Research article

Enhancing nighttime vehicle detection with day-to-night style transfer and labeling-free augmentation

Yunxiang Yang, Hao Zhen, Yongcan Huang and Jidong J. Yang[*]

Smart Mobility and Infrastructure Lab, College of Engineering, University of Georgia, Athens, GA 30602, USA

* **Correspondence:** Email: Jidong.Yang@uga.edu.

Academic Editor: Chih-Cheng Hung

Abstract: Deep learning-based object detection models perform well under daytime conditions but face significant challenges at night, primarily because they are predominantly trained on daytime images. Additionally, training with nighttime images presents another challenge: Even human annotators struggle to accurately label objects in low-light conditions. This issue is particularly pronounced in transportation applications, such as detecting vehicles and other objects of interest on rural roads at night, where street lighting is often absent, and headlights may introduce undesirable glare. In this study, we addressed these challenges by introducing a novel framework for labeling-free data augmentation, leveraging synthetic data generated by the *Car Learning to Act* (CARLA) simulator for day-to-night image style transfer. Specifically, the framework incorporated the efficient attention Generative Adversarial Network for realistic day-to-night style transfer and used CARLA-generated synthetic nighttime images to help the model learn the vehicle headlight effect. To evaluate the efficacy of the proposed framework, we fine-tuned the state-of-the-art object detection model with an augmented dataset curated for rural nighttime environments, achieving significant improvements in nighttime vehicle detection. This novel approach was simple yet effective, offering a scalable solution to enhance deep learning-based detection systems in low-visibility environments and extended the applicability of object detection models to broader real-world contexts.

Keywords: nighttime vehicle detection; low visibility; headlight glare; generative adversarial networks; image style transfer; rural environments; CARLA; synthetic data; data augmentation

1. Introduction

Accurate and reliable vehicle detection is essential for many transportation applications, such as traffic monitoring and incident management. However, large performance gaps exist for vehicle

detection between daytime and nighttime conditions. This especially true for rural environments, where streetlighting is often absent. The disproportionate nighttime fatalities in rural areas compared to urban areas have long been recognized [1]. Nighttime vehicle detection presents unique challenges, including limited visibility, unpredictable lighting conditions, and lower resolution from standard roadside cameras compared with daytime scenarios. In rural settings, where lighting infrastructure is often sparse or non-existent, these challenges are even more pronounced, coupled with the headlight glare that further aggravates the issue [2]. Researchers advancing in object detection technology must address these challenges to offer reliable performance in such undesirable conditions.

Computer vision techniques leverage appearance information like color, shape, or typical vehicle patterns to detect vehicles from different views to achieve good performance [3, 4], but most of these works address the problem during the daytime. At night, above appearance features become invalid, and headlights and taillights are almost the only obvious features. Efforts for nighttime vehicle detection have made significant strides in recent years, particularly in applications that utilize ego cameras and roadside cameras.

Ego cameras on vehicles offer a driver perspective and are primarily utilized in autonomous driving applications. Nighttime vehicle detection using ego cameras has been studied due to the relatively higher quality of data and less challenging nature of the tasks. Their primary focus is to accurately identify nearby vehicles, a task essential for real-time decision-making and navigation for the safety of the ego vehicle. For ego camera based nighttime vehicle detection tasks, researchers utilize generative adversarial networks (GAN) and image-to-image translation techniques have been used [5–7] to enhance object detection in challenging scenarios such as nighttime and adverse weather conditions.

Common approaches focus on translating images from a source domain (e.g., daytime) to a target domain (e.g., nighttime) and preserving critical object features during the process. Models like AugGAN [5] and CycleGAN [6–8] are popular approaches that leverage structure-aware mechanisms to maintain semantic and geometric consistency during style transfer. Techniques such as semantic segmentation and geometric attention maps [6] further ensure that essential object details are retained, enabling robust object detection performance in the target domain. These models generate high-quality synthetic datasets that mimic target domain characteristics, which are then utilized to train and fine-tune object detection models, resulting in improved accuracy and robustness.

In addition to domain translation, cross-domain learning techniques are integrated to bridge the performance gap between source and target domains. For instance, convolutional block attention mechanisms (CBAM) [9] enhance detection accuracy by focusing on salient image regions, while feature enhancement modules fuse daytime and nighttime data to mitigate ambient light interference. Furthermore, advanced loss functions and data augmentation strategies refine model training and address challenges like reduced visibility and occlusion. Collectively, these methodologies highlight the efficacy of GAN-based frameworks, feature enhancement, and domain adaptation in improving vehicle detection in low-visibility environments.

The aforementioned methods primarily address challenges in autonomous driving scenarios using data captured by ego cameras. Their applications are limited to the localized environment surrounding autonomous vehicles. Network-level traffic flow and incident monitoring are typically achieved by roadside cameras operated by state or local transportation agencies. These cameras, which are commonly mounted on roadside utility poles, capture a broader view for effective traffic monitoring and incident management. Such objectives are critical for enhancing the operational efficiency and

safety of entire road networks. These roadside cameras play a critical role in enhancing situational awareness by detecting incidents and disseminating alerts, such as speed warning or hazard notification, to drivers, thereby improving road safety. However, these camera are frequently characterized by low-resolution images and poor video quality. This challenge becomes even more severe under adverse conditions, such as nighttime or inclement weather, significantly complicating vehicle detection in these environments. Both roadside and ego cameras face common challenges in nighttime vehicle detection. These include difficulties in distinguishing vehicles from other objects under low illumination, glare from headlights, and insufficient detail captured by standard imaging sensors. Such limitations are particularly acute in rural settings, where minimal or inconsistent lighting exacerbates detection difficulties. It is critical to address these challenges to improve the effectiveness of both camera types in their respective operational contexts. In this paper, we focus on nighttime vehicle detection from roadside cameras in rural settings.

