Supplementary materials for the paper "SynDRA: Synthetic Dataset for Railway Applications"

Gianluca D'Amico, Federico Nesti, Giulio Rossolini, Mauro Marinoni, Salvatore Sabina, Giorgio Buttazzo
Scuola Superiore Sant'Anna, Italy

This document shall be considered as the supplementary material for the paper "SynDRA Dataset: Synthetic Segmentation Data for Railway Applications" submitted for WACV 2025. It includes statistics for the dataset (Section A), additional ablations and experiments (Section B), ethical concerns (Section C) and a few illustrations (Section D).

A. Dataset statistics

This section includes more complete statistics for the SynDRA dataset. Table 1 breaks down the number of pixels (and proportion) belonging to each class for each sequence. Please note while in the main paper and the other results in the supplementary materials the traffic-sign was omitted for compactness, here it is made explicit. The reason for this is that no traffic sign has been placed within the simulated environments, hence yielding all zeros in the statistics. Additional sequences that include traffic signs are going to be added, especially when considering the context of object detection, for which the traffic signs are of utmost importance (while for semantic segmentation it is not crucial).

Class	**Scenario 1** (2318 images)	**Scenario 2** (5683 images)	**Scenario 3** (2048 images)	**Scenario 4** (4288 images)	**Total** (13923 images)
sky	1378067842 [0.2867%]	3135553764 [0.2661%]	1282805614 [0.3021%]	2515389346 [0.2829%]	8311816566 [0.2796%]
road	35083073 [0.0073%]	87214502 [0.0074%]	89344250 [0.0210%]	39278624 [0.0044%]	250920449 [0.0084%]
sidewalk	26092838 [0.0054%]	54403633 [0.0046%]	96525180 [0.0227%]	23212845 [0.0026%]	200234496 [0.0067%]
construction	66851307 [0.0139%]	210688447 [0.0179%]	197159476 [0.0464%]	231192140 [0.0260%]	705891370 [0.0237%]
fence	0 [0.0000%]	10975458 [0.0009%]	2924403 [0.0007%]	54449571 [0.0061%]	68349432 [0.0023%]
pole	72812497 [0.0151%]	158342626 [0.0134%]	35665433 [0.0084%]	124315103 [0.0140%]	391135659 [0.0132%]
traffic-light	703659 [0.0001%]	1426205 [0.0001%]	0 [0.0000%]	0 [0.0000%]	2129864 [0.0001%]
traffic-sign	0 [0.0000%]	0 [0.0000%]	0 [0.0000%]	0 [0.0000%]	0 [0.0000%]
vegetation	656809620 [0.1366%]	1420008181 [0.1205%]	744324612 [0.1753%]	588230286 [0.0662%]	3409372699 [0.1147%]
terrain	1806350742 [0.3758%]	3744836826 [0.3178%]	1442711421 [0.3397%]	3894055398 [0.4379%]	10887954387 [0.3662%]
human	106049 [0.0000%]	261284 [0.0000%]	0 [0.0000%]	802227 [0.0001%]	1169560 [0.0000%]
ballast-ext-other	107217922 [0.0223%]	296727413 [0.0252%]	0 [0.0000%]	212083298 [0.0239%]	616028633 [0.0207%]
sleeper-ext-other	75060988 [0.0156%]	182047849 [0.0154%]	0 [0.0000%]	152125831 [0.0171%]	409234668 [0.0138%]
ballast-int-other	15767343 [0.0033%]	34664431 [0.0029%]	0 [0.0000%]	29097212 [0.0033%]	79528986 [0.0027%]
sleeper-int-other	100580234 [0.0209%]	229215932 [0.0195%]	0 [0.0000%]	194948415 [0.0219%]	524744581 [0.0177%]
railraised-other	53096553 [0.0110%]	143939209 [0.0122%]	0 [0.0000%]	118835263 [0.0134%]	315871025 [0.0106%]
ballast-ext-main	121019391 [0.0252%]	325102489 [0.0276%]	113130643 [0.0266%]	217850084 [0.0245%]	777102607 [0.0261%]
sleeper-ext-main	84304141 [0.0175%]	206624633 [0.0175%]	73650032 [0.0173%]	152773693 [0.0172%]	517352499 [0.0174%]
ballast-int-main	19372773 [0.0040%]	47240434 [0.0040%]	16383167 [0.0039%]	34459182 [0.0039%]	117455556 [0.0040%]
sleeper-int-main	115685420 [0.0241%]	283571695 [0.0241%]	101265107 [0.0238%]	211266845 [0.0238%]	711789067 [0.0239%]
railraised-main	26453832 [0.0055%]	64206919 [0.0054%]	23017662 [0.0054%]	49182131 [0.0055%]	162860544 [0.0055%]
car	1222737 [0.0003%]	3044831 [0.0003%]	5730828 [0.0013%]	5946368 [0.0007%]	15944764 [0.0005%]
truck	0 [0.0000%]	460723 [0.0000%]	13719323 [0.0032%]	120365 [0.0000%]	14300411 [0.0005%]
onrail	4991224 [0.0010%]	14790583 [0.0013%]	0 [0.0000%]	0 [0.0000%]	19781807 [0.0007%]
unlabelled	38954615 [0.0081%]	1128920733 [0.0958%]	8375649 [0.0020%]	41982573 [0.0047%]	1218233570 [0.0410%]

Table 1. Number of pixels and proportion (with respect to each column) for each class, for each of the 4 sequences of the dataset.

