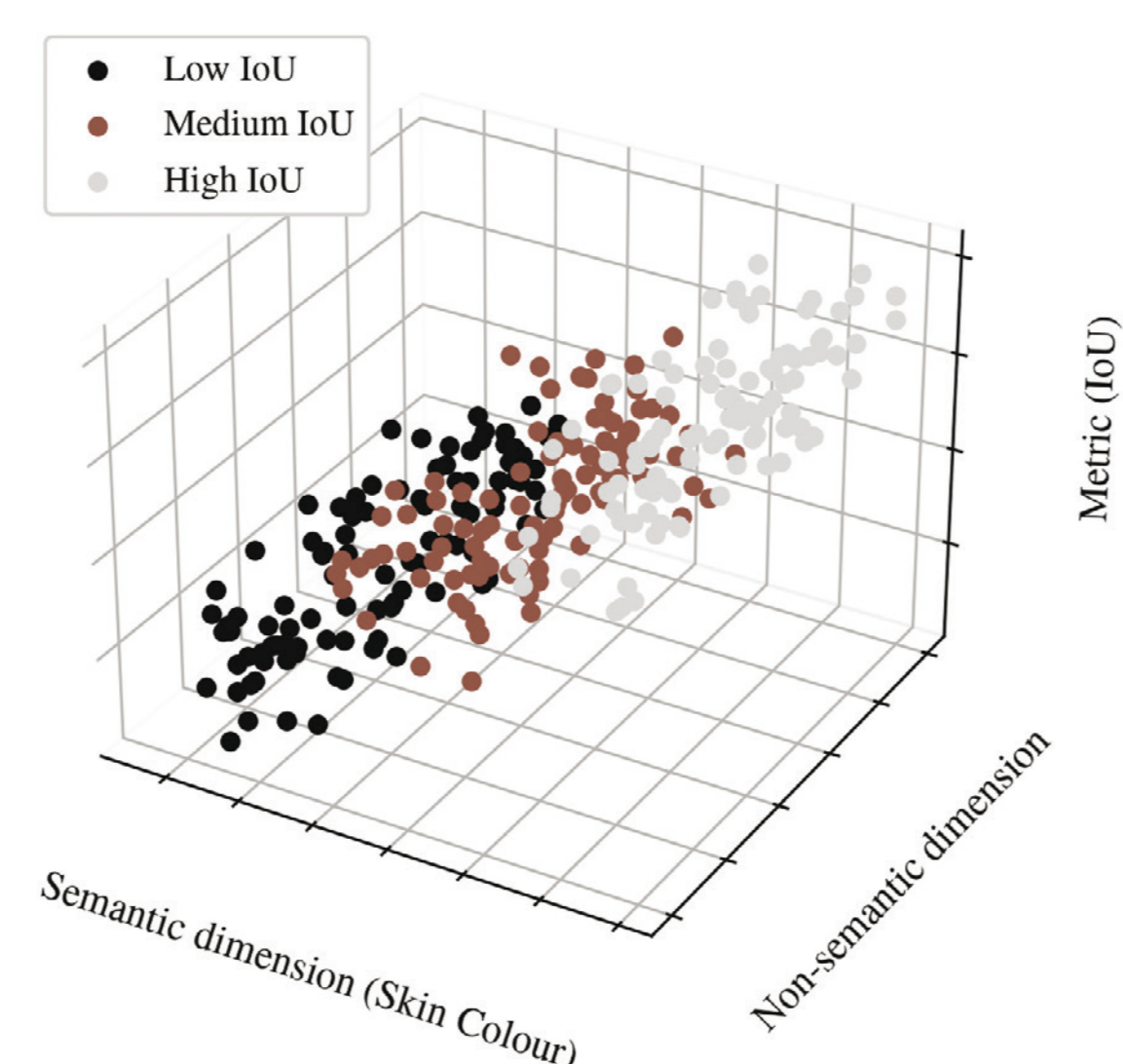


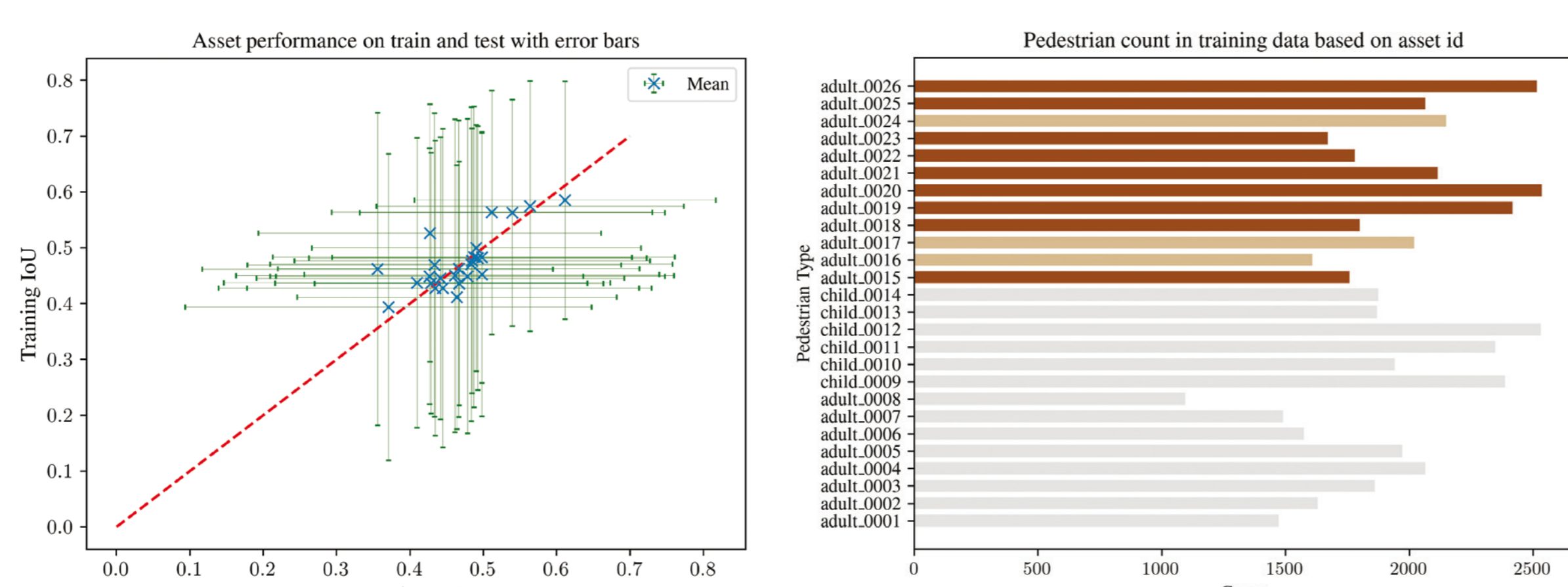
Motivation

Real-World data could have semantic dependencies which are not explicitly tested when using aggregated metrics. Combining existing metrics with semantic subsets of data, we detect such weaknesses, e.g., DNNs being consistently bad at detecting dark-skinned pedestrians.

