Enhancing Rural Autonomous Driving Performance with Diffusion-Augmented Synthetic Datasets

Siddharth Arun
siddhartharun306@gmail.com

Trisha Panchangmath
trishapan2411@gmail.com

Saanvi Celamkoti
saanvicelamkoti7@gmail.com

Vayden Wong
vaydenwong2@gmail.com

Charles Duong*

Vasu Sharma*

Sean O'Brien*

Kevin Zhu*

Abstract

Synthetic datasets are increasingly used to train autonomous vehicle (AV) models, providing large-scale, diverse, and realistic data for perception and decision-making tasks. However, their predominant focus on urban environments limits effectiveness in rural areas, creating potential safety risks, while collecting real-world rural driving data remains time-consuming and costly. In this work, we leverage diffusion models to enhance the realism of synthetic driving data, focusing on features critical to rural navigating such as curves, hills, and varied terrain. Quantitative metrics and qualitative evaluations demonstrate that diffusion-enhanced datasets improve the robustness and reliability of AV models in underrepresented rural scenarios. Our results show statistically significant improvements over both heuristic baselines and real-world trained models, highlighting the scalability of diffusion-based synthetic data to cover rare but critical driving situations. [1]

1 Introduction

Autonomous driving research relies on large, diverse datasets and realistic simulators to advance perception systems [14]. However, collecting and labeling real-world data is expensive, time-consuming, and often biased toward urban environments, leaving rural and rare scenarios underrepresented. Models trained solely on such limited datasets may fail to generalize in safety-critical situations [21]. Synthetic datasets offer a potential solution, but traditional approaches often lack realism and scenario diversity. Recent surveys emphasize the need for more scalable synthetic data to have better robustness and safety in autonomous driving [20, 31, 5].

Recent advances in generative AI, particularly diffusion-based methods, have significantly improved the realism and diversity of simulated traffic scenarios [16, 23]. Latent diffusion models lead to scalable generation of photorealistic, rare, and dangerous driving scenarios, addressing gaps in existing datasets [4, 15, 19, 28, 11]. Simulator-conditioned approaches further ensure alignment with realistic vehicle and environment dynamics [34]. Together, these methods provide high-quality synthetic data that can improve safety, generalization, and model robustness in autonomous driving

*Equal Senior Authorship

[1] Source Code: https://github.com/siddarun/RuralGen

39th Conference on Neural Information Processing Systems Workshop on Harnessing Artificial Intelligence for Smart Cities (UrbanAI).

systems. In this work, we make two main contributions. First, we generate synthetic data specifically targeting rural driving environments, addressing the under-representation of these scenarios. Second, we train and evaluate downstream autonomous vehicle models using the CARLA simulator [3], demonstrating improved performance and robustness in realistic driving conditions.

2 Related Works

High-quality datasets are critical for autonomous driving, yet real-world data collection is expensive, time-consuming, and biased toward urban environments, leaving rural and rare scenarios underrepresented [21, 27, 14, 20, 22, 7]. Generative modeling has emerged as a powerful solution, improving scenario realism, diversity, and controllability. Yu et al. (2025) [28] introduced a multi-guided diffusion framework with direct preference optimization to generate diverse, realistic traffic scenarios. Harvey et al. (2022) [6] demonstrated a flexible diffusion approach producing temporally coherent long videos, while Ho et al. (2022) [9] extended text-to-image diffusion to high-definition video synthesis with strong temporal consistency. In driving simulations, Kim et al. (2021) [13] developed DriveGAN, a controllable generative model producing high-quality, editable virtual environments. Pronovost et al. (2023) [18] proposed Scenario Diffusion, balancing realism and diversity for improved scenario coverage. Latent diffusion methods such as Syndiff-AD [4], DiffAD [24], and DiffusionDrive [15] advanced end-to-end AV data generation, while Epona [29] used auto-regressive diffusion models for complex traffic scenarios.

Simulator-conditioned approaches allowed for enhancement of the realism by aligning synthetic data with real-world dynamics. SimGen [34] and Waymax [5] offered scalable, high-fidelity scenario generation, whereas GenAD [32], OpenDriveVLA [33], and VLM-AD [26] integrated multimodal supervision to improve fidelity. ST-P3 [10] and Senna [12] leveraged spatial-temporal and vision-language learning to handle dynamic environments effectively. Diffusion-based methods also enhance dataset augmentation and high-resolution synthesis, as shown by Rombach et al. (2021) [19], Chen et al. (2024) [2], Lu et al. (2024) [16], and Wang et al. (2025) [23], with additional studies [31, 35, 11] demonstrating scalable generation of rare and safety-critical scenarios.

Although prior work has demonstrated advances in the use of synthetic data for autonomous driving and improving AV model performance, most of these works focus on urban scenarios, leaving rural driving settings underrepresented. Our work generates synthetic rural driving data and evaluates its impact on downstream AV models using the CARLA simulator [3], addressing the persistent under-representation of rural and safety-critical scenarios.

3 Methodology

In this paper, we introduce RuralGen, an enhanced Stable Diffusion XL (SDXL) model[17][1]. We employ a diffusion model designed to preserve structural details such as depth and edges to enhance CARLA simulation data. Three training datasets were constructed for evaluation: CARLA-only synthetic data, diffusion-enhanced synthetic CARLA data, and real-world data from the BDD100K dataset. To ensure a fair and unbiased comparison, all datasets were evaluated using the same network architectures and training protocols, enabling a clear assessment of the impact of diffusion-based augmentation on autonomous driving performance.

3.1 Dataset Generation & Augmentation

The baseline CARLA dataset consists of simulated images captured in various environments, including both rural and urban settings. To create the diffusion-enhanced dataset, these CARLA images are processed through a pretrained SDXL model, producing photorealistic images while preserving the structural features of the original dataset, such as edge density and depth perception. For evaluation, we compare the performance of perception control networks trained in the diffusion-enhanced dataset against those trained in real-world images from the BDD100K dataset, which provides a broad and diverse collection of driving scenarios.

3.2 Perception-Control Network Architecture

The experiments use a unified CNN-based perception-control network. Preprocessed RGB images are fed into a convolutional backbone to extract features that are then mapped to the steering, throttle, and brake output via fully connected layers. All datasets, including baseline CARLA, synthetic diffusion data, and real-world BDD100K, use the same architecture and hyper parameters, ensuring a fair evaluation of diffusion-based data augmentation.

3.3 Diffusion-Based Dataset Generation

To produce photorealistic rural road scenes under diverse environmental conditions, the SDXL model was tuned with both positive and negative prompts (over 1000 unique prompt combinations) that combined base scenarios with road conditions, lighting conditions, weather variations, and quality enhancers (our prompts can be seen in Appendix A.4) to generate 20,000 unique synthetic images at 1024×1024 resolution (sample images can be found in Figures 30,31,32). Realism was further improved through automated post-processing and quality filtering that evaluated image quality and rejected images that did not match the negative prompts (a detailed description of our postprocessing methodology can be found in Appendix A.2). Metrics such as FID [8], SSIM [25], and LPIPS [30] are calculated to assess fidelity relative to real-world datasets. The diffusion-enhanced images, generated in batches of over 10,000, are then used to train the perception-control network.

3.4 Real-World Dataset Integration

A subset of BDD100K with rural driving scenes is filtered and preprocessed (normalized and resized) to match the CNN input format. This real-world dataset serves as a benchmark to compare networks trained on baseline CARLA and SDXL diffusion-enhanced images. We chose BDD100K despite its urban bias because it reflects the real problem we're trying to solve. Like most autonomous driving datasets, BDD100K has limited rural coverage. This scarcity is exactly why synthetic data matters. By comparing against BDD100K's sparse rural samples, we demonstrate whether SDXL can fill the rural data gap that exists in real-world datasets, showing practical value for researchers facing the same data limitations.

4 Experimental Setup

4.1 Setup

We evaluated RuralGen by comparing its performance with a heuristic-based model included with CARLA and a real-world-data model (BDD100K). Experiments span multiple **rural driving scenarios** with varied weather, lighting, and maps. Models are assessed using safety, control, and efficiency metrics to compare performance across data sources and scales. A detailed explanation of our development environment can be found in Appendix A.1.

4.2 Driving Model Training on Rural Scenarios

Using 20,000 RuralGen images, we trained perception-control networks in CARLA and evaluated them on five **rural scenarios**: highway consistency, curves and hills, wet weather, night, and dawn conditions. Training used RGB, depth, and semantic segmentation data collected from CARLA's autopilot. A CNN-based imitation learning model processed 1024×1024 RGB images through convolutional layers, followed by fully connected layers outputting steering, throttle, and brake commands. Trained models were deployed in autonomous mode on predefined **rural routes**, assessing safety (collisions, lane invasions), efficiency (distance traveled, average speed), and control quality (steering smoothness).

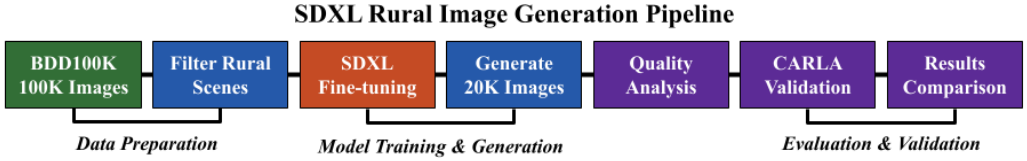

Figure 1: Data Pipeline for Training, Image generation and testing in CARLA AV Simulator. The data pipeline filters 10K BDD100K images to extract rural driving scenes, fine-tunes SDXL to generate 20K high-quality synthetic rural images, and validates their effectiveness through comprehensive quality analysis and CARLA autonomous driving simulation.

5 Results

5.1 Image Quality Analysis

RuralGen produced realistic rural driving images that are nearly indistinguishable from real photos to the naked eye. The FID score of RuralGen images compared with BDD100K, in rural scenarios was 7.1 (A detailed view of the image advanced quality metrics can be found in Figure 24), suggesting good match with BDD100K datasets. The synthetic images perfectly capture the lighting, colors, and visual characteristics of rural roads, making them highly effective for training self-driving car systems. This collection of 20,000 high-quality synthetic images provides excellent training data that closely mirrors real-world rural driving conditions. However, the LPIPS, SSIM and Dice scores suggested that there was still room for improvement with structural details and human perception differences. Descriptions of the metrics used can be found in A.3

5.2 Performance Analysis

RuralGen consistently outperforms both comparison models approaches across all evaluation metrics. Compared to the base model, it achieves a 22.5-point improvement (77.9% vs 55.5%, Cohen's $d = 3.35$), and against the BDD100K-trained model, it improves by 29.2 points (77.9% vs 48.8%, Cohen's $d = 2.23$), with all differences statistically significant ($p < 0.001$).

