the extent that you like to request removal of specific images from the dataset, please contact privacy@zenseact.com.

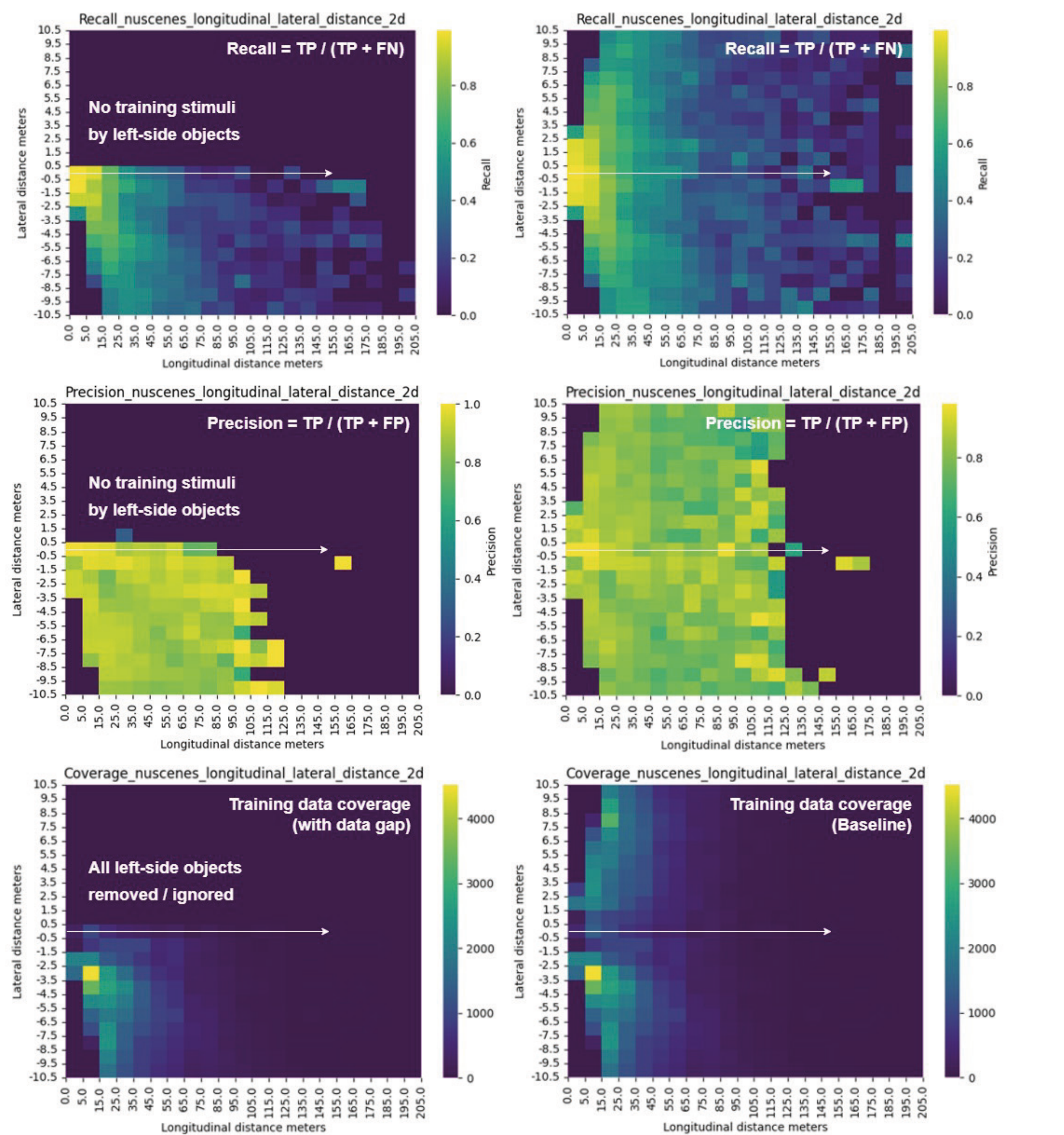


Figure 3: Recall and precision of a model with left-side gap (left) compared to a baseline model with no additional gap (right) using NanoDet [4] plus coverage of the original and the modified nuScenes [2] training dataset vs. long. / lat. obj. position from the ego vehicle (© Robert Bosch GmbH)