Several studies have been conducted to address the challenges of nighttime vehicle detection. Fu et al. [10] proposed a framework to improve nighttime object detection accuracy using a StyleMix-based method that generates day-night image pairs for training and a kernel prediction network (KPN) to enhance nighttime-to-daytime image translation. While this framework aims to adapt models trained on daytime images for nighttime detection, the data used in their study was captured from a top-down perspective, and the resulting augmented nighttime images fail to accurately represent real roadside conditions. Specifically, the images suffer poor qualities such as low illumination, poor contrast, and the presence of glare from headlights and reflections, which are common in real-world scenarios. Similarly, Guo et al. [11] employed CycleGAN to generate nighttime traffic images from daytime data, integrating these images with a dense traffic detection network (DTDNet) to enhance detection accuracy and address the scarcity of nighttime annotations. Nevertheless, their data was collected using phone cameras from specific angles, which is constrained by limited view perspectives. Consequently, the approach does not adequately account for real-world challenges such as low illumination and headlight glare, reducing its effectiveness in more complex and realistic nighttime environments.

The suboptimal performance of detection models in nighttime rural scenarios stems from several interconnected challenges. Nighttime environments are characterized by low illumination and poor contrast, which hinder models' ability to distinguish vehicles from the background and accurately delineate vehicle boundaries. Additionally, intense headlight glare and reflections often confuse models, as these bright spots can obscure objects of interest or be misinterpreted as vehicles. With these issues, nighttime images frequently suffer from noise, motion blur, and low resolution, resulting from reduced sensor performance in low-light conditions. This degradation in data quality further impacts model accuracy and reliability. Another significant limitation is the scarcity of large, diverse, and annotated nighttime datasets. Most datasets predominantly consist of daytime images, leading to an imbalance that prevents models from effectively generalizing to nighttime conditions. Furthermore, domain adaptation remains a critical challenge, as models trained on daytime images struggle to perform in nighttime environments due to the stark differences in visual features and environmental conditions. These challenges collectively underscore the need for innovative approaches to enhance nighttime data quality, increase dataset diversity, and improve model adaptability for rural nighttime detection scenarios.

Generative models, particularly generative adversarial networks (GAN), have emerged as powerful tools for augmenting datasets in scenarios where high-quality real-world data is scarce or difficult to

obtain. Traditional GAN-based approaches, however, rely heavily on the availability of paired data from two distinct domains (e.g., daytime and nighttime images). Several researchers, such as Fu et al. [10] and Guo et al. [11], utilized CycleGAN for day-to-night image style transfer, generating synthetic nighttime images to enhance dataset diversity and improve the performance of detection models. While using these methods can effectively achieve day-to-night transfer, they often fail to address more complex challenges, such as accurately replicating headlight effects, which is critical for vehicle detection in rural nighttime settings where roadside lighting is absent. Moreover, rapid advancement of generative pretrained transformers (GPTs) has led to the development of various AI tools capable of performing image style transfer tasks, including text-based image editing. Motivated by these advancements, we explored text-based image editing tools leveraging diffusion models [12]. Using a prompt such as "Given the daytime image, transfer it into a nighttime setting without ambient light and turn on the headlights of all the vehicles", we observed promising results for day-to-night style transfer. However the generated images exhibited poor and unrealistic headlight modeling, highlighting the limitations of GPT-like models in accurately simulating rural nighttime transportation conditions.

A key limitation of other approaches lies in their inability to accurately model headlight effects, as the distribution of headlight illumination is governed by complex physical principles that are challenging to replicate using simple domain mapping techniques. At night, headlights serve as the most prominent and reliable vehicle feature for detection. Several researchers have explored nighttime vehicle detection through headlight detection, tracking, and pairing methods [13, 14]. While these techniques perform well in low-light scenarios, they are impractical for roadside camera settings in rural areas. To address these challenges, we propose a novel framework that enables the augmentation of annotated nighttime images directly from daytime images, and realistically model headlight effects using the CARLA simulator for image style transfer. To the best of our knowledge, we are the first to leverage CARLA-generated synthetic data for both day-to-night image style transfer and headlight effect modeling. Our framework offers a novel and effective solution to enhance vehicle detection in challenging rural nighttime environments.

2. Proposed framework

Our proposed framework, illustrated in Figure 1, introduces a novel labeling-free data augmentation method that enables realistic day-to-night image style transfer using synthetic data generated by CARLA [15]. The framework comprises two major components: (1) Synthetic nighttime data generation under rural settings: This component leverages the CARLA simulator to generate synthetic nighttime images that incorporates realistic headlight effects and varying illumination conditions, as observed from roadside cameras in rural environments. The CARLA simulator is integral to this process, as it can faithfully model vehicle headlight effects at night, effectively addressing the limitation of existing AI models that often fail to capture headlight effect during day-to-night style transfer. (2) Day-to-night style transfer process: To address data scarcity of nighttime road scene images in rural environments, a CycleGAN model is trained to perform day-to-night image transfer. Daytime images are collected and processed using the state-of-the-art YOLO11 model [16] to perform vehicle detection and classification. The resulting annotations are directly mapped to the style-transferred nighttime images, enabling the creation of an augmented nighttime dataset without additional labeling effort.