Performance gains are particularly pronounced in challenging scenarios: the model achieves 81.7% in night driving compared to 34.1% (BDD100K) and 56.9% (base), 85.8% in wet weather versus 35.2% and 58.1%, and 79.2% on curves and hills relative to 46.9% and 55.2%, respectively. Safety metrics further highlight the advantages of the synthetic approach, showing 12.0% fewer collisions and 12.7% fewer lane departures than the heuristic baseline, while maintaining 26.1% higher average speed and covering 35.2% greater distance.O

Table 1: Performance Summary by Model

Metric	Heuristic	**Synthetic**	BDD100K
Mean Perf	55.5	**77.9**	48.8
Std Perf	5.2	**8.0**	16.7
Mean Coll	7014	**6923**	7218
Mean Lane Inv	2.0	**4.0**	4.0
Mean Steering	92.8	**94.3**	94.9
Tests	22	**22**	20

Table 2: Average Performance by Scenario and Model

Scenario	Heuristic	**Synthetic**	BDD100K
Highway	54.1	**69.7**	76.4
Curves	55.2	**79.2**	46.9
Wet	58.1	**85.8**	35.2
Night	56.9	**81.7**	34.1
Dawn	53.7	**77.4**	51.3

5.3 Scaling Law Analysis

We conducted a comprehensive scaling study comparing different sizes of our synthetic dataset against varying sizes of real-world data. As shown in Figure 23, RuralGen scales more effectively than real-world alternatives. For example, 5K synthetic samples achieve 76.1% performance compared to 47.0% for 10K BDD100K samples, 10K synthetic samples reach 77.0% versus 48.1% for 5K BDD100K, and 20K synthetic samples attain 82.3% compared to 51.9% for 10K BDD100K.This

behavior demonstrates that RuralGen provides a scalable solution, where increasing the size of the synthetic dataset consistently improves performance, whereas real-world data collection suffers from diminishing returns and higher costs.

6 Limitations and Conclusion

Limitations Although the BDD100K dataset provided valuable real-world grounding, its geographic and temporal coverage is limited, which constrains adaptability across diverse traffic cultures and infrastructural conditions. Additionally, our experiments were restricted to a relatively small set of rural and urban driving scenarios, meaning performance may differ in more complex environments such as dense urban intersections or other edge cases. Furthermore, while RuralGen enhances data scalability, fine structural details and perceptual nuances in the synthetic images remain areas that require further refinement.

While this work focuses on pure rural scenarios as a proof-of-concept for synthetic data generation in underrepresented driving environments, the modular prompt engineering framework is designed for straightforward extension to broader contexts. The current 1,000 prompt combinations can be augmented with rural-urban transition elements (approaching small towns, highway exits, suburban boundaries) and global diversity factors (unpaved roads, tropical highways, varied road markings and signage) by simply adding new base scenarios and regional characteristics to the existing architecture. The CARLA validation framework similarly supports additional maps and scenarios without structural changes. Future work will leverage this extensibility to systematically expand coverage to transition zones and global road diversity, building on the foundation established here where rural data scarcity is most acute.

Conclusion Despite these limitations, RuralGen demonstrates strong potential for improving autonomous driving performance, particularly in underrepresented rural environments where real-world data is scarce. By scaling synthetic data effectively, our approach enables coverage of rare but critical driving scenarios, resulting in statistically significant improvements over both heuristic baselines and real-world trained models. Future work will aim to enhance image realism, broaden scenario diversity, and validate performance with real-world autonomous systems to maximize safety, efficiency, and operational impact.

References

[1] Stability AI. Stable diffusion xl base 1.0, 2023. URL https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0.

[2] Dong Chen, Xinda Qi, Yu Zheng, Yuzhen Lu, Yanbo Huang, and Zhaojian Li. Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition. *Computers and Electronics in Agriculture*, 216:108517, 2024.

[3] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator, 2017. URL https://arxiv.org/abs/1711.03938.

[4] Harsh Goel, Sai Shankar Narasimhan, Oguzhan Akcin, and Sandeep Chinchali. Syndiff-ad: Improving semantic segmentation and end-to-end autonomous driving with synthetic data from latent diffusion models. *arXiv preprint arXiv:2411.16776*, 2024.

[5] Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, et al. Waymax: An accelerated, data-driven simulator for large-scale autonomous driving research. *Advances in Neural Information Processing Systems*, 36:7730–7742, 2023.

[6] William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, and Frank Wood. Flexible diffusion modeling of long videos. *Advances in neural information processing systems*, 35:27953–27965, 2022.

[7] S Hausler and M Milford. P1–021: Map creation. *Monitoring and Maintenance for Automated Driving–Literature Review*, 2021.

[8] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018. URL https://arxiv.org/abs/1706.08500.

[9] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. Imagen video: high definition video generation with diffusion models (2022). *arXiv preprint arXiv:2210.02303*, 3, 2022.

[10] Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision*, pages 533–549. Springer, 2022.

[11] Zhiyu Huang, Zixu Zhang, Ameya Vaidya, Yuxiao Chen, Chen Lv, and Jaime Fernández Fisac. Versatile scene-consistent traffic scenario generation as optimization with diffusion. *arXiv preprint arXiv:2404.02524*, 3, 2024.

[12] Bo Jiang, Shaoyu Chen, Bencheng Liao, Xingyu Zhang, Wei Yin, Qian Zhang, Chang Huang, Wenyu Liu, and Xinggang Wang. Senna: Bridging large vision-language models and end-to-end autonomous driving. *arXiv preprint arXiv:2410.22313*, 2024.

[13] Seung Wook Kim, Jonah Philion, Antonio Torralba, and Sanja Fidler. Drivegan: Towards a controllable high-quality neural simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5820–5829, 2021.

[14] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. V2x-sim: Multi-agent collaborative perception dataset and benchmark for autonomous driving. *IEEE Robotics and Automation Letters*, 7(4):10914–10921, 2022.

[15] Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, et al. Diffusiondrive: Truncated diffusion model for end-to-end autonomous driving. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 12037–12047, 2025.

[16] Jinxiong Lu, Shoaib Azam, Gokhan Alcan, and Ville Kyrki. Data-driven diffusion models for enhancing safety in autonomous vehicle traffic simulations. *arXiv preprint arXiv:2410.04809*, 2024.

[17] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023.

[18] Ethan Pronovost, Meghana Reddy Ganesina, Noureldin Hendy, Zeyu Wang, Andres Morales, Kai Wang, and Nick Roy. Scenario diffusion: Controllable driving scenario generation with diffusion. *Advances in Neural Information Processing Systems*, 36:68873–68894, 2023.

[19] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. arxiv 2022. *arXiv preprint arXiv:2112.10752*, 2021.

[20] Zhihang Song, Zimin He, Xingyu Li, Qiming Ma, Ruibo Ming, Zhiqi Mao, Huaxin Pei, Lihui Peng, Jianming Hu, Danya Yao, et al. Synthetic datasets for autonomous driving: A survey. *IEEE Transactions on Intelligent Vehicles*, 9(1):1847–1864, 2023.

[21] Ho Suk, Yerin Lee, Taewoo Kim, and Shiho Kim. Addressing uncertainty challenges for autonomous driving in real-world environments. In *Advances in computers*, volume 134, pages 317–361. Elsevier, 2024.

[22] Paolo Visconti, Giuseppe Rausa, Carolina Del-Valle-Soto, Ramiro Velázquez, Donato Cafagna, and Roberto De Fazio. Innovative driver monitoring systems and on-board-vehicle devices in a smart-road scenario based on the internet of vehicle paradigm: A literature and commercial solutions overview. *Sensors*, 25(2):562, 2025.

[23] Juanran Wang, Marc R Schlichting, Harrison Delecki, and Mykel J Kochenderfer. Diffusion models for safety validation of autonomous driving systems. *arXiv preprint arXiv:2506.08459*, 2025.

[24] Tao Wang, Cong Zhang, Xingguang Qu, Kun Li, Weiwei Liu, and Chang Huang. Diffad: A unified diffusion modeling approach for autonomous driving. *arXiv preprint arXiv:2503.12170*, 2025.

[25] Zhou Wang, Alan Bovik, Hamid Sheikh, and Eero Simoncelli. Image quality assessment: From error visibility to structural similarity, 05 2004.

[26] Yi Xu, Yuxin Hu, Zaiwei Zhang, Gregory P Meyer, Siva Karthik Mustikovela, Siddhartha Srinivasa, Eric M Wolff, and Xin Huang. Vlm-ad: End-to-end autonomous driving through vision-language model supervision. *arXiv preprint arXiv:2412.14446*, 2024.

[27] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning, 2020. URL `https://arxiv.org/abs/1805.04687`.

[28] Seungjun Yu, Kisung Kim, Daejung Kim, Haewook Han, and Jinhan Lee. Direct preference optimization-enhanced multi-guided diffusion model for traffic scenario generation. *arXiv preprint arXiv:2502.12178*, 2025.

[29] Kaiwen Zhang, Zhenyu Tang, Xiaotao Hu, Xingang Pan, Xiaoyang Guo, Yuan Liu, Jingwei Huang, Li Yuan, Qian Zhang, Xiao-Xiao Long, et al. Epona: Autoregressive diffusion world model for autonomous driving. *arXiv preprint arXiv:2506.24113*, 2025.

[30] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018. URL `https://arxiv.org/abs/1801.03924`.

[31] Haonan Zhao, Yiting Wang, Thomas Bashford-Rogers, Valentina Donzella, and Kurt Debattista. Exploring generative ai for sim2real in driving data synthesis. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 3071–3077. IEEE, 2024.

[32] Wenzhao Zheng, Ruiqi Song, Xianda Guo, Chenming Zhang, and Long Chen. Genad: Generative end-to-end autonomous driving. In *European Conference on Computer Vision*, pages 87–104. Springer, 2024.

[33] Xingcheng Zhou, Xuyuan Han, Feng Yang, Yunpu Ma, and Alois C Knoll. Opendrivevla: Towards end-to-end autonomous driving with large vision language action model. *arXiv preprint arXiv:2503.23463*, 2025.

[34] Yunsong Zhou, Michael Simon, Zhenghao Mark Peng, Sicheng Mo, Hongzi Zhu, Minyi Guo, and Bolei Zhou. Simgen: Simulator-conditioned driving scene generation. *Advances in Neural Information Processing Systems*, 37:48838–48874, 2024.

[35] Jun Zhu. Synthetic data generation by diffusion models. *National Science Review*, 11(8): nwae276, 2024.

A

A.1 Environment Setup

The system requires an NVIDIA GPU with 16GB+ VRAM (tested on A5000 24GB), 32GB+ system RAM, and Python 3.8+ with CUDA 11.0+ for GPU acceleration. Key dependencies include PyTorch 2.0+, diffusers 0.21.0+, transformers 4.25.0+, and CARLA 0.9.15+ for autonomous driving validation.

