Enhancing Nighttime Vehicle Detection Through Data Augmentation for Intelligent Traffic Monitoring

Luong Duc Hung; Nguyen Quoc Viet; Chu Minh Quan; Vu Tuan Anh

Advisor: Phan Duy Hung

Keywords: Vehicle Detection, Nighttime Surveillance, CycleGAN, YOLOv8, Low-Light Conditions, Data Augmentation, Traffic Monitoring, Image-to-Image Translation, Object Detection, Generative Adversarial Networks (GANs)

1. Introduction

1.1 Problem & Motivation

Effective traffic management and road safety are critical challenges in modern urban environments, where increasing congestion demands reliable monitoring systems. Ensuring that vehicles comply with lane regulations and traffic signals is essential for reducing accidents and optimizing traffic flow. However, accurately identifying vehicle types and assessing their behavior remains a complex task, particularly under varying environmental conditions such as nighttime or poor visibility.

A robust vehicle classification and monitoring system can enhance traffic surveillance by automatically detecting cars, motorcycles, buses, and trucks while evaluating their adherence to traffic laws. Leveraging advanced image processing and machine learning techniques enables real-time analysis of traffic data, allowing authorities to detect violations more efficiently and implement corrective measures. Furthermore, automated monitoring reduces reliance on manual enforcement, providing valuable insights for urban planning and law enforcement. This research aims to develop an enhanced vehicle detection system that improves recognition accuracy across different lighting conditions, particularly at night, to support safer and more efficient road management.

1.2 Related Works

Recent studies in vehicle detection and classification have predominantly explored deep learning frameworks to enhance accuracy and processing speed. For instance, Shekhar, J. Debadarshini, and S. Saha [1] developed LiVeR, an IoT-based solution for lightweight real-time vehicle detection and classification. Although LiVeR achieves impressive accuracies ranging from 91.3% to 98.8%, its current optimization for single-lane traffic scenarios suggests a need for further adaptation in more complex environments.

Complementing this approach, H. Ahn and Y.-H. Lee [2] introduced a YOLO-based system that achieved an average precision of 90.32%, demonstrating the efficacy of convolutional neural networks for real-time tracking and classification. Similarly, Yunxiang Yang, Hao Zhen, Yongcan Huang, and Jidong J. Yang [3] proposed an innovative strategy that combines day-to-night style transfer with labeling-free augmentation, effectively simulating adverse nighttime conditions and enhancing detection accuracy.

Other research efforts have sought to address the limitations of conventional methods under challenging environmental conditions. P. Mahto, P. Garg, P. Seth, and J. Panda [4] refined the YOLOv4 architecture through self-adversarial training, spatial attention modules, and customized anchor boxes, thereby significantly improving detection performance despite persistent challenges related to occlusions and low-light conditions.

Additional approaches have focused on alternative paradigms to complement deep learning techniques. M. Pemila, R. K. Pongiannan, R. Narayanamoorthi, E. A. Sweelem, E. Hendawi, and M. I. Abu El-Sebah [5] implemented a vehicle classification system on FPGA using machine learning algorithms, achieving an accuracy of 98.81% with exceptional processing speeds. Similarly, M. Liang, X. Huang, C.-H. Chen, X. Chen, and A. Tokuta [6] proposed a cascaded regression framework for the real-time counting and classification of highway vehicles, although its performance is closely tied to the accuracy of background subtraction methods.

Finally, the comprehensive review by Boukerche, A. J. Siddiqui, and A. Mammeri [7] synthesizes various models and techniques in automated vehicle detection, highlighting common challenges such as occlusions, fluctuating lighting conditions, and the demand for real-time processing. Together, these studies underscore the multifaceted nature of vehicle detection research and the ongoing need for robust, adaptive solutions that perform reliably across diverse operational scenarios.