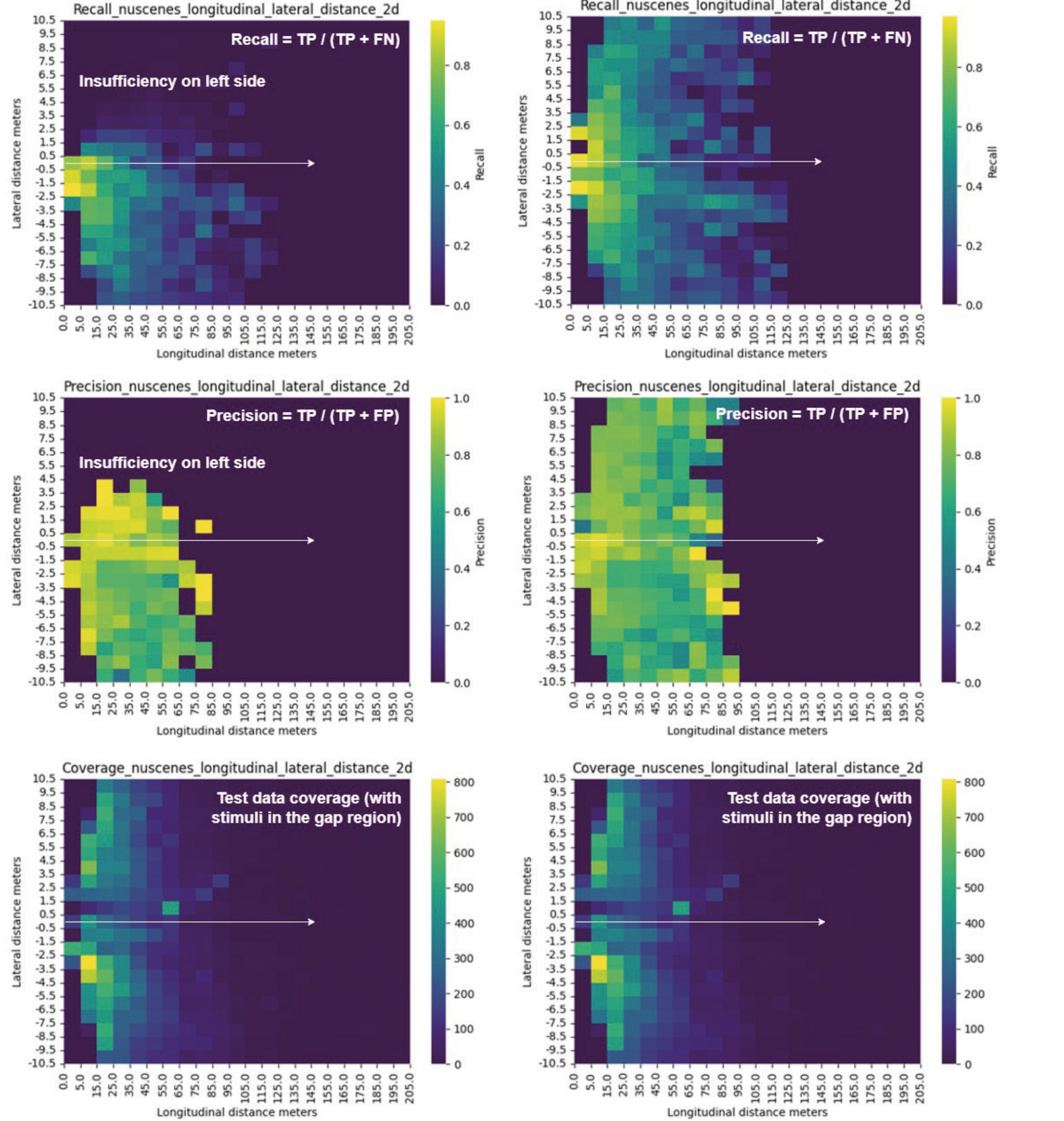


Figure 4: Recall and precision of a model with left-side gap (left) compared to a baseline model with no additional gap (right) using NanoDet [4] plus data coverage of the original nuScenes [2] test dataset vs. long. / lat. object position from the ego vehicle (© Robert Bosch GmbH)

References:

- [1] Lichtblick Suite, <https://github.com/Lichtblick-Suite/>.
- [2] nuScenes Dataset, <https://www.nuscenes.org/>.
- [3] Zenseact Open Dataset (ZOD), <https://zod.zenseact.com/>. *)
- [4] NanoDet-Plus, <https://github.com/RangilYu/nanodet/>.
- [5] Albrecht A. et al.: A Data-Driven Engineering (DDE) Process for Verification and Validation of AI Systems, In: M. Yilmaz et al. (Eds.): EuroSPI 2026, CCIS 3136, pp. 1-15, 2027. https://doi.org/10.1007/978-3-032-36813-3_7.

For more information regarding the poster, contact:

andreas.albrecht@de.bosch.com,
max.nestoriuc@avl.com



Safe AI Engineering is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Energy.



[linkedin.com/company/ki-familie](https://www.linkedin.com/company/ki-familie)
www.safe-ai-engineering.de