A.2 Postprocessing Details

Quality Assessment Pipeline: Each generated image undergoes a multi-stage validation process to ensure visual quality and numerical integrity. The system first checks for invalid values (NaN or infinite numbers) that could cause rendering issues, then evaluates brightness levels by rejecting images with mean brightness below 20 on a 0-255 scale to filter out dark or empty images. Contrast analysis follows, discarding images with standard deviation below 10 for being too flat or lacking visual detail. Finally, a composite quality score is calculated based on normalized brightness and contrast metrics, with images scoring below the 0.3 threshold being automatically rejected from the final dataset.

Technical Post-Processing: The pipeline implements several technical optimizations to ensure numerical stability and efficient processing. The Variational Autoencoder (VAE) is maintained in float32 precision to prevent numerical errors during image decoding, while memory optimization techniques like attention slicing and VAE slicing reduce GPU memory usage during generation. Images are processed with xformers memory-efficient attention when available on CUDA devices, and the final outputs are saved as PNG files with optimization enabled for better compression without quality loss.

Metadata Generation: Each accepted image receives comprehensive metadata tracking that includes generation timestamp, the specific prompt used, calculated quality score, generation time per image, batch index, and unique image ID. This metadata is saved as JSON files alongside each image, enabling detailed analysis of generation patterns and quality metrics. Additionally, rejected images are logged with specific rejection reasons (darkness, flatness, or invalid values) to maintain quality control statistics, while the system tracks overall success rates and average quality scores across the entire generation run for performance monitoring.

A.3 Metric Definitions

FID (Fréchet Inception Distance): Measures how similar the overall "style" of synthetic images is to real images (lower is better, <50 is good)

SSIM (Structural Similarity Index): Compares image structure and patterns (0-1 scale, higher is better, >0.7 is good)

LPIPS (Learned Perceptual Image Patch Similarity): Measures perceptual similarity as humans see it (lower is better, <0.3 is good)

A.4 Positive and Negative Prompts

Category	Positive Prompts
Base Scenarios	- winding rural country road through rolling green hills - straight farm road between golden wheat fields - mountain highway through dense pine forest - coastal rural road with ocean glimpses - prairie highway through vast grasslands - country lane with stone walls and hedgerows - forest service road through mixed woodland - rural highway through vineyard country - desert highway through sagebrush landscape - highland road through moorland and heather
Road Surfaces	- asphalt road with yellow center line - concrete highway with lane markings - well-maintained country road - weathered rural highway - freshly paved road surface
Lighting Conditions	- golden hour lighting, warm natural light - overcast day, soft diffused lighting - early morning light with long shadows - late afternoon sun, dramatic lighting - bright daylight, clear visibility - partly cloudy sky, dynamic lighting
Weather Conditions	- clear blue sky with wispy clouds - partly cloudy, dynamic sky - morning mist in the distance - crisp autumn air, perfect visibility - spring day, fresh atmosphere - adverse weather, heavy rain
Quality Enhancers	- professional automotive photography, DSLR quality - ultra-detailed, photorealistic, masterpiece - award-winning photography, crystal clear - high resolution, perfect focus, sharp details - professional grade, ultra-realistic

Table 3: Positive prompts for scenario generation.

Category	Negative Terms
Image Quality Issues	- low quality, blurry, out of focus - dark image, black image, completely black - noise, grain, artifacts, compression
Style or Realism Issues	- cartoon, anime, painting, sketch - artificial, fake, unrealistic, oversaturated
Geometric or Perspective Errors	- distorted geometry, impossible perspective - floating objects, unnatural lighting
Unwanted Elements	- city, urban, buildings, cars, vehicles - people, pedestrians, traffic signs - watermark, text, logo, signature, frame, border - multiple exposures
Anatomical or Structural Issues	- bad anatomy, deformed, mutated
Lighting and Exposure Issues	- night scene, darkness - overexposed, underexposed - neon colors

Table 4: Negative prompt terms for filtering undesirable images.

B Figures

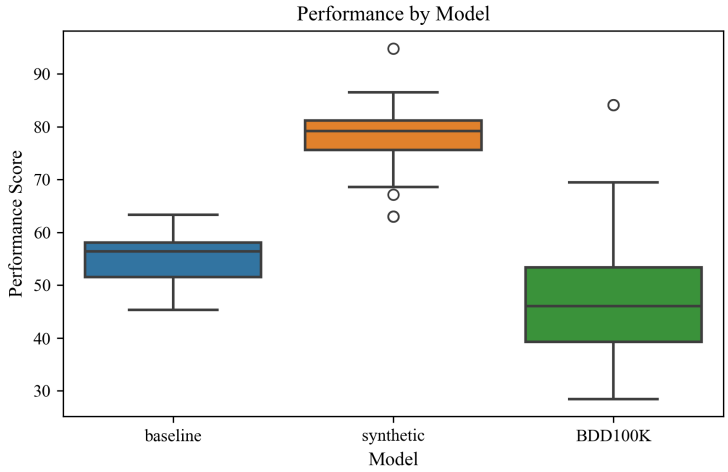

Figure 2: Performance of CARLA AV models trained on baseline, RuralGen, and BDD100K datasets across rural driving scenarios. Performance scores of the RuralGen model, as compared with the heuristic baseline and the BDD100K datasets showed that RuralGen consistently outperformed both comparison models.

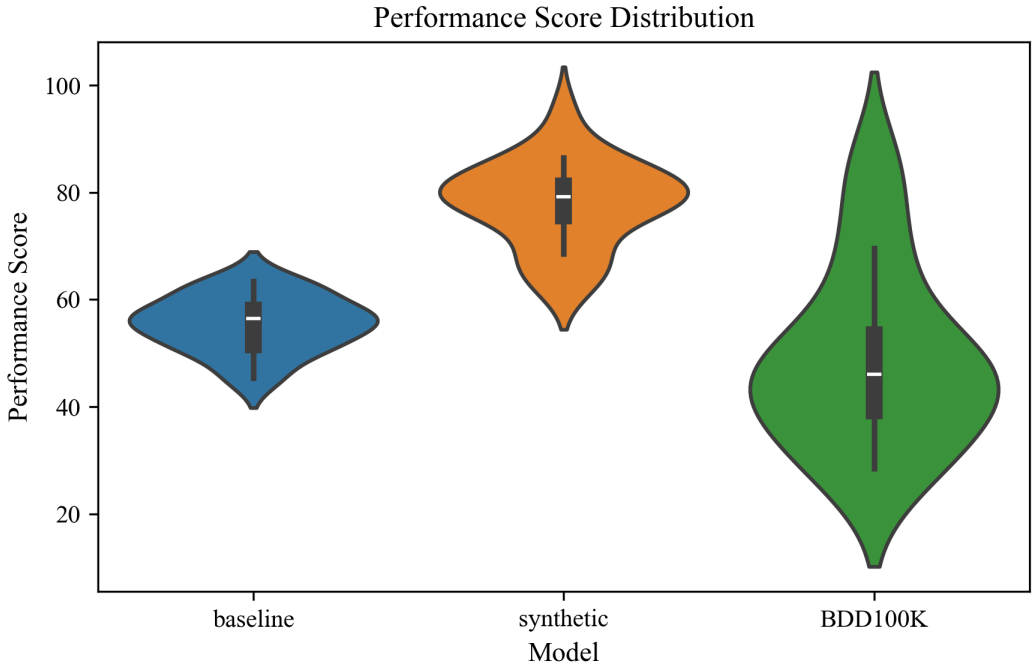

Figure 3: Performance distribution of CARLA AV models trained on baseline, diffusion-enhanced, and BDD100K datasets across rural driving scenarios.

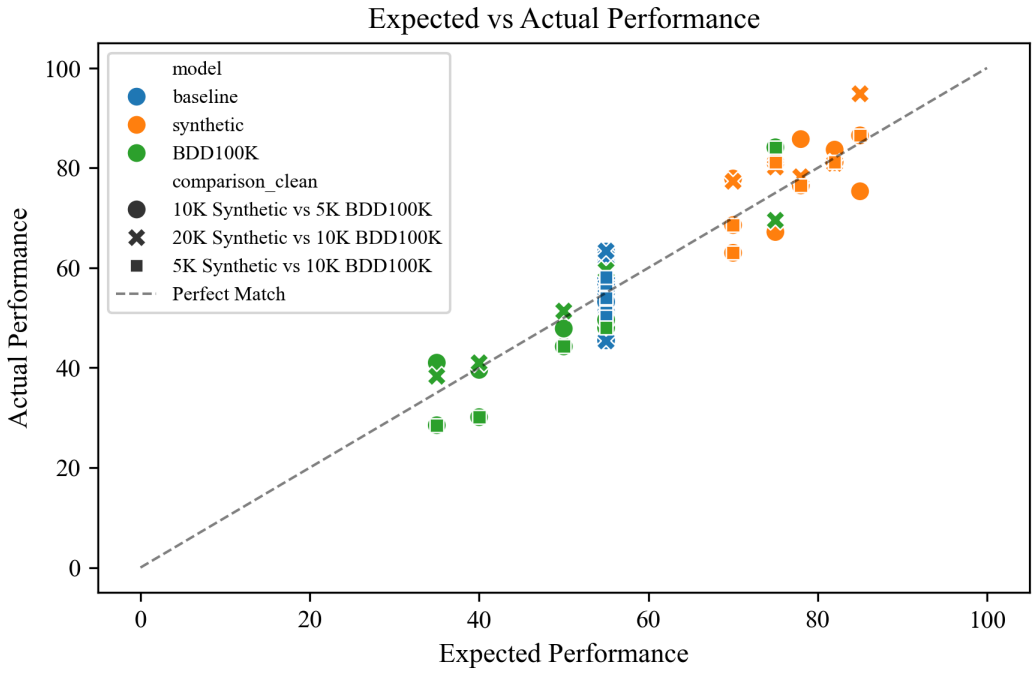

Figure 4: Expected vs Actual performance of the models showing that RuralGen matched or outperformed expectations when compared with BDD100K and the baseline heuristic model

Figure 5: RuralGen outperformed both the baseline heuristics model and the BDD100K model across all training sizes