Figure 9. RGB images (left) and binary mask (right) for SynDRA (top), RailSem19 (middle) and Osdar23 (bottom).

B. Ablations and Experiments

B.1. Labeling for Ego Track Discrimination

For the task of ego track discrimination, no dedicated binary masks are available for the three datasets used, i.e., SynDRA, RailSem19 and Osdar23. However, SynDRA masks are easily computed by considering the pixels belonging to classes sleepers-int-main, ballast-int-main, and railraised-main. Conversely, RailSem19 and Osdar23 do not specify which of the annotated tracks is the main one, hence, a heuristic algorithm is required to find it.

For both datasets, the tracks are annotated with a sequence of points defining the left and right rail polyline. These points can be concatenated and used as the input contour of the following OpenCV's function for each track.

```
ImageDraw.Draw(mask_img, 'L').polygon(contour_list, fill=(255))
```

While Osdar23 is structured enough to make it sufficient to select the most central mask as the ego track, RailSem19 includes images with wildly varying viewing angles, which requires a careful manual selection of the correct track. Figure 9 shows a sample from each dataset.

The actual masks used in all the training experiments can be provided upon written request to the authors.

B.2. How many SynDRA samples are enough?

In the experiments of the main paper the number of synthetic samples is 200, 50 for each sequence. This choice is motivated by the result we obtained training ego track discrimination model on SynDRA samples only, depicted in Figure 10. Using more samples is beneficial, but it is important to note that there is not a wide variation of performance, as using 1000 samples (5× with respect to 200) yields less than 1% overall accuracy, and less than 5% mIoU. Furthermore, more samples increase training time, and, most importantly, it is crucial to be cautious when mixing two datasets with different distributions. Since the test dataset has a real-world distribution, it is important not to skew the training set too much towards the synthetic distribution. We found that 200 samples is a good balance of improved overall performance and additional training time.

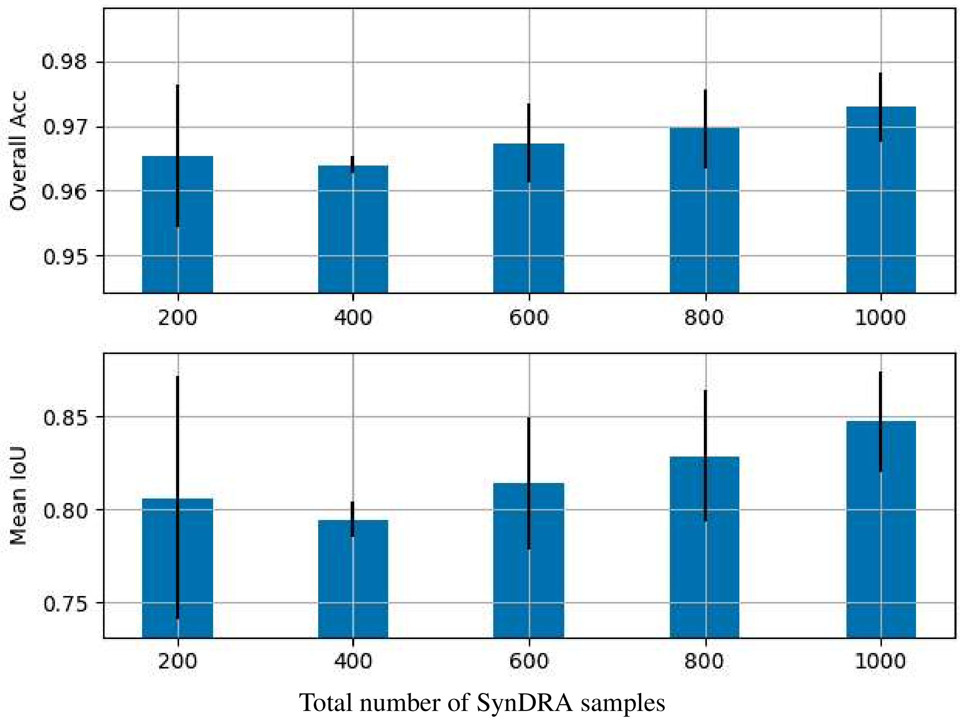


Figure 10. Results on Osdar23 for models trained only on SynDRA samples.

B.3. Additional experiments

Figure 11 depicts additional metrics from the same experiments of Section 4.2 in the main paper. Specifically, it shows overall accuracy, precision and F1 score as a function of the number of RailSem19 samples used in the training process. While precision values are much more noisy than the other metrics, both metrics show with a certain consistency the benefit of using additional synthetic images during training.

