Enhancing Nighttime Vehicle Detection via Transformer-based Data Augmentation

Min Young Lim[†§], Seong Hee Park[†§], Soo-Hyun Lee[†§], Tae Hyung Kim[†], Dongwoo Kang[‡], and Youn Kyu Lee[†*]
Department of Computer Engineering[†], *Department of Electronic & Electrical Engineering*[‡]
Hongik University
Seoul, Republic of Korea
minyounge17@g.hongik.ac.kr, tjdgml0401@g.hongik.ac.kr, lsh01223@g.hongik.ac.kr,
taehyung@hongik.ac.kr, dkang@hongik.ac.kr, younkyul@hongik.ac.kr

***Abstract*—In autonomous driving systems, vehicle detection technology typically relies on object detection models trained on driving image datasets. However, accurate vehicle detection becomes challenging during nighttime due to low-light conditions, necessitating a sufficient amount of nighttime driving images for training the model. Unfortunately, publicly available datasets lack an adequate amount of nighttime driving images, and collecting them directly is cost-ineffective. In this paper, we propose a novel augmentation method based on transformer to convert daytime driving images into realistic nighttime driving images. Our method analyzes the style case of the given daytime driving image, selects a tailored style image that corresponds to the analyzed style case, and transfers the daytime driving image into the realistic nighttime driving image using the selected style image. Our diverse range of evaluations demonstrates the effectiveness of our proposed method in augmenting realistic nighttime driving images.**

***Index Terms*—vehicle detection, style transfer, transformer, data augmentation**

I. Introduction

Autonomous driving systems rely on the detection of various objects in road images, such as vehicles and traffic lights, captured by the vehicle's forward camera. In particular, vehicle detection is a key technology in autonomous driving, typically accomplished using object detection models trained on a variety of driving image datasets [1], [2]. However, the majority of vehicle detection methods primarily focus on daytime driving images, posing a significant challenge for nighttime vehicle detection [3]. Distinguishing vehicles from the background at night becomes difficult due to factors like underexposure, noise, and low brightness, which obscure the vehicle's appearance [2]–[4]. Therefore, a sufficient quantity of nighttime driving images is necessary to train object detection models that support robust nighttime vehicle detection. However, current publicly available datasets lack nighttime driving images in comparison to their daytime driving images [5]–[7]. While collecting additional nighttime driving images could mitigate this problem, it would be cost-ineffective to gather diverse datasets that remain unbiased towards factors affecting vehicle detection (e.g., road shapes, surrounding landscapes, and weather).

[§]These authors contributed equally to the work.
[*]Corresponding Author

To address this issue, various methods have been proposed to augment nighttime driving images using Generative Adversarial Networks (GAN), a deep learning-based method for image generation [2], [8]. However, existing methods have a limitation in accurately presenting the "realistic" nighttime scenes in the generated images [8]. Specifically, in low-light conditions at night, the illumination information from the headlights and taillights of surrounding vehicles plays a vital role in detecting vehicles. However, the existing methods may generate light in improper locations (e.g., on trees or clouds) [8], and they can also produce incorrect colors of light on the front-lamp and rear-lamp of vehicles (e.g., red light on the headlamps) [9]. As a consequence, models trained on such unrealistic nighttime driving images may exhibit reduced performance in nighttime vehicle detection. Hence, it is required to design more advanced style transfer techniques to generate realistic nighttime driving images.

In this paper, we propose a novel method to augment nighttime driving datasets effectively using the transformer. The transformer utilizes attention mechanisms and is known for its superior performance in style transfer compared to GAN [10]. By leveraging the transformer's ability to extract key features and apply them to input images, we can generate highly realistic nighttime driving images. The transformer-based style transfer process involves extracting features from both the input image and the style image, resulting in significant variations in the transferred image's reality depending on the style image. To capture different lighting styles, we define three style cases based on the types of lamps present in the vehicle images: front-lamps only (*S1*: front-lamp), both front-lamps and rear-lamps (*S2*: front+rear lamp), and rear-lamps only (*S3*: rear-lamp). Our proposed method analyzes the style case of the input image and utilizes a tailored nighttime driving-style image (TND-style image) as the style image for realistic transfer. For example, when the input image falls under the *S3* type (=rear-lamp), using the corresponding TND-style image ensures a more accurate depiction of the rear-lamp, including improved clarity, correct light position, and correct color. By transferring daytime driving images into realistic nighttime driving images, our method effectively augments datasets, ultimately enhancing the nighttime vehicle detection performance of object detection models in autonomous driving systems.