To enhance dataset diversity and reality, the final augmented dataset combines human-labeled real nighttime low-light images (44%) with transferred images (56%). This dataset is subsequently used to fine-tune the YOLO11 model, which is evaluated against its raw counterpart on a real-world nighttime test dataset.

By combining realistic synthetic data generation with effective style transfer techniques and automated annotation mapping, our framework addresses critical challenges in rural nighttime vehicle detection, offering a novel and practical solution to improve model performance in real-world scenarios.
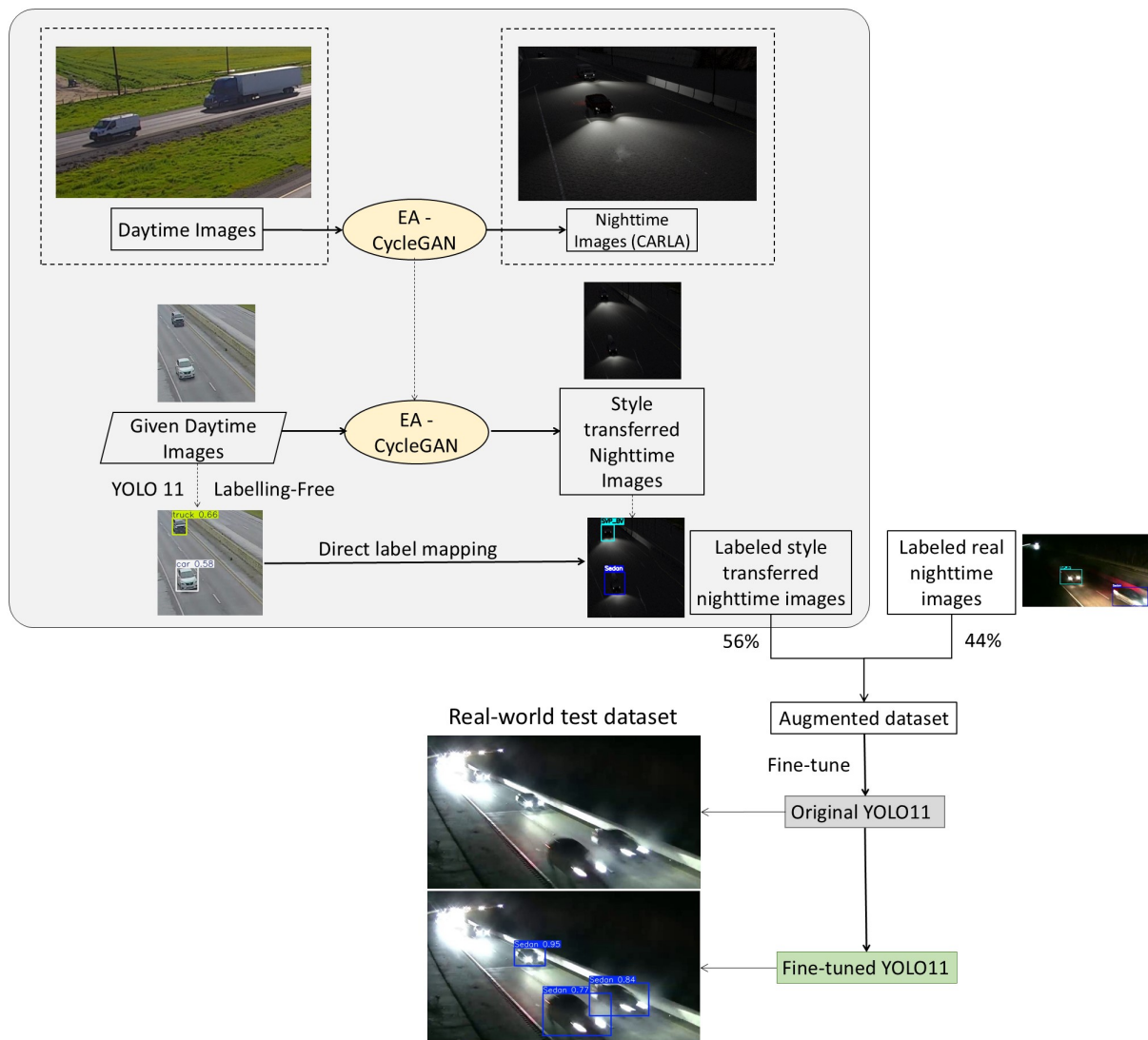

Figure 1. Framework overview.

3. Technical approach

In this section, we introduce our proposed method, which addresses the challenges of nighttime vehicle detection in rural environments through three key steps: (1) Synthetic Nighttime Data

Generation: The process of generating realistic nighttime images is described, where the CARLA simulator is utilized to incorporate critical features such as headlight effect and varying illumination conditions. (2) Day-to-night image style transfer: The model architecture employed for performing day-to-night image transfer is presented, enabling the creation of nighttime images that closely resemble real-world scenarios. (3) Labeling-free data augmentation for nighttime images: The approach for achieving labeling-free augmentation is described, where annotations from daytime images are directly mapped onto nighttime images, facilitating the development of a robust augmented dataset.

