A deep learning approach for realistic traffic scene video changes across lighting and weather conditions

Md Shadab Alam*
Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
0000-0001-9184-9963

Sagar Hitendra Parmar
Amazon Fire TV
Amazon
Bengaluru, India
0000-0008-1326-6525

Marieke Martens
Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
0000-0002-1661-7019

Pavlo Bazilinskyy
Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
0000-0001-9565-8240

***Abstract*—Recent advances in GAN-based architectures have led to innovative methods for image transformation. The lack of diverse environmental data, such as different lighting conditions and seasons in public data, prevents researchers from effectively studying the difference in driver and road user behaviour under varying conditions. This study introduces a deep learning pipeline that combines CycleGAN-turbo and Real-ESRGAN to improve video transformations of the traffic scene. Evaluated using dashcam videos from London, Hong Kong, and Los Angeles, our pipeline shows a 7.97% improvement in T-SIMM for temporal consistency compared to CycleGAN-turbo for night-to-day transformation for Hong Kong. PSNR and VPQ scores are comparable, but the pipeline performs better in DINO structure similarity and KL divergence, with up to 153.49% better structural fidelity in Hong Kong compared to Pix2Pix and 107.32% better compared to ToDayGAN. This approach demonstrates better realism and temporal coherence in day-to-night, night-to-day, and clear-to-rainy transitions.**

***Index Terms*—Traffic scenes, Generative adversarial networks, Diffusion models, Environmental conditions**

I. Introduction

Human behaviour in traffic has been studied extensively. Garay et al. [1] used a simulator task to compare the scanning behaviour of drivers in night and daytime driving situations. They determined that all 48 participants were less likely to scan for risks at night, with 24 novice drivers (aged 16–17 years) and 24 experienced drivers (aged 40–50 years) participating in the study. This study revealed a significant overall effect of lighting conditions on risk prediction accuracy ($F(1, 46) = 18.74$, $p < 0.001$). The novice drivers correctly identified the risks 46% of the time during daytime conditions, compared to 37% at night. In contrast, experienced drivers demonstrated higher accuracy, predicting risk 67% of the time during the day and 55% during the night. This indicates that although experienced drivers outperformed novices in both lighting conditions, both groups were adversely affected by reduced visibility at night. Trankle et al. [2] found that young male drivers (aged 18–21) rated traffic situations as less risky than older males (aged 35–75 years), particularly in situations involving darkness, curved roadways and rural environments. Siddique et al. [3] studied the severity of pedestrian injuries, including differences in crossing locations and light conditions. They showed that the chances of pedestrians sustaining a fatal injury are 49% lower at intersections than at midblock locations under daylight conditions, 24% lower under dark-with-street-light conditions, and 5% lower under dark-without-street-light conditions. Compared to dark conditions without street lighting, daylight reduces the odds of a fatal injury by 75% at the mid-block locations and 83% at intersections, while street lighting reduces the odds by 42% at midblock locations and 54% at intersections. Evan et al. [4] compared risk and difficulty from the driver's point of view at day and night by showing 14 different videos in which 12 videos were in pairs from the same location, but the traffic conditions were different. In addition, they were unable to collect the videos in time for the remaining two videos. Campbell et al. showed in their review that fatal pedestrian crashes are more likely to occur during nighttime hours, and nonfatal pedestrian crashes are more likely to occur during daytime hours [5]. The probability of a pedestrian being killed increases at least three times when the person is involved in a night-time crash compared to a daytime crash [6], [7]. A large body of research determines the reasons for the high risk of death in traffic at night [8]–[11].

The limitations of existing public traffic datasets present a challenge in this area. However, data sets, such as CamVid [12], Cityscapes [13], KITTI [14], LYFT L5 [15], and H3D [16], have traffic scenes only from daylight and a clear sky.