The contributions of this paper are as follows: (1) We propose a novel method for augmenting nighttime driving datasets to enhance the performance of vehicle detection models; (2) We introduce an effective data augmentation approach utilizing transformer-based style transfer, employing suitable styles for generating realistic nighttime driving images; (3) We validate our proposed method using real-world datasets and evaluate its effectiveness in improving the performance of nighttime vehicle detection models.

This paper is organized as follows. Related work is presented in Section 2. Section 3 describes our proposed method. Evaluation is presented in Section 4 and the conclusion is discussed in Section 5.

II. Related Work

A. Neural Style Transfer

Neural style transfer is a computer vision technique that projects the style of one image onto another, to transfer only the style while preserving the content of the image [11].

A number of style transfer methods using GAN have been proposed. Zhu et al. [11] proposed CycleGAN, which transfers the style of the source image into the style of the target image without requiring a set of image pairs. Karras et al. [12] proposed StyleGAN, which generates realistic high-resolution images by distinguishing styles using the disentangled latent space.

Recently, style transfer methods using transformer have been proposed, outperforming GAN in terms of generating high-quality images. Deng et al. [13] proposed the first baseline for a style transfer approach using transformer. Their method generates high-quality images with specific style patterns by leveraging semantic features regardless of the image size. Zhang et al. [9] proposed a style transfer method using transformer that incorporated a strips windows attention mechanism. This mechanism performs attention by combining image features in both horizontal and vertical directions, resulting in the generation of more natural images with smoother connections between image regions.

B. Day-to-night

Recently, a number of research studies have focused on transferring daytime driving images into nighttime driving images using the style transfer mechanism. Arruda et al. [14] proposed a method utilizing CycleGAN to transfer daytime driving images into nighttime driving images. Lin et al. [15] proposed a method utilizing AugGAN to recognize the structure in daytime driving images and transform them into nighttime driving images. Previous studies have primarily focused on utilizing GANs to transform daytime images into nighttime images. However, GAN-based studies have the drawback of potentially introducing unrealistic artifacts in the transferred images. To improve the quality of the day-to-night transferred images, it is required to incorporate advanced style transfer algorithms.

III. Proposed Method

In this paper, we propose a novel method to effectively augment nighttime driving datasets by applying a tailored style for nighttime driving images to daytime driving images. Fig. 1 shows an overview of our proposed method, which comprises three main stages: *Style Case Analysis*, *TND-Style Selection*, and *Style Transfer*. In the *Style Case Analysis* stage, our method analyzes the style case of the input daytime driving image based on the types of lamps present in the image. Then, in the *TND-Style Selection* stage, our method selects a tailored nighttime driving-style image (TND-style image) that corresponds to the analyzed style case. Next, in the *Style Transfer* stage, our method transfers the input daytime driving image into a nighttime driving image using the selected TND-style image. Our proposed method enables effective augmentation of nighttime driving datasets by transferring daytime driving images into "realistic" nighttime driving images. By utilizing the augmented nighttime driving images as a training dataset for vehicle detection models, it has the potential to significantly improve the performance of nighttime vehicle detection. Detailed explanations of each stage are provided as follows.

A. Style Case Analysis

Autonomous driving images may contain various objects, including vehicles, traffic lights, and pedestrians. However, during nighttime, distinguishing the appearance of objects from the background becomes challenging due to low-light. Therefore, the light emitted from objects becomes a crucial feature in identifying and differentiating objects. In particular, the light emitted from the front-lamps and rear-lamps is a significant feature for detecting vehicles at night [4]. In driving images, there are cases where only the front or rear lights of a vehicle are visible. Moreover, in situations where two lanes coexist on the road, there are also instances where both front and rear lights are visible simultaneously.