3.1. Synthetic nighttime data generation

The primary challenges in improving nighttime vehicle detection arise from the low quality of roadside camera images and the difficulty of collecting sufficiently large and diverse datasets. To address these issues, synthetic nighttime images are generated using CARLA [15], a widely used open-source platform primarily designed for autonomous driving research. CARLA offers extensive control over various environmental and operational parameters, such as weather conditions, lighting, vehicle types, headlight settings (e.g., low-beam, high-beam), as well as camera positions and viewing angles. These customizable options enable the creation of a comprehensive and diverse dataset that accurately reflect real-world rural transportation settings. In particular, for rural highway safety research, the simulator enables the strategic placement of cameras in critical locations, such as curves and ramps, where lower speed limits are often imposed [17].

To closely mimic realistic rural environments, synthetic images were collected under the following scenarios: (1) Departing and approaching vehicles relative to cameras, (2) side-view and top-view perspectives, and (3) scenes with multiple vehicles and single vehicles. Several representative examples are presented in Figure 2. It is important to note that, in this study, all synthetic images were generated under clear weather conditions with no environmental modifications.
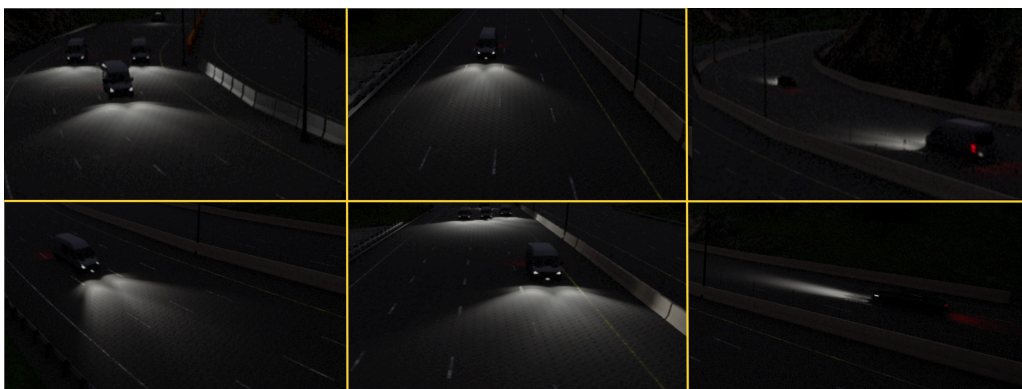

Figure 2. CARLA examples (from left to right, first column: Side-view approaching; second column: center-view approaching; and third column: Side-view departing).

3.2. Day-to-night image transfer

The efficient attention GAN (EAGAN) [18] builds upon the CycleGAN framework by integrating efficient attention blocks into the generator networks while enabling attention sharing between corresponding encoder and decoder blocks. This mechanism enables the re-utilization of the long-range dependencies computed from the source-domain images during the reconstruction of their target-domain counterparts. This design makes EAGAN a robust choice for high-quality image-to-image (I2I) translation tasks, particularly in scenarios where maintaining consistency between domains is critical.

In this study, the EAGAN architecture is adopted to perform day-to-night style transfer in rural environments. The model is trained using datasets from two domains: Real-world daytime images and CARLA virtual nighttime images.

The I2I translation task generally considers transforming image x from domain X (daytime) to image y in domain Y (nighttime), represented as mappings: $G : x \rightarrow y$, $F : y \rightarrow x$, where G and F are generator networks. The objective is to ensure that the distributions $G(X)$ and $F(Y)$ are indistinguishable from X and Y, respectively, while preserving semantic information and cycle consistency.

To train the EAGAN for the data augmentation purpose from daytime images and CARLA nighttime images, the model input consists of {Real X, Real Y}, where Real X, and Real Y are from the domains X and Y. The detailed information flow is shown in Figure 3. Following the standard training process for GANs, the discriminators and generators are trained simultaneously by optimizing a min-max adversarial objectives. Instead of the traditional adversarial loss, the least-square adversarial loss proposed by [19] is used due to improved stability, which also encourages the generator to produce realistic images indistinguishable by the discriminator:

$$\min_G L_{\text{GAN}}(G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[(D(G(x)) - 1)^2 \right], \tag{1}$$

where $p_{\text{data}}(x)$ denotes the true data distribution of domain X, comprising real daytime images. $D(G(x))$ represents the output of the discriminator, which assigns a score between 0 and 1. Ideally, $D(G(x))$ should approach 1 if the generated nighttime image appears realistic. The term $(D(G(x)) - 1)^2$ imposes a penalty on the generator if the discriminator fails to classify $G(x)$ as realistic, i.e., when the score deviates from 1.

$$\min_D L_{\text{GAN}}(D) = \mathbb{E}_{y \sim p_{\text{data}}(y)} \left[(D(y) - 1)^2 \right] + \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[D(G(x))^2 \right], \tag{2}$$

where $p_{\text{data}}(y)$ denotes the true data distribution of domain Y, i.e., CARLA nighttime images. The term $\mathbb{E}_{y \sim p_{\text{data}}(y)} \left[(D(y) - 1)^2 \right]$ ensures that the discriminator assigns a score close to 1 to authentic CARLA nighttime images from domain Y. Conversely, $\mathbb{E}_{x \sim p_{\text{data}}(x)} \left[D(G(x))^2 \right]$ penalizes the discriminator if it assigns a high score to the nighttime images $G(x)$ generated from real daytime images x. This dual mechanism maintains the discriminator's reliability in distinguishing real and generated samples.

