# MORDA: A Synthetic Dataset to Facilitate Adaptation of Object Detectors to Unseen Real-target Domain While Preserving Performance on Real-source Domain

Hojun Lim*, Heecheol Yoo*, Jinwoo Lee, Seungmin Jeon, Hyeongseok Jeon†
MORAI Inc.
Republic of Korea
hjlim@morai.ai, hcyoo@morai.ai, jwlee@morai.ai, smjeon@morai.ai, hsjeon@morai.ai

## 1. Summary of the Work

In our accepted work [2] at IEEE ICRA 2025, we aim to alleviate an indispensable dataset-creation cost in deep-learning-based training pipelines for autonomous vehicles (AVs) by leveraging simulation environments. In many cases, costs to build datasets are not a one-time concern as AVs, which are trained on one domain ($D^{Src}_{Real}$) but need to be deployed to another domain ($D^{Trg}_{Real}$), often require a novel dataset that represents the features of target region $D^{Trg}_{Real}$. To mitigate the mentioned reliance on real-world datasets, we propose to utilize a synthetic dataset constructed in a cost-effective simulator ($D^{Src+Trg}_{Syn}$) which provides a virtual preview of the driving environments of $D^{Trg}_{Real}$, captured by the sensor configuration of $D^{Src}_{Real}$. As a concrete demonstration of our method, we create a novel synthetic dataset, MORDA: ***M**ixture **O**f **R**eal-domain characteristics for synthetic-data-assisted **D**omain **A**daptation*.

The purpose of MORDA is to provide virtual features to neural networks that they can indirectly learn about unforeseen $D^{Trg}_{Real}$ without compromising performance on $D^{Src}_{Real}$. To achieve this goal, we reproduced pivotal characteristics of $D^{Src}_{Real}$ and $D^{Trg}_{Real}$ in a simulator $D^{Src+Trg}_{Syn}$, as depicted in Fig. 1. Then, $D^{Src+Trg}_{Syn}$ is leveraged to create MORDA dataset. In our research, MORDA takes nuScenes [1] and South Korea for $D^{Src}_{Real}$ and $D^{Trg}_{Real}$, respectively. Hence, MORDA mimics a dataset that contains driving scenes of South Korea collected by the sensor suite of nuScenes.

In our experiments, we observe that including MORDA in the training of 2D/3D object detectors can assist adaptation to $D^{Trg}_{Real}$ without performance degradation on $D^{Src}_{Real}$. For trained domain ($D^{Src}_{Real}$), Faster-RCNN [3] and Center-Point [4] with MORDA report +0.5% and +1.57% mAP gains. For unforeseen target domain ($D^{Trg}_{Real}$), the same models show +6.35% and +18.16% mAP gains, indicating the simulated features of MORDA can lessen the dependency on real sensor data from $D^{Trg}_{Real}$.

DOI: 10.1109/ICRA55743.2025.11127978
*: These two authors contributed equally. †: Corresponding author.

## 2. Relevance to the Workshop

Domain gap is a well-known challenge that numerous Sim-to-Real methods have been proposed to mitigate for further simulation utilization. Our work proposes a novel method to address the gap and showcases simulator-assisted adaptation to improve performance on unseen real-world domains.

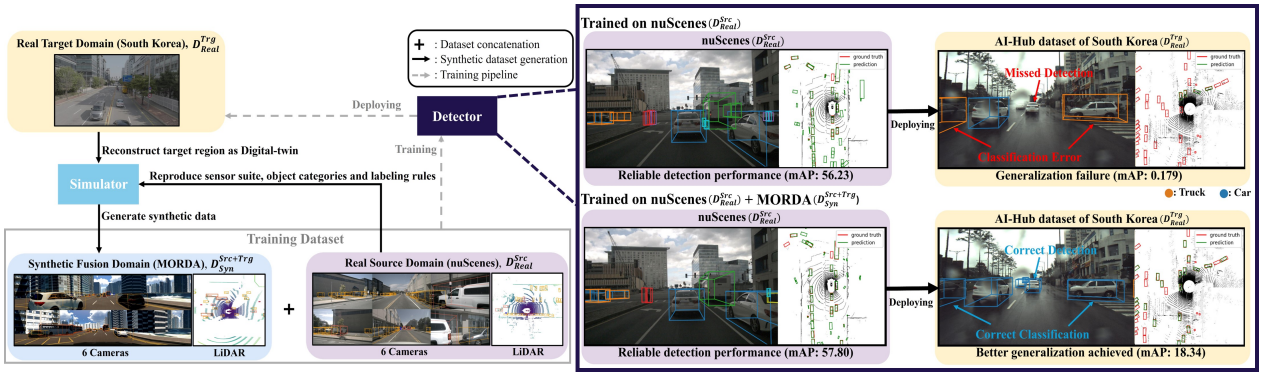


Figure 1. Object categories, labeling policies, and sensor suite (multiview cameras, LiDAR) of nuScenes ($D^{Src}_{Real}$) are reproduced in a simulator ($D^{Src+Trg}_{Syn}$). To reflect the geographical features of South Korea ($D^{Trg}_{Real}$), digital-twin maps constructed from HD map and geographical survey data are employed during data generation of MORDA. In our research, AI-hub dataset (http://www.aihub.or.kr) represents driving scenes of South Korea. Image is adapted from our accepted work [2].

## 3. Acknowledgements

This research was supported by a grant(code RS-2023-00233952) from R&D Program funded by Ministry of Land, Infrastructure and Transport of Korean government. This research (paper) used datasets from High-precision data collection vehicle daytime city road data. All data information can be accessed through AI-Hub (http://www.aihub.or.kr).

## References

[1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020. 1

[2] Hojun Lim, Heecheol Yoo, Jinwoo Lee, Seungmin Jeon, and Hyeongseok Jeon. MORDA: a synthetic dataset to facilitate adaptation of object detectors to unseen real-target domain while preserving performance on real-source domain, 2025. 1

[3] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1

[4] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. *CVPR*, 2021. 1