Generating Realistic Traffic Scenarios: A Deep Learning Approach Using Generative Adversarial Networks (GANs)

MD SHADAB ALAM*, MARIEKE MARTENS, and PAVLO BAZILINSKYY, Eindhoven University of Technology (TU/e), The Netherlands

Diverse and realistic traffic scenarios are crucial for testing systems and human behaviour in transportation research. Leveraging Generative Adversarial Networks (GANs), this study focuses on video-to-video translation to generate a variety of traffic scenes. By employing GANs for video-to-video translation, the study accurately captures the nuances of urban driving environments, enriching realism and breadth. One advantage of this approach is the ability to model how road users adapt and behave differently across varying conditions depicted in the translated videos. For instance, certain scenarios may exhibit more cautious driver behaviour, while others may involve heavier traffic and faster speeds. Maintaining consistent driving patterns in the translated videos improves their resemblance to real-world scenarios, thereby increasing the reliability of the data for testing and validation purposes. Ultimately, this approach provides researchers and practitioners with a valuable method for evaluating algorithms and systems under challenging conditions, advancing transportation models and automated driving technologies.

Additional Key Words and Phrases: Generative Adversarial Networks(GANs), Future traffic, Deep Learning, Traffic modelling, Diurnal Traffic Behavior

ACM Reference Format:

Md Shadab Alam, Marieke Martens, and Pavlo Bazilinskyy. 2024. Generating Realistic Traffic Scenarios: A Deep Learning Approach Using Generative Adversarial Networks (GANs). 1, 1 (April 2024), 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

1.1 Traffic scenarios in transportation research

In contemporary research projects, data collection efforts encompass diverse traffic scenarios under various conditions, often involving instrumented vehicles equipped with costly sensors such as cameras and LiDARs. Datasets like KITTI [20], NuScenes [7], One Thousand and One Hours [26], Pedestrian Intention Estimation (PIE) [31], Waymo Open Dataset [33], ApolloScape Auto [37], Cityscapes [10], A*3D dataset [30] and Argoverse [8] are benchmarks for numerous computer vision and automated driving-related tasks. These datasets have been used in various studies to learn different insights, such as Kooijman [23], which addresses a research gap regarding the impact of objective in-scene features on driver perceptions during interactions with pedestrians, utilising crowdsourced data and annotations from the Pedestrian Intention Estimation (PIE) dataset to analyse factors such as pedestrian behaviour, vehicle speed, and visual clutter. De Winter conducted a study in which he developed a predictive model for human risk perception in driving scenarios using KITTI Vision Benchmark data and validated it against an online survey, revealing non-linear risk

*Corresponding Author

Authors' address: Md Shadab Alam, m.s.alam@tue.nl; Marieke Martens, m.h.martens@tue.nl; Pavlo Bazilinskyy, p.bazilinskyy@tue.nl, Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, 5612DS.

Manuscript submitted to ACM

perception trends and highlighting the importance of factors like road users' information, vehicle velocity, and road type for model accuracy [14].

Spanning approximately 39.2 km of driving, the KITTI dataset comprises more than 200k 3D object annotations captured in cluttered scenarios captured in urban environments using a vehicle outfitted with cameras, LiDAR, and GPS and IMU sensors. Identifiable individuals are usually absent in these scenes, emphasising a focus on traffic and environmental factors rather than human behaviour analysis. KITTI's restricted focus on traffic scenes and absence of identifiable individuals limit its applicability in studying human behaviour patterns and human factors research.

While comprehensive datasets are mentioned above, they entail significant costs and time investments for data collection and annotation, particularly concerning diverse environmental and climatic conditions. Moreover, these datasets provide limited coverage of nighttime scenarios, hindering the development of robust models for low-light conditions. Furthermore, many datasets, like One Thousand and One Hours, Cityscapes, etc, do not contain any data for night conditions.

