An Overview of the SVIRO Dataset and Benchmark

Extended Abstract

Steve Dias Da Cruz
IEE S.A.
DFKI - German Research Center for Artificial Intelligence
steve.dias-da-cruz@iee.lu

Oliver Wasenmüller
DFKI - German Research Center for Artificial Intelligence
oliver.wasenmueller@dfki.de

Hans-Peter Beise
Trier University of Applied Sciences
h.beise@inf.hochschule-trier.de

Thomas Stifter
IEE S.A.
thomas.stifter@iee.lu

Didier Stricker
DFKI - German Research Center for Artificial Intelligence
didier.stricker@dfki.de

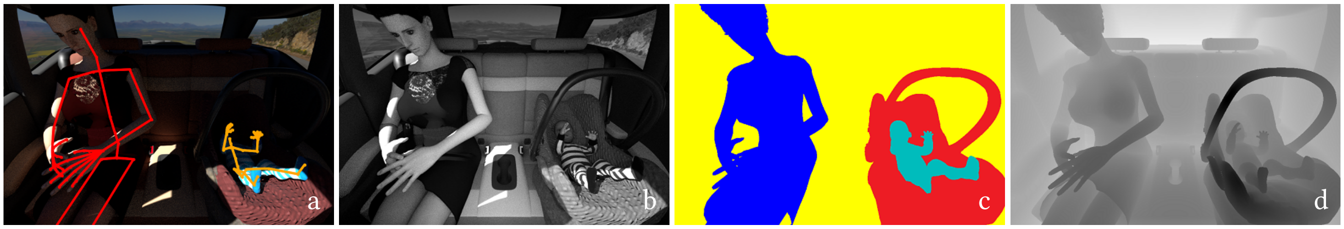

Figure 1: Example data of our SVIRO dataset for occupancy detection together with the provided ground truth information. Left seat: an adult passenger. Middle seat: empty. Right seat: an infant seat with a infant. a) RGB image with keypoints for human pose estimation. b) Simulated infrared image. c) Position and class based instance segmentation. d) Depth map.

ABSTRACT

In this extended abstract, we provide an overview of SVIRO, a recently generated synthetic dataset for sceneries in the passenger compartment of ten different vehicles. We showed that SVIRO can be used to analyze machine learning-based approaches for their generalization capacities and reliability across several tasks when trained on a limited number of variations (e.g. identical backgrounds and textures, few instances per class). This is in contrast to the intrinsically high variability of common benchmark datasets and as a result SVIRO allows investigations under novel circumstances.

1 INTRODUCTION

With SVIRO we focus on rear seat occupant detection and classification using a camera system and different ground truth data, as illustrated in Figure 1. Interior vehicle sensing has gained increased attention in the research community, in particular due to challenges and developments related to automated vehicles [8, 16]. It will be important to understand the overall scenery in the car interior [20], e.g. for handover situations [14], but also to adjust the strength of airbag deployment [7, 19] in case of an accident. However, one has to ensure that trained machine learning models will be capable of classifying new types of child seats correctly while not being mislead by arbitrary everyday objects or through the window background sceneries. Machine learning-based models, and specifically neural networks, trained in a single environment take non-relevant characteristics of the specific environmental conditions into account in an uncontrolled way [22] and therefore data must be recorded repetitively for different environments. Acquiring images in various lightning conditions and accounting for different seat textures, car interior features, or changing camera poses make the data acquisition even more difficult. Consequently, the means for generating a real training dataset with the corresponding annotations are limited and need to be repeated for each additional new car model and automotive manufacturer. Therefore, theoretically founded means to overcome the limitations of datasets collected for many real world applications have to be developed or advanced.

We present a summary of the key-features of SVIRO [4] and highlight its advantageous to serve as a starting point for investigating the aforementioned challenges. SVIRO can be used to benchmark common machine learning tasks under new circumstances while allowing the investigation of theoretical questions due to its intrinsically more tractable environment.