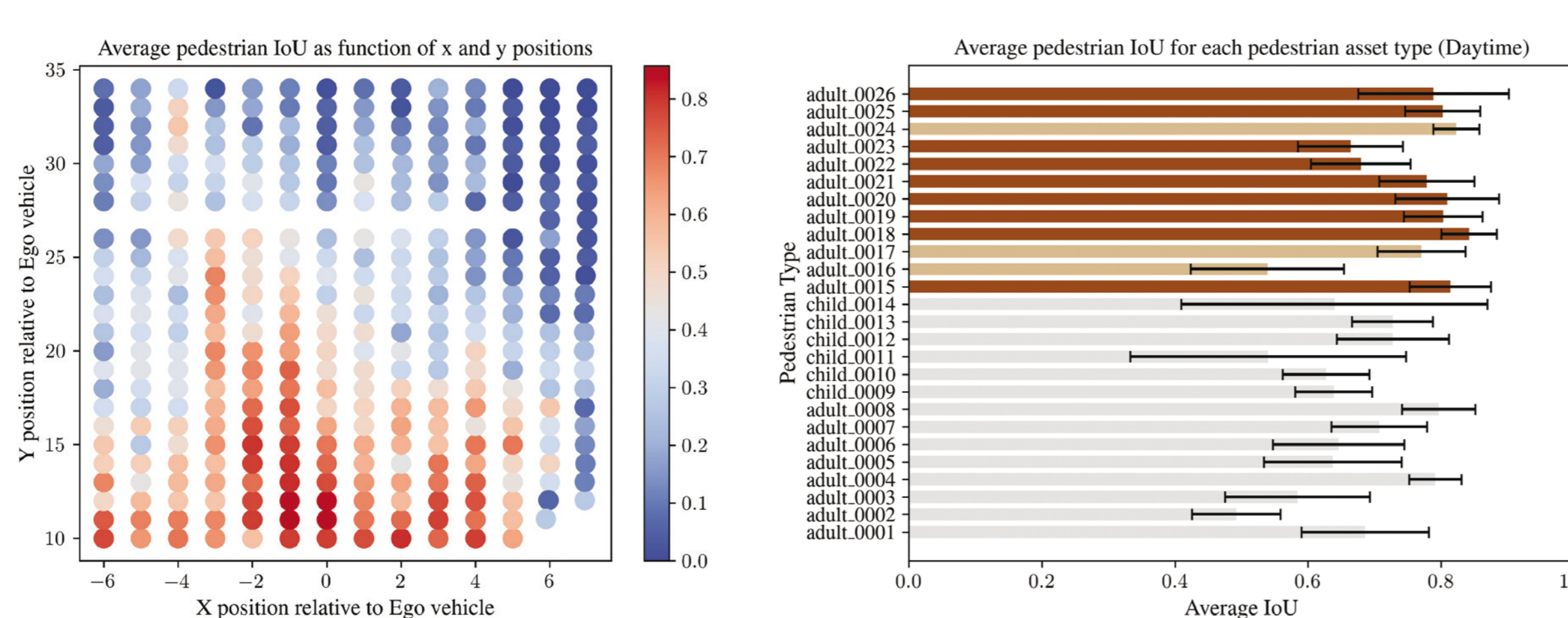


Semantic testing of Carla data

To obtain initial metadata, we developed a Carla extension. Training DeepLabv3+ on the resulting data showed performance variations w.r.t. semantic dimensions.

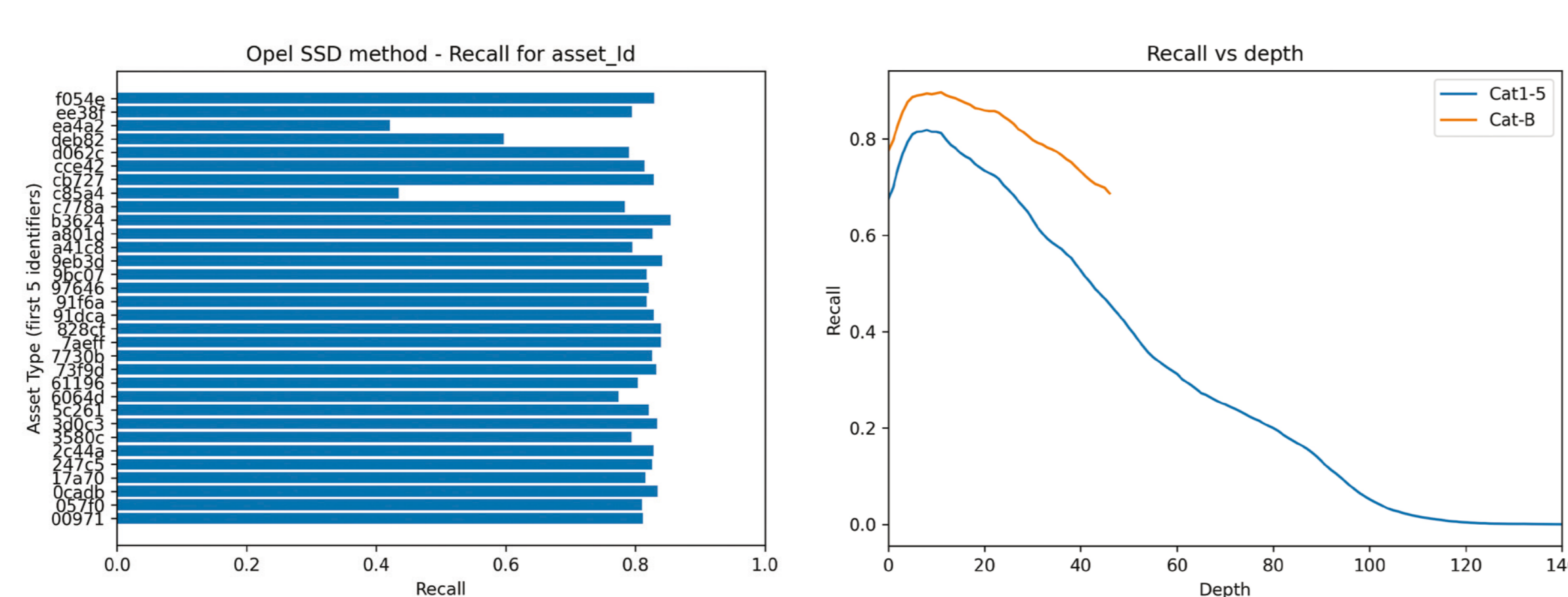


Significant asset level variations that could not be explained by asset distribution in training data were detected. Further analysis with special scenario test datasets shows other weaknesses of the DNN.



Semantic testing of KIA data

Investigations into SSD performance trained on KIA data show that asset-level performance is more uniform than DeepLabv3+, with deviations mostly on out-of-domain assets. In addition, weaknesses w.r.t. depth of pedestrian were also uncovered.



Safety Hypothesis:

DNNs might suffer from weaknesses along semantic dimensions leading to correlated risks. Our test uncovers such weaknesses providing more significant evidence for safety argumentations.