*Corresponding author

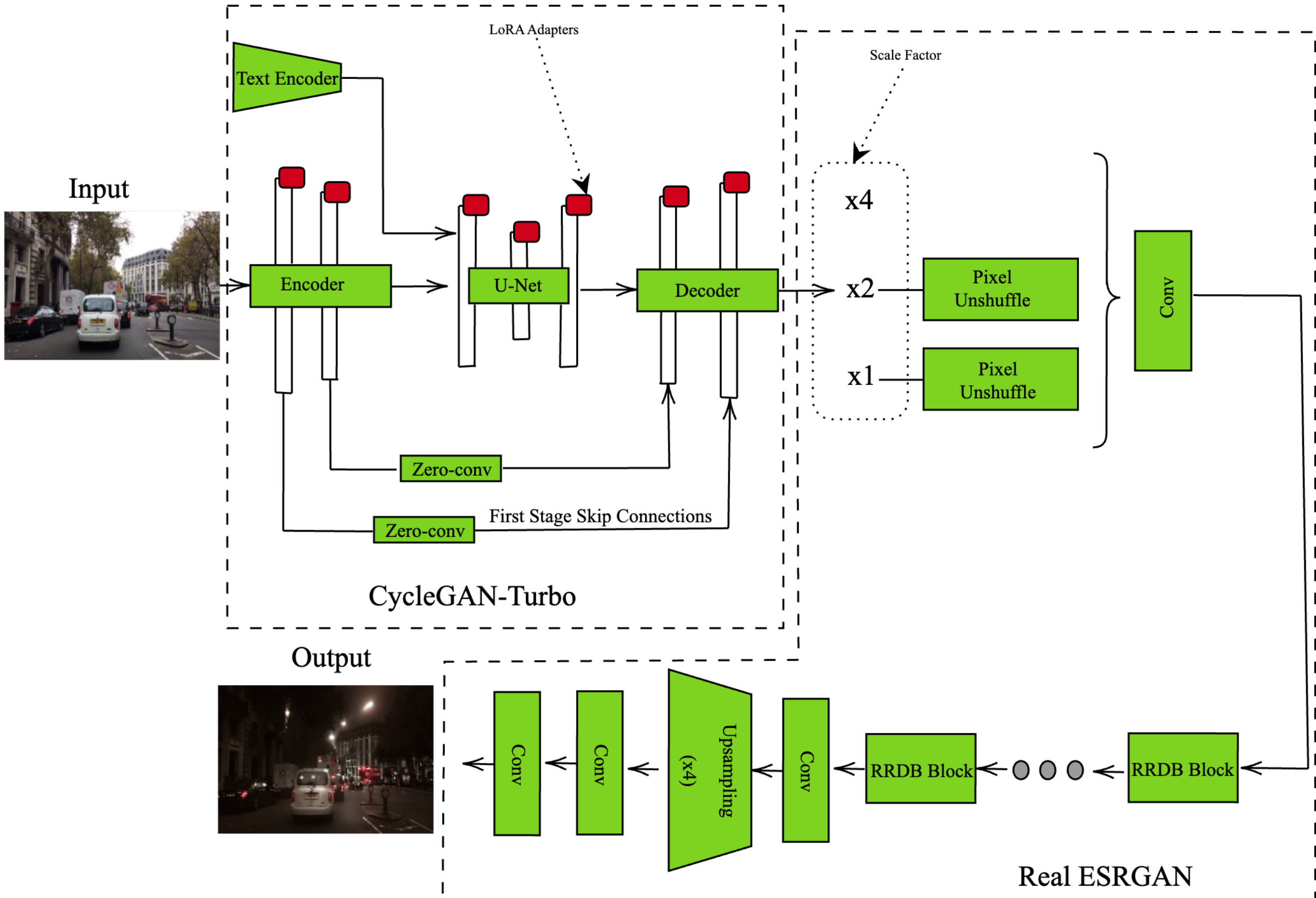

Fig. 1: Representation of the neural network architecture.

Some datasets, such as AppoloScape [17] and KAIST [18], do have scenes of rainy conditions, but do not have scenes of driving at night.

Since the advent of deep learning frameworks such as Generative Adversarial Networks (GANs) and Diffusion models, both of the architectures have shown the potential to translate one image from one environment to another. Zhao et al. [19] used Deep Convolutional Generative Adversarial Networks (DCGAN) [20] to generate images and videos of a new traffic scene. The training videos and images were taken from the KITTI dataset [14], and traffic scenes were collected while driving from Xi'an to Changshu, China, of overtaking scenes. Similarly, Tan et al. have attempted to create realistic scenes using GANs [21]. GANs were used in multiple projects to translate images from day to night and vice versa [22]–[27] and diffusion models [28], [29]. Similarly, different weather conditions were simulated with GANs, such as rainy, snowy, and overcast [30]. Cheng et al. [31] used GANs to address the problem of low resolution and blurred details in highway images caused by factors such as rain and fog, illumination interference, and nighttime lighting. These translations were used to improve all-day vehicle detection [32], traffic perception [33], object detection [34] and the functionality of Advanced Driver Assistance Systems [35].

