A Concept for Highly Automated Pre-Labeling via Cross-Domain Label Transfer for Perception in Autonomous Driving

Maarten Bieshaar[0000−0002−6471−6062], Marek Herde[0000−0003−4908−122X], Denis Huseljic[0000−0001−6207−1494], and Bernhard Sick[0000−0001−9467−656X]

University of Kassel, Wilhelmshöher Allee, 34121 Kassel, Germany
{mbieshaar,marek.herde,dhuseljic,bsick}@uni-kassel.de

Abstract. This article proposes a novel concept to leverage the time-consuming labeling process for training object detectors in automated driving. The approach uses pre-trained probabilistic, well-calibrated object detectors for different sensor modalities. Based on the knowledge about the sensor extrinsics, the probabilistic detections are transformed from one sensor modality into another. These transformed detections are then used as pre-labels for the respective sensor modality. However, these pre-labels are error-prone, such that we propose an additional dedicated labeling quality assessment. The latter allows us to attach a quality seal to automatically pre-labeled data sets and is the starting point for interactive human-in-the-loop learning.

Keywords: Highly Automated Pre-Labeling · Object Detection · Human-In-The-Loop Learning · Autonomous Driving · Imperfect Labels

1 Introduction

Artificial intelligence and, in particular, *machine learning* (ML) are the enabling technologies in autonomous driving. In this context, ML and deep learning techniques are already successfully used for perception, i.e., sensory environment and object recognition [4]. Training and validating these mostly deep neural networks, e.g., *convolutional neural networks* (CNN), requires vast amounts of labeled data. However, labeling, especially for object detection, is a time-consuming and, therefore, costly task [17]. In this article, we present an approach to significantly reduce the labeling effort in the particular application domain of ML-based object detection for highly automated driving. Our approach considers that many modern vehicles are equipped with various sensors, including cameras, LiDAR, and RADAR. We exploit this sensor diversity (i.e., the strengths and weaknesses of the respective sensors [12]) in our approach by transferring labels between different sensor modalities. First, we train object detectors for the single sensors. We further use these predictions as so-called pre-labels (i.e., imperfect, potentially error-prone labels). These can, in turn, be used to improve the object detectors of the other sensor modality. Our approach can be understood as semi-supervised cross-domain learning [2], whereas the object detectors are interpreted as multiple error-prone annotators [9]. However, in safety-critical applications such as highly automated driving, the created labelings must be quality-checked to ensure that no incorrect concepts are learned.

Contributions: We address this problem by proposing a detailed concept to automatically generate pre-labelings via cross-domain label transfer for perception in autonomous driving. Therefore, we identify four major research questions arising within our concept's stages and provide ideas for targeting each of them. We envision our concept as an application-driven starting point for human-in-the-loop learning. In this context, our concept provides methods leveraging interactive learning techniques in object detection, e.g., probabilistic object detectors, improving the labeling quality, coping with imperfect labels, and decreasing annotation effort. As another major contribution, we see the quality assessment of pre-labelings to support subsequent human-in-the-loop learning processes. In this sense, we aim to provide a quality seal to pre-labeled data sets, e.g., pre-labels having an expected mean average precision of 90%.

2 Highly Automated Pre-Labeling

This section describes the four stages of our concept (cf. Fig. 1) and formulates research questions. In the first stage, we develop probabilistic object detectors. The second stage aims to improve the calibration of these detectors further. Subsequently, the probabilistic predictions are interpreted as pre-labels of the different sensor modalities and optionally fused in the third stage. Finally, the concept is concluded by the fourth stage, including a labeling quality assessment based on probabilistic outputs. It serves as a starting point for human-in-the-loop learning to refine the pre-labels and release them for further model training.

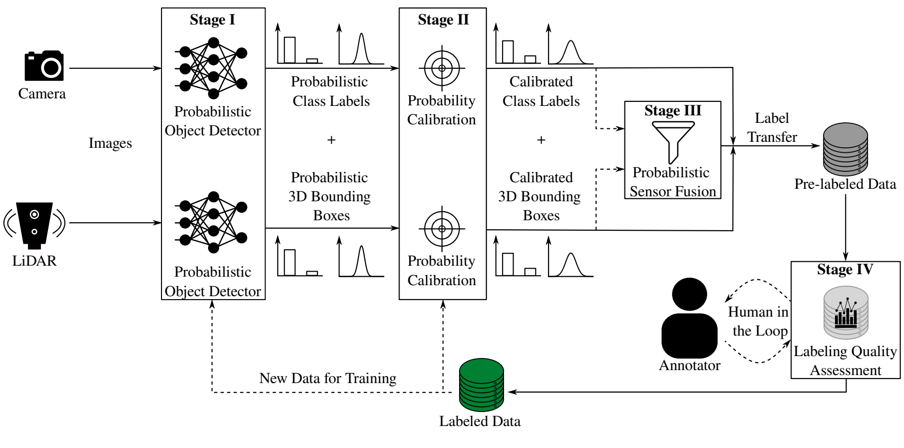

Fig. 1. Illustration of the proposed highly automated labeling process exemplary shown for two sensors. Dashed arrows represent optional processes.