In addition to the least square adversarial loss, cycle consistency is enforced between the two generators to ensure the reversibility of the translation process. Specifically, when an image is passed sequentially through both generators, the reconstructed image should closely resemble the original. To achieve this, a cycle consistency loss term is added to the objective function alongside the adversarial loss:

$$L_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim P_X} \left[\|F(G(x)) - x\|_1 \right] + \mathbb{E}_{y \sim P_Y} \left[\|G(F(y)) - y\|_1 \right]. \tag{3}$$
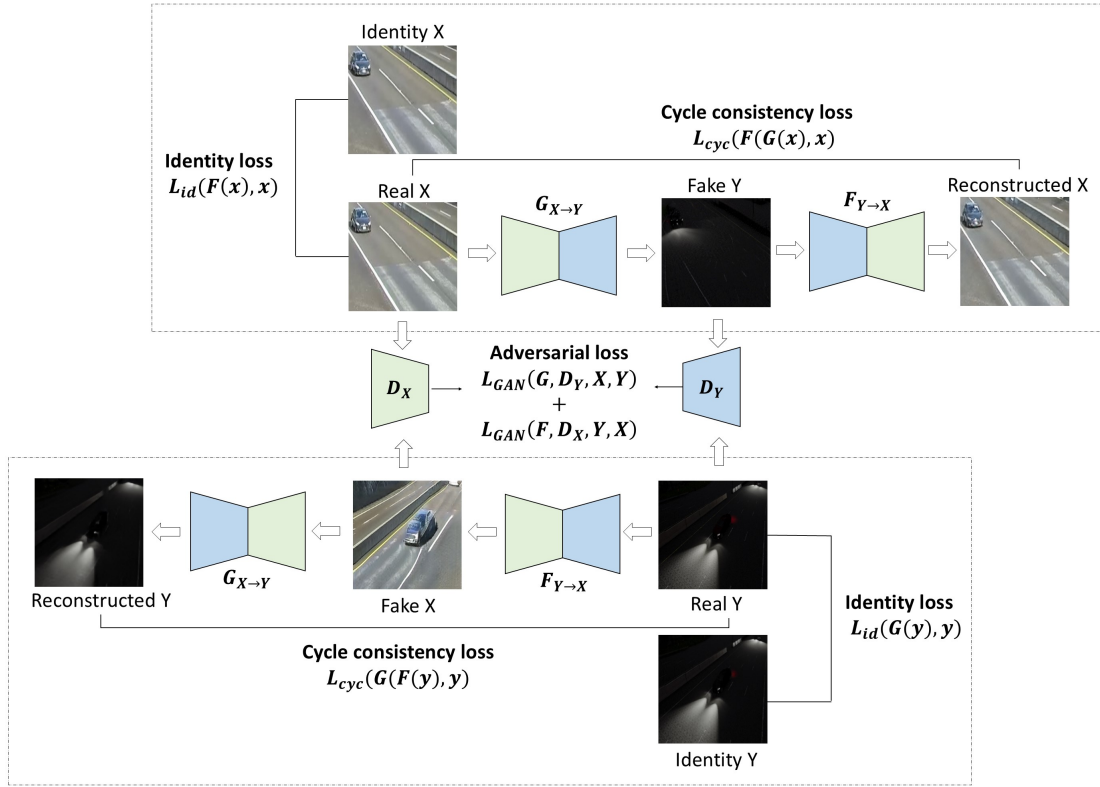
Figure 3. The information flow of EAGAN.

Furthermore, an identity loss term, as proposed by [20], is also included. By leveraging identity mappings for domains X and Y, the generators are encouraged to make minimal alterations to the input images *x* and *y* when they already belong to the target domain. This constraint helps the generators to better preserve the original tint and coloration of the input images. The identity loss is expressed as:

$$L_{\mathrm{id}}(G, F) = \mathbb{E}_{y \sim P_Y} \left[\|G(y) - y\|_1 \right] + \mathbb{E}_{x \sim P_X} \left[\|F(x) - x\|_1 \right]. \qquad (4)$$

As a result, the overall training objective for EAGAN combines adversarial loss, cycle consistency loss, and identity loss, and is written as:

$$L(G, F, D_X, D_Y) = L_{\mathrm{GAN}}(G, D_Y, X, Y) + L_{\mathrm{GAN}}(F, D_X, Y, X) + \lambda_{\mathrm{cyc}} L_{\mathrm{cyc}}(G, F) + \lambda_{\mathrm{id}} L_{\mathrm{id}}(G, F), \qquad (5)$$

where λ_{cyc} and λ_{id} are weighting parameters for the respective loss terms.