In this paper, we defined three style cases based on the types and positions of lamps in the vehicle: front-lamp (*S1*), front+rear lamp (*S2*), and rear-lamp (*S3*). *S1* represents the case where only the front-lamps of the vehicle (e.g., head-lamp and fog-lamp), are present in the image, predominantly appearing in a natural white color. *S3* represents the case where only the rear-lamps of the vehicle (e.g., tail-lamp and brake-lamp) are present in the image, predominantly appearing in a red color. *S2* represents the case where both the front-lamps and rear-lamps of the vehicle are present in the image, with the natural white color of the front-lamps and the red color of the rear-lamps visible on both sides. Based on the defined style cases, our method analyzes the input daytime driving image to determine the appropriate style case in the *Style Case Analysis* stage.

B. TND-Style Selection

Since the transformer model employed in this paper applies the features extracted from a style image to an input image, the reality of the output image varies depending on the suitability

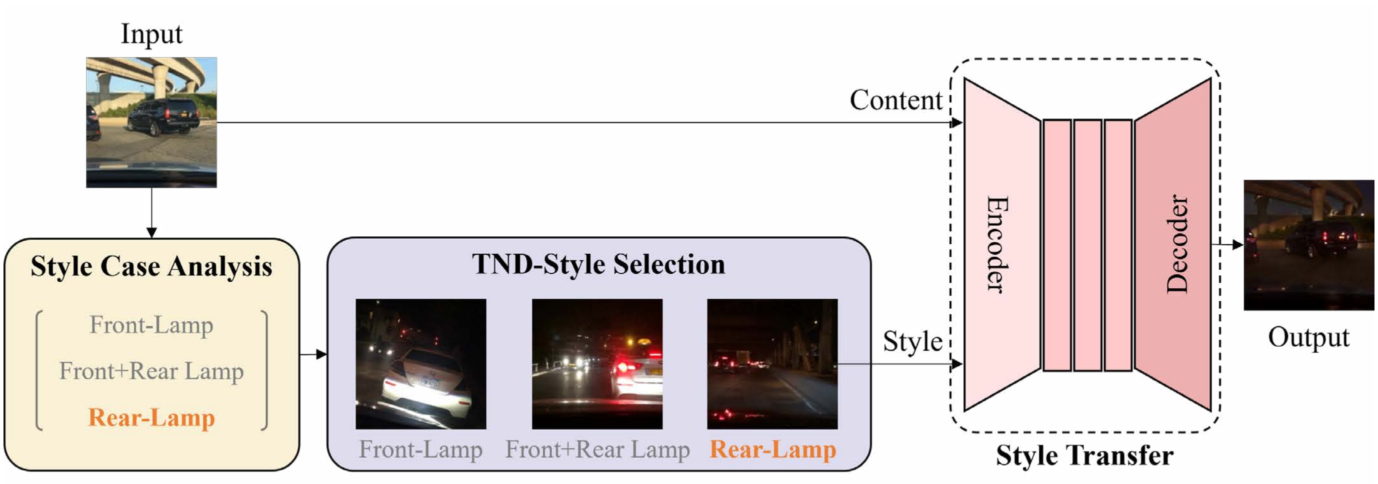

Fig. 1. Overview of Our Proposed Method

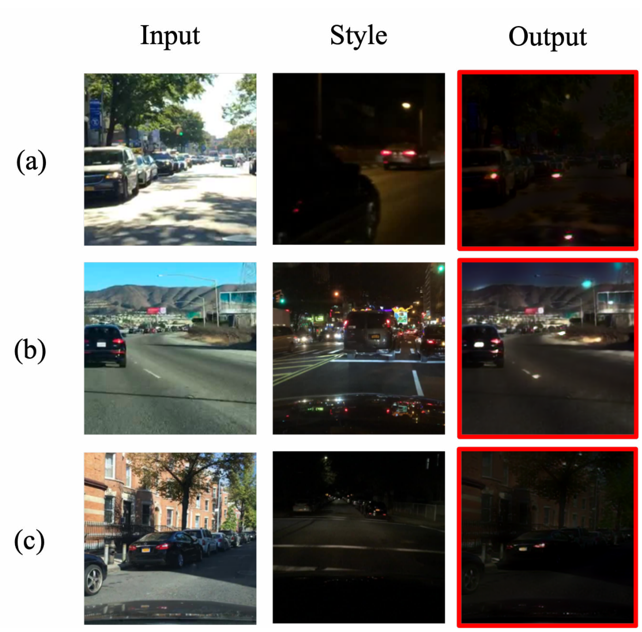

Fig. 2. Examples of Style Transfer: (a) Negative Example of Color Match, (b) Negative Example of Realistic Image, and (c) Negative Example of Color Visibility