Both academic and industrial projects often rely on video content generated by software platforms like Unity3D or Unreal Engine or use footage from public sources on the Internet to conduct experiments. For instance, Bazilinskyy et al. (2023) presented participants with simulations wherein they assumed the roles of cyclists navigating roads alongside automated or non-automated vehicles [4]. Rasouli et al. (2017) curated a dataset prompting participants to predict pedestrian intentions when crossing the road [32]. Evans et al. (2020) investigated driving behaviour disparities between daytime and nighttime conditions [18]. Chen et al. (2019) utilised videos to assess perceived risk under varying weather conditions such as snow or rain [9]. Bazilinskyy et al. (2020) employed crowdsourced YouTube dashcam footage from India, Venezuela, the United States, and Western Europe to assess perceived risk [5]. Oxley et al. (2005) conducted studies wherein videos depicting road-crossing scenarios were presented to participants spanning different age demographics, analysing their hesitation levels in crossing roads [29].

1.2 Generative Adversarial Networks (GANs)

The availability of large datasets and powerful processing components has propelled the advancement of artificial intelligence (AI) in recent years. AI has found application in diverse fields such as drug discovery [6, 24, 25], autonomous controllers [1, 2, 15], and humanities [19, 27], among others. Specifically, within computer vision [34], AI algorithms can analyse and interpret images or videos, aiding in tasks like identification, tracking, and classification, which are integral to daily life. Notably, the introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. (2014) revolutionised data synthesis, enabling the generation of synthetic images and videos that closely mimic real-world data distributions [21]. This breakthrough has further expanded the horizons of AI applications, particularly in domains such as medical imaging, remote sensing, and social media analysis, where access to diverse, high-quality datasets may be limited.

A GAN comprises two neural networks: (1) *generator*, a network that inputs random noise and endeavours to create new data mirroring the actual data, and (2) *discriminator*, a network that evaluates both real data and the data generated by the generator (see Figure 1), aiming to distinguish between them as authentic or synthetic. These networks engage in adversarial training, wherein they vie against each other. The generator strives to enhance its capacity for generating lifelike data, while the discriminator endeavours to refine its skill in discerning fabricated data. This competitive dynamic fosters iterative improvement in both networks over time. Additionally, the core equation of a GAN is defined as follows:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

where G represents the generator, D the discriminator, $p_{\text{data}}(x)$ the distribution of real data, and $p_z(z)$ the distribution of random noise. This equation encapsulates the adversarial training process, wherein the generator aims to minimise this objective function while the discriminator seeks to maximise it, leading to the iterative refinement of both networks.

GANs present a promising alternative as a data source for creating traffic scenarios. GANs can generate realistic traffic scenes through unsupervised learning, offering a cost-effective and efficient means of augmenting existing datasets with diverse scenarios, including nighttime environments and varying traffic densities. Furthermore, the synthetic nature of GAN-generated data helps preserve individual privacy by eliminating identifiable elements and ensuring ethical compliance in research endeavours. This addresses the need for data diversity and scalability in traffic analysis, ultimately enhancing the development and evaluation of machine learning algorithms for real-world applications.

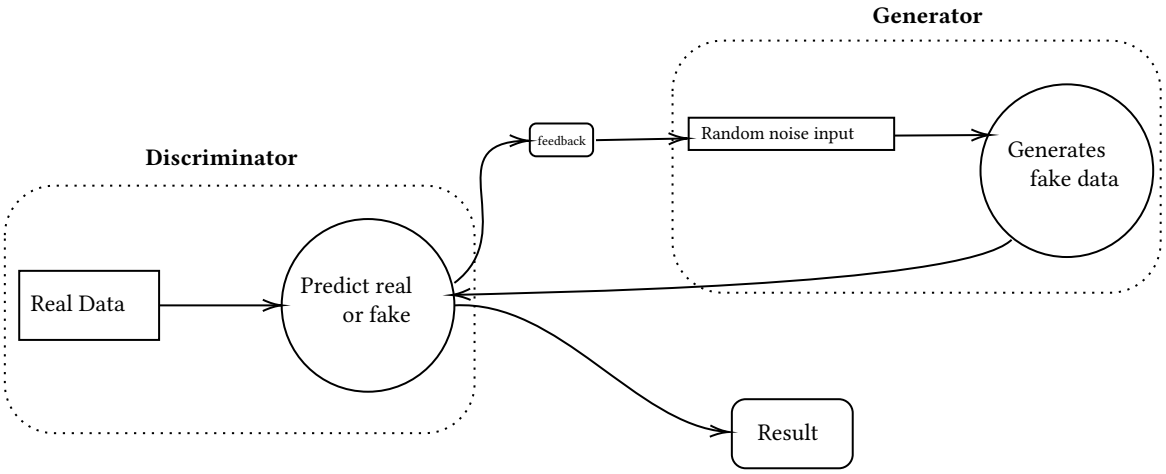

Fig. 1. GANs architecture.