Recent research has made significant strides in vehicle detection and classification through the use of deep learning, IoT, and regression-based techniques, yielding high accuracy and efficient processing. However, several challenges still persist, including issues with occlusions, variations in lighting and weather conditions, dependence on labeled datasets, and limited adaptability to dynamic environments. Most existing approaches rely heavily on vision-based methods, with less attention given to multi-modal fusion or self-learning techniques.

To address these challenges, we propose a two-stage approach that leverages generative image transformation and object detection. First, daytime images are converted into nighttime images using CycleGAN, creating a synthetic dataset that simulates real-world nighttime conditions. This transformed dataset is then used to train YOLOv8, a state-of-the-art object detection model optimized for vehicle recognition in low-light environments.

During the testing phase, the trained YOLOv8 model is evaluated on nighttime images that have been further modified to represent various challenging conditions, such as increased noise, blurring, and contrast adjustments. After detection, the system undergoes an evaluation phase to assess performance metrics, ensuring robust vehicle recognition under diverse nighttime scenarios. This approach enhances detection accuracy while minimizing the need for large-scale real-world nighttime datasets.

1.3 Contribution

In this study, our team developed a traffic monitoring system with a primary focus on nighttime vehicle recognition. Acknowledging the challenges posed by low-light conditions, we adopted a data-centric approach to enhance model performance without requiring modifications to real-world infrastructure.

To improve nighttime recognition capabilities, we first converted daytime images into nighttime images using CycleGAN. This transformation created a synthetic nighttime dataset, which was then used for model training. In addition, we applied data augmentation techniques such as brightness reduction, contrast adjustment, noise injection, and blurring to further simulate diverse nighttime conditions (e.g., foggy, glare-affected, and noisy environments). This augmentation process enriched the training data, ensuring the model's robustness against real-world lighting variations.

Our system detects and classifies vehicles into basic categories, including passenger cars, buses, trucks, and motorcycles. By training YOLOv8 on the CycleGAN-generated nighttime dataset, we improved detection accuracy under low-light conditions. During the testing phase, the trained model was evaluated using nighttime images modified with additional augmentations, ensuring its effectiveness across various challenging scenarios.

Through these contributions, our research highlights the effectiveness of CycleGAN-based nighttime transformation and data augmentation as practical solutions to enhance nighttime traffic monitoring. This approach enables continuous and reliable surveillance while improving the efficiency of traffic management systems without requiring specialized hardware or external interventions.

2. Data Preparation

We used two public datasets for this study [8, 9], each designed for different models. The first dataset [8], used for the YOLOv8 model, consists of 1,391 daytime and nighttime images, each with a resolution of 640×640 pixels. Out of these, 1,235 images are used for training, and 156 images are used for testing. The second dataset [9], used for the CycleGAN model, contains 2,004 images with a resolution of 1280×720 pixels. Of these, 1,776 images are used for training, and 228 images are used for testing. This dataset is intended for converting daytime images into nighttime images to improve vehicle detection in low-light conditions.

After analyzing the dataset, we found that most images contain between 5 and 15 objects, with an average of 8.4 objects per image. The distribution of object counts per image is right-skewed, indicating that while most images have relatively few objects, a smaller subset contains a high vehicle density. This pattern aligns with real-world traffic conditions, where congestion is often concentrated in specific areas. The dataset consists of approximately 1,500 images, which are used for training, while three videos are allocated exclusively for testing. Table 1 provides an overview of the dataset.

Table 1: Example of review dataset for vehicle detection.