A. *Aim of the study*

This study aims to develop a deep learning-based pipeline that enhances the realism of traffic scene videos by transforming environmental conditions, specifically focussing on lighting and weather condition variations such as day-to-night, night-to-day and clear-to-rainy transitions. By combining CycleGAN-turbo [36] for domain translation with Real-ESRGAN [37] for resolution enhancement. Although significant progress has been made in image domain translation, relatively few studies have emphasised the realism of the generated videos. To assess its effectiveness, the generated scenes will be compared with both the output of CycleGAN-turbo and other state-of-the-art neural network architectures designed for similar tasks, such as translating day-to-night images. The comparison will be based on several performance metrics, including Temporal Consistency, Structural Fidelity, Visual Quality (PSNR, SSIM), and KL Divergence.

Fig. 2: Comparison of images of traffic scenes from different cities.

II. Method

The dashcam videos were collected from YouTube from three cities, namely London (UK) (https://www.youtube.com/watch?v=QI4_dGvZ5yE; https://www.youtube.com/watch?v=mEXVBiT1eAM) Hong Kong (https://www.youtube.com/watch?v=ULcuZ3Q02SI; https://www.youtube.com/watch?v=XaR6qEt-BIY) and Los Angeles (CA, USA) (https://www.youtube.com/watch?v=4uhMg5na888; https://www.youtube.com/watch?v=eR5vsN1Lq4E) in two different lighting conditions: day and night. These cities were chosen to represent diverse geographic locations on three continents (Europe, Asia, and North America), offering a wide range of traffic conditions and environmental contexts to test the effectiveness of the proposed enhancement pipeline. In addition, London and Hong Kong both have left-sided traffic systems, while Los Angeles follows a right-sided system. This choice of cities introduces an interesting variation in driving behaviour, although differences in traffic flow, such as the direction of travel, were not explicitly controlled in this study. In each of these videos, a 20 second video was extracted and passed through the pipeline, which encompassed CycleGAN-turbo and Real ESRGAN as illustrated in Figure 1. The 20 second segments contained traffic for a better evaluation of the proposed enhancement pipeline. This selection ensured that the processing stages were applied to scenarios that are representative of real-world traffic conditions, including various vehicle movements and lighting variations. The day videos were used to translate from day to night and from clear to rainy scenarios, while the night videos from these locations were converted into daytime videos. Figure 2 shows some of the translated images from the pipeline. The complete translated and original videos are available in the supplementary material (section V).

The generated videos from the pipeline was compared with the CycleGAN-turbo across different metrics used to assess the quality of video processing, compression, or generation tasks, namely the Temporal Structural Similarity Index Metric (T-SIMM), Peak Signal-to-Noise Ratio (PSNR), Fréchet Video Distance (FVD) and Video Perceptual Quality (VPQ).

The extracted frames from the videos were used to compare with other neural network architecture models such as CycleGAN-turbo [36], CycleGAN [38], HEDNGAN [27], Pix2Pix [39] and ToDayGAN [40]. The weights are provided