3.3. Labeling-free data augmentation for nighttime images

The you only look once (YOLO) family of models has revolutionized object detection, offering real-time detection and high accuracy. YOLO11, the latest version [16], builds upon this legacy with attention mechanisms, deeper feature extraction layers, and an anchor-free detection approach. It was designed to address challenges such as detecting small, occluded, or fast-moving vehicles. By integrating strengths of CNNs and self-attention mechanism, YOLO11 improves both detection accuracy and computational efficiency, making it well-suited for real-world applications. In our study, YOLO11 is applied as an "annotator" to automatically annotate daytime images. The obtained

labels from daytime images are directly applied to the style-transferred nighttime images as the objects of interest (i.e., vehicles) will remain in the same locations. This enables us to leverage the accurate vehicle detection capability of the YOLO11 model to automatically obtain labels for its style-transferred nighttime counterparts.

The original YOLO11 model was pretrained on the COCO dataset [21], which includes vehicle classes of car, bus, and truck that are relevant to rural settings. For this study, our focus is on classifying two vehicle categories: Class 0 (sedan) and class 1 (SVP-BV). The SVP-BV category includes SUV, Van, Pick-UP, and Bigger Vehicles. To align with the new vehicle categories, the COCO vehicle classes are remapped as follows: (1) cars $->$ sedan; (2) bus and truck $->$ SVP-BV.

The augmented dataset is obtained by assembling the style-transferred nighttime images generated by the EAGAN model with the corresponding labels predicted by YOLO11.

4. Experiments and results

4.1. Data

Our data was gathered from multiple public traffic cameras in California, which includes both daytime and nighttime images, serving distinct purposes for training and testing. The data is organized into three categories: (1) Training datasets for the EAGAN model, (2) fine-tuning datasets for the YOLO11 model, and (3) an evaluation dataset for comparing the performance of the original YOLO11 model and the fine-tuned version.

A detailed summary of each dataset is provided in Table 1.

Table 1. Dataset details.

Dataset	Split	Description	Image #
EAGAN Training	Train A	Real daytime images	239
	Train B	CARLA nighttime images	413
	Test A	Real daytime images	82
	Test B	CARLA nighttime images	50
YOLO11 Fine-Tuning	Train	124 real nighttime images + 163 augmented images	287
	Validation	43 real nighttime images + 20 augmented images	63
	Test	20 real nighttime images + 10 augmented images	30
Evaluation	Test	Real nighttime images	38

4.2. Image style transfer

For the image style transfer with EAGAN, we target two domains: Domain X, which consists of daytime images in real-world settings, and domain Y, comprising CARLA-generated nighttime images. The EAGAN is trained for 200 epochs with the scheduled learning rate in Eq (6) that is initialized at 0.0002 and begins to decrease linearly after the 100th epoch. This decay strategy of learning rate

ensures a smooth model convergence, leading to better generalization and performance [22].

$$lr(t) = \begin{cases} lr_0, & \text{if } t \leq n_{\text{e}}, \\ lr_0 \cdot \left(1 - \frac{t - n_{\text{e}}}{n_{\text{d}}}\right), & \text{if } n_{\text{e}} < t \leq n_{\text{e}} + n_{\text{d}}, \\ 0, & \text{if } t > n_{\text{e}} + n_{\text{d}}, \end{cases} \tag{6}$$

where lr_0 = initial learning rate t = current epoch n_{e} = the number of epochs that the learning rate decay starts n_{d} = the number of epochs that the learning rate decay ends.

Table 2 shows the detailed parameter settings for EAGAN training.

Table 2. Training parameter settings for EAGAN.

Parameter	Value	Description
n_e	100	Epochs with constant learning rate
n_d	200	Epochs for linear learning rate decay
lr_0	0.0002	Initial learning rate
β_1	0.5	Momentum term of Adam optimizer
input_size	256×256	Size of input images
λ_{cyc}	10.0	Weight for cycle consistency loss
λ_{id}	0.5	Weight for identity loss

Each epoch takes approximately 150 seconds, with the training process completed in about 8 hours on a single NVIDIA A6000 GPU. Figure 4 showcases test examples from our trained EAGAN model. The results confirm successful day-to-night translation, including effective addition of headlight features. Notably, the model accurately places headlights in the correct locations of the vehicles, demonstrating its ability to reliably locate the vehicles and identify their positions. Interestingly, some shadow-related effects are observed:

(1) Vehicle shadows under sunlight: For shadows cast by vehicles in sunny conditions (e.g., rows 1 and 2 in columns 5 and 6 of Figure 4), the model tends to interpret the shadow in front of the car's bumper as part of the vehicle. This results in a slight angular misalignment between the illuminated headlights and the front of the car in the transferred image. However, this minor deviation does not impair the model's ability to recognize vehicles at night.

(2) Vehicles in shadowed areas beneath trees: When vehicles are passing tree-shaded areas (e.g., row 3 in columns 5 and 6 of Figure 4), the blending of vehicle features with the blotchy tree shadows create challenges for the model. These shadowed regions act as noise, degrading the quality of the transferred images and negatively impacting downstream tasks.

While these shadow effects introduce some artifacts, the overall performance of the EAGAN model remains robust in generating high-quality day-to-night image translation.

Figure 4. Test examples of the trained EAGAN (From left to right: the first, third, and fifth columns are original daytime images; the second, fourth, and sixth columns are the corresponding style-transferred nighttime images, respectively).

