Mixed data set training

Andrea Kraus, Martin Simon | Valeo

ML Function Development

Challenge & experiment setup

DNNs need large amounts of training data to ensure good accuracy and robustness. While the synthetic images produced do not seem to require annotation effort, their effectiveness in properly feeding DNNs remains to be proven. The aim of this work is about exploring the potential of synthetic data, in particular the capacity to overcome the lack of real data.

In order to limit the scope of our study, we have chosen to focus on instance segmentation. To perform this task, we use a Mask R-CNN architecture to which we add the PointRend module [1]. We use a **real** (Cityscapes [2]) and a **synthetic** (Synscapes [3]) data set for the network training.
The Synscapes data set was designed to be similar in structure and content to the Cityscapes data set. Nevertheless a pre-processing in label harmonization and non-valuable instance removing from synthetic images is needed, before training and evaluation is possible on both data sets. Cityscapes contains only 3475 images, while Synscapes contains 25000 images.

		Evaluation Set	
		Synscapes	Cityscapes
Training Set	Synscapes	**31.6%**	12.9%
	Cityscapes	11.8%	**35.7%**

Table 1: Baseline performance of the network, measured in mean Average Precision

Data set mixing strategy

The principle of our approach is simple. As we have so far made our two data sets as similar as possible, there is no more concrete incompatibility, for the network, to learn both synthetic and real images at the same time. Although domain shift is still a reality, each image, real or synthetic, contains valuable features for the network performance and robustness. For these reasons, we mix real and synthetic images together, to create a single training data set. During the training, we do not make any process distinction between real and synthetic images.

	Amount of additional synthetic data				
	0	3000	6000	9000	20000
AP	35.7%	37.9%	38.9%	39.3%	**40.9%**

Table 2: Average precision obtained on real images against amount of synthetic images

Table 2 shows that the network obtains a performance up to 40.9% for a 50 epoch training on a mixed data set containing Cityscapes and Synscapes in their entirety. This data set mixing strategy overcomes the current performance of the PointRend module on Cityscapes.

References:

[1] A. Kirillov, Y. Wu, K. He and R. Girshick, "PointRend: Image Segmentation As Rendering," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9796-9805, doi: 10.1109/CVPR42600.2020.00982.
[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[3] M. Wrenninge and J. Unger, "Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing," arXiv preprint arXiv:1810.08705, 2018.

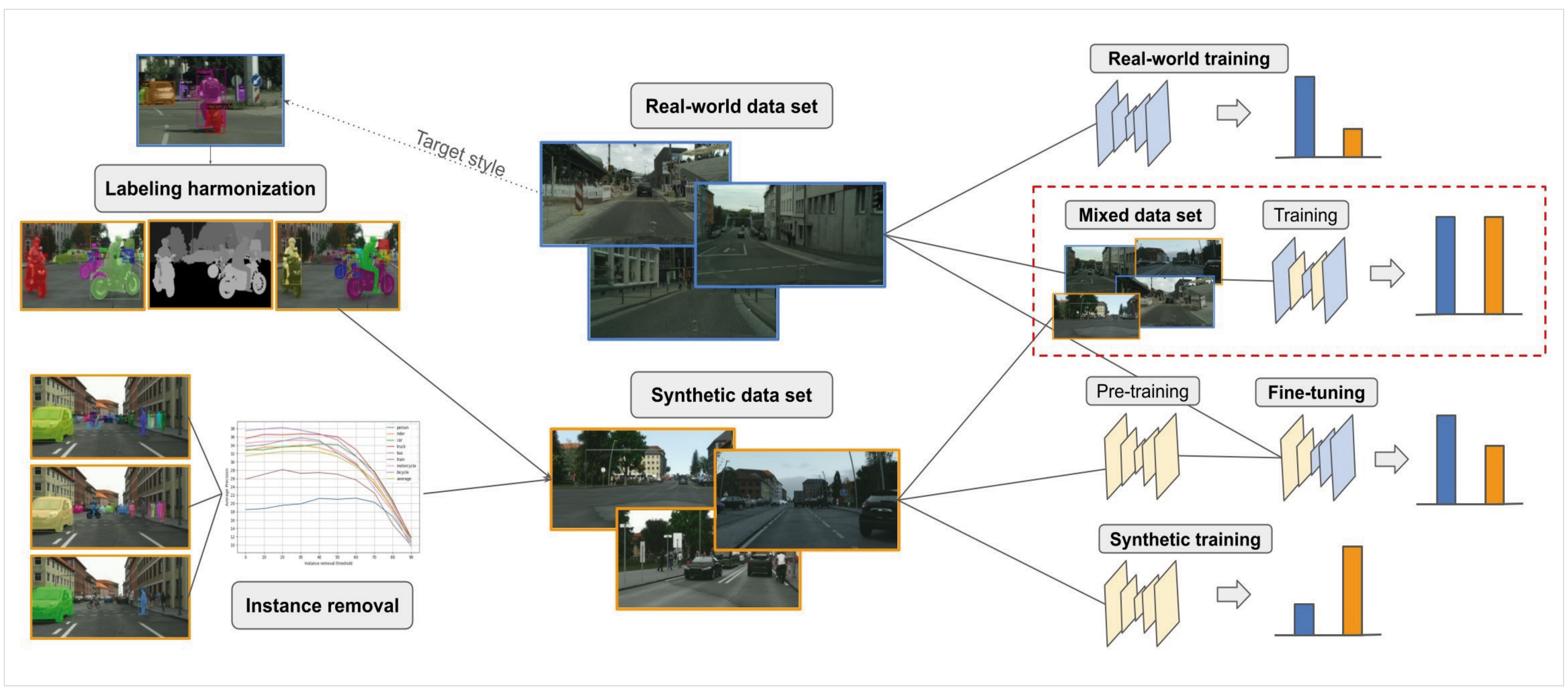

*Figure 1: **Processing steps for the use of synthetic data.** On the left, the pre-processing steps: labeling harmonization regarding class definition and non-valuable synthetic instance removal. On the right, different training approaches (real-world training, mixed data set training, fine-tuning, synthetic training) and their performance obtained on Cityscapes (blue) and Synscapes (orange)*

Partners

External partners

For more information contact:

andrea.kraus@valeo.com
martin.simon@valeo.com

KI Data Tooling 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 Climate Action.

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

www.ki-datatooling.de @KI_Familie KI Familie