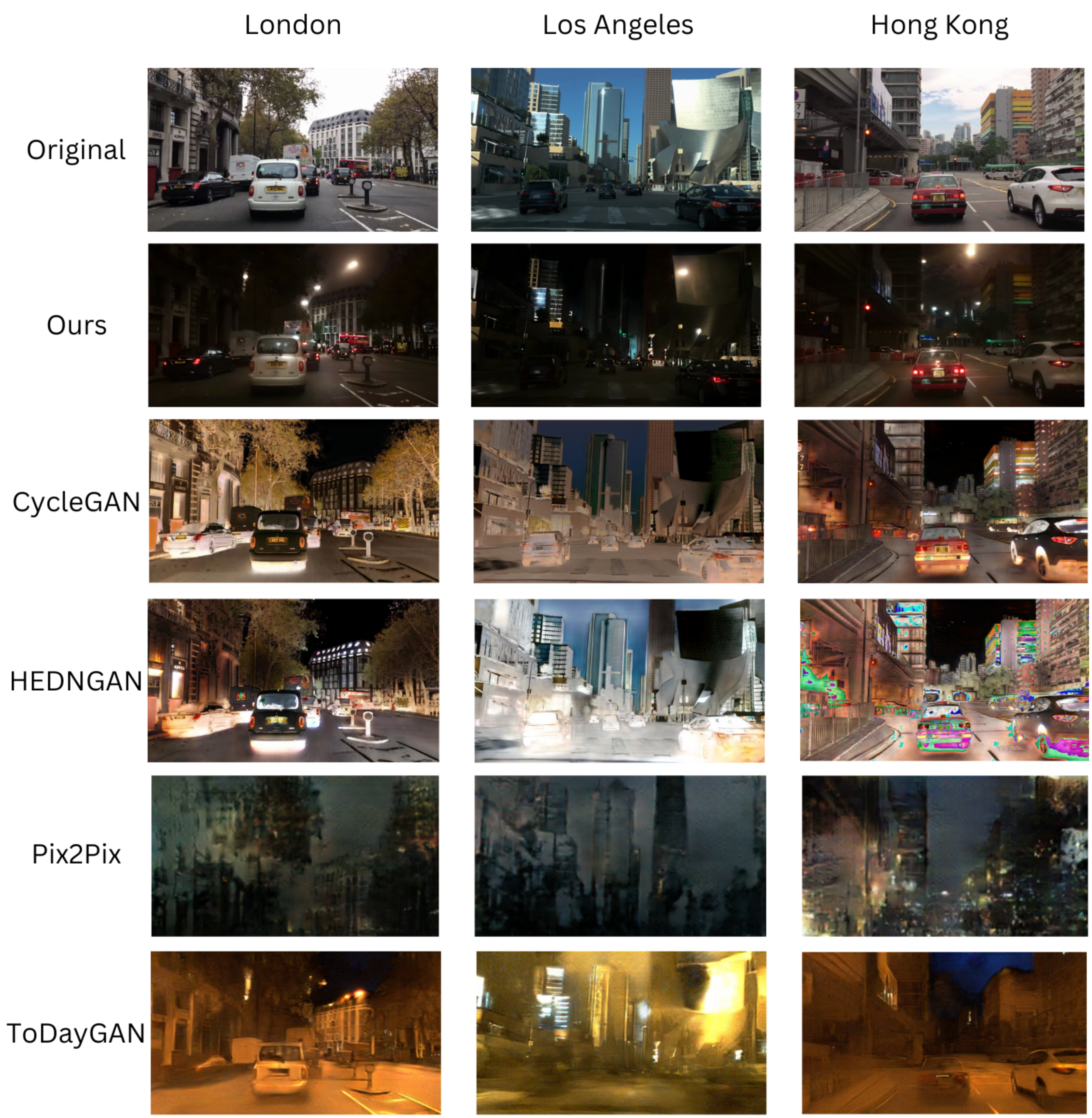

Fig. 3: Comparison of the translated image from day to night by different GANs architecture.

by the authors in the repository linked with their publication. We used CycleGAN [38] weights from HEDNGAN [27] as they have trained them to translate images from day to night.

III. Results

The Figure 2 shows the images of the traffic scenes generated by the CycleGAN-turbo [36] and the proposed pipeline. The generated videos are available in the supplementary material (section V). The images generated by our architecture are clearer, and the lights are less scattered. The scenes from our pipeline present improved clarity and realistic light distribution, highlighting the enhanced quality achieved by our method for the day-to-night transition.

The calculated values for T-SIMM, PSNR, and VPQ are listed in Table I. For the "day-to-night" transition in

London, our pipeline achieved a T-SIMM score of 0.9275 (0.0006), compared to 0.9204 (0.0024) for CycleGAN-turbo. In Hong Kong's "night-to-day" transition, the T-SIMM score for our method was 0.8540 (0.0080), while CycleGAN-turbo scored 0.7909 (0.0027). The PSNR for the same transition in HongKong was 29.7831 (7.6484) with our model, while CycleGAN-turbo reached 29.9508 (8.9165). In the "night-to-day" transition for Los Angeles, the VPQ values were 14.1763 for our pipeline and 14.1614 for the CycleGAN turbo, showing minimal variation between the methods.