of the style image [9]. For example, as depicted in Fig. 2 (a), if the input image is a daytime driving image featuring only front-lamps, and the style image is a nighttime driving image showing only rear-lamps, the front-lamp in the output image may be transformed into red color instead of the expected natural white color, which is not realistic. In the *TND-Style Selection* stage, our method selects a TND-style image specific to each style case in order to transfer the daytime driving image into a realistic nighttime driving image based on the following criteria.

(1) External lights: As shown in Fig. 2 (b), using style images that include both external lights (e.g., traffic lights, street lamps, and building lights) and vehicle lights may result in unnecessary artifacts (e.g., traffic light beams generated in mid-air and street lamp lights appearing in the middle of the road) in the output image. To generate a realistic nighttime driving image by minimizing such artifacts, style images with external lights are excluded. (2) Colorimetry: As shown in Fig. 2 (c), if the contrast of the style image is insufficient, indistinct vehicle lights (i.e., front-lamp and rear-lamp) may be generated in the output image. To generate clear vehicle lights in the output image, style images with a standard deviation of pixel value (value of HSV) ranging from 40 to 80, which provides sufficient contrast in terms of visual quality [16], [17], are selected. (3) Vehicle lights: As shown in Fig. 2 (a), if the vehicle lights in the input image and style image have different colors, it may result in generating inappropriate colors for the vehicle lights in the output image. Furthermore, if the key feature of nighttime images, such as the glare, is not incorporated into the style image, it may result in an unrealistic nighttime image generation. Therefore, to generate realistic vehicle lights in the output image, style images that meet the following conditions for the included vehicle lights are selected: ensuring that the light colors correspond to the respective style cases, verifying that the light sizes are visually distinguishable, and confirming the presence of the glare in the lights.

C. Style Transfer

Since the nighttime driving image includes local features (i.e., front-lamp and rear-lamp) and global features (i.e., low-brightness background), it is essential to transfer both of these elements realistically. S2WAT (Strips Window Attention Transformer), a *Style Transfer* method that utilizes a transformer with a strips window attention mechanism, enables the realistic application of both local and global features in generating images [9]. Therefore, we employ S2WAT for *Style Transfer* on the given driving image. The trained S2WAT model generates an image that preserves the form of the given content image while incorporating only the style of the style image.

To transfer a daytime driving image into the style of a nighttime driving image, the S2WAT model is trained using a clear weather daytime driving image as the content image and a clear weather nighttime driving image as the style image. This choice is made because the clear weather images effectively capture the distinct differences between daytime and nighttime conditions. As depicted in Fig. 1, when a trained model is provided with a daytime driving image as the content image and a TND-style image as the style image, it effectively transforms the image to realistically depict the nighttime scene while preserving the structure of the daytime image. The model extracts both local features (i.e., the vehicle lamp) and global features (i.e., glare and dark sky) commonly observed during nighttime from the TND-style image, and applies these extracted features to the content image.

IV. Evaluation

To assess the effectiveness of our proposed method, we addressed the following research questions:

- **RQ#1**. Does our proposed method generate appropriate lamp light in the transformed images?
- **RQ#2**. Does the TND-style image improve nighttime vehicle detection performance compared to the uni-style image?
- **RQ#3**. Does our proposed method improve nighttime vehicle detection performance compared to the existing augmentation method?

To evaluate the effectiveness of our method, we selected the BDD100K dataset [5], which comprises diverse driving images captured under various geographical, environmental, and weather conditions. To mitigate the influence of weather, we curated the dataset by including only daytime and nighttime driving images captured in clear weather conditions. We then divided the dataset into a training set and a transferring set, consisting of 36,614 and 4,239 images, respectively. The images were center-cropped to 720 $\times$ 720 and subsequently resized to 224 $\times$ 224. For style transfer, we carefully selected a TND-style image for each style case and performed the style transfer accordingly. To assess the effectiveness of our augmentation method, we employed the YOLOv5 model [18] to evaluate nighttime vehicle detection performance.

