Approaches focused on synthetic Data in Domain Transfer for Street View Understanding

Ayhan Can Erdur,
Technical University of Munich

KI DATA TOOLING
The Data Kit for Automotive AI

From synthetic to real

Deep learning models thrive with more data. While this is beneficial in most cases it sometimes creates bottlenecks as the data collection and especially labeling processes can become expensive. Street view understanding is one of the fields where both, data collection and pixel-wise labeling is hard. Synthetic simulator data can be used instead of or alongside with real data. However, the domain gap mostly confuses the object detection models. Domain transfer explores methods to make synthetic data more realistic.

Method

We use our modified version of CycleGAN [1] as the domain transfer network. It is enhanced to process auxiliary data and to keep the semantic consistency between inputs and outputs. Our dataset consists of CARLA as synthetic domain and Cityscapes as the real domain. Following suggestions of [4], we extract intermediate rendering products, G-Buffers, to use them as additional inputs.

Dataset	FID Score
Random 300^2 Patches	133.26
Paired 300^2 Patches	108.55
Paired 300^2 Patches + G-Buffers	100.1
Paired 300^2 Patches, Entropy Filtering	126.45
No Patches, Whole Scene	62.33
No Patches + G-Buffers	**58.16**

Table 1: FID scores of domain transfer by models with different training settings

Figure 2: Results from "No Patches" (above) and "Paired Patches + G-buffers" (below) settings

The work on this report focuses on effects of different data sampling and cropping techniques.

Getting the best out of data

Instead of feeding the generator and discriminator randomly, we follow a patch-pairing sampling strategy. Given a source patch, we find its pair in the target domain by computing the distance between their feature vectors.
We execute a profound analysis of pair-sampling and patch filtering techniques. We examine different patch sizes, different distance thresholds on pairing and patch filtering.

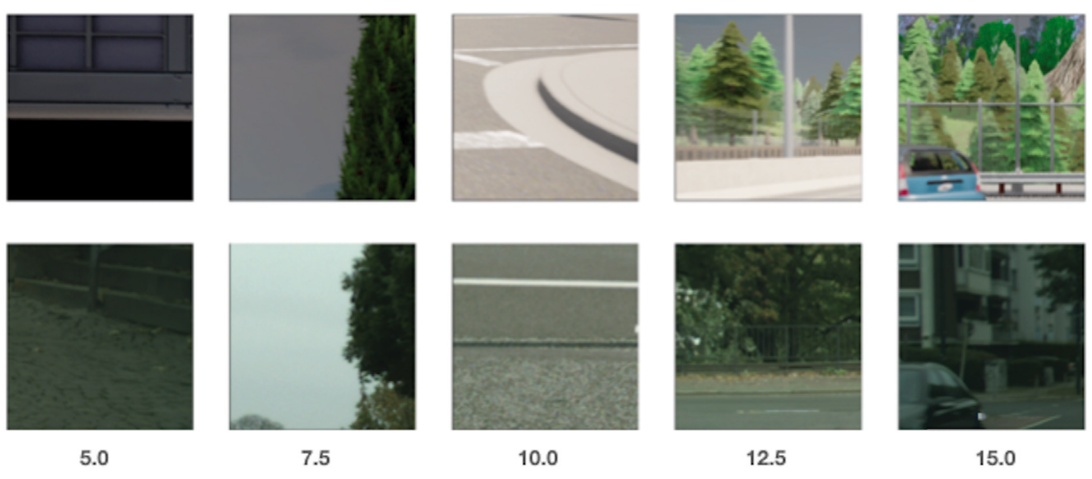

Figure 1: Source (up) and target (below) patches in different entropy regions

In some cases, the patches lack semantic diversity and are plain surfaces (e.g., sky, asphalt), which is ambiguous for the discriminator. We compute the entropy of patches to represent its semantic richness and remove the patches below a certain threshold.

Evaluation

We use Fréchet inception distance (FID) to compare if a set of generated images is similar with the target domain. A lower score indicates higher similarity. We conclude that despite our efforts, using the whole image instead of local crops works best for translating CARLA into Cityscapes.

References:

[1] Zhu, Jun-Yan, et al. „Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision (2017).
[2] Richter, Stephan R., Hassan Abu AlHaija, and Vladlen Koltun. „Enhancing Photorealism Enhancement." arXiv preprint arXiv:2105.04619 (2021).

Technische Universität München

For more information contact:
can.erdur@tum.de
dejan.azinovic@tum.de

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.

KI FAMILIE

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