One pivotal aspect contributing to the widespread adoption of GAN frameworks is their ability to address certain limitations inherent in other generative models, such as Variational Autoencoders (VAEs) [22]. Notably, GANs demonstrate superiority in generating high-fidelity images compared to VAEs. While VAEs rely on pixel-wise similarity metrics for reconstruction loss, GANs leverage semantic loss functions [17]. Pixel-wise measures often fail to align with human visual perception, as they may prioritise trivial discrepancies that humans overlook or vice versa. GANs circumvent this issue by implicitly integrating reconstruction loss into the training process via the discriminator's gradient feedback mechanism, which guides the generator towards generating images that are indistinguishable from real ones.

1.2.1 Recycle GAN. Recycle GAN [3], an extension of the vanilla Generative Adversarial Network (GAN) framework, introduces a novel approach to improve the quality and diversity of generated samples while addressing the issue of mode collapse. Mode collapse occurs when a GAN fails to capture the entire data distribution, generating limited and repetitive samples.

In Recycle GAN, the key idea is to recycle previously generated samples to encourage the generator to explore different modes of data distribution. This is achieved through a feedback loop mechanism, where samples generated by

the generator are fed back into the system as input data. By reusing these samples, the generator learns to refine its output distribution iteratively, resulting in a more diverse and realistic set of generated samples.

By incorporating the recycle mechanism, Recycle GAN effectively mitigates mode collapse and encourages the generator to explore diverse regions of the data distribution. Experimental results have demonstrated that Recycle GAN outperforms vanilla GAN regarding sample quality, diversity, and training stability across various datasets.

1.2.2 Unsupervised Recycle GANs. In unsupervised Recycle GANs, Wang et. al[36] modified the Recycle GANs by introducing the unsupervised recycle loss and the unsupervised spatial loss to conduct more accurate and efficient spatiotemporal consistency regularisation. The objective function becomes:

$$Loss = L_{adv} + \lambda_{ur} L_{ur} + \lambda_{us} L_{us} \tag{1}$$

where λ_{ur} and λ_{us} are the weights for the unsupervised losses. L_{adv} is the adversarial loss, L_{ur} is the unsupervised recycle loss and L_{us} is the unsupervised loss.

1.3 Aim of study

This study aims to investigate the potential efficacy of Unsupervised Recycle GANs for transforming traffic scenes, focusing specifically on the conversion of scenes between daytime and nighttime environments. We aim to assess the capability of Recycle GANs to produce realistic traffic scenarios under different lighting conditions through training and evaluation. This research seeks to address two primary challenges: (1) the scarcity of diverse datasets suitable for training machine learning models, particularly in low-light settings, and (2) the potential insights derived from analysing human behavioural patterns across diurnal cycles. Additionally, we endeavour to evaluate the feasibility of Recycle GANs in generating customised traffic videos featuring varied densities of vehicles and pedestrians, with the overarching objective of enhancing the authenticity and cross-cultural relevance of simulated traffic environments. Moreover, we will employ GPT-4V for evaluation, soliciting comments on the generated scenes' realism, thus aiming to enhance the authenticity of simulated traffic environments.

2 METHOD

2.1 Live webcam footage from YouTube used to train GANs

To comprehensively study the interaction between traffic dynamics and environmental variables throughout the day, we employed a dataset sourced from live footage available on YouTube[1] captured on Gangnam Street in Seoul, South Korea. The research was approved by the Human Research Ethics Committee of the Eindhoven University of Technology. The footage included one hour of daytime and one hour of nighttime scenes, recorded on 5 April 2024, 16:00–17:00 (GMT+9) and 5 April 2024, 20:00–21:00 (GMT+9), respectively. We strategically selected these times to maximize pedestrian activity, as the late afternoon typically sees increased foot traffic, while choosing too late in the evening might result in fewer people on the streets, potentially impacting the richness and diversity of the dataset. The videos are available in the supplementary material. To ensure thorough coverage of both daylight and nighttime conditions, we divided the footage into training and validation datasets, dedicating 80% to training and 20% to validation. This dataset served as the foundation for robust training and evaluation of our proposed models, enabling a detailed exploration of traffic behaviour under varying lighting conditions and environmental settings.