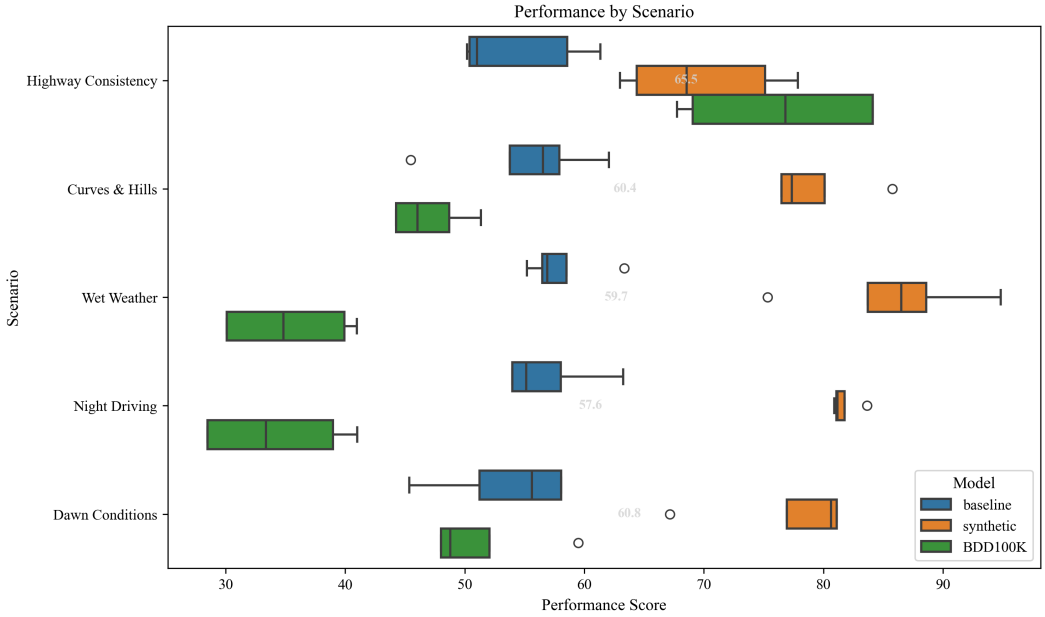

Figure 6: RuralGen had the best performance in all scenarios except Highway consistency, where BDD100K performed better in some cases

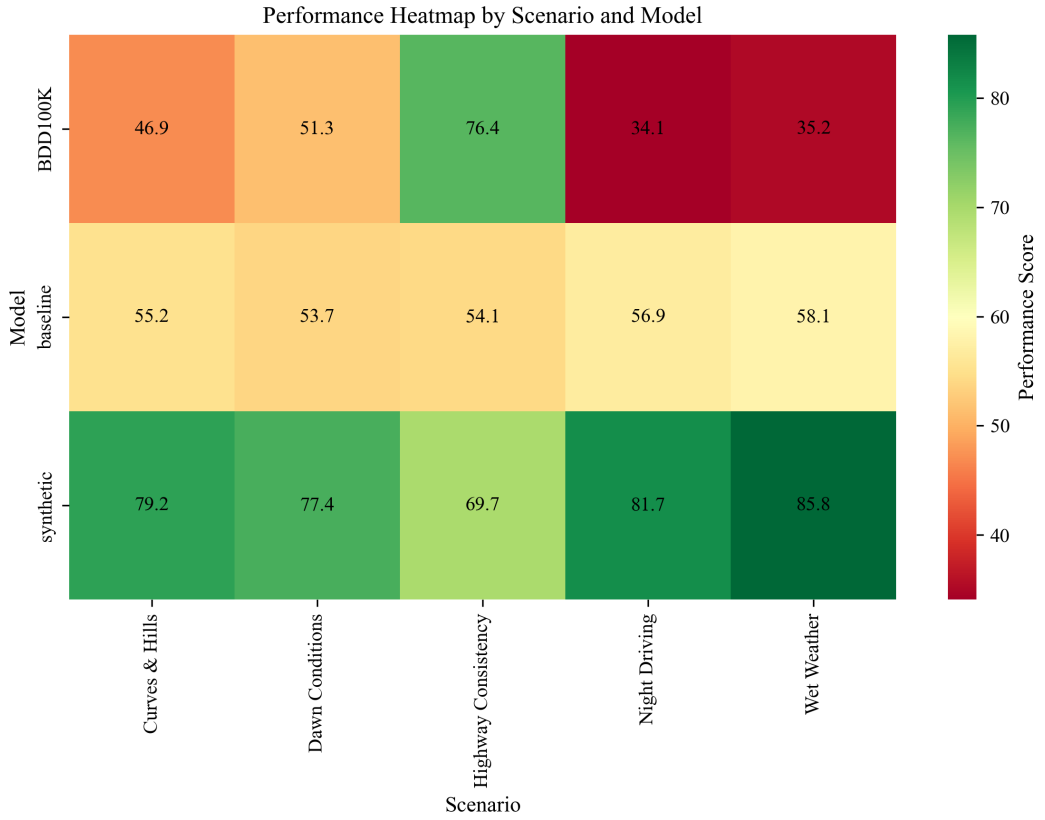

Figure 7: The figure shows a Heatmap of the model performance across scenarios showing how RuralGen performed when compared with baseline heuristics model and BDD100K

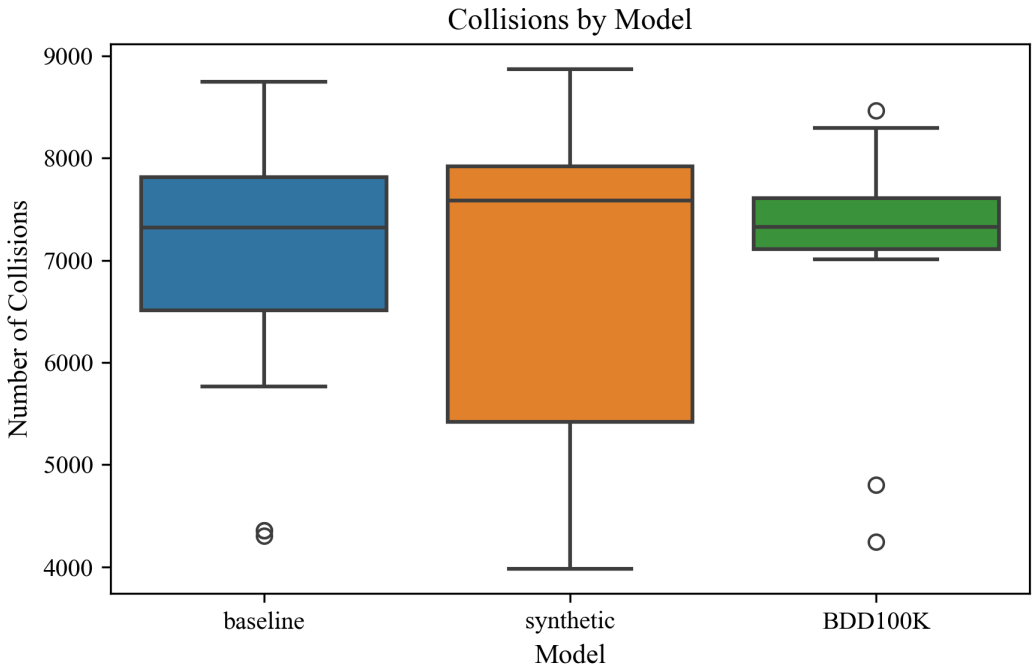

Figure 8: Figure shows the distribution of collisions in CARLA, by model. BDD100K performed better in this metric compared with the baseline heuristic model and RuralGen

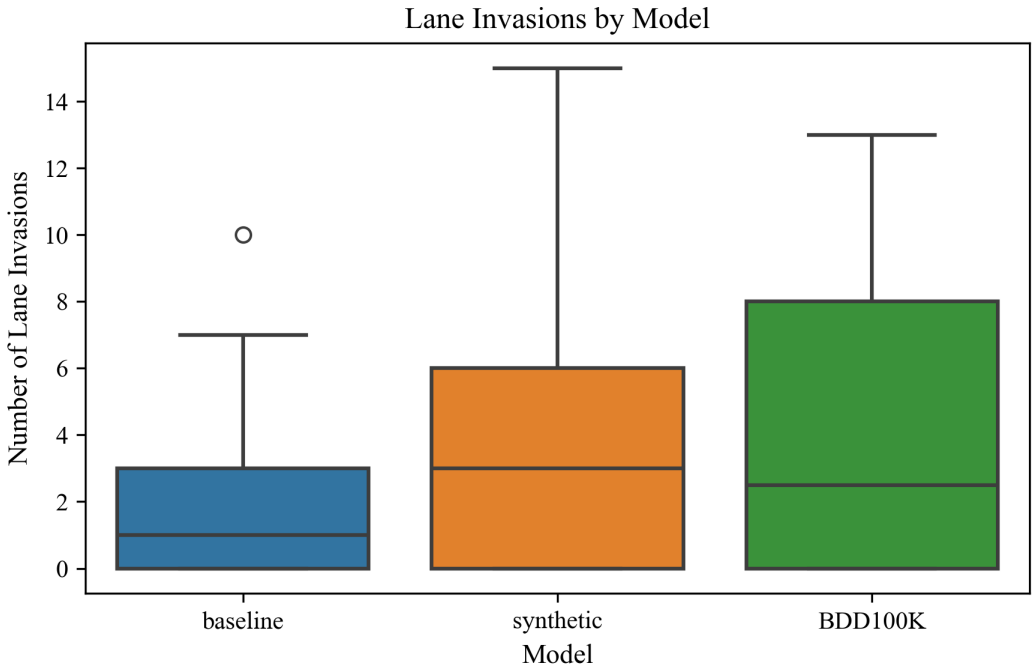

Figure 9: Figure shows the distribution of lane invasions in CARLA, by model. The aseline heuristics model performed better in this metric compared with the BDD100K model and RuralGen

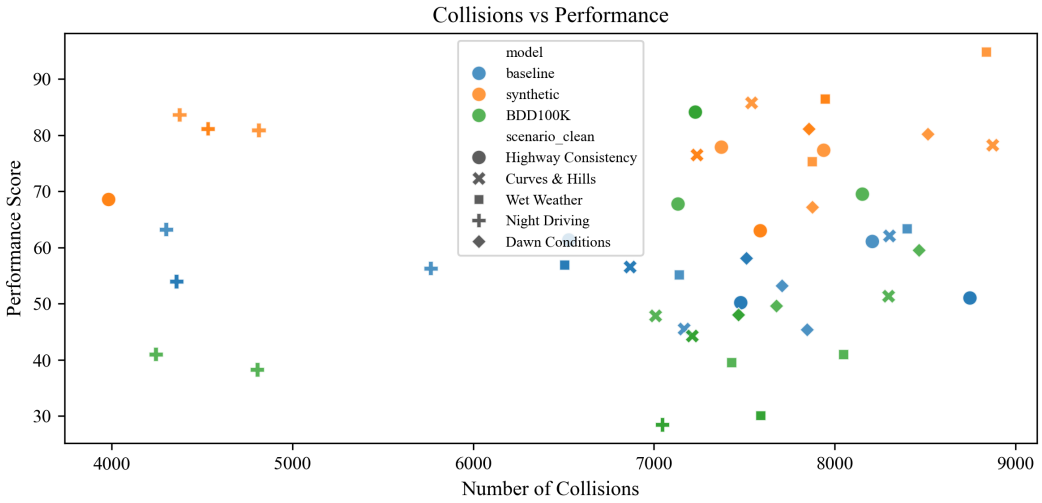

Figure 10: Plot showing number of collisions vs the performance of each model to account for lower collision numbers from a lower performance of a model due to less distance traveled.

