Proceedings of the DataNinja sAIOnARA 2024 Conference 69-71 DOI: 10.11576/dataninja-1175

Study on the Influence of Texture Variation on the Validation Performance of a Synthetically Trained Object Detector

Alexander Moriz A.MORIZ@WZL-IQS.RWTH-AACHEN.DE
Dominik Wolfschläger D.WOLFSCHLAEGER@WZL-IQS.RWTH-AACHEN.DE
WZL-IQS at RWTH Aachen University, Aachen, Germany
Robert H. Schmitt R.SCHMITT@WZL-IQS.RWTH-AACHEN.DE
WZL-IQS at RWTH Aachen University, & Fraunhofer Institute for Production Technology IPT, Aachen, Germany

Abstract

In recent years, the utilization of synthetic data for the training of Deep Learning (DL) approaches has emerged as a valid alternative to the costly process of real data acquisition. Yet, the influence of the sim-to-real gap on the model performance still poses an obstacle to the broader usage of synthetic data. To investigate the major contributing factors, this study focuses on the influence of texture variation as a first step. Examining different strategies for generating synthetic validation sets for the training process of an object detector, the results of this study indicate that the sole influence of textures is insufficient to cause the observable performance gap alone.

Keywords: synthetic data, object detection, textures

1. Introduction

Although increasingly employed for the training of Deep Learning (DL) models, the broad utilization of synthetic data is still impeded by the sim-to-real gap, appearing as a performance gap of synthetically trained DL models when evaluated on real data. While strategies to reduce the impact of the sim-to-real gap are available, the potential benefit in terms of improved performance is usually reported for a dedicated test set. Following best practice, the DL model used for such an evaluation is thereby chosen based on the performance on a separate validation set, monitored during the training process. However, if the validation set has been generated synthetically following the same strategy as for the training set, this choice might be misleading. The authors hypothesize that for such cases, the sim-to-real gap affects the performance already during the training process since the optimal model for the synthetic validation set might not be well suited for real data.

The study presented in this work focuses on the influence of texture variation on the performance of DL models as one potential critical factor. Taking the detection of a custom-designed object as a typical use case, the validation set performance of an exemplary DL model is evaluated over its training process on different validation sets. In particular, three different strategies for the generation of synthetic validation sets are examined, comparing their performance with a baseline approach and the performance on a small dataset of real images.

2. Related Work

In general, there are several ways to generate synthetic data, such as crop-out-based, 3D-modelling-based, or game-engine-based approaches [1]. The main challenge for all these approaches constitutes the sim-to-real gap [1–3]. Typically, domain adaptation or domain randomization strategies are applied to minimize its influence [1, 4]. Domain adaptation focuses on the generation of photorealistic images [5], creating a realistic scene of the target environment with e.g. physics-based rendering (PBR) [4]. The domain randomization approach pursues the opposite strategy, randomizing the simulated scene strongly to achieve a better generalization of the trained DL models directly [1]. In [6], the authors propose a framework for an end-to-end realization of DL models based on task-specific synthetic data generation. To reduce the impact of the sim-to-real gap as much as possible without requiring substantial manual design effort, the proposed framework utilizes a PBR-based domain randomization approach, varying multiple simulation parameters, such as object position, lighting, and (object) textures [6]. For the latter, the publicly available CC texture dataset is used [6].

© Moriz, D. Wolfschläger & R.H. Schmitt. Licensed under CC BY 4.0.

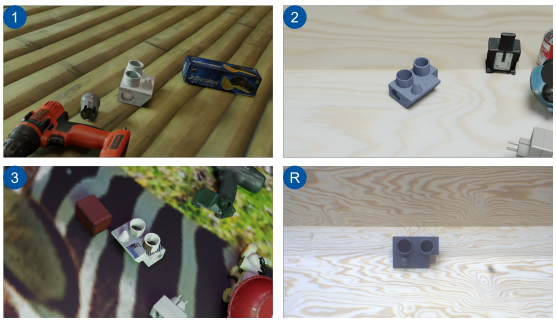

Figure 1: Examples from the considered validation sets: 1) CC textures, 2) realistic textures, 3) MS COCO textures, and R) real images.