[1] https://www.youtube.com/watch?v=JbnJAsk1zII

2.2 Implementation of GANs for traffic scene generation

We used Unsupervised Recycle GANs architecture [36] to train the network. The contemporary difference between recycle GANs and unsupervised recycle GANs is that they incorporate tonal constraints in the learning process, specifically focusing on enhancing the visual quality and realism of the generated images. This distinction is crucial for generating traffic scenes with high fidelity, as it ensures that the synthetic images closely resemble the characteristics of real-world traffic scenarios. Additionally, the utilisation of Recycle-GANs facilitates the preservation of essential features such as vehicle shapes, colours, and movement patterns during the generation process, thus contributing to the overall effectiveness of the framework in simulating realistic traffic dynamics.

2.3 Hyperparameters of the network

Table 1. Hyperparameters

Parameter	Value	Parameter	Value
`-loadSizeW`	542	`-loadSizeH`	286
`-resize_mode`	rectangle	`-crop_mode`	rectangle
`-fineSizeW`	512	`-fineSizeH`	256
`-no_dropout`	True	`-pool_size`	0
`-lambda_spa_unsup_A`	10	`-lambda_spa_unsup_B`	10
`-lambda_unsup_cycle_A`	10	`-lambda_unsup_cycle_B`	10
`-lambda_content_A`	1	`-lambda_content_B`	1
`-batchSize`	1	`-noise_level`	0.001
`-niter_decay`	0	`-niter`	1
`-which_model_netG`	resnet_6blocks		

The training process hinges on a set of hyperparameters delineating the experiment's crucial aspects. Enumerated in Table 1, these parameters encompass key configurations essential for effective training. After this enumeration, each hyperparameter is described in detail to elucidate its role in shaping the training procedure. The source code used to train the network is available in the supplementary material.

(1) `-loadSizeW` and `-loadSizeH`: Specify the width and height of the input images to be loaded during training, respectively.
(2) `-resize_mode` and `-crop_mode`: Determine the resizing and cropping modes employed during preprocessing. In this configuration, both modes are set to `rectangle`.
(3) `-fineSizeW` and `-fineSizeH`: Define the dimensions of the final input images after resizing.
(4) `-which_model_netG`: Specifies the architecture of the generator model. Here, it is set to `resnet_6blocks`, indicating a ResNet-based generator with six residual blocks.
(5) `-no_dropout`: Controls the utilisation of dropout regularisation during training. In this instance, dropout is disabled.
(6) `-pool_size`: Sets the size of the image pool used for storing previously generated images to aid in training stability. It is configured to `0`, indicating no image pool usage.
(7) `-lambda_spa_unsup_A` and `-lambda_spa_unsup_B`: Determine the weights assigned to the spatial unsupervised loss for domains A and B, respectively.

(8) `–lambda_unsup_cycle_A` and `–lambda_unsup_cycle_B`: Specify the weights for the unsupervised cycle consistency loss for domains A and B.
(9) `–lambda_cycle_A` and `–lambda_cycle_B`: Control the weights assigned to the cycle consistency loss for domains A and B when supervision is provided.
(10) `–lambda_content_A` and `–lambda_content_B`: Determine the weights for the content loss for domains A and B, respectively.
(11) `–batchSize`: Specifies the batch size used for training. Here, it is set to 1 for single-image processing per iteration.
(12) `–noise_level`: Defines the level of noise added to input images during training to enhance robustness.
(13) `–niter_decay` and `–niter`: Determine the number of epochs before starting learning rate decay and the total number of training epochs, respectively.

3 RESULTS

After completing the training process, the neural network underwent rigorous testing using footage captured on 13 April 2024. The resulting video showcases a comprehensive comparison between the daytime and nighttime scenarios, capturing a 10-minute sequence from each. This footage was selected randomly from daytime and nighttime sequences to bolster the network's generalisation capabilities, enabling robust performance across diverse environmental conditions. The training utilised 60 minutes of footage, whereas testing employed separate 10-minute sequences captured on distinct days. This approach ensures that the network's performance is evaluated on unseen data, enhancing its ability to generalise to new scenarios and environmental conditions. The videos generated by the GANs are available in the supplementary material. To visualise the efficacy of the trained model, Figure 2 presents a single frame comparison between the daytime and nighttime conditions alongside its transformation through the GANs. Notably, the GANs adeptly transpose scenes from one lighting condition to another, as evidenced by the seamless transition in the transformed images.

