ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

From Simulation to Deployment: Curriculum-Based Domain Adaptation for Semantic Segmentation in Autonomous Forklifts

Christof Schützenhöfer*† Patrick Rechberger* Thomas Ulz* Christian Steger†
*KNAPP Industry Solutions GmbH, Dobl, Austria
†Graz University of Technology, Institute of Technical Informatics, Graz, Austria,
{christof.schuetzenhoefer, patrick.rechberger, thomas.ulz}@knapp.com, {schuetzenhoefer, steger}@tugraz.at

***Abstract*— Deploying semantic segmentation models for autonomous forklifts in industrial environments is challenging because visual conditions vary across sites, leading to poor cross-domain generalization and costly re-annotation efforts. We propose a curriculum-based domain adaptation framework that progressively transfers a segmentation model from simulation to real-world industrial deployment. The model is first pretrained on synthetic datasets with increasing complexity, then fine-tuned on a labeled real source domain to reduce the sim-to-real gap and adapt to camera-specific characteristics. Finally, it is adapted to a new target domain using pseudo-label-based self-training. To reduce drift during target adaptation, pseudo-labeled target samples are combined with labeled samples from the source-real domain, while a replay buffer improves robustness to class imbalance by oversampling rare classes. Preliminary experiments with DDRNet demonstrate improved performance under both moderate and hard domain shifts, with mIoU gains from 67.37 to 71.36 and from 49.57 to 57.22, respectively. The results highlight the potential of progressive multi-domain adaptation for scalable industrial robotic perception.**

semantic segmentation, synthetic data, pseudo labeling

I. Introduction

Robotic systems are becoming increasingly prevalent in industrial environments. Among these systems, autonomous forklifts play an important role in factories and warehouses by transporting goods between production stations, storage areas and delivery points. To operate safely and efficiently in such dynamic environments, these platforms require a robust understanding of their surroundings. In this context, vision-based perception has emerged as a key enabler. In particular, semantic segmentation of RGB images allows the assignment of a semantic label to each pixel, providing detailed scene understanding. Despite the recent progress of deep learning-based perception systems, deploying semantic segmentation models in real industrial settings remains challenging. Industrial scenes often vary across sites due to differences in layout, illumination, object appearance and operational conditions. As a result, a model trained for one environment often does not generalize well to another. In practice, this typically requires the collection and manual annotation of large amounts of data for each new deployment scenario. Such a process is costly, time-consuming and prone to human labeling errors, which limits the scalability. A promising alternative is the use of synthetic data generated with simulation engines. Synthetic datasets can be created with pixel-perfect annotations at low cost and can be rapidly adapted to new object classes, scene layouts and environmental conditions. Our recent work has shown that synthetic data can be a key factor in achieving flexible perception for objects and environments [1]. More broadly, the use of synthetic data to address the scarcity of annotated real-world samples has become increasingly common in semantic segmentation and related perception tasks [2], [3]. However, a domain gap between simulated and real images still limits direct transfer, especially in industrial applications where scene-specific appearance plays a critical role.

To address the challenge of scalable deployment across changing industrial environments, we propose a curriculum-based domain adaptation for semantic segmentation. It consists of an offline stage, in which the model is pretrained on synthetic data and fine-tuned on a labeled real source domain and an online stage, in which it is adapted to a new target domain using pseudo-labeling. To reduce drift during self-training, labeled replay samples from the source-real domain are retained as an anchor, which is particularly important under class imbalance, where dominant classes may otherwise overwhelm rare but operationally relevant categories. This progressive adaptation strategy reduces annotation effort while improving robustness across diverse industrial settings.

II. Methodology

We propose a curriculum-based training pipeline for semantic segmentation that progressively adapts a model from simulation to real-world industrial environments.

A. Problem Formulation

The semantic labels consist of five classes: obstacle, floor, pallet, forklift and person. Among these, pallet, person and forklift are relatively rare compared to floor and obstacles, resulting in a strongly imbalanced class distribution.

Let D_S denote a synthetic source domain with dense semantic annotations, D_A a first real domain with labeled images and D_B a second real target domain with no manual annotations. Our goal is to learn a semantic segmentation model f_θ that performs reliably on D_B while minimizing annotation effort in the target environment.