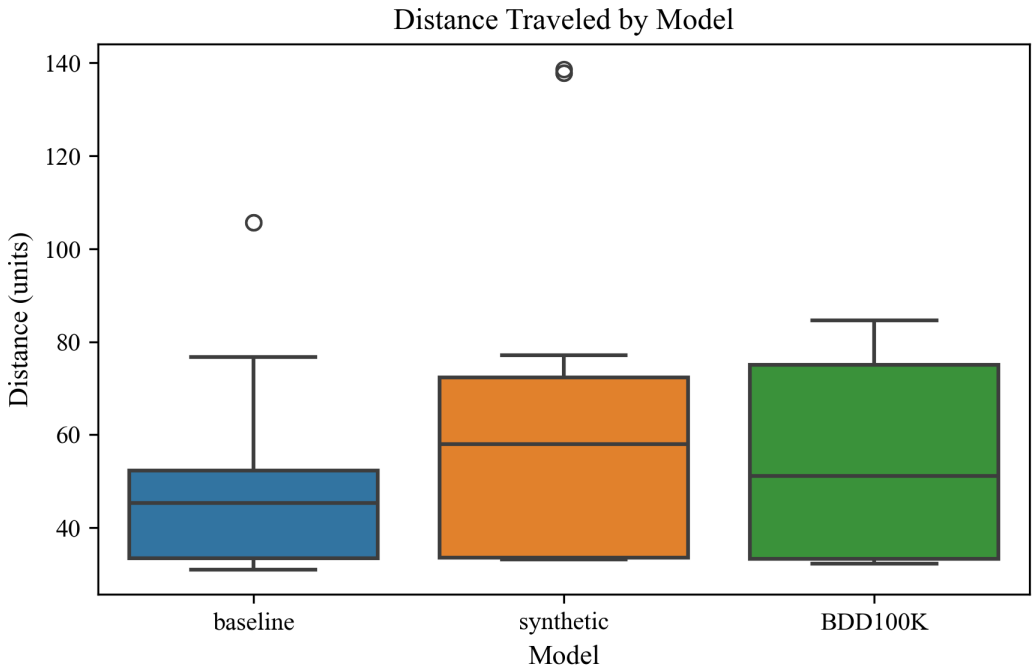

Figure 11: Distance traveled for each model. This shows that RuralGen closely matched distance traveled by BDD100K, while the baseline heuristics model lagged behind.

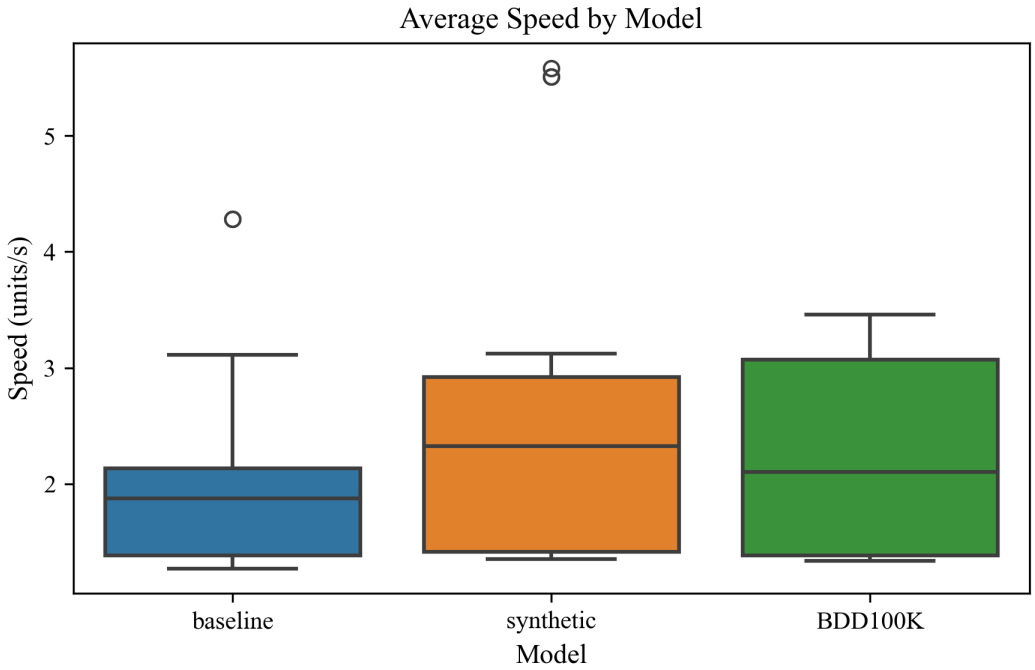

Figure 12: Average speed achieved with each model. This shows that RuralGen closely matched the average speed of BDD100K, while the baseline heuristics model lagged behind.

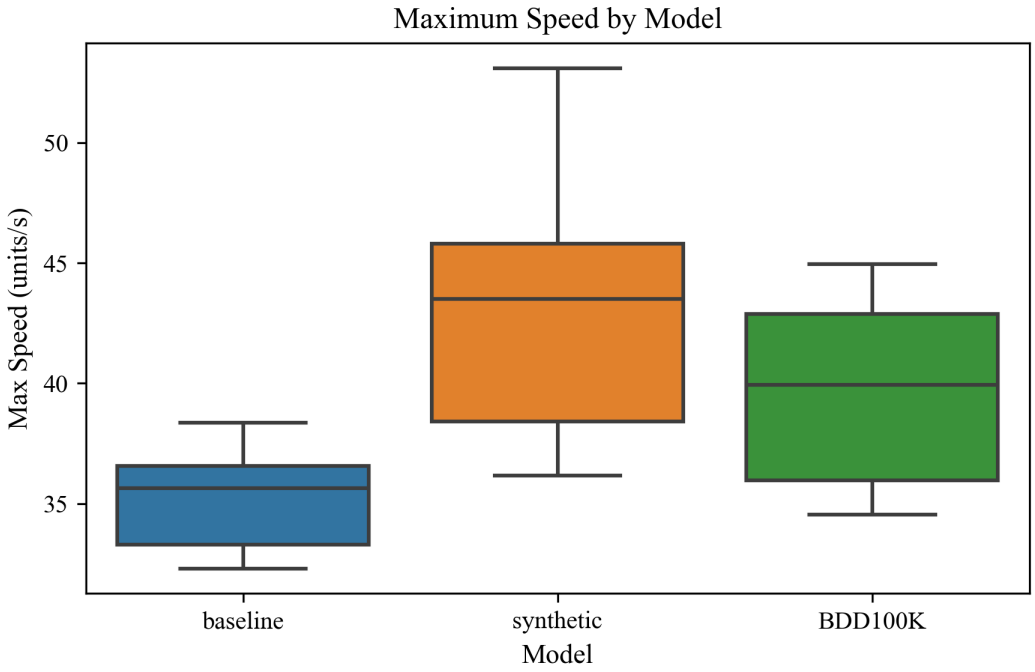

Figure 13: Maximum speed achieved with each model. This shows that RuralGen outperformed both the baseline heuristics model and BDD100K

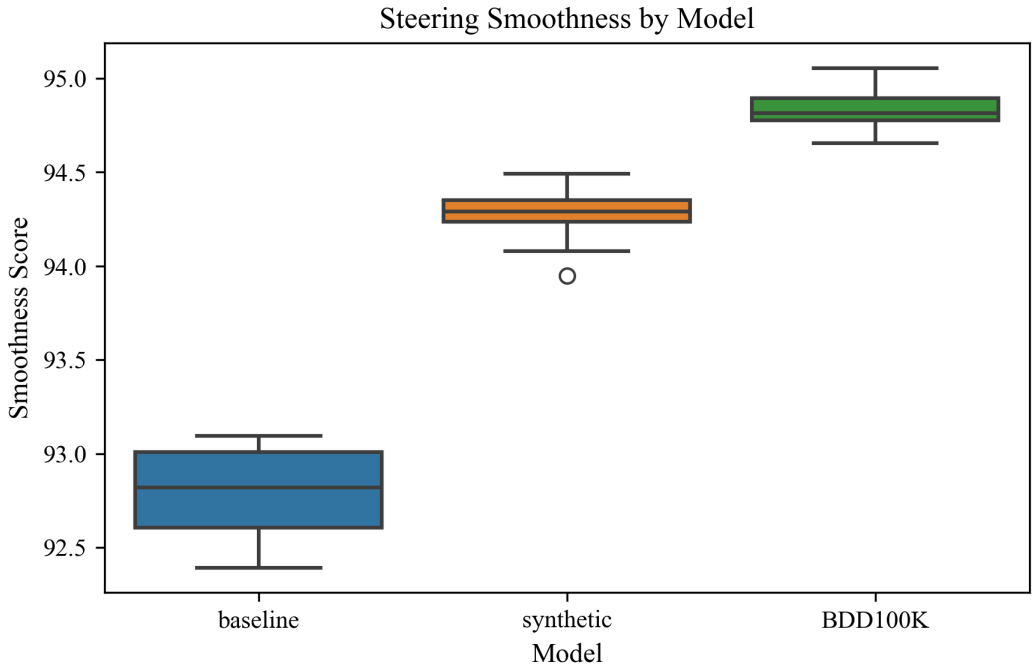

Figure 14: Steering Smoothness by model, showing that BDD100K had the best steering smoothness, followed by RuralGen.