4.3. Nighttime vehicle detection and classification

For this experiment, the YOLO11-small model is employed. Figure 5 showcases sample predictions generated by the original YOLO11-small model, which serve as labels for their transferred nighttime images.

Figure 5. Original YOLO11-small model predictions for auto-labeling.

Although CARLA can generate realistic nighttime road scene images, there are still subtle differences in appearance compared to real-world nighttime road scenes. To address this domain adaption gap, we incorporate a selection of manually annotated real-world nighttime images into the training dataset. This approach enables the model to learn relevant features from both CARLA-generated and real-world nighttime images, enhancing its overall performance and robustness.

To fine-tune the model, a learning rate scheduling strategy is implemented for different components of the model. Initially, the backbone is fine-tuned with a learning rate of 0.0001 for 50 epochs. Subsequently, the backbone network is frozen, and a learning rate of 0.00005 is applied exclusively to the neck network for another 50 epochs. Finally, both the backbone and neck networks are frozen, and a learning rate of 0.00001 is applied to the head network for an additional 50 epochs. This block-wise adaptation strategy, distinct from the approach used in EAGAN training, facilitates enhanced convergence and improved generalization.

For evaluation, representative real-world nighttime images captured by roadside traffic cameras in rural areas are analyzed, as shown in Figures 6–8. These nighttime images present various challenges, including low ambient light, poor image quality, and issues caused by headlight glare. The original YOLO11 model frequently struggles to distinguish vehicles from the background, even under relatively

favorable lighting conditions, and often produces low confidence scores when vehicles are detected. In contrast, the fine-tuned YOLO11 model, trained on the augmented dataset, achieves a 100% detection success rate, with significantly higher confidence scores, demonstrating the effectiveness of the proposed framework.

Figure 6. Comparison of predictions for single-vehicle images with ambient light (Top row: Original YOLO11 model; and bottom row: fine-tuned YOLO11 model).

Figure 7. Comparison of predictions on multiple-vehicle images without ambient light (Top row: Original YOLO11 model; and bottom row: fine-tuned YOLO11 model).
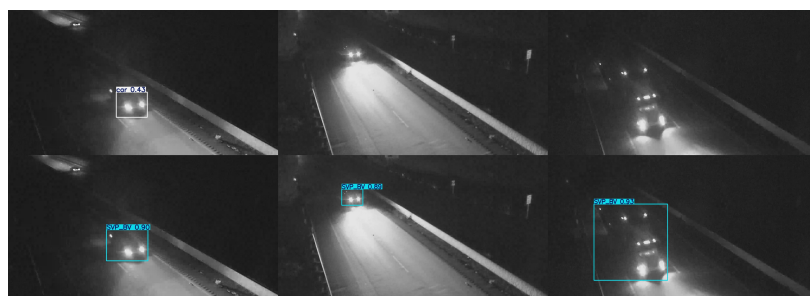

Figure 8. Comparison of predictions on gray-scale images without ambient light (Top row: Original YOLO11 model; and bottom row: fine-tuned YOLO11 model).

Table 3 presents detailed classification results across metrics for the original and fine-tuned YOLO11 models. It reveals significant improvements across classes, indicating that the fine-tuned model effectively captures most vehicles in the nighttime scenes, addressing the key limitation of state-of-the-art object detection models. The consistent gains in mAP metrics further highlight the fine-tuned model's robustness in detecting and localizing vehicles under challenging nighttime conditions. Class-

specific refinements enhance detection for both smaller vehicles (sedan) and larger ones (SVP-BV). Notably, the fine-tuned model shows a slightly lower bounding box precision for the SVP-BV class, largely due to the diverse mix of vehicle types in this newly defined class.

Table 3. Performance comparison of the original and fine-tuned YOLO11 models.

Model	Class	Bounding Box Precision	Recall	mAP50	mAP50-95
	all	0.56	0.25	0.26	0.16
Original YOLO11	car	0.21	0.39	0.20	0.12
	truck	0.91	0.11	0.32	0.20
	all	0.63	0.88	0.76	0.56
Fine-tuned YOLO11	sedan	0.51	0.85	0.59	0.40
	SVP_BV	0.75	0.92	0.93	0.72

5. Conclusions and discussion

In this work, we proposed a novel framework for enhancing nighttime vehicle detection, featuring a labeling-free method that leverages EAGAN for day-to-night image transfer and CARLA for realistic headlight modeling. We created an augmented dataset for fine-tuning object detection models, resulting in improved performance for nighttime conditions. Additionally, we adopted different learning rate scheduling strategies during EAGAN training and YOLO11 fine-tuning to ensure smooth convergence and enhanced generalization. A performance comparison between the original YOLO11 model and the fine-tuned version demonstrated that the YOLO11 model fine-tuned with the augmented dataset significantly outperformed the original YOLO11 model for nighttime vehicle detection. Its ability to detect and localize the vehicles with high confidence highlights the effectiveness of fine-tuning with properly augmented data, making it a more reliable solution for real-world applications.