2 SVIRO

During the data generation process we tried to simulate the conditions of a realistic application. We partitioned the available human models, child seats, everyday objects (e.g. backpack, pillows, cardbox) and backgrounds such that one part is only used for the training images (for all the vehicles) and the other part is used for the test

ACM Computer Science in Cars Symposium, 2019, Kaiserslautern, Germany

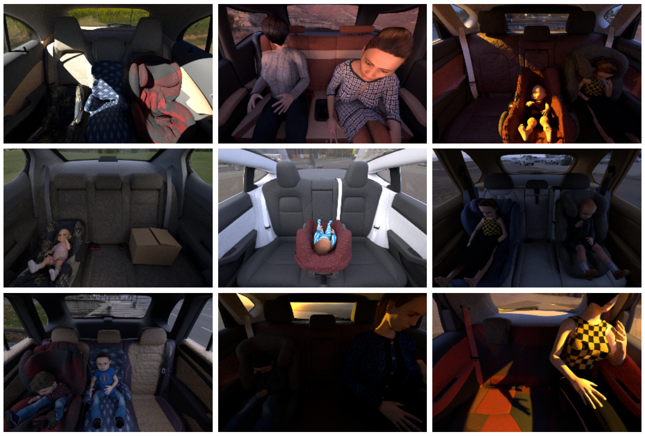

Figure 2: Examples of our dataset for different car models.

images. For each of the different vehicle passenger compartments and available child seats, we fixed the texture as if real images had been taken. Consequently, the machine learning models need to generalize to previously unknown variations of humans, child seats and environments. Further, we can train models in one (or several) car environment(s) and test them on a different one, for solving the same task. This is an advantage compared to common domain adaptation datasets [15, 17, 18, 21, 23] which usually focus on the transfer from synthetic to real images. Further, the photo-realistic rendering and close-to-real models introduce a high visual complexity which makes them more challenging than toy examples [2, 11]. The dataset has an intrinsic dominant background and texture bias: all of the images are taken in a few passenger compartments, but generalization to new, unseen, passenger compartments and child seats should be achieved. This evaluation is currently not possible by state-of-the-art datasets [1, 3, 5, 6, 9, 12, 13].

Our dataset consists of ten different vehicles. The perspective in the different vehicles changes and the number of windows varies, which causes different lightning conditions. Further, some cars have only two rear seats instead of three. SVIRO consists of 16000 training and 4000 test sceneries. The number and constellation of appearances of the different classes varies between the vehicles, because all the sceneries were generated randomly. Examples for the different vehicles are shown in Figure 2.

For each scenery, we provide a set of images and ground truth data: 1) An RGB image (Figure 1.a), 2) a grayscale image (Figure 1.b, physically non accurate infrared simulation), 3) an instance segmentation map (Figure 1.c), 4) bounding boxes, 5) keypoints for all the human poses (Figure 1.a) and 6) a depth map (Figure 1.d).

3 BASELINE EVALUATION

We showed in our baseline evaluation [4] that SVIRO provides the means to analyze the performance of common machine learning methods under new conditions. Specifically, we showed that state-of-the-art models cannot generalize well to new environments and textures when trained on limited number of variations only. For this extended abstract, we limit ourselves to the classification baseline, but additional results can be found in our paper [4]. We considered two training data versions: 1) the standard X5 training data with fixed textures and backgrounds 2) half of the standard X5 training data was replaced by randomly textured X5 training data with random backgrounds. We used the provided grayscale images (infrared simulation), split them into three rectangles (one for each seat position) and trained a single classifier for all seats. The resulting models were then tested on the X5, Tucson (three seats) and i3 (two seats). A comparison of the deep learning and the support vector (SVM) machine performances is reported in Table 1.

Table 1: Classification results of a ResNet-18 model and a SVM trained on HOG features. The models were trained on standard X5 data or we replaced half of it with randomly textured images. The models were evaluated on the test dataset of the X5, Tucson and i3 and we report the total accuracy.

Method	Training data (X5)	X5	Tucson	i3
ResNet-18	Fixed texture	71%	32%	39%
ResNet-18	Fixed and random texture	87%	41%	54%
HOG+SVM	Fixed texture	69%	35%	41%
HOG+SVM	Fixed and random texture	70%	53%	52%

CNN: We fine-tuned the last residual block and the classification layer of a pre-trained ResNet-18 model. The resulting model has problems to generalize to the test set, especially for new cars. The randomized backgrounds and textures help to improve the accuracy on the same car, which gives hint that the model mostly used the texture as a classification criterion. This observation seems to be in line with recent results by Geirhos et al. [10]. However, the model can still not generalize well to new vehicle interiors, probably because of the different interior structures.

HOG+SVM: We computed the histogram of oriented gradients (HOG) features and used them to train a SVM. This approach has similar problems as the deep learning approach when the standard X5 data is used, but can sometimes even generalize better. However, it cannot exploit the additional information of random textures and backgrounds to improve the accuracy in the car it was trained on.