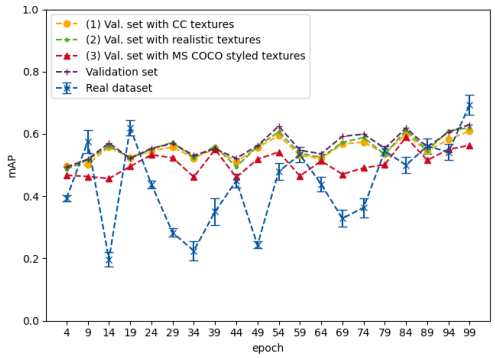

Figure 2: Performance of the individual model checkpoints for different validation sets.

3. Methodology

In this study, 6000 synthetic images have been generated utilizing the framework proposed in [6], split into 5000 images for the training- and 1000 images for the validation set. The chosen object detection model (RetinaNet, [7]) is trained for 100 epochs, storing the current model state as a checkpoint every five epochs. Afterward, the performance of the checkpoints is evaluated as individual models by determining the mean Average Precision (mAP) for all considered datasets, mimicking the performance monitoring during training for the investigated datasets.

To examine the influence of texture variation, three synthetic datasets have been generated as potential validation sets, each following a different strategy. Figure 1 visualizes an example from each dataset (sets 1 to 3) in addition to a real image (R-set). The first (1) dataset exhibits similar textures as the original synthetic validation set, utilizing two different, disjoint subsets of the CC textures for the generation of both datasets. The second (2) dataset features realistic textures, extracted from a real image, such as visualized in Figure 1, for the background and the object, respectively. The last investigated dataset (3) utilizes a small subset of plain MS COCO images [8], used randomly as textures for both the background and the object as displayed in Figure 1.

4. Results and Discussion

Figure 2 shows the performance of the individual checkpoints (models) for the different datasets. As assumed in Section 1, the performance of the real dataset (blue) exhibits a different behavior during the training process than the synthetic validation set (violet), showing distinct, not corresponding local minima and maxima. Considering the performance of the first synthetic dataset (1, yellow), no major difference to the validation set performance can be observed, indicating a good generalization capability to similar textures. Surprisingly, the performance of the second dataset (2, green) is also in agreement with the validation set, showing thus no benefit compared to the usage of the regular validation set (CC textures). Finally, the performance of the third dataset (3, red) deviates more strongly from the performance of the regular validation set. Showing on average a lower detection performance, it also features minima and maxima, which do not correspond with the validation set or the real dataset. The authors presume that this behavior might be linked to the resulting complexity of the considered textures, exhibiting patterns and artifacts, such as the zebra pattern observable in Figure 1, that are not present in the training set.

5. Conclusion and Outlook

The results of this study indicate that the sim-to-real gap, observable as the performance difference between the synthetic validation set and the real dataset, cannot be explained by the variation of texture properties alone. Future work will examine the presented results with a focus on other factors of influence such as object size or illumination in more detail. Also, additional use cases will be evaluated to support the observed findings.

References

[1] Hannah Schieber, Kubilay Can Demir, Constantin Kleinbeck, Seung Hee Yang, and Daniel Roth. Indoor synthetic data generation: A systematic review. *Computer Vision and Image Understanding*, 240:103907, 2024.

[2] Chafic Abou Akar, Jimmy Tekli, Daniel Jess, Mario Khoury, Marc Kamradt, and Michael Guthe. Synthetic object recognition dataset for industries. 1:150–155, 2022.

[3] Stefan Hinterstoisser, Olivier Pauly, Hauke Heibel, Martina Marek, and Martin Bokeloh. An annotation saved is an annotation earned: Using fully synthetic training for object instance detection. *CoRR*, abs/1902.09967, 2019.

[4] Leon Eversberg and Jens Lambrecht. Generating images with physics-based rendering for an industrial object detection task: Realism versus domain randomization. *Sensors (Basel, Switzerland)*, 21(23), 2021.

[5] Johannes Dümmel, Valentin Kostik, and Jan Oellerich. Generating synthetic training data for assembly processes. 633:119–128, 2021.

[6] Alexander Moriz, Dominik Wolfschläger, Robert H. Schmitt. Concept for a machine vision framework for production environments based on task-specific synthetic data generation. 2024. Paper accepted.

[7] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, abs/1708.02002, 2017.

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2014.