B.3.1 Results on Semantic Segmentation Finetuning

To extend the results reported in Section 4.1, we increased the range of settings by varying the number of training RailSem19 samples to 0, 15, 25, 50, 100, and 250. As shown in Figure 12, the trend described in Section 4.1 is confirmed: the inclusion of SynDRA samples is beneficial only when there are few RailSem19 samples (i.e., 15, 25, plots also reported in the main paper). However, when starting with 50 RailSem19 samples, the IoU scores of different classes become similar with or without the introduction of SynDRA. Additionally, we evaluated the scenario without any RailSem19 samples (i.e., 0). In this case, using only SynDRA data results in lower, but not zero, scores for different classes, showing a high risk of overfitting due to simulated textures. Despite the vast size and numerous images in our synthetic dataset, it cannot match the variance of a real-world dataset like RailSem19. We plan to address this issue in future research to contribute to establishing the use

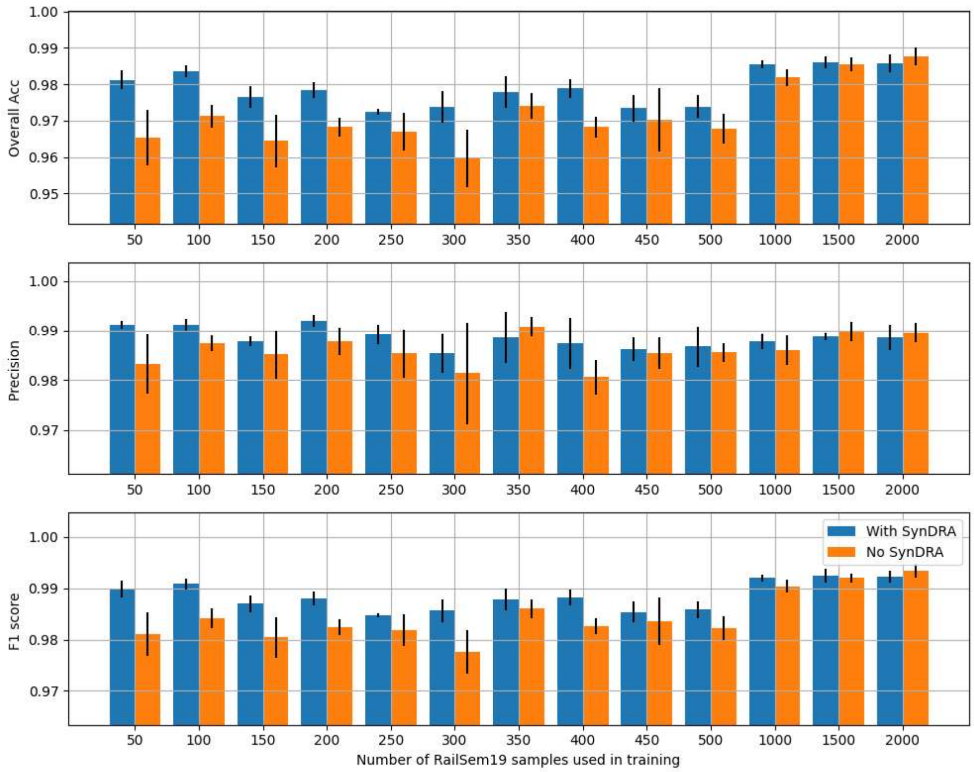


Figure 11. Ego track discrimination results in terms of overall accuracy, precision and F1 score (for binary classification).

of simulated environments in railway computer vision applications. For completeness, we also reported the overall accuracy of the same tests as an additional metric in Figure 13.

B.3.2 Impact of synthetic sample quantity in semantic segmentation

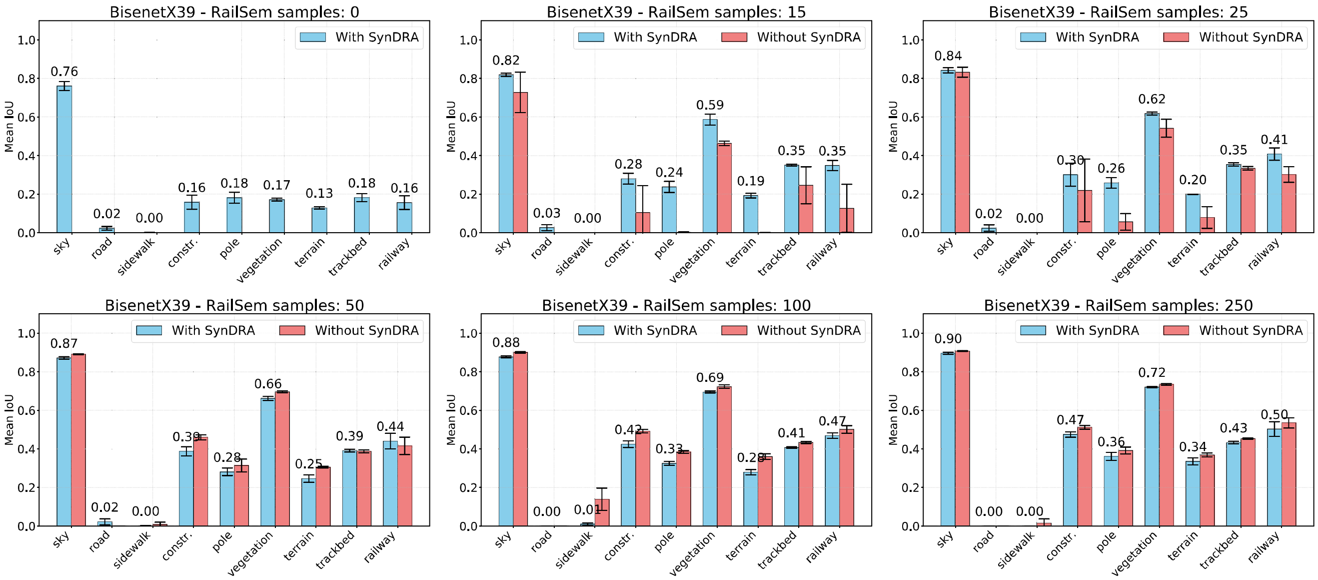
We conducted also ablation studies to understand the impact of varying the number of synthetic samples from SynDRA during the fine-tuning process, while keeping the number of samples from the real-world dataset, Railsem19, fixed. As shown in Figure 14, we recognize that using a small number of synthetic samples per scenario yields good benefits in terms of MIoU and accuracy, where the number of RailSem19 samples was fixed at 25. Beyond 50 SynDRA samples, the effectiveness of adding more synthetic samples diminishes and can even lead to overfitting for certain classes (e.g., vegetation, railway, and trackbed) when considering larger quantities, such as 100 and 250. This is likely because an excessive number of samples from simulated scenarios can cause the model to learn features too specific to the synthetic environment, even under the use of augmentation techniques.