Our experiments were conducted using a GPU (NVIDIA GeForce RTX 3090), Python 3.8.10, and PyTorch 2.0.0. Each model was trained using the following hyperparameters: (1) S2WAT: Adam optimizer, a learning rate of 0.0001, and a batch size of 4, and (2) YOLOv5: Momentum optimizer, a learning rate of 0.01, and a batch size of 16.

A. Evaluating the visual quality of transformed images

Case-based analysis: To verify whether the lamp light was properly generated in the transformed nighttime driving image, we conducted evaluations for each style case. We performed style transfer on all style cases using the TND-style images and compared the resulting transferred images. The transferred images were evaluated based on the following criteria. (1) Whether the position of the generated light matched the position of the *front* or rear-lamp in the original image, and (2) Whether the color of the generated light corresponded to each style case (i.e., front-lamp: natural white and rear-lamp: red).

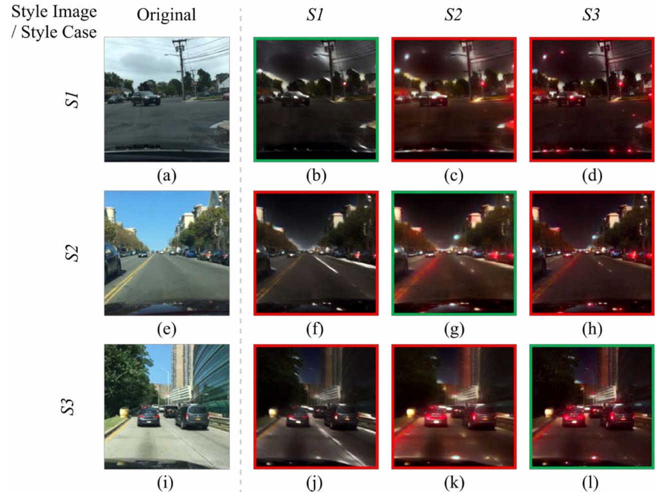

Fig. 3. Style Transferred Results for All Style Cases using Fit Style Images

Fig. 3 presents the selected nighttime driving images from the transfer dataset, along with the results of their transformation using TND-style images for all style cases. When the daytime driving image (a) of Style 1 (*S1*) is transferred into a nighttime driving image using the TND-style image of *S1*, as shown in (b), it can be observed that the natural white light is correctly generated at the front-lamp position of the vehicle. However, when transferred using the TND-style image of *S3*, as shown in (d), it can be seen that the red light is erroneously generated at the front-lamp position, which is incorrect.

Similarly, for Style 2 (*S2*), when the daytime driving image (e) of *S2* is transferred into a nighttime driving image using the TND-style image of *S2*, as shown in (g), it can be observed that the natural white light is correctly generated at the front-lamp position and the red light is correctly generated at the rear-lamp position. However, when transferred using the TND-style image of *S1*, as shown in (f), it can be seen that the natural white light is erroneously generated at the rear-lamp position, which is incorrect. Furthermore, when transformed using the TND-style image of Style 3 (*S3*), as shown in (h), it can be observed that the red light is erroneously generated at the front-lamp position, which is incorrect.

For *S3*, when the daytime driving image (i) of *S3* is transformed into a nighttime driving image using the TND-style image of *S3*, as shown in (l), it can be observed that the red light is appropriately generated at the rear-lamp position. However, when transferred using the TND-style image of *S1*, as shown in (j), it can be seen that no light is generated at the rear-lamp position. The results demonstrate that when performing style transfer using the TND-style image corresponding to each style case, the light is generated more accurately at the correct positions and with the appropriate colors. This indicates that our method enables the generation of realistic nighttime driving images with specific styles.