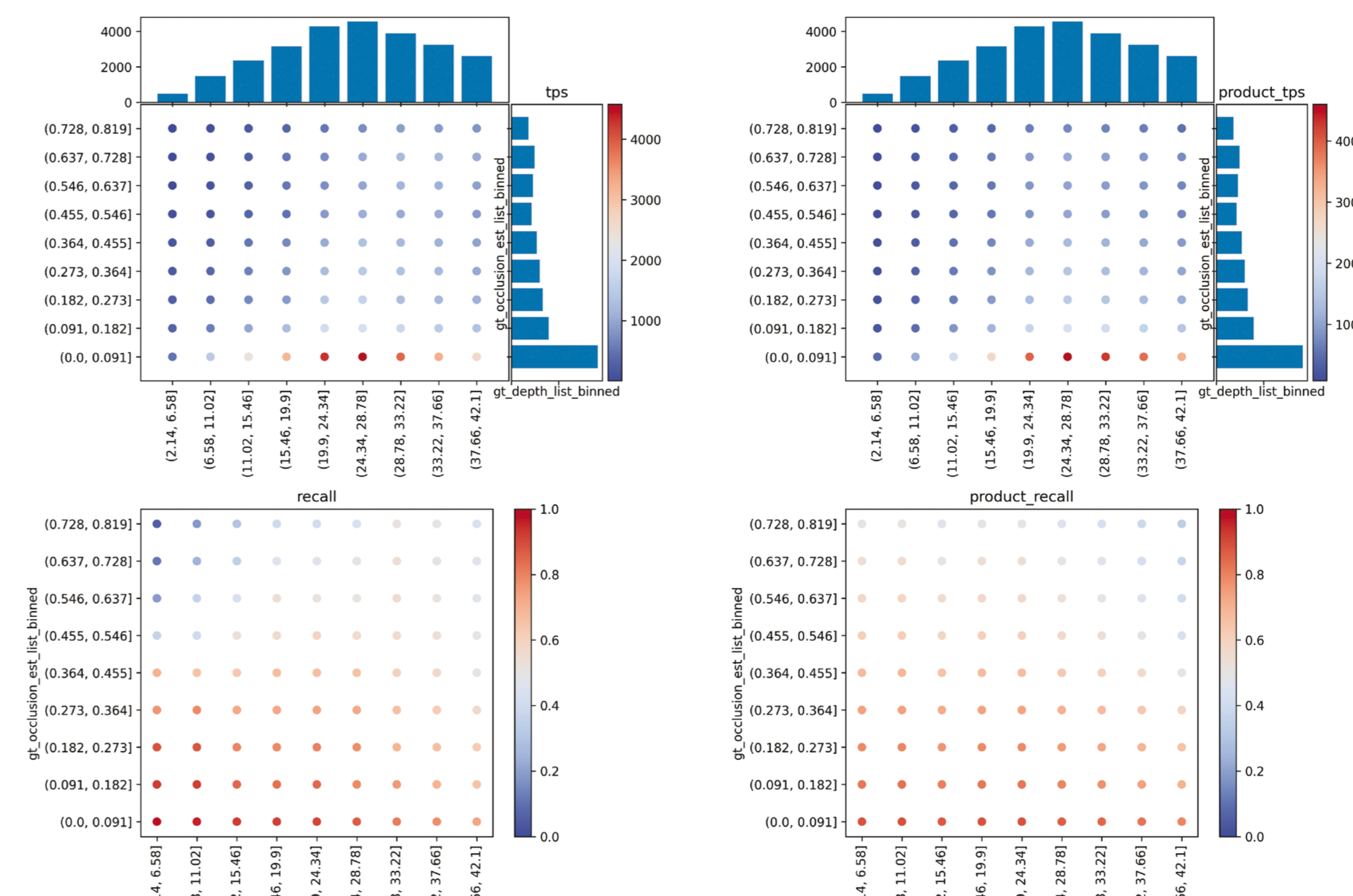
Investigations in the uncertainty work stream showed that specific mechanisms improve base models only on specific semantic dimensions. However, an overall improvement over base models was not observed.

Proxy models

Statistical modelling of DNN performance w.r.t. metadata could provide further insights into the presence of weaknesses. Using the marginals, we build such a simple proxy model.

$$TruePositives(depth, occl.) \approx Imp_{depth}(depth) * Imp_{occl.}(occl.)$$

Notice that while structural similarity is maintained, there are deviations at high occlusions between real and proxy model.



Outlook

Closed-loop testing with predefined or on the fly generated scenarios could potentially be a requirement for the systematic evaluation of DNNs. Our extension to Carla and work in semantic testing and proxy models provide the necessary tools for performing closed-loop testing.

References:

Gannamaneni et al. „Semantic Concept Testing in Autonomous Driving by Extraction of Object-Level Annotations from CARLA.“ *Proc. of the IEEE/CVF Int. Conf. on CV. 2021.*