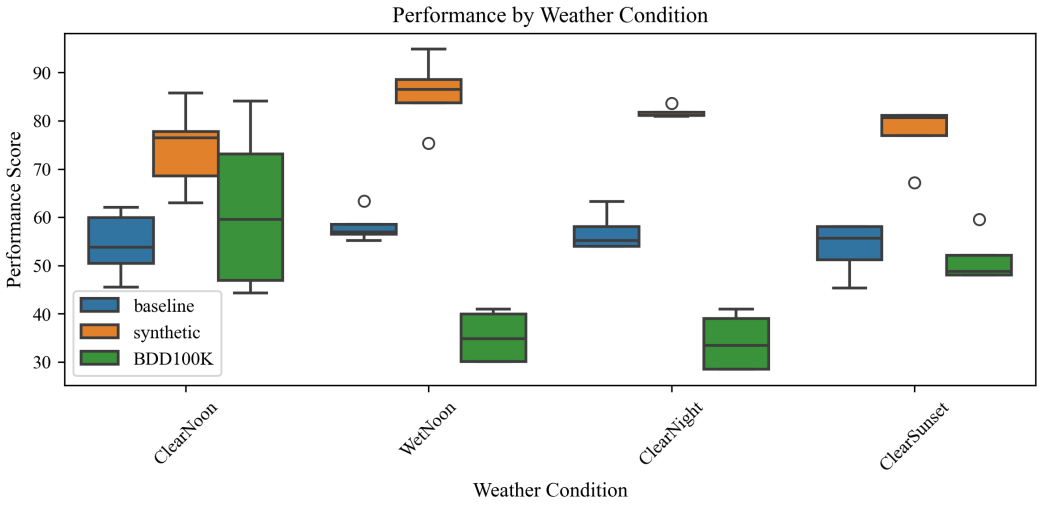

Figure 15: Model performance by weather conditions showing how RuralGen outperformed both the baseline heuristics model and BDD100K across conditions such as Clear/Noon, Wet/Noon, Clear/Night and Clear/Sunset

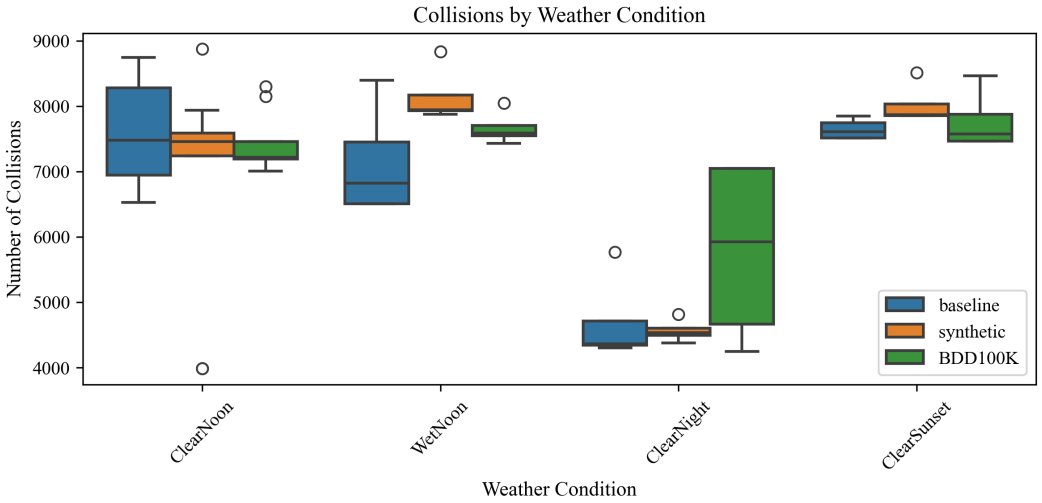

Figure 16: Number of collisions by weather conditions, showing how RuralGen had fewer collisions across all weather conditions except Clear/Sunset compared with the baseline heuristics model and BDD100K.

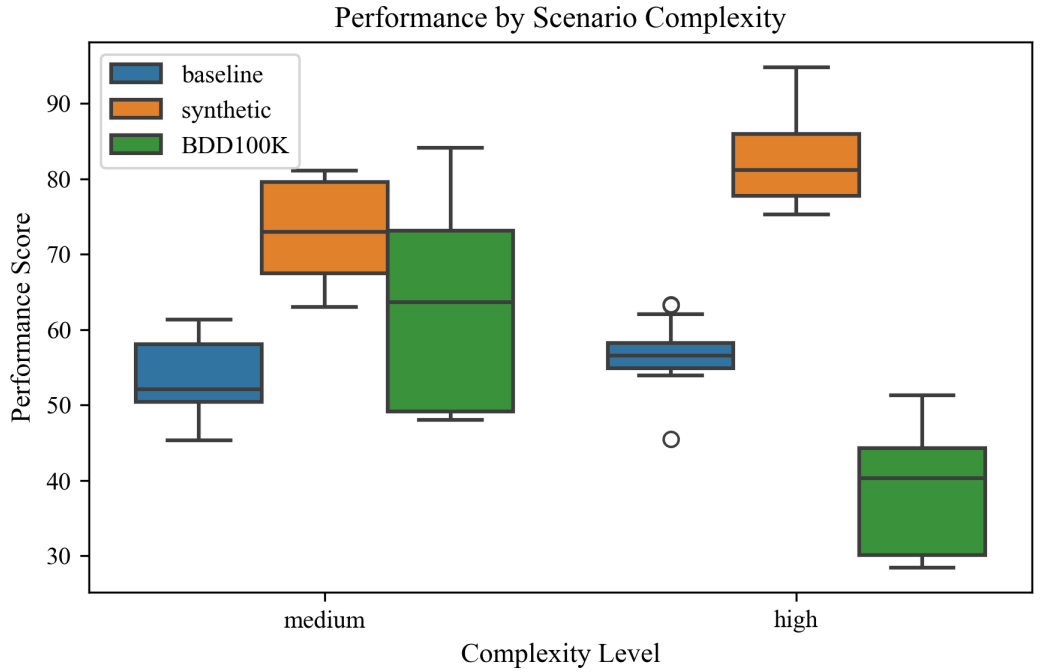

Figure 17: Model performance by scenario complexity, showing that RuralGen performed better as the complexity of scenarios increased.

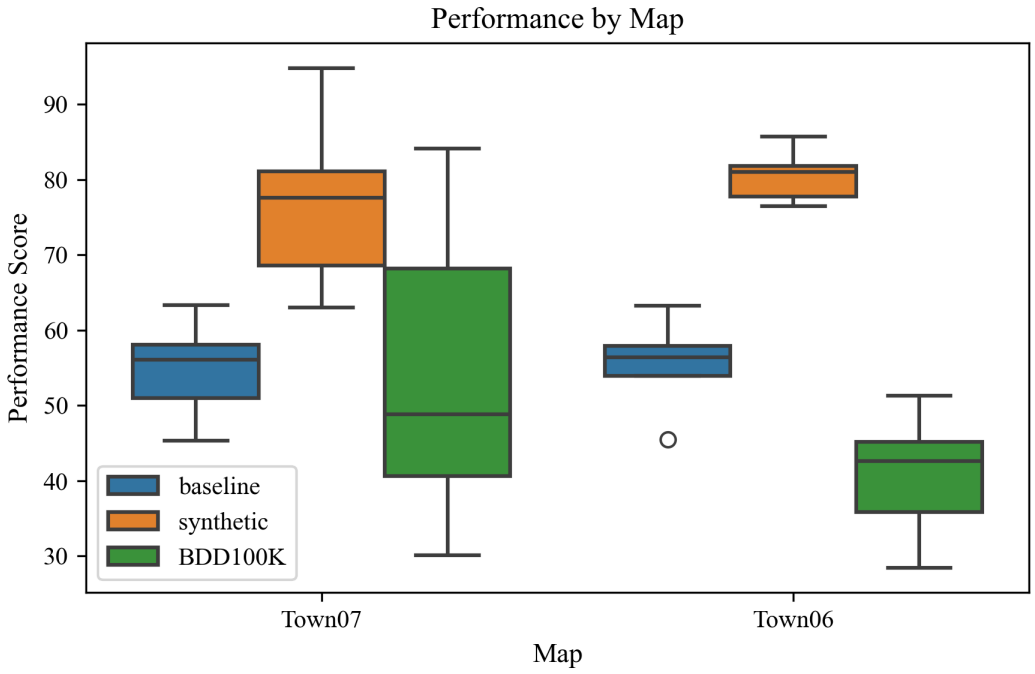

Figure 18: There were two town maps used for testing in CARLA. The figure shows how RuralGen performed with each map, in comparison with the baseline heuristics model and BDD100K

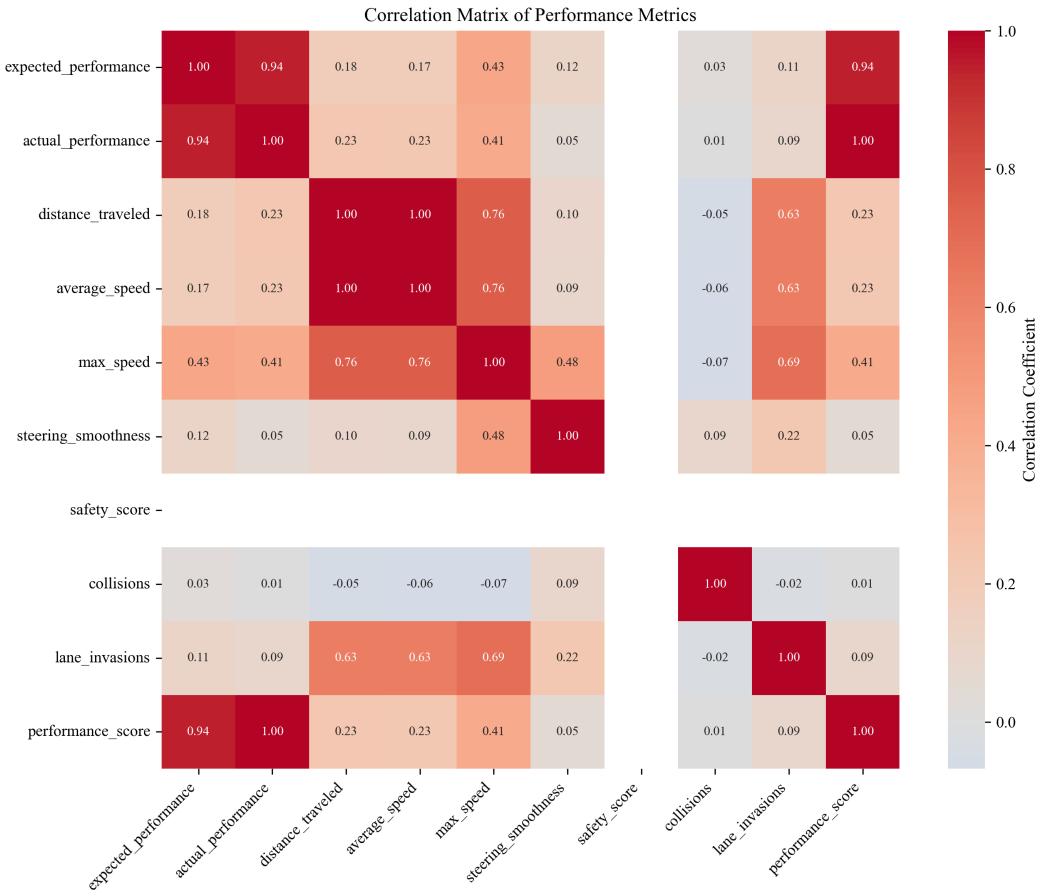

Figure 19: Correlation matrix of model performance metrics and safety scores

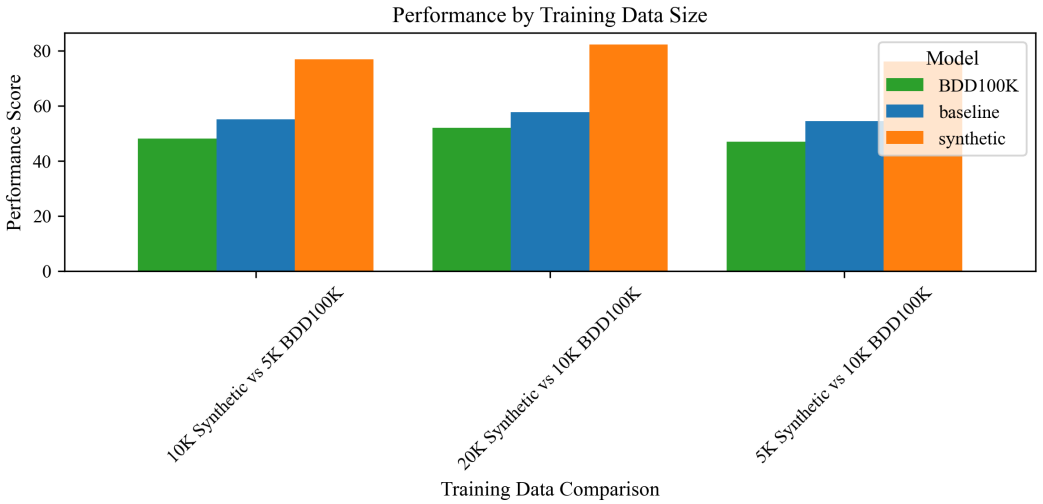

Figure 20: Model performance by the size of the training data, showing that the training data size had less impact to RuralGen compared with the baseline heuristics model and BDD100K