Upon obtaining the results, the subsequent step involved assessing the veracity of the images. To achieve this, we utilised GPT-4V [28]. This has been done in numerous research projects (e.g., [11–13, 16, 35]). Driessen et al. (2024) assessed GPT-4V's ability to predict human-perceived risk levels in traffic images, utilising 210 static images rated by approximately 650 individuals[16]. They found that repeating prompts under identical conditions, varying prompt text, and incorporating object detection features alongside GPT-4V-based risk ratings significantly enhance model validity. This resulted in a high correlation coefficient of r = 0.83 between AI predictions and human risk scores, indicating accurate population-level risk prediction and emphasising the importance of prompting GPT-4V similar to human multi-item questionnaire responses.

We inputted 12 distinct scenes into the model and requested its assessment regarding the authenticity of the images. To ensure a comprehensive evaluation and to check the consistency of the response from GPT-4V's we divided the 12 scenes into 2 batches comprising 6 scenes, each containing 3 day and 3 night scenarios. In both cases, GPT-4V was prompted with *"Based on the following criteria, could you determine if these images are artificially created? 1. Uniformity of lighting 2. Shadow behavior 3. Perspective and scale 4. Texture and detail 5. Presence of edge artefacts"*.

These criteria are widely recognised as signs of digital image manipulation or creation. Typically, genuine photographs captured by cameras exhibit consistent lighting, shadows that align with light sources, accurate perspectives and scales, and natural textures. Deviations from these elements within an image usually indicate that it has been edited or completely synthesised. While these indicators are not solely characteristic of fabricated images, they represent