4 CONCLUSION

Our dataset and baseline evaluation addresses real-world engineering obstacles regarding the robustness and generalization of machine learning models. Using SVIRO, we showed that traditional and deep learning approaches drastically decrease classification performance when trained in a setting with limited variations without taking additional precautions. The models cannot generalize well to the new intra-class variations, even in the car they were trained on and perform even worse in unknown vehicles. Both presented approaches do not fully grasp the underlying task, although the environment and the objects are similar. In order to be applicable in real world applications additional (theoretical) improvements need to be investigated and developed. SVIRO provides a starting point to perform these investigations. For more details check our original paper [4]. **Acknowledgement:** The first author is supported by the Luxembourg National Research Fund (FNR) under the grant number 13043281. This work was partially funded by the European Union’s Horizon 2020 Program in the project VIZTA (826600).

REFERENCES

[1] Markus Braun, Sebastian Krebs, Fabian B. Flohr, and Dariu M. Gavrila. 2019. EuroCity Persons: A Novel Benchmark for Person Detection in Traffic Scenes. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)* (2019).

[2] Christopher P Burgess, Loic Matthey, Nicholas Watters, Rishabh Kabra, Irina Higgins, Matt Botvinick, and Alexander Lerchner. 2019. MONet: Unsupervised Scene Decomposition and Representation. *arXiv preprint arXiv:1901.11390* (2019).

[3] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[4] Steve Dias Da Cruz, Oliver Wasenmüller, Hans-Peter Beise, Thomas Stifter, and Didier Stricker. 2020. SVIRO: Synthetic Vehicle Interior Rear Seat Occupancy Dataset and Benchmark. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. 2010. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision (IJCV)* (2010).

[6] Matteo Fabbri, Fabio Lanzi, Simone Calderara, Andrea Palazzi, Roberto Vezzani, and Rita Cucchiara. 2018. Learning to Detect and Track Visible and Occluded Body Joints in a Virtual World. In *European Conference on Computer Vision (ECCV)*.

[7] Michael E Farmer and Anil K Jain. 2003. Occupant classification system for automotive airbag suppression. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[8] Lex Fridman, Daniel E Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Julia Kindelsberger, Li Ding, Sean Seaman, et al. 2017. Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation. *arXiv preprint arXiv:1711.06976* (2017).

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[10] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. 2018. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231* (2018).

[11] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. 2017. CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[12] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Tom Duerig, and Vittorio Ferrari. 2018. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. *arXiv preprint arXiv:1811.00982* (2018).

[13] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European Conference on Computer Vision (ECCV)*.

[14] Roderick McCall, Fintan McGee, Alexander Meschtscherjakov, Nicolas Louveton, and Thomas Engel. 2016. Towards a taxonomy of autonomous vehicle handover situations. In *International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI)*.

[15] Boris Moiseev, Artem Konev, Alexander Chigorin, and Anton Konushin. 2013. Evaluation of Traffic Sign Recognition Methods Trained on Synthetically Generated Data. In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*.

[16] E. Ohn-Bar and M. M. Trivedi. 2016. Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles. *Transactions on Intelligent Vehicles (T-IV)* (2016).

[17] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. 2018. Moment Matching for Multi-Source Domain Adaptation. *arXiv preprint arXiv:1812.01754* (2018).

[18] Xingchao Peng, Ben Usman, Kuniaki Saito, Neela Kaushik, Judy Hoffman, and Kate Saenko. 2018. Syn2Real: A New Benchmark forSynthetic-to-Real Visual Domain Adaptation. *arXiv preprint arXiv:1806.09755* (2018).

[19] Toby Perrett and Majid Mirmehdi. 2016. Cost-based feature transfer for vehicle occupant classification. In *Asian Conference on Computer Vision (ACCV)*.

[20] Erwin Jose Lopez Pulgarin, Guido Herrmann, and Ute Leonards. 2017. Drivers' Manoeuvre classification for safe HRI. In *Conference Towards Autonomous Robotic Systems*.

[21] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks* (2012).

[22] Maoqing Tian, Shuai Yi, Hongsheng Li, Shihua Li, Xuesen Zhang, Jianping Shi, Junjie Yan, and Xiaogang Wang. 2018. Eliminating background-bias for robust person re-identification. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

[23] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. 2017. Deep Hashing Network for Unsupervised Domain Adaptation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.