Dataset	Size	Train	Test	Total
Daytime and Nighttime Images for YOLOv8	640*640	1235	156	1391
Daytime and Nighttime Images for CycleGAN	1280*720	1776	228	2004

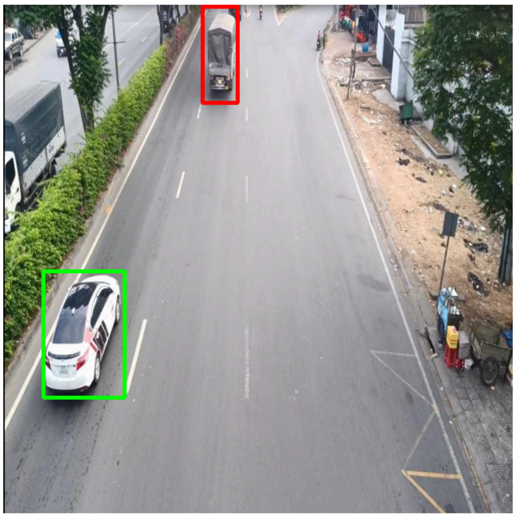

Figure 1: Example of vehicle detection dataset

Each vehicle in the dataset [8] is represented by essential geometric and positional features. The bounding box dimensions (width and height) provide an estimate of vehicle size, crucial for distinguishing between different vehicle types. The normalized coordinates (X, Y) define the position of the vehicle within the image, which is useful for spatial analysis and object tracking. Additionally, the class label categorizes the vehicle, facilitating further classification tasks. These extracted features serve as the foundation for vehicle detection, classification, and movement analysis in real-world applications.

From the dataset annotations (Figure 1), we observe two detected objects, one labeled as “2”, the other one labeled as “0”. The extracted features for each detected object consist of:

- Object 1 (Class 2 - Truck): Center at (0.426, 0.098), size (0.070, 0.188)
- Object 2 (Class 0 - Car): Center at (0.159, 0.648), size (0.159, 0.249).

These features are crucial for training object detection models, as they define the spatial location and size of each detected entity. Further preprocessing and feature engineering, such as histogram equalization or edge detection, may enhance detection performance. Additionally, cross-validation techniques should be applied to ensure the generalization of the detection model.

3. Methodology

In this study, we leverage generative image transformation and augmentation techniques to enhance vehicle detection performance in low-light conditions. Initially, daytime images are transformed into nighttime counterparts using CycleGAN, creating a synthetic nighttime dataset. This dataset serves as the primary training set for the YOLOv8 object detection model, enabling it to learn vehicle features under nocturnal lighting conditions.

To further improve the model's generalization and robustness, additional transformations, including noise injection, blurring, and contrast adjustments, are applied to nighttime images. These augmentations simulate real-world environmental challenges such as fog, motion blur, and varying illumination. During the evaluation phase, the trained YOLOv8 model is tested on these augmented nighttime images to assess its detection accuracy across diverse low-light scenarios.

The overall workflow of our methodology is illustrated in Figure 2, highlighting the integration of generative nighttime image synthesis, targeted augmentations, and object detection to enhance vehicle recognition in challenging nighttime environments.

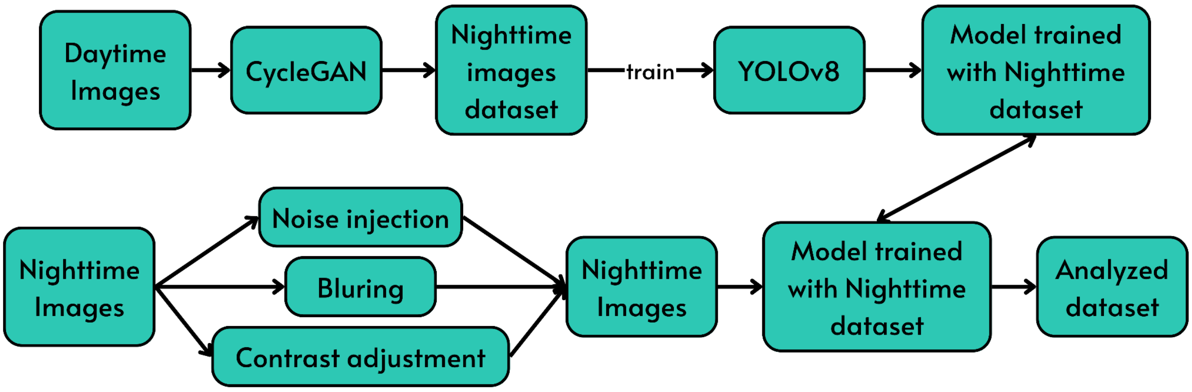

Figure 2: Example of the process (Enhancing Vehicle Detection in Low-Light Conditions)