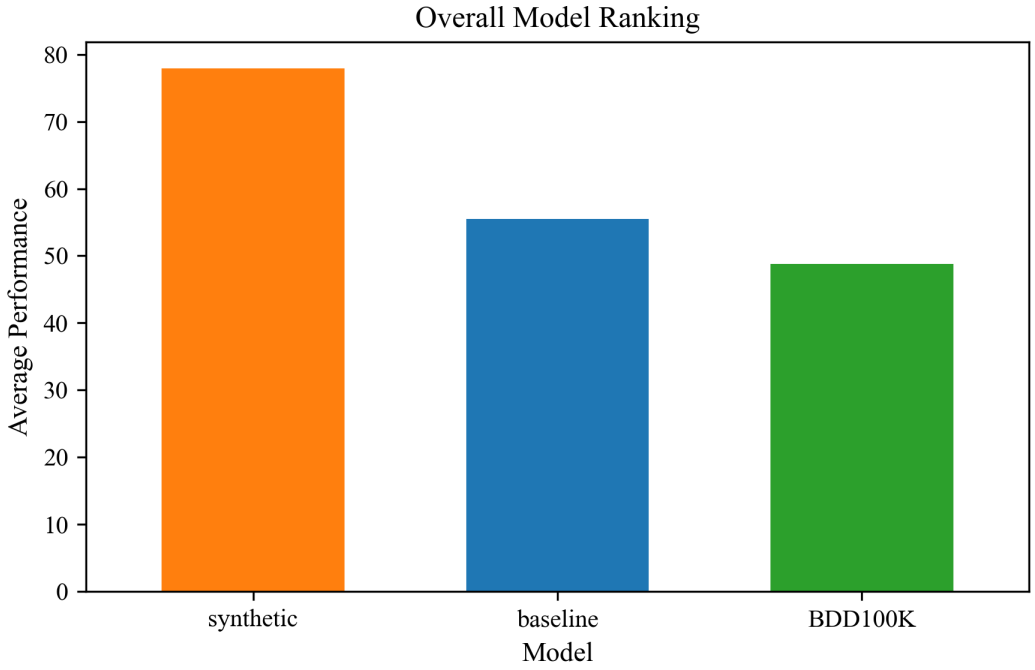

Figure 21: Overall Ranking of models based on performance scores achieved by each model

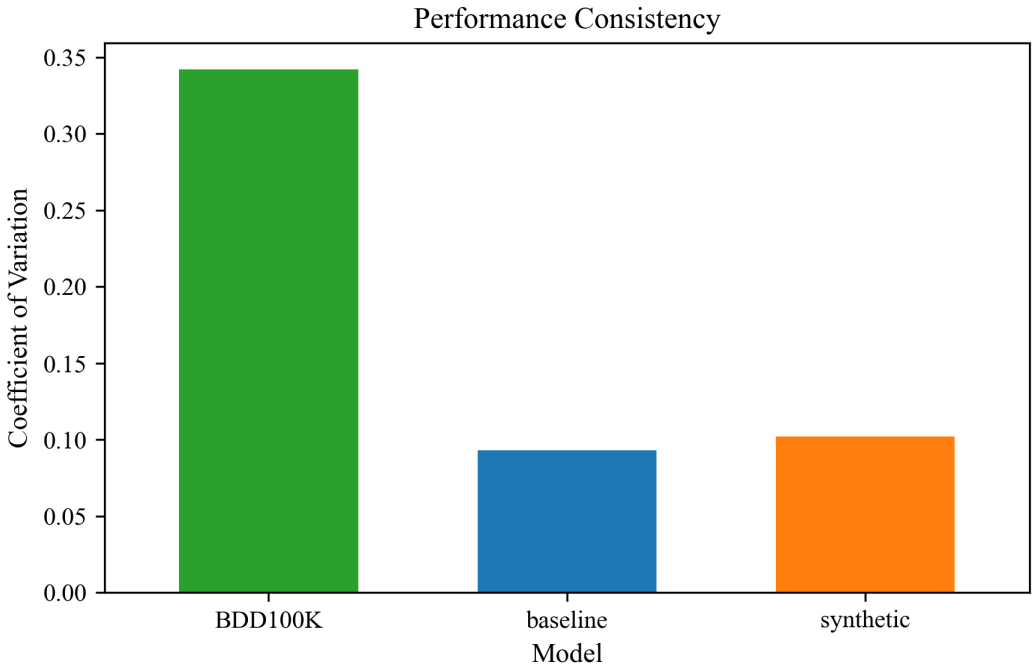

Figure 22: Performance consistency of each model, showing that BDD100K had the most consistent performance, even though RuralGen outperformed it in most scenarios.

Training Data	Model	Mean Perf	Std Perf	Mean Coll	Mean Dist	Tests
10K Synthetic vs 5K BDD100K	Baseline	55.0	5.1	6756	55.9	11
10K Synthetic vs 5K BDD100K	Synthetic	77.0	7.8	6744	56.7	11
10K Synthetic vs 5K BDD100K	Bdd100K	48.1	16.8	7004	55.2	10
20K Synthetic vs 10K BDD100K	Baseline	57.6	7.4	7705	34.7	5
20K Synthetic vs 10K BDD100K	Synthetic	82.3	7.2	7796	77.3	5
20K Synthetic vs 10K BDD100K	Bdd100K	51.9	13.0	7554	61.5	5
5K Synthetic vs 10K BDD100K	Baseline	54.5	3.3	6912	52.4	6
5K Synthetic vs 10K BDD100K	Synthetic	76.1	8.8	6524	59.9	6
5K Synthetic vs 10K BDD100K	Bdd100K	47.0	22.4	7309	46.7	5

Figure 23: Table showing the impact of training data on the model performance

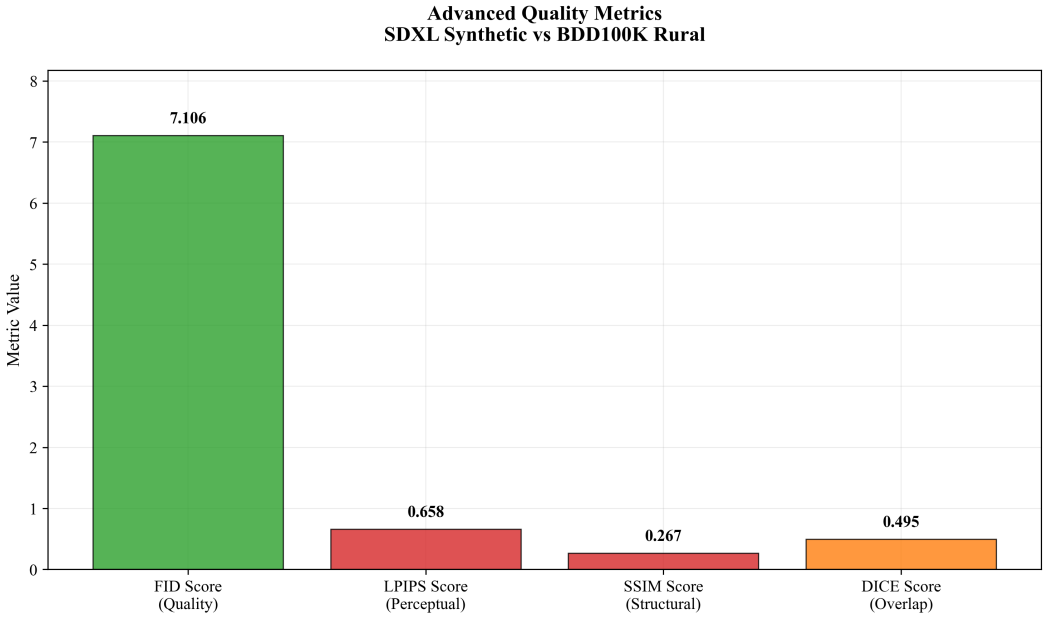

Figure 24: Advanced Image Quality Metrics (FID, LPIPS, SSIM and DICE) showing how RuralGen performed compared with BDD100K

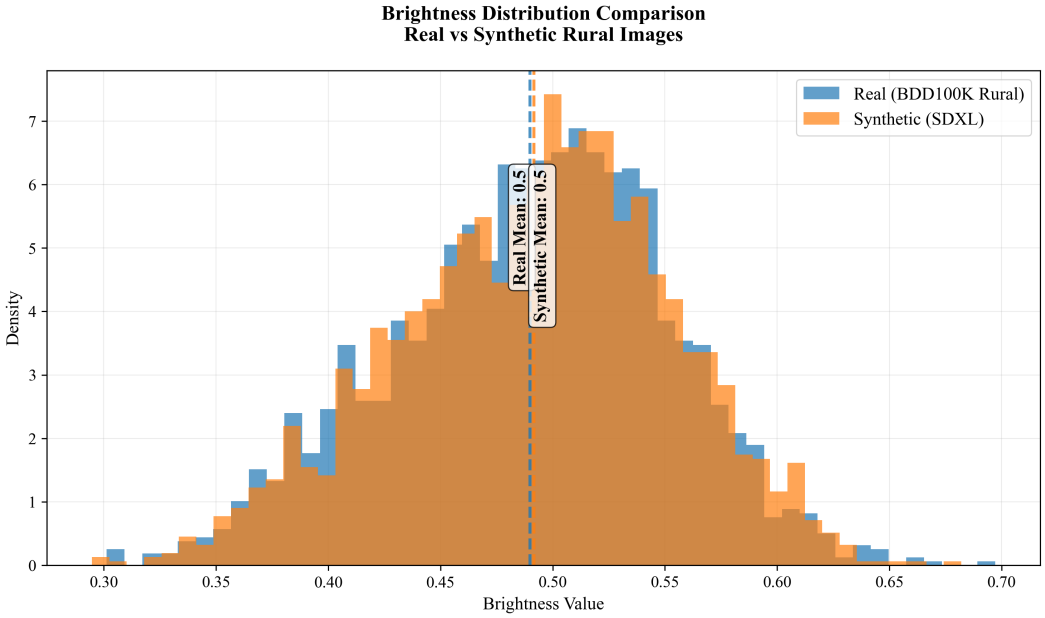

Figure 25: Brightness distribution comparison between RuralGen and BDD100K datasets