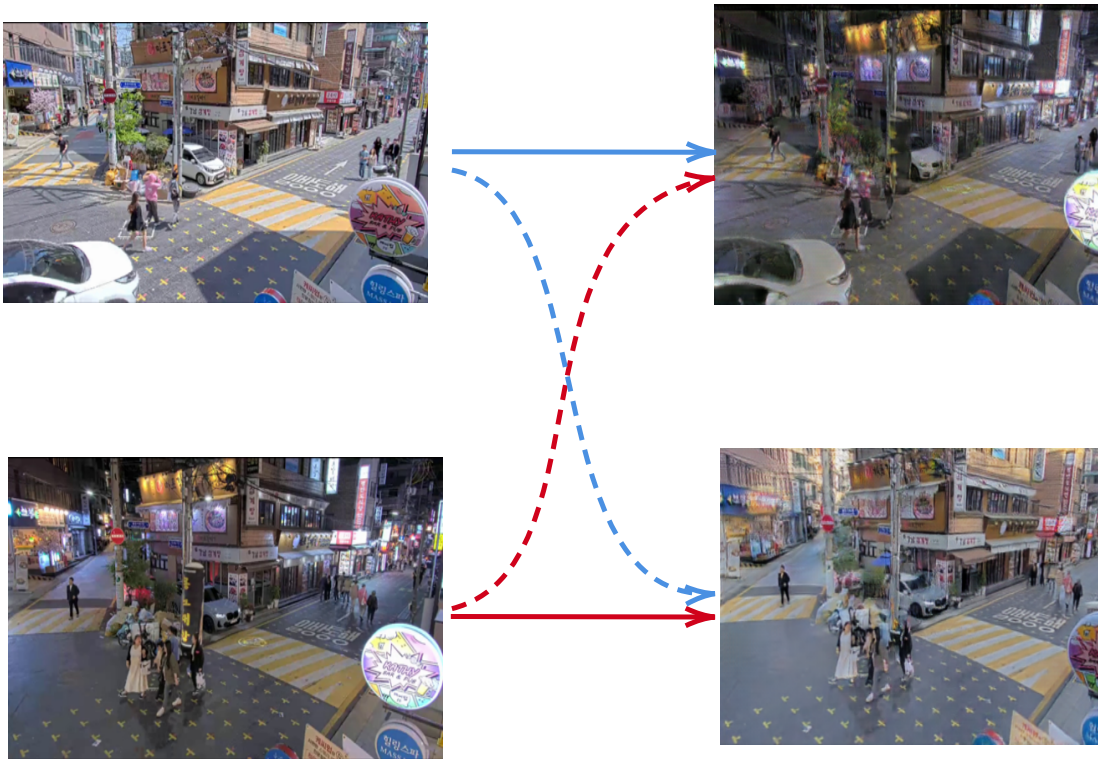

Fig. 2. Dynamic Scene Translation with Recycle GANs: This frame exemplifies the successful application of Recycle GANs in seamlessly transforming day scenes into night and vice versa. The images on the left depict the original footage, while those on the right are generated via GANs. We achieve a compelling reversal of time and illumination through innovative generative techniques, showcasing the potential of GANs in real-world scene adaptation.

anomalies that are generally absent in authentic, unaltered photographs of real-world scenes. In authentic images, all elements are expected to align cohesively with the physics of light and space. During our assessment, we searched for discrepancies from this natural coherence to determine if an image appeared fabricated.

For the initial set of six images, GPT-4V's feedback, as shown in Figure 3a, indicated: "*Based on these observations, there are no definitive indications that these images are artificially created*". Interestingly, GPT-4V identified each image and gave comments on daylight images and nighttime as can be seen in the Figure 3.

Subsequently, we fed another batch of six images—three from daylight and three from nighttime—was analysed. GPT-4V commented on these images stating, "*Overall, these images also do not exhibit clear signs of artificial creation upon visual inspection. They appear to maintain consistent lighting, shadow behaviour, perspective, and detail that one would expect from unaltered photos*". The detailed response is shown in Figure 3b.

4 CONCLUSION AND FUTURE STUDIES

In this study, we have demonstrated the effectiveness of Unsupervised Recycle GANs for traffic scene transformation across different times of the day. Our approach addresses the challenge of the lack of datasets for training machine learning algorithms for transportation research, particularly in low-light conditions such as nighttime scenes. By

1. **Uniformity of Lighting**:
   - In the night images, the lighting does appear to have a consistent source, with illuminated signs and streetlights casting appropriate brightness on nearby objects.
   - In the day images, the sunlight appears natural, casting shadows that match the direction of light typically seen during daytime.
2. **Shadow Behavior**:
   - Shadows in the daytime images are consistent with sunlight direction, and the shadows cast by people and objects in the street appear correct.
   - In the nighttime images, shadows are less visible but where they are seen, they seem to correspond to the artificial light sources.
3. **Perspective and Scale**:
   - Buildings, people, and street elements are in perspective, with no apparent distortion that would suggest artificial manipulation.
   - The scale of objects seems consistent; larger objects are closer, smaller objects are further away, suggesting a correct perspective.
4. **Texture and Detail**:
   - There is a noticeable granularity to the images, especially visible in the road surface and building facades, which suggests that texture is well represented.
   - Details such as street signs, window patterns, and the texture of clothing on people appear natural and varied without obvious repetition.
5. **Presence of Edge Artifacts**:
   - No obvious edge artifacts, such as halos or unnaturally sharp edges, are visible at this resolution, indicating the images might be authentic or well edited.

(a) Response for 1st set of scenes.

1. **Uniformity of Lighting**:
   - There is consistency in the lighting, especially visible in the day images with the sunlight.
   - The nighttime images have varied lighting sources which seem to reflect naturally on the surfaces around them.
2. **Shadow Behavior**:
   - Shadows cast by pedestrians and objects in the day images align with the sun's position.
   - At night, the shadows that are visible seem to be cast by the streetlights appropriately, although it's more challenging to discern due to lower light levels.
3. **Perspective and Scale**:
   - Perspective and scale look consistent. The buildings and street elements decrease in size appropriately as they recede into the distance.
   - People and objects closer to the camera are larger, as would be expected in a three-dimensional space.
4. **Texture and Detail**:
   - The textures, such as those on the road surface and the building facades, are clear and varied.
   - Details are consistent across different elements within each image, suggesting a natural scene.
5. **Presence of Edge Artifacts**:
   - No clear edge artifacts are visible upon this inspection. However, without zooming in and examining at a higher resolution, it's possible to miss subtle signs of digital manipulation.