Image Transformation Using CycleGAN [10]

To generate realistic nighttime images from daytime inputs, we employ CycleGAN, a generative adversarial network (GAN) specifically designed for unpaired image-to-image translation. CycleGAN consists of two generators: one that converts daytime images into nighttime images and another that performs the inverse transformation. To ensure that the fundamental structure and content of the images remain unchanged, CycleGAN utilizes a cycle consistency loss. This loss function enforces the constraint that when an image is transformed from one domain to another and then back, it should closely resemble the original image. Mathematically, the cycle consistency loss is expressed as:

$$L_{\{cyc\}(G,F)} = E_{\{x \sim P_{\{data\}(x)}\}}[||F(G(x)) - x||] + E_{\{y \sim P_{\{data\}(y)}\}}[||F(G(y)) - y||]$$

where G represents the generator that maps daytime images to nighttime images, while FFF is the inverse generator. Additionally, an adversarial loss is employed to ensure that the transformed images closely resemble real nighttime images. This is achieved by training a discriminator that differentiates between real and generated nighttime images, while the generator learns to produce images that can successfully deceive the discriminator. The adversarial loss is formulated as:

$$L_{\{GAN\}(G,D_Y,X,Y)} = E_{\{y \sim P_{\{data\}(y)}\}}[logD_{Y(y)}] + E_{\{x \sim P_{\{data\}(x)}\}}[log(1-logD_{Y(G(X))})]$$

where D_Y represents the discriminator responsible for distinguishing real nighttime images from generated ones. Through this adversarial training strategy, CycleGAN is capable of producing high-quality nighttime images that preserve the structural integrity of the original daytime inputs.

Simulating Various Low-Light Conditions

Beyond nighttime transformation, we introduce additional image modifications to simulate real-world low-light conditions. These include blurring, noise injection, and contrast adjustments, each designed to replicate specific environmental challenges encountered in nighttime or adverse weather scenarios.

One critical transformation is blurring, which simulates reduced visibility caused by fog, rain, or motion. We apply Gaussian Blur [11] to create a smooth blurring effect by convolving the image with a Gaussian kernel, which is defined as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where σ determines the degree of blurring. Additionally, we apply Motion Blur [12, 13] to mimic the effects of vehicle movement or camera shake, creating linear streaks in the image that replicate real-world motion distortions.

To further enhance the realism of low-light conditions, we introduce noise injection, which simulates sensor noise commonly found in nighttime photography. We apply Gaussian Noise [14, 15], where pixel values fluctuate according to a normal distribution:

$$N(\mu, \sigma^2)$$

where μ and σ^2 control the mean and variance of the noise. Additionally, we incorporate Salt-and-Pepper Noise [16, 17], which introduces random white and black pixel variations, effectively simulating sensor defects and signal distortions in low-light environments.

In addition to blurring and noise, we apply contrast adjustments to replicate the impact of uneven illumination, such as headlight glare or insufficient lighting. One method used for contrast enhancement is Gamma Correction [18, 19], which adjusts the overall brightness of the image using the transformation:

$$I' = I^{\gamma}$$

where γ controls the degree of brightness adjustment. A higher gamma value enhances bright regions, while a lower value emphasizes darker areas. Additionally, Histogram Equalization [20, 21] is applied to redistribute pixel intensities, improving visibility in underexposed or overexposed regions of the image. These transformations ensure that the detection model is trained on a dataset that captures a wide range of real-world lighting conditions.