Nevertheless, we acknowledge several limitations that should be addressed in future research: (1) While CARLA offers many vehicle types, it lacks coverage of all vehicle types on the road, particularly the tractor-trailers and RVs, which limits the diversity of synthetic data. Additionally, the headlights in CARLA need to be refined to better replicate glare effects observed in real-world settings. (2) Although the EAGAN model incorporates an attention-sharing mechanism in the generators of CycleGAN, future researchers could explore alternative mechanisms to more effectively address the observed shadow effects. (3) For proof of concept, the training and testing datasets utilized in this study were relatively small. Thus, researchers should consider significantly expanding the datasets using our proposed data augmentation approach, which is expected to enhance model performance and robustness.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

Conflict of interest

Jidong J. Yang is an editorial board member for Applied Computing and Intelligence and was not involved in the editorial review and the decision to publish this article.

References

1. *National Highway Traffic Safety Administration, Daytime and nighttime seat belt use by fatally injured passenger vehicle occupants*, U.S. Department of Transportation, 2010. Available from: `https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/811281.pdf`.

2. *National Highway Traffic Safety Administration, Report to congress: nighttime glare and driving performance*, U.S. Department of Transportation, 2007. Available from: `https://www.nhtsa.gov/sites/nhtsa.gov/files/glare_congressional_report.pdf`.

3. U. Mittal, P. Chawla, R. Tiwari, Ensemblenet: a hybrid approach for vehicle detection and estimation of traffic density based on faster r-cnn and yolo models, *Neural Comput. Applic.*, **35** (2023), 4755–4774. https://doi.org/10.1007/s00521-022-07940-9

4. M. Bie, Y. Liu, G. Li, J. Hong, J. Li, Real-time vehicle detection algorithm based on a lightweight you-only-look-once (yolov5n-l) approach, *Expert Syst. Appl.*, **213** (2023), 119108. https://doi.org/10.1016/j.eswa.2022.119108

5. S. Huang, C. Lin, S. Chen, Y. Wu, P. Hsu, S. Lai, Auggan: cross domain adaptation with gan-based data augmentation, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, 718–731.

6. G. Bang, J. Lee, Y. Endo, T. Nishimori, K. Nakao, S. Kamijo, Semantic and geometric-aware day-to-night image translation network, *Sensors*, **24** (2024), 1339. https://doi.org/10.3390/s24041339

7. X. Shao, C. Wei, Y. Shen, Z. Wang, Feature enhancement based on cyclegan for nighttime vehicle detection, *IEEE Access*, **9** (2020), 849–859. https://doi.org/10.1109/ACCESS.2020.3046498

8. J. Zhu, T. Park, P. Isola, A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, 2223–2232.

9. H. Xu, S. Lai, X. Li, Y. Yang, Cross-domain car detection model with integrated convolutional block attention mechanism, *Image Vision Comput.*, **140** (2023), 104834. https://doi.org/10.1016/j.imavis.2023.104834

10. L. Fu, H. Yu, F. Xu, J. Li, Q. Guo, S. Wang, Let there be light: improved traffic surveillance via detail preserving night-to-day transfer, *IEEE Trans. Circ. Syst. Vid.*, **32** (2022), 8217–8226. https://doi.org/10.1109/TCSVT.2021.3081999

11. F. Guo, J. Liu, Q. Xie, H. Chang, Improved nighttime traffic detection using day-to-night image transfer, *Transport. Res. Rec.*, **2677** (2023), 711–721. https://doi.org/10.1177/03611981231166686

12. B. Kawar, S. Zada, O. Lang, O. Tov, H. Chang, T. Dekel, et al., Imagic: text-based real image editing with diffusion models, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, 6007–6017.

13. Q. Zou, H. Ling, S. Luo, Y. Huang, M. Tian, Robust nighttime vehicle detection by tracking and grouping headlights, *IEEE Trans. Intell. Transp.*, **16** (2015), 2838–2849. https://doi.org/10.1109/TITS.2015.2425229

14. S. Parvin, L. Rozario, M. Islam, Vision-based on-road nighttime vehicle detection and tracking using taillight and headlight features, *Journal of Computer and Communications*, **9** (2021), 107619. https://doi.org/10.4236/jcc.2021.93003

15. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, V. Koltun, Carla: an open urban driving simulator, *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, 1–16.

16. R. Khanam, M. Hussain, YOLOv11: an overview of the key architectural enhancements, arXiv: 2410.17725. https://doi.org/10.48550/arXiv.2410.17725

17. S. Malik, M. Khan, H. El-Sayed, Carla: car learning to act—an inside out, *Procedia Computer Science*, **198** (2022), 742–749. https://doi.org/10.1016/j.procs.2021.12.316

18. *J. Zhu, T. Park, T. Wang, Efficient-attention-GAN: unsupervised image-to-image translation with shared efficient attention mechanism*, GitHub, Inc., 2023. Available from: `https://github.com/jccb15/efficient-attention-GAN`.

19. X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, S. Smolley, Least squares generative adversarial networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, 2794–2802.

20. Y. Taigman, A. Polyak, L. Wolf, Unsupervised cross-domain image generation, arXiv: 1611.02200. https://doi.org/10.48550/arXiv.1611.02200

21. T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft coco: common objects in context, In: *Computer vision—ECCV 2014*, Cham: Springer, 2014, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

22. K. You, M. Long, J. Wang, M. Jordan, How does learning rate decay help modern neural networks? arXiv: 1908.01878. https://doi.org/10.48550/arXiv.1908.01878