C. Ethical considerations

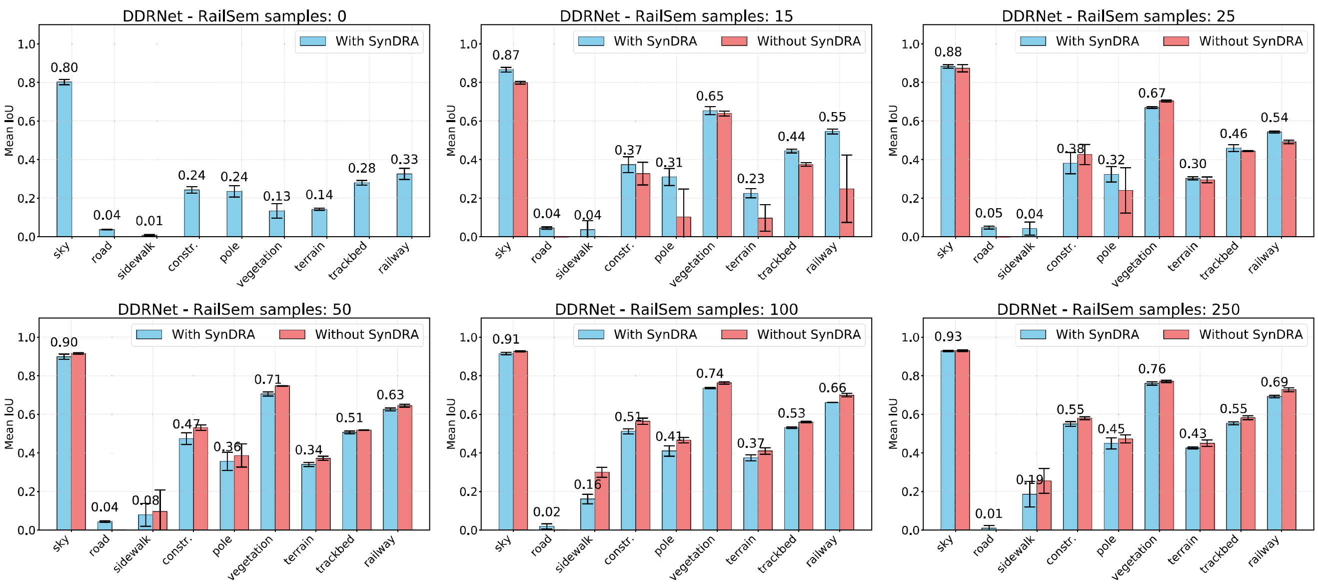
The environments in the four scenarios were generated using a variety of data sources and assets, including geospatial data from OpenStreetMap (OSM) and BayernAtlas, as well as 3D models and assets from Megascans, TurboSquid, Sketchfab, and plugins from the Epic Games Marketplace. This section addresses the ethical, privacy, and copyright considerations associated with the use of these resources.

OpenStreetMap (OSM): The data from OSM is available under the Open Database License (ODbL), which permits sharing, modifying, and using the data provided that OSM is credited. This allows for the incorporation of OSM data in our scenarios in compliance with the community's licensing requirements.

BayernAtlas: Data from BayernAtlas is governed by the usage terms set by the Bavarian State Office for Survey and Geoinformation (LDBV). This data can be used for non-commercial and educational purposes, as long as proper attribution is provided, following the CC BY-ND 4.0 license.



(a) Finetuning results for Bisenet.