Vehicle Detection with YOLOv8

After generating a dataset consisting of CycleGAN-transformed nighttime images, we use this dataset to train YOLOv8, a state-of-the-art real-time object detection model. YOLOv8 is designed for high-performance vehicle recognition and is optimized using three key loss functions.

During the testing phase, we evaluate the trained YOLOv8 model on nighttime images that have been further modified to simulate various challenging conditions, such as fog, blur, noise, and glare. This allows us to assess the model's robustness under diverse low-light scenarios.

The classification loss ensures accurate vehicle type identification based on learned features. The bounding box regression loss refines the localization of detected vehicles, ensuring precise bounding box placements. Finally, the objectness loss differentiates between vehicles and background objects, enhancing detection accuracy in complex nighttime scenes.

By integrating CycleGAN-based nighttime transformation with targeted low-light augmentations and YOLOv8-based detection, our methodology significantly improves vehicle detection accuracy in challenging nighttime environments. This approach enhances the reliability of vehicle recognition systems, making them more robust for real-world traffic monitoring under varying lighting and weather conditions.

4. Experiment & Analysis

In this study, we trained and evaluated two models: CycleGAN for image-to-image translation and YOLOv8 for object detection. The CycleGAN model was trained for 40 epochs over 20 hours, with evaluation metrics including an average PSNR of 7.57, an average SSIM of 0.2942, an average LPIPS of 0.4157, and an FID Score of 219.63. The YOLOv8 model was trained for 50 epochs over 1 hour and 30 minutes, yielding overall metrics of Precision at 0.0286, Recall at 0.2879, mAP@0.5 at 0.0319, and mAP@0.5:0.95 at 0.0112, alongside detailed per-class performance metrics.

For the CycleGAN model, the average PSNR of 7.57 is notably low compared to typical thresholds for high-quality image reconstruction tasks (often above 30), indicating significant pixel-wise differences between the generated and target images. The average SSIM of 0.2942 reflects low structural similarity, while the average LPIPS of 0.4157 suggests moderate perceptual similarity. The high FID Score of 219.63 indicates a substantial distributional difference between generated and real images, implying that the model has not achieved the desired quality in image translation. These results suggest that CycleGAN may require extended training or architectural adjustments to enhance its performance.

Regarding the YOLOv8 model, the overall metrics highlight limited object detection performance. The low Precision (0.0286) indicates a high rate of false positives, while the Recall of 0.2879 suggests that many true objects are missed. The mAP@0.5 (0.0319) and mAP@0.5:0.95 (0.0112) scores are both low, confirming the model's poor accuracy in localizing and classifying objects. Per-class analysis reveals significant variation: the "bicycle" class achieves an mAP@0.5 of 0.1056 and the "car" class 0.0199, whereas the "person" class scores only 0.0020, and the "motorcycle" class fails entirely with an mAP@0.5 of 0.0000. This indicates particular difficulties with certain object classes, possibly due to insufficient training data or class imbalance.

Table 2: Performance Metrics of CycleGAN and YOLOv8

Model	Metrics	Values
CycleGAN	PSNR	7.57
	SSIM	0.2942
	LPIPS	0.4157
	FID	219.63
YoloV8	Precision	0.0286
	Recall	0.2879
	mAP@0.5	0.0319
	mAP@0.5:0.95	0.0112

In summary, both models exhibit suboptimal performance in their respective tasks. CycleGAN requires improvements in generated image quality, while YOLOv8 needs enhancements in detection accuracy. Future research directions may involve augmenting training data, optimizing hyperparameters, or exploring alternative model architectures.