(b) Response for 2nd set of scenes.

Fig. 3. GPT-4V evaluations of twelve images for signs of artificial creation, examining criteria such as uniformity of lighting, shadow behaviour, perspective and scale, texture and detail, and presence of edge artefacts. Panels (a) and (b) respectively show the model's responses to the first and second sets of images, both during day and night conditions, indicating no clear evidence of artificial manipulation.

leveraging Recycle GANs, we bridge the gap between data availability during day and night scenarios, enhancing the robustness and applicability of traffic analysis models.

In addition to our study, we sought external validation of the generated images by leveraging GPT-4V for qualitative assessment. The feedback from GPT-4V affirmed the high quality and authenticity of the generated images, with no discernible indications of artificial generation. This external validation underscores the robustness and realism of our approach, as the generated images closely resemble real-world counterparts. Such confirmation bolsters confidence in the fidelity and effectiveness of our model, positioning it as a valuable tool for generating authentic traffic scenarios for various applications in transportation engineering and automated driving research.

In future research, this study can be extended to generate custom videos featuring diverse scenarios with varying densities of cars and pedestrians. Additionally, there is scope for integrating cross-cultural perspectives into the training process, encompassing traffic conditions from different regions worldwide. Furthermore, a crucial aspect of enhancement lies in incorporating background sounds, such as traffic honks and ambient noise, into the generated videos. This integration would enhance the realism and immersion of the simulated traffic scenes, offering a more comprehensive dataset for analysis and training of machine learning algorithms. Additionally, exploring techniques for fine-tuning the generated videos to specific cultural and geographical contexts can further enhance the utility and accuracy of the generated traffic simulations. Comparing our approach with OpenAI's SORA[2] would provide valuable insights and contribute to advancing traffic simulation technology.

[2]https://openai.com/sora

5 SUPPLEMENTARY MATERIAL

Videos used to test, train and validate the network, and a tag of the source code in Python can be found at https://www.dropbox.com/scl/fo/wikk927sitse4dc0iwm9v/AAM5hWlKjlCOHkjfr7w587g?rlkey=j4yw3q6q5oipq7gawu0ic5pgl&st=0nyw3yy6&dl=0. A maintained version of the source code is available at https://github.com/Shaadalam9/gans-traffic. A video demonstration of the results is available at https://www.dropbox.com/scl/fo/pli63rbd8bb0z5nv1xtjv/AED5m6pdSf3VgwtKZAl2NO8?rlkey=n3zkml8vp4ch20cvqnceo5h5z&st=o5xz676d&dl=0.

REFERENCES

[1] Alam, M. S. and Carlucho, I. (2023). Harnessing traditional controllers for fast-track training of deep reinforcement learning control strategies.

[2] Alam, M. S., Sudha, S. K. R., and Somayajula, A. (2023). Ai on the water: Applying drl to autonomous vessel navigation. *arXiv*.

[3] Bansal, A., Ma, S., Ramanan, D., and Sheikh, Y. (2018). Recycle-gan: Unsupervised video retargeting. In *Proceedings of the European Conference on Computer vision (ECCV)*, pages 119–135.

[4] Bazilinskyy, P., Dodou, D., Eisma, Y. B., Vlakveld, W., and De Winter, J. C. F. (2023). Blinded windows and empty driver seats: The effects of automated vehicle characteristics on cyclists' decision-making. *IET Intelligent Transport Systems*, 17(1):72–84.

[5] Bazilinskyy, P., Eisma, Y. B., Dodou, D., and De Winter, J. C. F. (2020). Risk perception: A study using dashcam videos and participants from different world regions. *Traffic Injury Prevention*, 21(6):347–353.

[6] Blanco-Gonzalez, A., Cabezon, A., Seco-Gonzalez, A., Conde-Torres, D., Antelo-Riveiro, P., Pineiro, A., and Garcia-Fandino, R. (2023). The role of ai in drug discovery: Challenges, opportunities, and strategies. *Pharmaceuticals*, 16(6):891.

[7] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O. (2020). Nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11621–11631.

[8] Chang, M.-F., Lambert, J., Sangkloy, P., Singh, J., Bak, S., Hartnett, A., Wang, D., Carr, P., Lucey, S., Ramanan, D., et al. (2019). Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757.