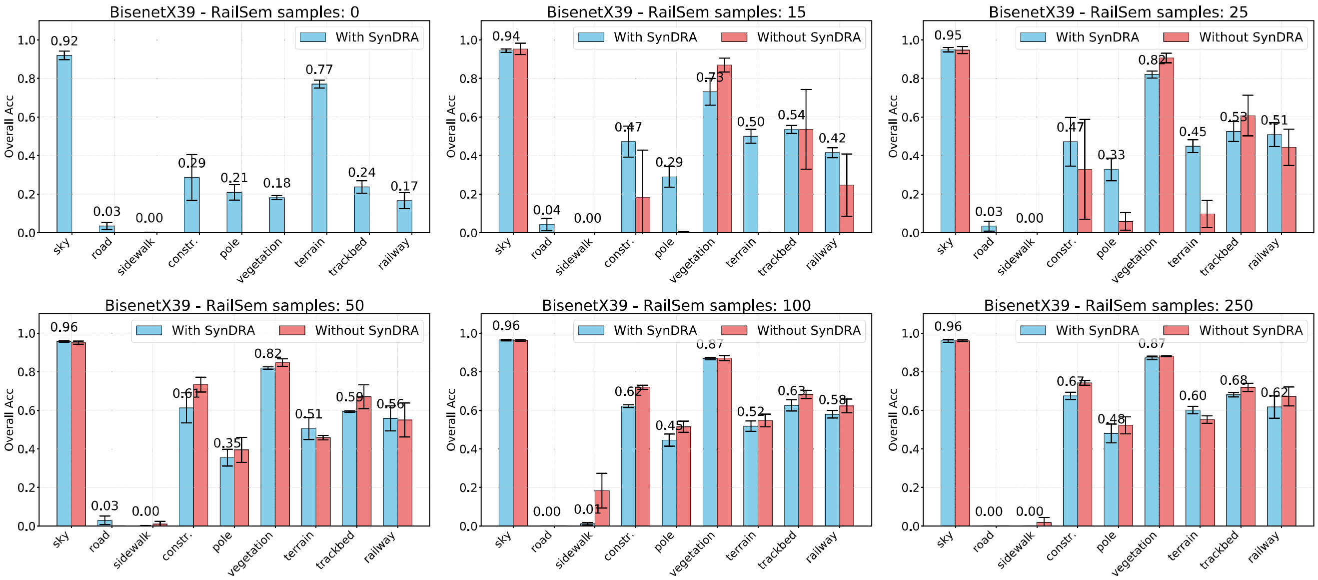


(b) Finetuning results for DDRNet23Slim.

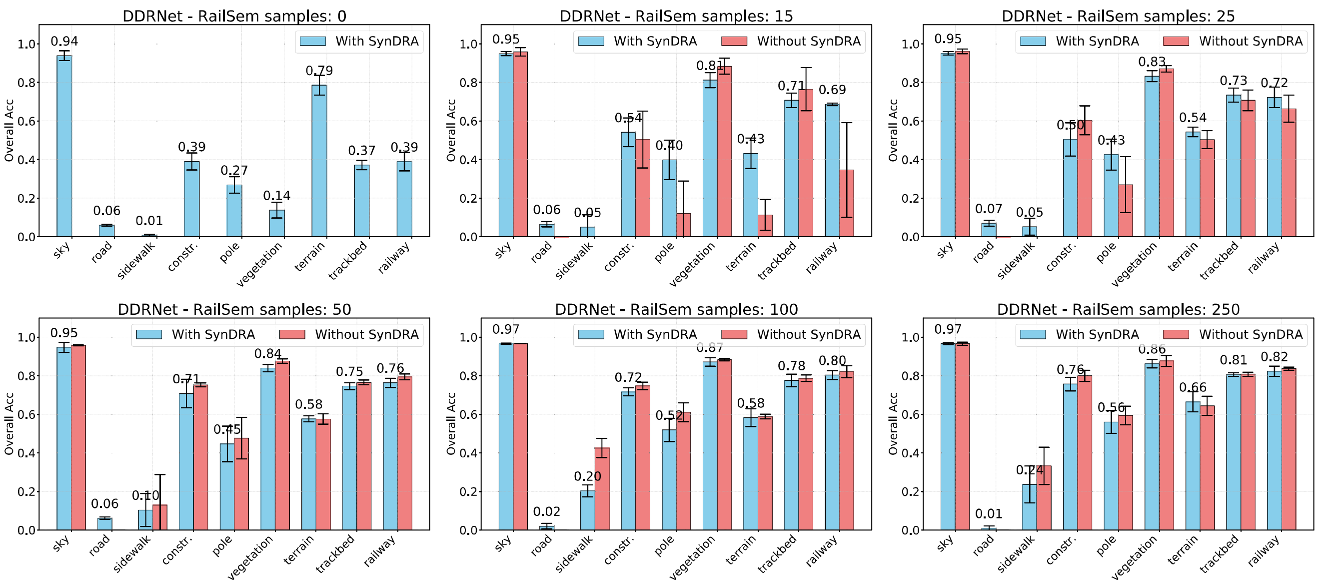
Figure 12. Finetuning MIoU results for Bisenet (a) and DDRNet (b) using different amounts of RailSem samples (0, 15, 25, 50, 100, 250) with (blue) and without (red) SynDRA samples. The tests were conducted over 3 runs, with the variation shown by error bars. In these runs, the selection of samples from SynDRA and RailSem, as well as the seeds for applying data augmentation transformations, were randomized.

Megascans by Quixel (Epic Games): The Megascans assets, available through Epic Games, are licensed for use under the Epic Games End User License Agreement (EULA). These assets provide high-quality textures and 3D models that enhance the realism of the synthetic environments. The use of these assets is compliant with the EULA, which allows their use in projects associated with UE5.