5. Conclusion and Perspectives

Our study proposes a two-stage method for nighttime vehicle detection, combining CycleGAN to convert daytime images into nighttime images and YOLOv8 for object detection. The results indicate that while CycleGAN can generate nighttime images from daytime ones, the quality remains suboptimal (PSNR: 7.57, SSIM: 0.2942, LPIPS: 0.4157, FID: 219.63), directly impacting YOLOv8's performance. The model demonstrates low accuracy (Precision: 0.0286, Recall: 0.2879, mAP@0.5: 0.0319, mAP@0.5:0.95: 0.0112), particularly struggling with detecting certain vehicle types such as motorcycles. These limitations stem from data quality issues, the realism of synthetic images, and class imbalance. Additionally, the constraints of GPU resources on Google Colab prevent the model from fully utilizing its learning potential.

To enhance performance, we propose several improvements, including increasing the quality of image translation by expanding the nighttime dataset, experimenting with more advanced GAN models such as StarGAN v2 or Attention-GAN, and applying data augmentation techniques. Additionally, optimizing YOLOv8 through transfer learning with a larger nighttime dataset and addressing class imbalance can further improve detection accuracy. Since limited GPU resources restrict the model's training capacity, utilizing more powerful hardware would also be beneficial. Finally, evaluating the method in real-world conditions, such as adverse weather and varying lighting environments, will be essential to validate its feasibility.

REFERENCES

1. H. . Ahn and Y.-H. . Lee, "Vehicle Classification and Tracking Based on Deep Learning", JWE, vol. 21, no. 04, pp. 1283–1294, Apr. 2022.
2. Shekhar, J. Debadarshini, and S. Saha, "LiVeR: Lightweight vehicle detection and classification in real-time," ACM Trans. Internet Things, vol. 5, no. 4, Art. no. 20, pp. 39, Nov. 2024.
3. Yunxiang Yang, Hao Zhen, Yongcan Huang, Jidong J. Yang. Enhancing nighttime vehicle detection with day-to-night style transfer and labeling-free augmentation[J]. Applied Computing and Intelligence, 2024.
4. P. Mahto, P. Garg, P. Seth, and J. Panda, "Refining YOLO v4 for vehicle detection," Int. J. Adv. Res. Eng. Technol., vol. 11, no. 5, pp. 409-419, May 2020.
5. M. Pemila, R. K. Pongiannan, R. Narayanamoorthi, E. A. Sweelem, E. Hendawi, and M. I. Abu El-Sebah, "Real-time classification of vehicles using machine learning algorithm on the extensive dataset," IEEE Access, vol. 12, pp. 12345-12356, Jun. 2024.
6. M. Liang, X. Huang, C.-H. Chen, X. Chen, and A. Tokuta, "Counting and classification of highway vehicles by regression analysis," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 10, pp. 1-11, Oct. 2015.
7. Boukerche, A. J. Siddiqui, and A. Mammeri, "Automated vehicle detection and classification: Models, methods, and techniques," ACM Comput. Surv., vol. 50, no. 5, pp. 1-39, Oct. 2017.
8. https://universe.roboflow.com/car-classification/vietnamese-vehicle/dataset/3
9. https://www.kaggle.com/datasets/solesensei/solesensei_bdd100k
10. J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2223–2232.
11. D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B*, vol. 207, no. 1167, pp. 187–217, 1980.
12. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.
13. M. Ben-Ezra and S. K. Nayar, "Motion-based motion deblurring," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 689-698, Jun. 2004.

14. J. S. Lim, *Two-Dimensional Signal and Image Processing*, Englewood Cliffs, NJ, USA: Prentice Hall, 1990.
15. A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ, USA: Prentice Hall, 1989.
16. W. K. Pratt, *Digital Image Processing*, 4th ed. New York, NY, USA: Wiley, 2007.
17. T. A. Clarke and J. G. Fryer, "The development of camera calibration methods and models," *Photogrammetric Record*, vol. 16, no. 91, pp. 51-66, Apr. 1998.
18. E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267-276, Jul. 2002.
19. S. Pizer, E. Amburn, and J. Austin, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355-368, Sep. 1987.
20. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.
21. K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphics Gems IV*, San Diego, CA, USA: Academic Press, 1994, pp. 474-485.