The scenes of the original video were processed through different architectures, namely CycleGAN [38], HEDNGAN [27], Pix2Pix [39], and ToDayGAN [40] for comparison for day-to-night translation. Figure 3 shows the output of the different architecture. These outputs were compared with each other using DINO structure similarity [41] and KL divergence and reported in Table II.

IV. Discussion

In this study, we proposed a novel pipeline to generate realistic traffic scene transformations across varying environmental conditions, combining CycleGAN-turbo for domain translation and Real-ESRGAN for resolution enhancement. Our method showed improved performance in producing realistic visuals in multiple scenarios, as reflected in key metrics.

Our proposed model demonstrates superior performance compared to CycleGAN-turbo in various transformation scenarios, as evidenced by Figure 2 and Table II. Qualitatively, our model consistently generates more realistic outputs with enhanced clarity, sharper structural details, and natural lighting transitions, particularly evident in day-to-night transformations, where the CycleGAN-turbo exhibits scattered lighting effects and less realistic shadows. Quantitatively, our model achieves higher T-SIMM scores in most scenarios, such as a 1.37% improvement for day-to-night transitions in Los Angeles (0.9729 vs. 0.9597) and a 7.97% improvement for night-to-day transitions in Hong Kong (0.8540 vs. 0.7909). For clear-to-rainy transitions, the T-SIMM score for London is nearly on par, with our model achieving 0.8205 compared to 0.8235, a minor 0.36% reduction that is offset by better perceptual realism and structural integrity.

Although PSNR values for our model are slightly lower in certain cases (e.g., 27.6784 vs. 27.7023 for day-to-night in London, a 0.09% difference), this is attributed to the trade-off between pixel-level fidelity and perceptual improvements. PSNR focusses on pixel-wise similarity and may not fully capture perceptual quality enhancements that result from subtle adjustments in texture and lighting. Similarly, VPQ scores, which reflect perceptual quality, show minimal variations, such as 14.0138 versus 14.0362 for day to night in London, a negligible 0.16% reduction, highlighting that perceived quality remains consistent while structural preservation and realism improve. For clear-to-rainy transformations, our model achieves slightly higher VPQ scores, such as 14.2089 in London compared to 14.1931, reflecting a 0.11% improvement, further showcasing its ability to handle diverse environmental conditions effectively.

The results presented in Table II and Figure 3 underscore the advantages of our proposed pipeline over existing methods. Table II demonstrates that our model consistently achieves better structural preservation (DINO Structure Similarity) and KL divergence compared to models. For example, in the night-to-day transition, our method achieves a DINO similarity score compared to Cycle GAN (0.2805 vs. 0.3605) for London, 28.52%, indicating better structural fidelity.

V. Limitations and future work

The current study has several limitations. It relies heavily on established neural network architectures, such as CycleGAN-Turbo and Real-ESRGAN, which restrict the adaptability of the approach to specific traffic scene transformations. Furthermore, the evaluation process lacks direct human feedback, which is crucial to validate the realism and quality of the generated videos and frames. Then, the pipeline only translates the frames from one environment to another. It does not change the density of traffic, such as the number of cars and pedestrians.

Transition	Location	Metric	CycleGAN turbo	Ours
Day to night	London	T-SIMM	0.9204(0.0024)	0.9275(0.0006)
		PSNR	27.7023(0.0018)	27.6784(0.0016)
		VPQ	14.0362	14.0138
	HongKong	T-SIMM	0.9777(0.0005)	0.9724(0.0004)
		PSNR	27.6930(0.0003)	27.6818(0.0002)
		VPQ	14.0502	14.0372
	Los Angeles	T-SIMM	0.9729(0.0001)	0.9597(0.0001)
		PSNR	27.7803(0.0060)	27.7274(0.0049)
		VPQ	14.0551	14.0106
Night to day	London	T-SIMM	0.8677(0.0251)	0.8973(0.0080)
		PSNR	29.4572(3.8110)	29.3145(3.1562)
		VPQ	15.1004	15.0912
	HongKong	T-SIMM	0.7909(0.0027)	0.8540(0.0011)
		PSNR	29.9508(8.9165)	29.7831(7.5484)
		VPQ	15.3165	15.2251
	Los Angeles	T-SIMM	0.7925(0.0012)	0.8508(0.0003)
		PSNR	27.8157(0.0124)	27.8067(0.0091)
		VPQ	14.1619	14.1552
Clear to rainy	London	T-SIMM	0.8235(0.0075)	0.8205(0.0014)
		PSNR	27.6295(0.0014)	27.6618(0.0011)
		VPQ	14.1931	14.2089
	HongKong	T-SIMM	0.9636(0.0020)	0.9459(0.0022)
		PSNR	27.6244(0.0001)	27.6579(0.0001)
		VPQ	14.2044	14.2161
	Los Angeles	T-SIMM	0.9380(0.0002)	0.9093(0.0002)
		PSNR	27.5918(0.0053)	27.6259(0.0063)
		VPQ	14.1614	14.1763