Stage I: Probabilistic Object Detection – *How to train probabilistic object detectors for different sensor modalities?* In stage I, we consider pre-trained object detectors for different sensor domains, i.e., camera- and LiDAR-based object detectors [20, 21]. Common object detectors provide point estimates for the classification probabilities of the object and the coordinates of its 3D bounding box, i.e., position in space, the yaw angle, and its size [13]. In contrast, probabilistic object detectors provide predictive distributions for all quantities. Starting from pre-trained CNNs for 3D bounding box detection, e.g., for camera [21] and LiDAR [20], the main challenge will be a meaningful separation between aleatoric and epistemic uncertainty [10] without massively increasing computational complexity during training and inference. Therefore, we aim to leverage the approach proposed in [14], which enforces specific properties, i.e., smoothness and sensitivity in the feature space learned by a deep neural network. In this way, we can capture epistemic uncertainty by distributing that features and aleatoric uncertainty by evaluating the entropy of its predictive distribution.

Stage II: Probability Calibration – *How to further improve the calibration of the probabilistic detectors?* In stage II, we aim at improving probabilistic outputs by the object detectors as a foundation for the estimation of the labeling quality of the pre-labeling. For example, if the detector outputs a probability for a car with 90%, this statement should also be true in exactly 90% of the cases. The same applies to the probabilistic estimation of continuous target values, e.g., the coordinates of a 3D bounding box. However, deep neural networks tend to frequently output overconfident predictions, which can be alleviated through probability calibration methods [3, 15]. We intend to investigate post-hoc calibration methods, such as temperature scaling [5], and proper scoring rules to optimize the probabilistic object detectors [8].

Stage III: Label-Transfer and Probabilistic Sensor Fusion – *How to transfer labels between different sensor modalities and combine probabilistic predictions originating from different sensor modalities?* In stage III, we use the extrinsic sensor parameters to transfer pre-labels from one sensor domain to another. Therefore, we assume that the sensor extrinsics are known in advance. The transferred pre-labels can be used as labels for the other sensor modality and vice-versa. Moreover, we also investigate the fusion of probabilistic pre-labels (i.e., detections) originating from different sensor modalities (cf. stage I and II). The fusion is realized employing a Bayesian approach (cf. [6]). Therefore, we aim at examining using a joint probabilistic data-association filter [1] for the assignment of 3D bounding box detection from each sensor modality. Furthermore, we investigate the usage of Kalman and particle filters for object tracking [19].

Stage IV: Labeling Quality Assessment – *How to assess the label quality of probabilistic 3D bounding box predictions?* In stage IV, the aim is to assess the labeling quality [7, 16] of the obtained pre-labeled data set. For this purpose, we use the probabilistic predictions and determine expected values, e.g., with respect to the number of expected false classifications, the undetected objects, or the bounding box error. At this point, we want to explore the extent to which these expectations hold. Based on this, we aim to derive a quality seal for the pre-labeled data set. In this context, a starting point for the pre-labeling quality estimation is the ML-based online performance estimation using multiple sensors [11]. Moreover, we expect to assess whether a camera can be used for labeling LiDAR and vice-versa and when such a label transfer is useful. Ideally, the final labeling quality assessment supports human experts to decide whether individual bounding boxes need to be re-labeled (cf. active learning [18]) or whether the pre-labeling is of sufficient quality to release the pre-labeled dataset for further processing such as model training.

3 Conclusion

This article presents a novel concept for highly automated pre-labeling via cross-domain label transfer for perception in autonomous driving. The novelty of our concept lies in the label transfer exploiting the strengths and weaknesses of different sensor modalities for object detection. The use of multiple sensors to improve perception is not new. However, the use for pre-labeling (in the context of 3D bounding box detection) in combination with an explicit quality assessment component under consideration of calibrated probabilistic predictions represents a novel approach. It allows us to attach a quality seal to pre-labeled data sets. The quality assessment is the starting point for human-in-the-loop learning and iterative model improvement. Although our concept focuses on the autonomous driving domain with LiDAR and camera sensors, it can be extended toward multiple sensors and possibly different applications involving data from multiple sensors. Moreover, the presented ideas form a foundation for further investigations in the area of interactive adaptive learning. For example, the uncertainty estimates of the developed probabilistic object detectors might be used to derive novel utility measures for active learning in object detection.