3D Models from TurboSquid and Sketchfab: Some 3D models used in the scenarios were sourced from TurboSquid and Sketchfab, platforms that offer assets under various licenses. Each model was selected and used in accordance with its specific license agreement, ensuring legal and ethical usage. Attribution was provided where required, and models were used



(a) Finetuning results for Bisenet.



(b) Finetuning results for DDRNet23Slim.

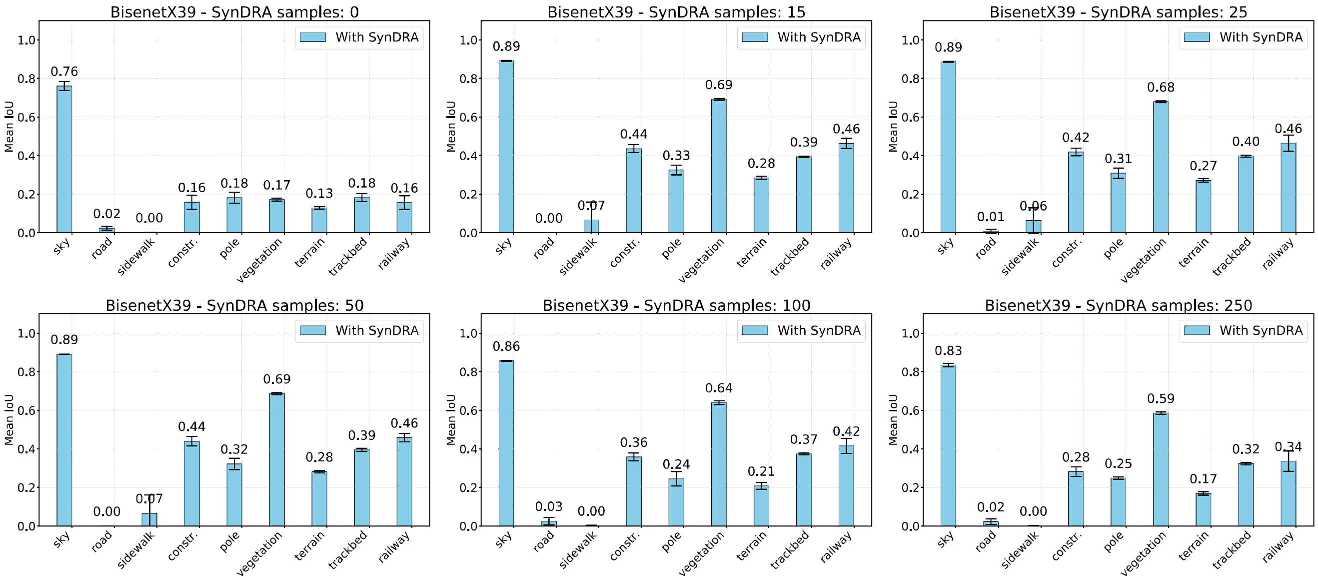
Figure 13. Finetuning Overall Accuracy results for Bisenet (a) and DDRNet (b) using different amounts of RailSem samples (0, 15, 25, 50, 100, 250) with (blue) and without (red) SynDRA samples. The results corresponds to the same runs of Figure 12.

within the scope permitted by their creators.

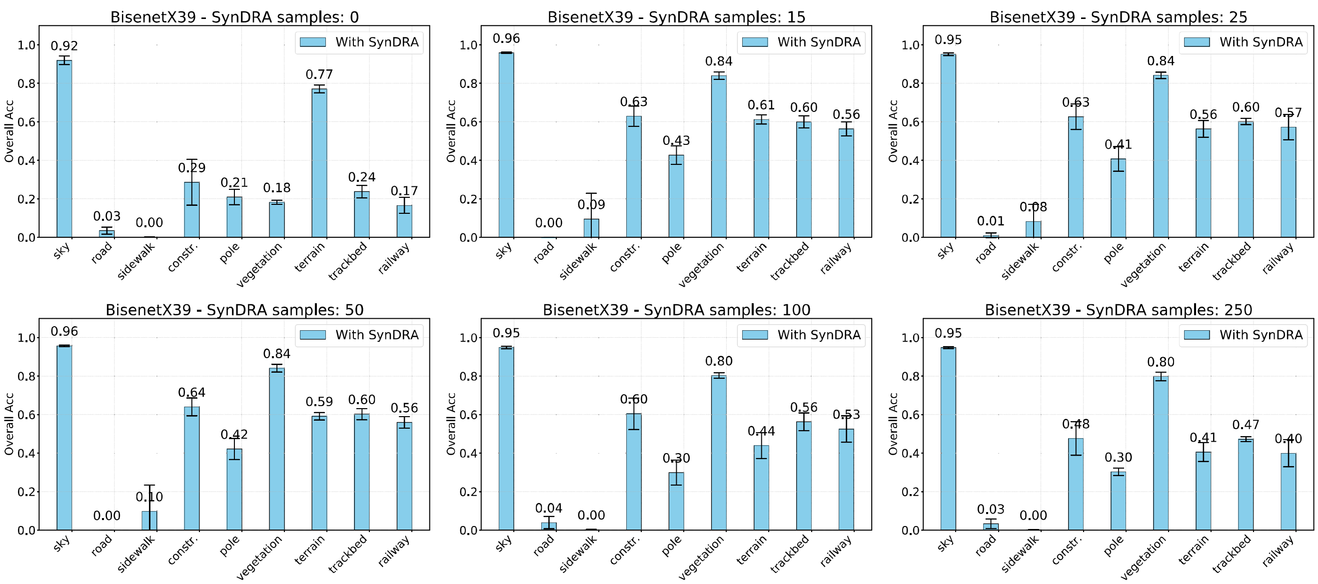
Plugins from the Epic Games Marketplace, in particular TrainTemplate and RailwaySystem: various plugins from the Epic Games Marketplace were employed to enhance the functionality and visual fidelity of the scenarios. These plugins were used in accordance with their respective licenses, as provided by the creators and governed by the Epic Games Marketplace terms of service.