TABLE I
EVALUATION CRITERIA AND AVERAGE SCORES OF TRANSFORMED IMAGES

No.	Evaluation Criteria	Description	Average Scores	
			TND-style	**Uni-style**
1	The presence of front-lamp or rear-lamp light	Participants were asked to evaluate whether the vehicle lights are generated at the front-lamp or rear-lamp position.	**4.62**	4.19
2	The color of front-lamp or rear-lamp light	Participants were asked to evaluate whether the color of the vehicle lights corresponded to the style case (front-lamp: natural white, rear-lamp: red).	**4.41**	4.23
3	The size of front-lamp or rear-lamp light	Participants were asked to evaluate whether the size of the generated light resembled the size of the front-lamp or rear-lamp.	**4.15**	4.07
4	The perception of the transformed images as nighttime scenes	Participants were asked to evaluate whether the transformed images were perceived as nighttime scenes.	**4.17**	3.91

User-based analysis: To verify the accurate generation of lamp light in the transferred nighttime driving images, we conducted a user study. The study involved ten computer science students and followed the procedure outlined below. Firstly, the participants were presented with two series of 15 images each. One series includes the transferred images obtained by selecting images from the transferring set and applying TND-style images, while the other series consists of images transferred using uni-style images. To minimize participant bias, the order of the images was randomly assigned, and the participants were not informed whether the style image used was a TND-style image or a uni-style image. For each image, the participants were asked to rate it on a 5-point scale according to the provided evaluation criteria as shown in Table I, which ranged from 1 (strongly disagree) to 5 (strongly agree). Table I presents the description of each evaluation criterion and the average scores rated by the users for both style images. The results demonstrate that our proposed method outperforms the uni-style-based transformation method by achieving the highest average scores across all evaluation criteria. This indicates the superiority of our proposed method over the existing method in accurately capturing the appropriate position, color, and size of the front-lamp and rear-lamp light.

B. Validating the effectiveness of TND-style images

To evaluate the effectiveness of TND-style images, we compared the nighttime vehicle detection performance among four YOLOv5 models: one model trained on images transferred using TND-style images and three models trained on images transferred using each type of uni-style images (i.e., *S1*, *S2*, and *S3*). Each YOLOv5 model was trained using images that were transformed by the transformer from the 4,239 images in the transferring set. Each YOLOv5 model was tested using 1,413 images randomly selected from the nighttime driving images in clear weather conditions, which also served as the training set for the transformer. The entire dataset was divided into a training set and a testing set in a 3:1 ratio.

Table II presents a comparison of nighttime vehicle detection performance between a model trained on images transferred using TND-style images and models trained on images transferred using uni-style images. As shown in Table II, the model trained on TND-style images exhibited the highest detection performance with an mAP@50 of 69.3%, demonstrating a 2.9% improvement compared to the model trained on *S1*-style images. Furthermore, the model trained on TND-style images also achieved the highest performance with an mAP@50-95 of 38.8%. These results demonstrate that our proposed TND-style images enables a superior improvement in nighttime vehicle detection performance.

TABLE II
PERFORMANCE COMPARISON BETWEEN DIFFERENT STYLE MODELS (TND, *S1*, *S2*, AND *S3*)

Model	TND	*S1*	*S2*	*S3*
mAP@50	**69.3%**	66.4%	64.1%	63.8%
mAP@50-95	**38.8%**	37.9%	36.1%	35.8%

C. Validating the augmentation performance of our proposed method

To assess the augmentation effect of our proposed method, we conducted a comparison of the improvement in nighttime vehicle detection performance among three YOLOv5 models: a model without augmentation, a model with our augmentation, and a model with an existing augmentation technique. The model without augmentation was trained on 1,000 nighttime driving images carefully selected from the training dataset used for training our transformer. The models with augmentation (specifically, a model with our augmentation and a model with an existing augmentation) were trained using a combined dataset of 5,239 images. A model without augmentation was trained using 1,000 nighttime driving images in clear weather conditions, which were selected from the BDD100K dataset. On the other hand, models with augmentation (i.e., a model with our augmentation and a model with existing augmentation) were trained using a total of 5,239 images, comprising 1,000 selected images and 4,239 images transferred by the transformer from the transferring set. All models were tested using 1,746 images randomly selected from nighttime driving

TABLE III
PERFORMANCE COMPARISON BETWEEN A MODEL WITHOUT AUGMENTATION, A MODEL WITH OUR AUGMENTATION, AND A MODEL WITH EXISTING AUGMENTATION

Model	W/O Augmentation	W/ Our Augmentation	W/ Existing Augmentation
# of Training Data (Real / Aug)	1,000 (1,000 / 0)	5,239 (1,000 / 4,239)	5,239 (1,000 / 4,239)
mAP@50	86.5%	**87.5%**	84%
mAP@50-95	54.3%	**56.2%**	53.2%

images in clear weather conditions, which also served as the training set for the transformer. The entire dataset for the models with augmentation was divided into a training set and a testing set in a 3:1 ratio.