Acknowledgments

This work results from the project KI Data Tooling (19A20001O) funded by German Federal Ministry for Economic Affairs and Energy (BMWI) and the DeCoInt^2-project financed by the German Research Foundation (DFG) within the priority program SPP 1835: “Kooperativ interagierende Automobile”, grant number SI 674/11-2. We, in particular, thank Kristina Dingel, Alice Moallemy-Oureh, and Daniel Kottke for their insightful feedback.

References

1. Bar-Shalom, Y., Daum, F., Huang, J.: The probabilistic data association filter. IEEE Control Systems Magazine **29**(6), 82–100 (2009)
2. van Engelen, J.E., Hoos, H.: A survey on semi-supervised learning. Machine Learning **109**, 373–440 (2019)
3. Feng, D., Rosenbaum, L., Gläser, C., Timm, F., Dietmayer, K.: Can We Trust You? On Calibration of a Probabilistic Object Detector for Autonomous Driving. In: IEEE/RSJ IROS Workshops. Macau, China (2019)
4. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Gläser, C., Timm, F., Wiesbeck, W., Dietmayer, K.: Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE T-ITS **22**(3), 1341–1360 (2021)
5. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On Calibration of Modern Neural Networks. In: ICML. pp. 1321–1330. Sydney, NSW, Australia (2017)
6. Gustafsson, F.: Statistical Sensor Fusion. Stud.Lit. AB, Lund, Sweden (2015)
7. Haase-Schütz, C., Hertlein, H., Wiesbeck, W.: Estimating labeling quality with deep object detectors. In: IV. pp. 33–38. Paris, France (2019)
8. Harakeh, A., Waslander, S.L.: Estimating and Evaluating Regression Predictive Uncertainty in Deep Object Detectors. In: ICLR (2021)
9. Herde, M., Kottke, D., Huseljic, D., Sick, B.: Multi-Annotator Probabilistic Active Learning. In: ICPR. pp. 10281–10288. Milan, Italy (2021)
10. Huseljic, D., Sick, B., Herde, M., Kottke, D.: Separation of aleatoric and epistemic uncertainty in deterministic deep neural networks. In: ICPR. pp. 9172–9179. Milan, Italy (2021)
11. Klingner, M., Bär, A., Mross, M., Fingscheidt, T.: Improving Online Performance Prediction for Semantic Segmentation. In: CVPR Workshop Safe Artificial Intelligence for Automated Driving. p. 8 (2021)
12. Kowol, K., Rottmann, M., Bracke, S., Gottschalk, H.: YOdar: Uncertainty-based Sensor Fusion for Vehicle Detection with Camera and Radar Sensors. In: ICAART. pp. 177–186 (2021)
13. Mousavian, A., Anguelov, D., Flynn, J., Košecká, J.: 3D Bounding Box Estimation Using Deep Learning and Geometry. In: CVPR. pp. 5632–5640 (2017)
14. Mukhoti, J., Kirsch, A., van Amersfoort, J., Torr, P.H., Gal, Y.: Deterministic neural networks with appropriate inductive biases capture epistemic and aleatoric uncertainty. arXiv preprint arXiv:2102.11582 (2021)
15. Niculescu-Mizil, A., Caruana, R.: Predicting good probabilities with supervised learning. In: ICML. pp. 625–632. New York, NY (2005)
16. Northcutt, C., Jiang, L., Chuang, I.: Confident Learning: Estimating Uncertainty in Dataset Labels. Journal of Artificial Intelligence Research **70**, 1373–1411 (2021)
17. Papadopoulos, D.P., Uijlings, J.R.R., Keller, F., Ferrari, V.: Extreme Clicking for Efficient Object Annotation. In: ICCV. pp. 4940–4949 (2017)
18. Settles, B.: Active Learning Literature Survey. Tech. rep., University of Wisconsin-Madison (2010)
19. Thrun, S., Burgard, W., Fox, D. (eds.): Probabilistic Robotics. Intelligent Robotics and Autonomous Agents, MIT Press, Cambridge, MA (2005)
20. Yin, T., Zhou, X., Krähenbühl, P.: Center-based 3D Object Detection and Tracking. CVPR (2021)
21. Zhou, X., Wang, D., Krähenbühl, P.: Objects as Points. arXiv preprint arXiv:1904.07850 (2019)