D. Additional Illustrations

Figure 16 shows a frame from Scenario 1 in different lighting and weather conditions. Figure 17 illustrates the structure of the dataset folder.



(a) MIoU results.



(b) Acc.

Figure 14. Ablation studies performed on the BisenetX39 model, analyzing the impact of the number of synthetic samples per scenario. The plot shows the results of fine-tuning in terms of MIoU and overall accuracy, with 25 samples from RailSem kept constant, while varying the number of synthetic samples from SynDRA.

Figure 18 compares the virtual landscape generated in UE5 using the BayernAtlas heightmap and its real-world counterpart captured from Google Earth.

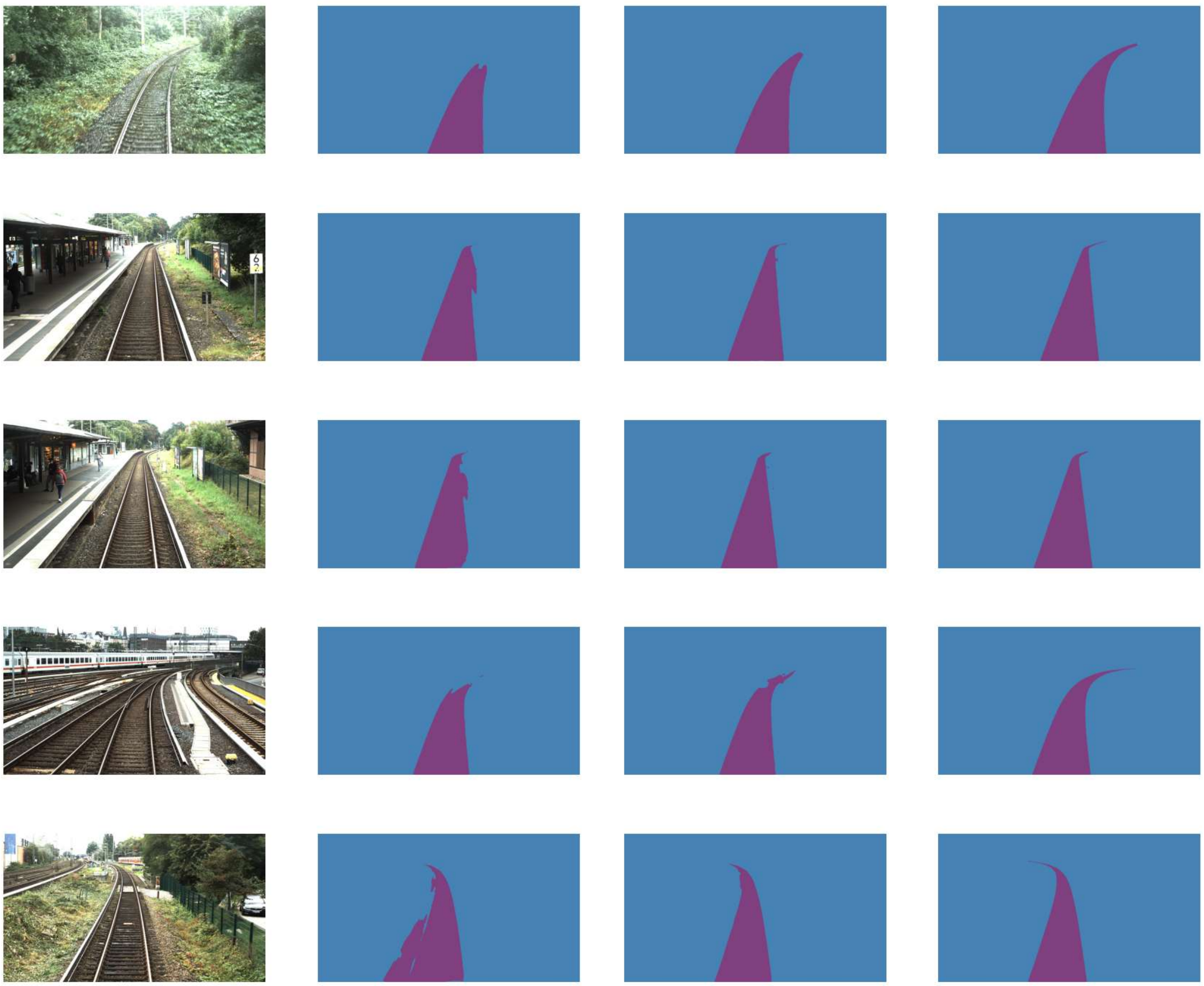


Figure 15. From left to right: Osdar23 RGB image, prediction from model trained without SynDRA, prediction from model trained with SynDRA, ground truth. Both models are trained with 1000 samples from Railsem19.