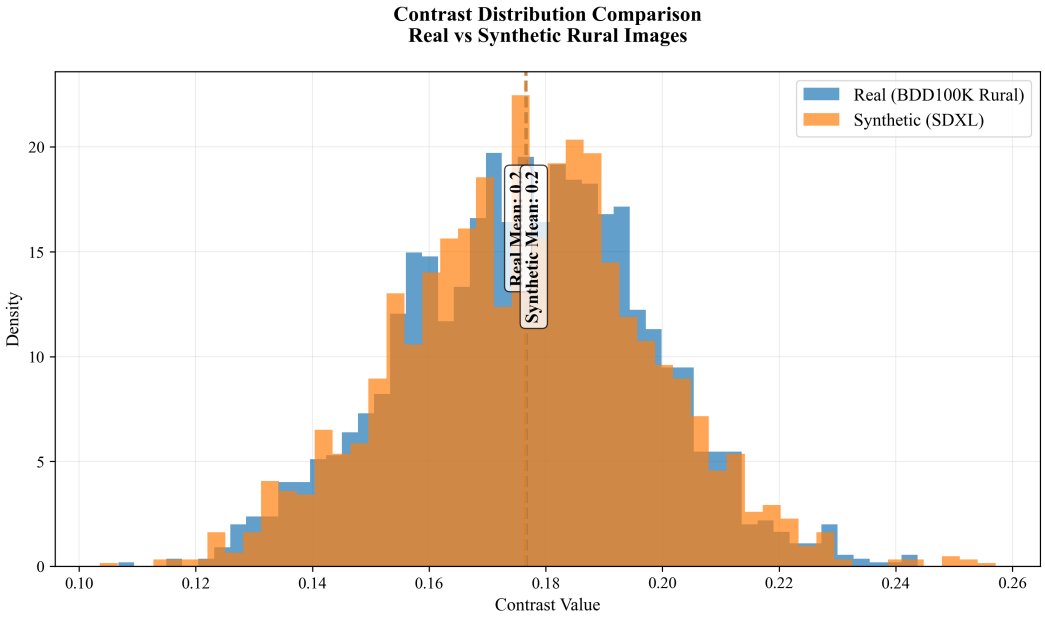

Figure 26: Contrast distribution comparison between RuralGen and BDD100K datasets

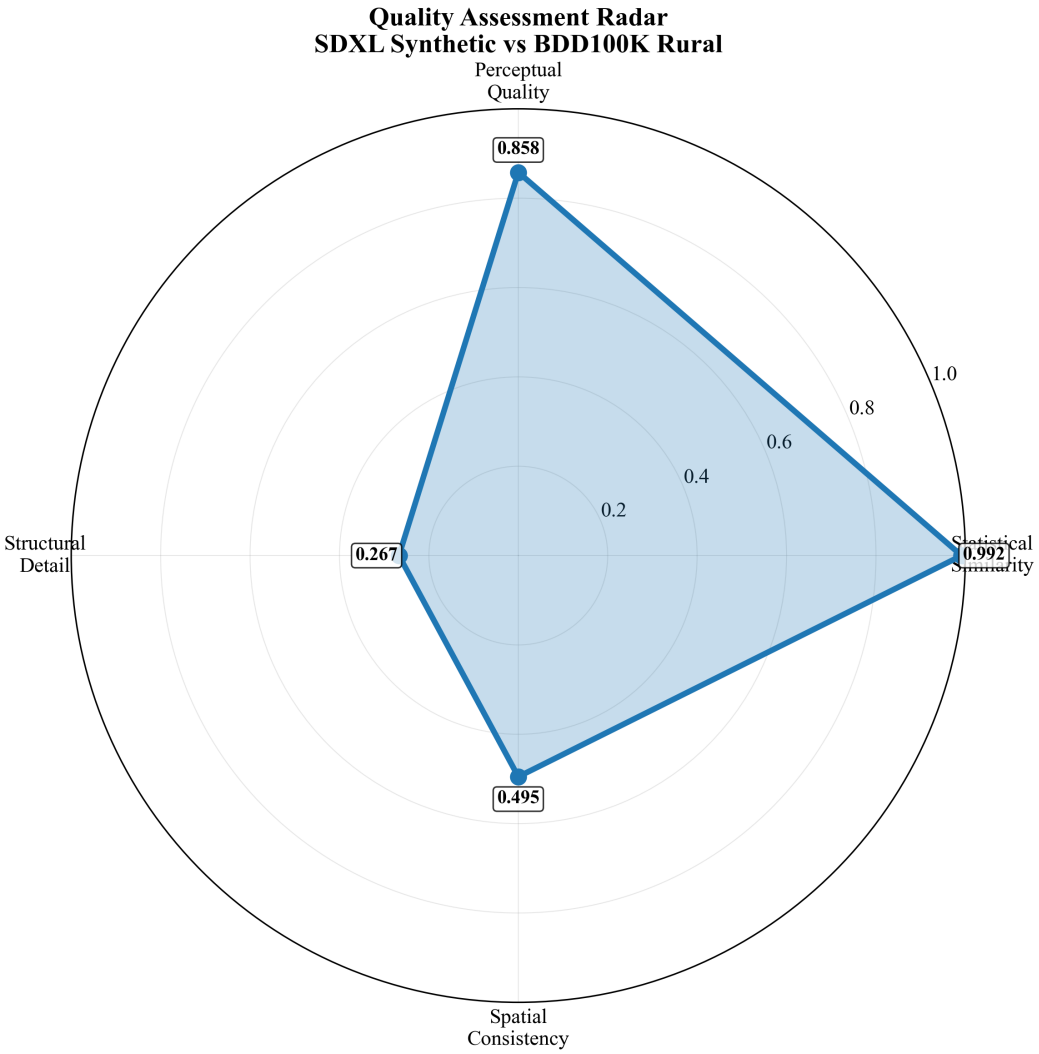

Figure 27: Quality assessment radar for generated Synthetic data, showing room for improvement in structural detail and spatial consistency

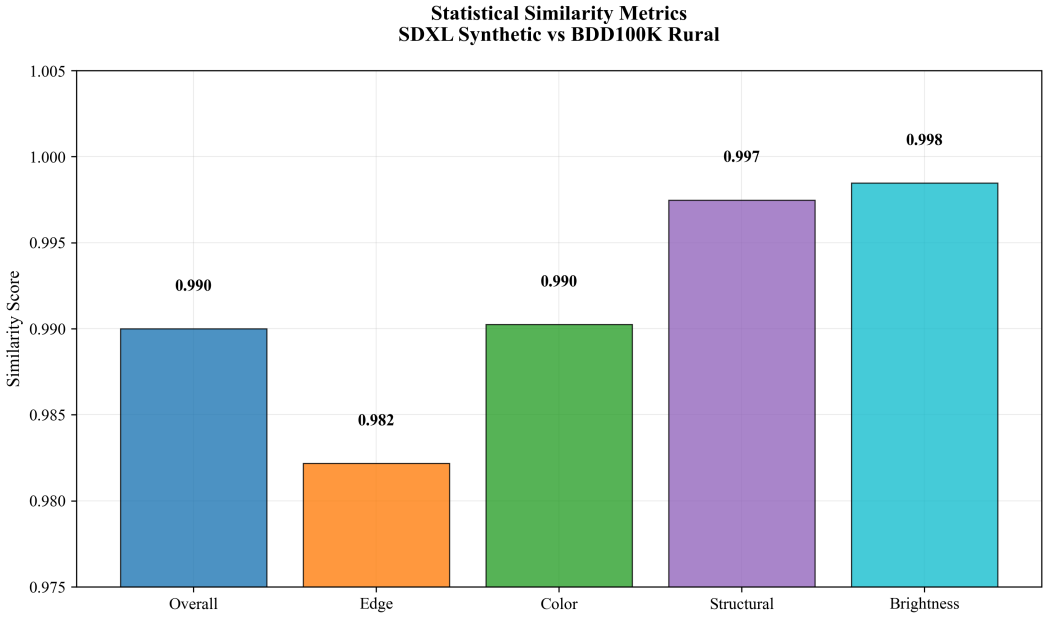

Figure 28: Statistical similarity metrics between Synthetic and BDD100K datasets, showing a 99% similarity with BDD100K when considering Edge, Color and Brightness scores.

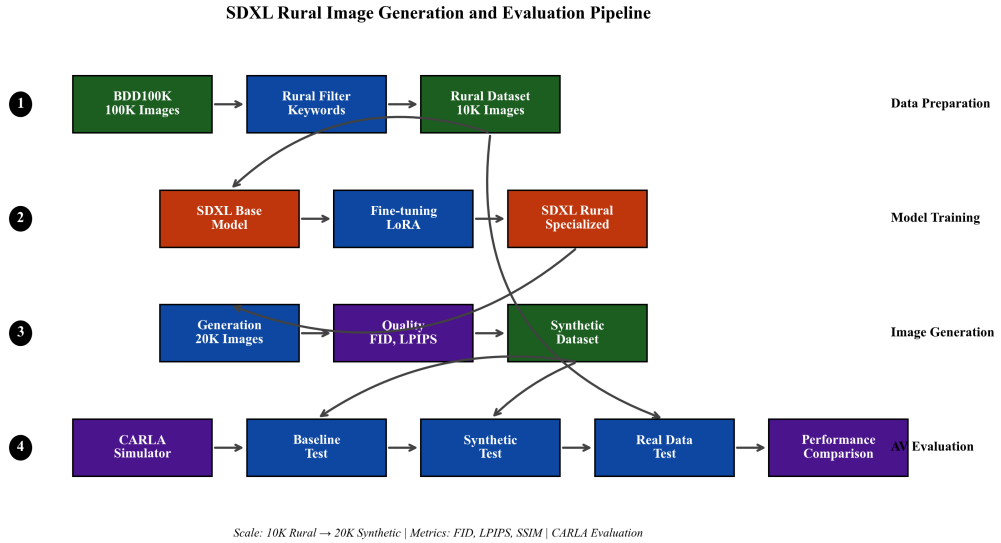

Figure 29: Visual depicting the research data pipeline to prepare, train, generate and test images using an SDXL diffusion model

Figure 30: Sample Image 1

Figure 31: Sample Image 2

Figure 32: Sample Image 3