The proposed pipeline, as illustrated in Figure 1, follows a staged curriculum. First, the model is pretrained on synthetic data to learn general semantic representations. Second, it is fine-tuned on Domain A to adapt to real-world appearance

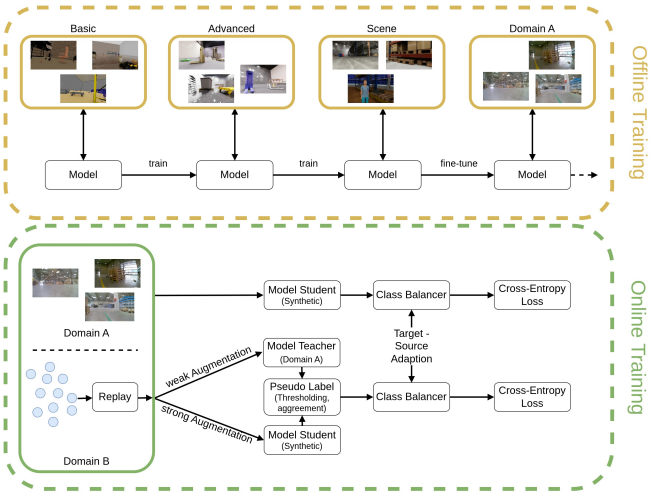

Fig. 1. Overview of the proposed two-stage training pipeline. During offline training, the model is first pretrained on multiple synthetic datasets of increasing complexity and subsequently fine-tuned on labeled real-world data from Domain A to reduce the sim-to-real gap. During online adaptation, pseudo-labeled samples from Domain B are combined with a labeled subset of Domain A, which acts as an anchor to stabilize training and bootstrap adaptation. Pseudo-labels are generated by the teacher model and filtered using class-aware confidence thresholds and teacher-student agreement.

and camera-specific characteristics. Third, the model is transferred to Domain B using pseudo-labeling, while a small labeled subset from Domain A is retained during training to stabilize optimization and bootstrap the adaptation process.

B. Stage 1: Offline Training

In the first stage, the segmentation network is trained on the synthetic dataset D_S. The purpose of this stage is to learn task-relevant visual features and class-specific priors from a large amount of automatically annotated data. Training on D_S serves as the initial step of the curriculum, where the model learns the segmentation task in a controlled setting before being exposed to the appearance variations of real industrial environments. After synthetic pretraining, the model is fine-tuned on Domain A, which consists of real images captured in an operational environment. This step reduces the sim-to-real gap by adapting the learned features to real textures, illumination conditions and scene statistics. In addition, this stage allows the model to implicitly adapt to camera-specific imaging characteristics, such as intrinsic parameters, viewpoint, lens distortion and image formation properties, which are typically not fully matched by the simulator. By fine-tuning on Domain A, the model learns a representation that is no longer only semantically meaningful, but also aligned with the physical sensing characteristics of the real platform. This is particularly important in industrial applications, where camera setup and mounting position can vary between deployments.

C. Stage 2: Online Training

To transfer the model from Domain A to a new target environment, Domain B, we employ a pseudo-labeling strategy. First, the model adapted on Domain A is used to generate pixel-wise predictions for unlabeled images from Domain B. These predictions are then treated as pseudo-labels and used as supervisory signals for further training. Pseudo-labels may contain noise, especially in the early stages of adaptation, therefore, we combine Domain B pseudo-labeled samples with a small labeled subset of Domain A. The inclusion of Domain A data acts as an anchor that regularizes the training process and prevents the model from drifting toward erroneous target predictions. In this way, Domain A provides a bootstrap signal that stabilizes learning while the model gradually adapts to the visual characteristics of Domain B. In addition, we employ a teacher-student setup with weak and strong augmentations to improve prediction consistency under image perturbations [4]. To mitigate the class imbalance issue, a replay buffer is used to sample the rare classes more frequently [5]. This design enables real-to-real domain transfer with reduced manual annotation effort. Instead of densely labeling Domain B from scratch, the model leverages knowledge acquired from synthetic data and Domain A, while pseudo-labels provide task-specific supervision in the target domain.

TABLE I
PRELIMINARY RESULTS

Domain Shift Level	Model	mIoU	mAcc
Moderate	Pretrained	67.37	74.84
Moderate	Domain Adaptation	71.36	82.56
Hard	Pretrained	49.57	67.60
Hard	Domain Adaptation	57.22	79.67

III. DISCUSSION

In Table I, preliminary results obtained with a slim DDRNet are reported [6]. The proposed adaptation strategy improves segmentation performance under both moderate and hard domain shift. Here, moderate shift denotes target scenes with similar floor appearance, human clothing and forklift types, whereas hard shift denotes substantial changes in floor appearance, human clothing and obstacle types. In both settings, domain adaptation leads to clear gains in mIoU and mAcc.

REFERENCES

[1] C. Schützenhöfer *et al.*, "Domain-randomized pointcloud simulation and label filtering for industrial 3d object detection," in *IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2025.
[2] W. Armstrong *et al.*, "Synthetic data for semantic image segmentation of imagery of unmanned spacecraft," in *IEEE Aerospace Conference*, 2023.
[3] A. E. B. Zekri *et al.*, "Towards using synthetic data in aerial image segmentation," in *Joint Urban Remote Sensing Event (JURSE)*, 2025.
[4] L. Yang, L. Qi *et al.*, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
[5] A. Rangnekar *et al.*, "Semantic segmentation with active semi-supervised learning," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.
[6] H. Pan *et al.*, "Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes," *IEEE Transactions on Intelligent Transportation Systems*, 2022.