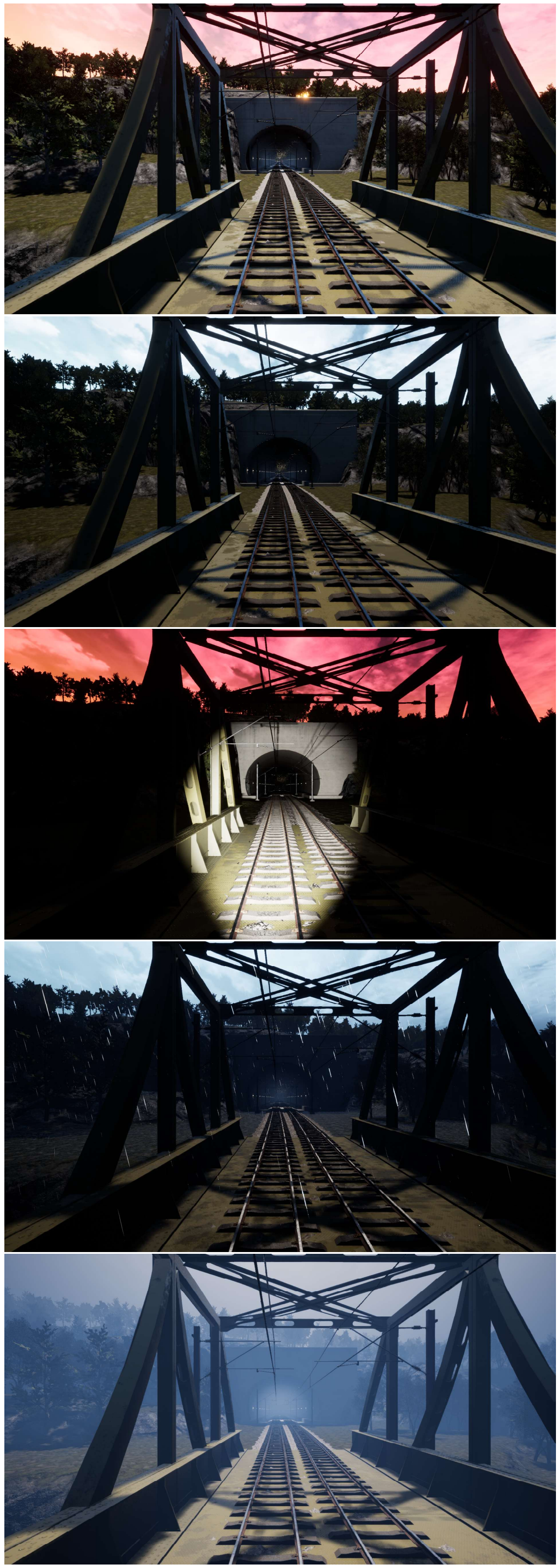


Figure 16. Comparison of the different lighting and weather conditions in SynDRA, taken from Scenario 1. From top to bottom: morning, afternoon, evening, rain, fog.

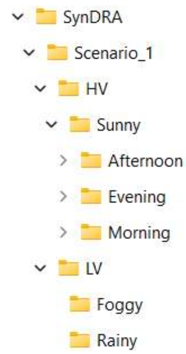
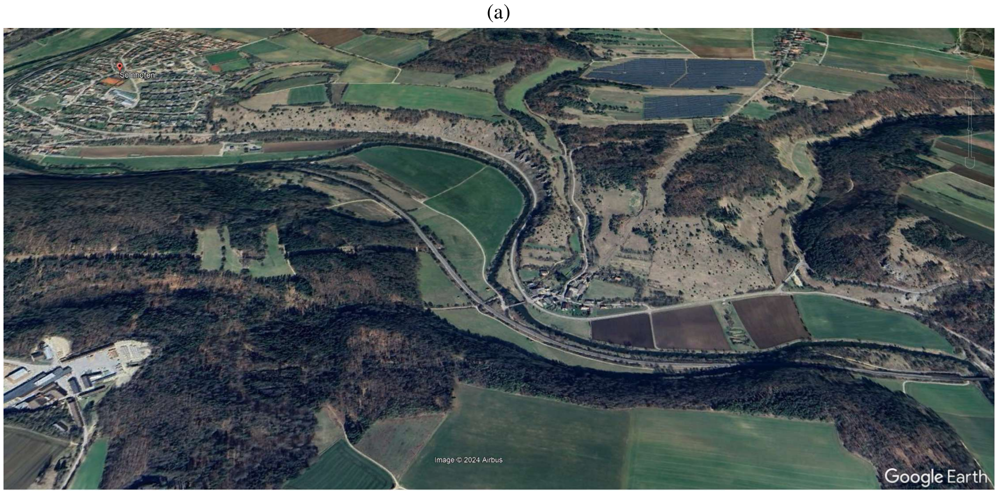


Figure 17. SynDRA folder structure.

(a)



(b)

Figure 18. Comparison between the virtual terrain generated in UE5 via the BayernAtlas web application (a) and its real-world counterpart captured from Google Earth (b).