TABLE I: Comparison of T-SIMM, PSNR, and VPQ values.

To address these limitations, future work will focus on collecting human feedback to enhance the evaluation and refinement of the generated outputs. Custom neural network architectures specifically designed for traffic scene transformations will be developed to improve adaptability and performance. Moreover, the network will be extended to simulate traffic scenarios with varying densities of vehicles and pedestrians, increasing its utility for dynamic traffic environments.

Method	City	DINO-Struct	KL Divergence
Ours	London	0.2805	1.0860
	Hongkong	0.2677	1.3936
	Los Angeles	0.3290	1.6416
CycleGAN	London	0.3611	0.6170
	Hongkong	0.4029	0.3995
	Los Angeles	0.3645	0.9067
HEDNGAN	London	0.3840	0.2638
	Hongkong	0.4841	0.4695
	Los Angeles	0.4283	0.6883
Pix2Pix	London	0.6432	4.3092
	Hongkong	0.6786	0.4995
	Los Angeles	0.6211	2.9395
ToDayGAN	London	0.4755	0.6773
	Hongkong	0.5550	2.1409
	Los Angeles	0.5775	1.9629

TABLE II: Comparison of DINO-Struct-Distance and KL Divergence for Different Methods

Supplementary material

Original videos used for network translation, translated videos, frames used for comparison, and a tag of the code are available at https://www.dropbox.com/scl/fo/khvfx0bg4ndh4ktg7eqe8/AO7KhDXHE75OldhHDIBQHMA?rlkey=t7h7np4kmvzz3shyaupkv2zpu. The maintained version of the code is available at https://github.com/Shaadalam9/traffic-pipeline.

References

[1] L. Garay, D. L. Fisher, and K. L. Hancock, "Effects of driving experience and lighting condition on driving performance," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 19. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 2290–2294.

[2] U. Tränkle, C. Gelau, and T. Metker, "Risk perception and age-specific accidents of young drivers," *Accident Analysis & Prevention*, vol. 22, no. 2, pp. 119–125, 1990.

[3] N. A. Siddiqui, X. Chu, and M. Guttenplan, "Crossing locations, light conditions, and pedestrian injury severity," *Transportation research record*, vol. 1982, no. 1, pp. 141–149, 2006.

[4] T. Evans, R. Stuckey, and W. Macdonald, "Young drivers' perceptions of risk and difficulty: Day versus night," *Accident Analysis & Prevention*, vol. 147, p. 105753, 2020.

[5] B. J. Campbell, C. V. Zegeer, H. H. Huang, M. J. Cynecki *et al.*, "A review of pedestrian safety research in the united states and abroad," 2003.

[6] R. Miles-Doan, "Alcohol use among pedestrians and the odds of surviving an injury: evidence from florida law enforcement data," *Accident Analysis & Prevention*, vol. 28, no. 1, pp. 23–31, 1996.

[7] J. M. Sullivan and M. J. Flannagan, "The role of ambient light level in fatal crashes: inferences from daylight saving time transitions," *Accident Analysis & Prevention*, vol. 34, no. 4, pp. 487–498, 2002.

[8] M. J. Allen, "Vision and highway safety," 1970.

[9] W. Ackaah, B. A. Apuseyine, and F. K. Afukaar, "Road traffic crashes at night-time: characteristics and risk factors," *International journal of injury control and safety promotion*, vol. 27, no. 3, pp. 392–399, 2020.

[10] T. Åkerstedt, G. Kecklund, and L.-G. Hörte, "Night driving, season, and the risk of highway accidents," *Sleep*, vol. 24, no. 4, pp. 401–406, 2001.