Table III presents a performance comparison among a model without augmentation, a model with our augmentation, and a model with existing augmentation. The model with our augmentation achieved an mAP@50 of 87.5% and an mAP@50-95 of 56.2%. In contrast, the model without augmentation exhibited an mAP@50 of 86.5% and an mAP@50-95 of 54.3%. These results indicate that our augmentation improves vehicle detection performance. Notably, the model with existing augmentation displayed even lower performance, with an mAP@50 of 84% and an mAP@50-95 of 53.2%, compared to the model without augmentation. These findings demonstrate that while augmenting the training data does not guarantee performance improvement, our proposed method successfully enhances the detection performance.

V. Conclusion

In this paper, we propose a novel augmentation method for generating realistic nighttime driving images using a transformer-based approach. Our proposed method analyzes three style cases from the input daytime driving images and utilizes TND-style images that represent each style case as the style image, enabling the generation of realistic nighttime driving images. The experimental results demonstrate that the images transferred using our proposed method are not only visually realistic but also significantly enhance the performance of the nighttime vehicle detection model compared to existing data augmentation methods.

Our future work includes incorporating additional factors that may impact nighttime vehicle detection, such as traffic lights and street lamps, into the style cases. Furthermore, we plan to fully automate the analysis of style cases and selection of TND-style images to facilitate efficient augmentation.

Acknowledgment

This work was supported partly by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2022R1F1A1074786) and partly by 2023 Hongik University Research Fund.

References

[1] A. Asvadi, L. Garrote, C. Premebida, P. Peixoto, and U. J. Nunes, "Multimodal vehicle detection: fusing 3d-lidar and color camera data," *Pattern Recognition Letters*, vol. 115, pp. 20–29, 2018.

[2] 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.

[3] L. Fu, H. Yu, F. Juefei-Xu, J. Li, Q. Guo, and S. Wang, "Let there be light: Improved traffic surveillance via detail preserving night-to-day transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8217–8226, 2021.

[4] J. Li, D. Xiao, and Q. Yang, "Efficient multi-model integration neural network framework for nighttime vehicle detection," *Multimedia Tools and Applications*, vol. 81, no. 22, pp. 32675–32699, 2022.

[5] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2636–2645, 2020.

[6] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11621–11631, 2020.

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

[8] V. F. Arruda, T. M. Paixao, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos, "Cross-domain car detection using unsupervised image-to-image translation: From day to night," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2019.

[9] C. Zhang, J. Yang, L. Wang, and Z. Dai, "S2wat: Image style transfer via hierarchical vision transformer using strips window attention," *arXiv preprint arXiv:2210.12381*, 2022.

[10] K. Lee, H. Chang, L. Jiang, H. Zhang, Z. Tu, and C. Liu, "Vitgan: Training gans with vision transformers," *arXiv preprint arXiv:2107.04589*, 2021.

[11] 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*, pp. 2223–2232, 2017.

[12] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4401–4410, 2019.

[13] Y. Deng, F. Tang, W. Dong, C. Ma, X. Pan, L. Wang, and C. Xu, "Stytr2: Image style transfer with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11326–11336, 2022.

[14] V. F. Arruda, T. M. Paixao, R. F. Berriel, A. F. De Souza, C. Badue, N. Sebe, and T. Oliveira-Santos, "Cross-domain car detection using unsupervised image-to-image translation: From day to night," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2019.

[15] 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.

[16] M. Chen, L. Jin, Y. Jiang, L. Gao, F. Wang, X. Xie, *et al.*, "Study on leading vehicle detection at night based on multisensor and image enhancement method," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[17] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "Statistics of visual representation," in *Visual information processing XI*, vol. 4736, pp. 25–35, SPIE, 2002.

[18] G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, L. Diaconu, J. Poznanski, L. Yu, P. Rai, R. Ferriday, *et al.*, "ultralytics/yolov5: v3.1 - bug fixes and performance improvements," *Zenodo*, 2020.