[9] Chen, C., Zhao, X., Liu, H., Ren, G., and Liu, X. (2019). Influence of adverse weather on drivers' perceived risk during car following based on driving simulations. *Journal of Modern Transportation*, 27:282–292.

[10] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3213–3223.

[11] De Winter, J. C. F. (2023). Can chatgpt pass high school exams on english language comprehension? *International Journal of Artificial Intelligence in Education*, pages 1–16.

[12] De Winter, J. C. F. (2024). Can chatgpt be used to predict citation counts, readership, and social media interaction? an exploration among 2222 scientific abstracts. *Scientometrics*, pages 1–19.

[13] De Winter, J. C. F., Dodou, D., and Stienen, A. H. A. (2023a). Chatgpt in education: Empowering educators through methods for recognition and assessment. In *Informatics*, volume 10, page 87. MDPI.

[14] De Winter, J. C. F., Hoogmoed, J., Stapel, J., Dodou, D., and Bazilinskyy, P. (2023b). Predicting perceived risk of traffic scenes using computer vision. *Transportation Research Part F: Traffic Psychology and Behaviour*, 93:235–247.

[15] Deraj, R., Kumar, R. S., Alam, M. S., and Somayajula, A. (2023). Deep reinforcement learning based controller for ship navigation. *Ocean Engineering*, 273:113937.

[16] Driessen, T., Dodou, D., Bazilinskyy, P., and De Winter, J. C. F. (2024). Putting chatgpt vision (gpt-4v) to the test: Risk perception in traffic images. *Royal Society Open Science*.

[17] El-Kaddoury, M., Mahmoudi, A., and Himmi, M. M. (2019). Deep generative models for image generation: A practical comparison between variational autoencoders and generative adversarial networks. In *Proceedings of 5th International Conference on Mobile, Secure, and Programmable Networking*, pages 1–8. Springer.

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

[19] Gefen, A., Saint-Raymond, L., and Venturini, T. (2021). Ai for digital humanities and computational social sciences. *Reflections on Artificial Intelligence for Humanity*, pages 191–202.

[20] Geiger, A., Lenz, P., and Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361.

[21] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

[22] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv*.

[23] Kooijman, B. (2021). The identification of factors affecting drivers' perceived risk in pedestrian-vehicle interaction: A crowdsourcing study.

[24] Mak, K.-K. and Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and future prospects. *Drug Discovery Today*, 24(3):773–780.

[25] Mak, K.-K., Wong, Y.-H., and Pichika, M. R. (2023). Artificial intelligence in drug discovery and development. *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays*, pages 1–38.

[26] Mao, J., Niu, M., Jiang, C., Liang, H., Chen, J., Liang, X., Li, Y., Ye, C., Zhang, W., Li, Z., et al. (2021). One million scenes for autonomous driving: Once dataset. *arXiv*.

[27] Nourbakhsh, I. R. and Keating, J. (2020). *AI and Humanity*. MIT press.

[28] OpenAI (2023). Gpt-4 technical report. *arXiv*.

[29] Oxley, J. A., Ihsen, E., Fildes, B. N., Charlton, J. L., and Day, R. H. (2005). Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, 37(5):962–971.

[30] Pham, Q.-H., Sevestre, P., Pahwa, R. S., Zhan, H., Pang, C. H., Chen, Y., Mustafa, A., Chandrasekhar, V., and Lin, J. (2020). A* 3d dataset: Towards autonomous driving in challenging environments. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2267–2273. IEEE.

[31] Rasouli, A., Kotseruba, I., Kunic, T., and Tsotsos, J. K. (2019). Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6262–6271.

[32] Rasouli, A., Kotseruba, I., and Tsotsos, J. K. (2017). Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 206–213.

[33] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al. (2020). Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454.

[34] Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.

[35] Tabone, W. and De Winter, J. C. F. (2023). Using chatgpt for human–computer interaction research: a primer. *Royal Society Open Science*, 10(9):231053.

[36] Wang, K., Akash, K., and Misu, T. (2022). Learning temporally and semantically consistent unpaired video-to-video translation through pseudo-supervision from synthetic optical flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2477–2486.

[37] Wang, P., Huang, X., Cheng, X., Zhou, D., Geng, Q., and Yang, R. (2019). The apolloscape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.