KI DATA TOOLING
The Data Kit for Automotive AI

Real-synthetic data mix training

Thomas Stone | BMW
Tianming Qiu | fortiss

ML Function Development

Objective:

- Synthetic data is typically more accessible and cost-effective for annotation
- Combine real and synthetic data for robustness and generalization

Batch-wise mix training:

- Mainstream mix training starts with pre-training using synthetic data, followed by fine-tuning the model with real-world image, often referred to as “epoch-wise” mixing.
- This work package primarily focuses on “batch-wise” mixing, which involves combining real and synthetic images within the same training batch.
- Non-convex optimization is challenging in high-dimensional spaces. Even minor adjustments can lead to significant variations in results.
- We have the flexibility to adjust the mixing ratio as shown in Fig. 1.
- We can also tune the weights of the loss function from both data sources:

$$Loss(\mathcal{X}, \mathcal{Y}) = \alpha \cdot Loss_{real} + \beta \cdot Loss_{synth}$$

Working model:

- We work on multi-modal 3D bounding box detection for pedestrian, the pipeline can be seen in Fig. 2.
- Bird-Eye-View (BEV) feature: both camera input and Lidar input are encoded into BEV features.
- CenterNet: 3D bounding boxes will be predicted from the BEV features using CenterNet architecture.[1]

Results on KI-DT + KI-A data:

We use KI-DT data as real-world input and KI-A data as synthetic input.
Here are three training strategies:

- Pure: only use real-world dataset training
- Constant ratio: constant 0.5 mixing ratio between synthetic and real-world data
- Increasing ratio: increasing mixing ratio from 0.2 to 1 between real-world data and synthetic data

We are focusing on pedestrian detection, where the common Intersection over Union (IoU) thresholds are 0.3 and 0.5. Consequently, we compare the results in the following table:

IoU	Pure	Constant ratio	Increasing ration
0.3	0.26344	0.23813	0.33197
0.5	0.09862	0.08525	0.13570

References:

[1] Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." arXiv preprint arXiv:1904.07850 (2019).
[2] Roddick, Thomas, Alex Kendall, and Roberto Cipolla. "Orthographic feature transform for monocular 3d object detection." arXiv preprint arXiv:1811.08188 (2018).
[3] Ku, Jason, et al. "Joint 3d proposal generation and object detection from view aggregation." 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.
[4] Ros, German, et al. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

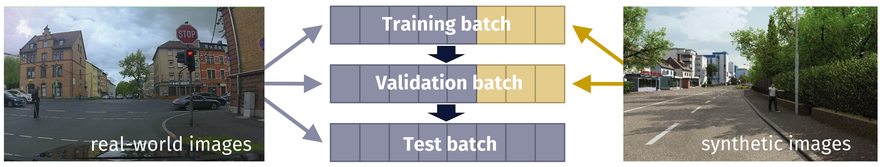

Figure 1: We randomly combine various real-world and synthetic images into each batch during training. The mixing ratio is a hyperparameter that can be adjusted. The validation batch can adhere to the same strategy as the training batch. However, our primary focus is on the test performance with real-world data, so the test batch exclusively contains real-world data. (© BMW)

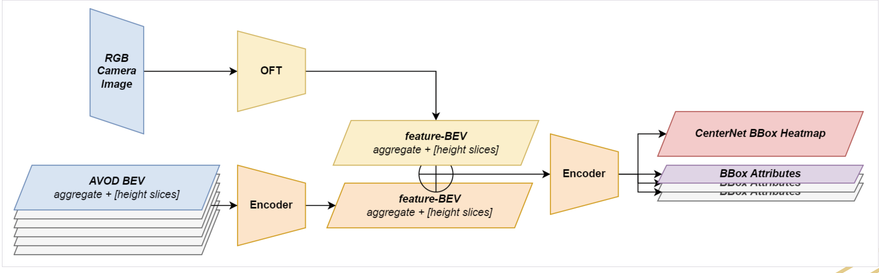

Figure 2: Multi-modal 3D pedestrian detection using Lidar and camera input (© BMW)

Partners

External partners

For more information contact:
Thomas.Stone@bmw.de
qiu@fortiss.org

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