[11] K. Zhang and M. Hassan, "Crash severity analysis of nighttime and daytime highway work zone crashes," *PloS one*, vol. 14, no. 8, p. e0221128, 2019.

[12] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I 10*. Springer, 2008, pp. 44–57.

[13] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[14] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[15] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," in *Conference on Robot Learning*. PMLR, 2021, pp. 409–418.

[16] A. Patil, S. Malla, H. Gang, and Y.-T. Chen, "The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9552–9557.

[17] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The apolloscape open dataset for autonomous driving and its application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2702–2719, 2020.

[18] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1037–1045.

[19] D. Zhao, J. Weng, and Y. Liu, "Generating traffic scene with deep convolutional generative adversarial networks," in *2017 Chinese Automation Congress (CAC)*. IEEE, 2017, pp. 6612–6617.

[20] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.

[21] S. Tan, K. Wong, S. Wang, S. Manivasagam, M. Ren, and R. Urtasun, "Scenegen: Learning to generate realistic traffic scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 892–901.

[22] C.-T. Lin, S.-W. Huang, Y.-Y. Wu, and S.-H. Lai, "Gan-based day-to-night image style transfer for nighttime vehicle detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 951–963, 2020.

[23] G. Bang, J. Lee, Y. Endo, T. Nishimori, K. Nakao, and S. Kamijo, "Semantic and geometric-aware day-to-night image translation network," *Sensors*, vol. 24, no. 4, p. 1339, 2024.

[24] D. Shiotsuka, J. Lee, Y. Endo, E. Javanmardi, K. Takahashi, K. Nakao, and S. Kamijo, "Gan-based semantic-aware translation for day-to-night images," in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, 2022, pp. 1–6.

[25] A. Somanna, S. M, A. Kodipalli, and T. Rao, "Day and night city view image translations using cycle gan," in *2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2024, pp. 1–5.

[26] A. Punnappurath, A. Abuolaim, A. Abdelhamed, A. Levinshtein, and M. S. Brown, "Day-to-night image synthesis for training nighttime neural isps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 769–10 778.

[27] A. Mohwald, T. Jenicek, and O. Chum, "Dark side augmentation: Generating diverse night examples for metric learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 11 153–11 163.

[28] J. Li, B. Li, Z. Tu, X. Liu, Q. Guo, F. Juefei-Xu, R. Xu, and H. Yu, "Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 205–15 215.

[29] M. Ishikawa, N. Okada, and T. Ikeda, "Image processing methods and the utility of diffusion models for architectural exterior inspection by day and night," in *2024 Australian & New Zealand Control Conference (ANZCC)*. IEEE, 2024, pp. 37–42.

[30] V. Mușat, I. Fursa, P. Newman, F. Cuzzolin, and A. Bradley, "Multi-weather city: Adverse weather stacking for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2906–2915.

[31] X. Cheng, J. Zhou, J. Song, and X. Zhao, "A highway traffic image enhancement algorithm based on improved gan in complex weather conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8716–8726, 2023.

[32] W. Zhou, C. Wang, Y. Ge, L. Wen, and Y. Zhan, "All-day vehicle detection from surveillance videos based on illumination-adjustable

generative adversarial network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3326–3340, 2024.

[33] K. Li, Z. Dai, X. Wang, Y. Song, and G. Jeon, "Gan-based controllable image data augmentation in low-visibility conditions for improved roadside traffic perception," *IEEE Transactions on Consumer Electronics*, 2024.

[34] I. Teeti, V. Musat, S. Khan, A. Rast, F. Cuzzolin, and A. Bradley, "Vision in adverse weather: Augmentation using cyclegans with various object detectors for robust perception in autonomous racing," *arXiv preprint arXiv:2201.03246*, 2022.

[35] H. Lakmal and M. B. Dissanayake, "Improving the visibility at night for advanced driver assistance systems using night-to-day image translation," in *2022 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, 2022, pp. 1–6.

[36] G. Parmar, T. Park, S. Narasimhan, and J.-Y. Zhu, "One-step image translation with text-to-image models," *arXiv preprint arXiv:2403.12036*, 2024.

[37] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1905–1914.

[38] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[39] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[40] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool, "Night-to-day image translation for retrieval-based localization," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5958–5964.

[41] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9650–9660.