Theory and Practice of Science and Technology
2025, VOL. 6, NO. 9, 125-130
DOI: 10.47297/taposatWSP2633-456925.20250609

A Review of the Research on Pedestrian Detection of Autonomous Driving Under Complex Lighting Conditions

Hanmou Sun

Chongqing University of Posts and Telecommunications, Chongqing, 400046, China

ABSTRACT
In the future, it is necessary to break through the physical limits of sensors, improve the real-time robustness of algorithms, build a multi-level collaborative perception system, and combine social governance to finally realize safe pedestrian detection under all-climate conditions.

KEYWORDS
Autonomous driving; Pedestrian detection; Complex lighting; multimodal fusion; Light robustness

I Introduction

As the core technology of the next generation of intelligent transportation, autonomous driving is moving towards application at an unprecedented speed: market research shows that the global "autonomous driving/autonomous driving-related" market size has great prospects (from tens of billions of dollars to hundreds of billions of dollars or even higher) according to different calibers: for example, some institutions estimate that the market size will be about $20.74 billion in 2024. The calculation of the broader industrial chain caliber gives different valuations of nearly 100 billion yuan. Still, these data reflect the importance of autonomous driving in future markets. From cruise assistance on highways to intelligent navigation in urban congestion, vehicle perception of the environment relies on the collaboration of multi-source sensors, among which visual information has attracted much attention due to its low cost and rich information. However, as the most vulnerable subject in road traffic, pedestrians' detection accuracy directly determines the safety of driving decisions. Numerous studies have shown that drastic changes in lighting conditions—such as high contrast in strong midday light, low brightness in shadowed areas, dynamic gradients at dusk, and even no lighting or spot lighting at night—can lead to missed detections, false detections, or even complete failures of traditional detection algorithms.

In autonomous driving scenarios, pedestrian detection must not only meet high accuracy, but also take into account real-time and robustness. The impact of lighting changes on imaging quality includes increased noise, dynamic range compression, color distortion, etc., which can interfere with feature extraction and classification decisions. With the rise of deep learning, researchers have continuously improved the performance of algorithms by improving network structure, designing lighting adaptive modules, and using synthetic data enhancement. Multimodal fusion further uses thermal imaging, infrared or event cameras to compensate for the shortcomings of visible light imaging, and significantly improves the detection effect in low-light environments.

Although a large number of work has made breakthroughs in a single dimension, there is a lack of systematic comparison and induction. This paper aims to deeply analyze the performance differences and underlying mechanisms of pedestrian detection under different lighting conditions from the perspective of "traditional → deep learning → multimodal fusion", summarize the main optimization strategies, and verify the advantages and disadvantages of each method through horizontal experimental comparison. Finally, the existing bottlenecks and future trends are prospected, in order to provide panoramic reference and enlightenment for the research of autonomous driving visual perception.

2 Technological evolution

As mentioned in the introduction, the author divides the development of pedestrian detection technology into three stages, namely traditional, deep learning and multimodal fusion. The following will briefly describe these three stages:

2.1 Traditional method phase

(1) HOG+SVM, manual feature extraction

Taken by N. The HOG (directional gradient histogram) proposed by Dalal et al. in 2005 has become a milestone in pedestrian detection by statistically statistically distributing the gradient direction of the local area of the image, combined with the SVM classifier. HOG is good for pedestrian profile detection, but it relies on manual design and is acceptable under simple light, but there is a problem of greatly increasing the missed detection rate in the face of complex light conditions.

* **Corresponding Author:** Hanmou Sun, 2586618121@QQ.com

Limitations of this method: it relies on manual design, lacks adaptive ability, and has low computational efficiency. At present, it is rarely used independently in academic research, and is still used for rapid deployment or tuning in some resource-constrained embedded systems.

（2）DPM model, preliminary exploration of attitude adaptation

The variable deformation partial model (DPM) can improve the robustness of pedestrian detection by decomposing human parts (such as head, limbs, and torso), but the improvement is still limited in the face of complex lighting environments.

（3）Optical Flow and Background Modeling

Optical flow algorithms such as Lucas-Kanade detect pedestrians by analyzing pixel motion over successive frames, but rely on constant brightness assumptions, which are susceptible to light sudden changes or shadow coverage. Background subtraction is effective in static scenes, but it is difficult to cope with dynamic complex environments such as flashing lights.

2.2 Deep learning stage

Core breakthrough: Convolutional neural networks (CNNs) significantly improve the robustness of detection under complex lighting by automatically learning multi-level feature expressions.

（1）Two-stage model:

Faster R-CNN: Introduce the Regional Suggestion Network (RPN) to generate candidate boxes and combine the RoI Pooling alignment features to achieve high accuracy of day and night scenes on the KAIST dataset, but the real-time performance is limited.

Mask R-CNN: Segmentation branches are added on the basis of detection to improve the boundary localization ability of obstructed pedestrians.

（2）Single-stage model:

YOLO series: YOLOv3/v4 enhances the detection of low-light small targets through multi-scale prediction and feature pyramid network (FPN), with frame rates of up to 60 FPS to meet real-time requirements.

Lighting improvements:

（1）Data improvements: GANs are used to generate synthetic data under different lighting conditions (e.g., low-light enhancement, backlight simulations) to alleviate the lack of real data.

（2）Attention mechanism: Introduce a light attention module to dynamically assign the weight of detection to mitigate the interference of overexposure or low light.

2.3 Multimodal fusion

Technology core: Integrate infrared, ultrasonic and other detectors to compensate for detection defects under extreme light conditions.

（1）Early Fusion (Data Fusion):

RGB-thermal fusion: Aligning visible and infrared thermal imaging features through a dual-stream network improves the detection accuracy at night by more than 5%[].

LIDAR-camera fusion: The LIDAR point cloud is projected onto the image plane to generate enhanced 3D positioning with depth maps, solving the problem of inaccurate detection under extreme light conditions.

（2）late fusion (decision-level fusion);

Probabilistic fusion: weighted multimodal detection data based on Bayesian framework. For example, in foggy weather, the geometric information of LIDAR is trusted first.

Typical architecture:

RODNET: Cross-supervised learning between radar and cameras greatly reduces the false positive rate in rain and fog scenes at night.

2.4 Technology evolution map and key turning points

Table 1 Summary of technology by category in history

stage	Representative technology	Improved light adaptability	Detection Performance (mAP)
Traditional methods	HOG+SVM, DPM	Uniform light only	40%-60%(Caltech)
Deep learning	Faster R-CNN, YOLOv4	Multi-scale feature fusion and adversarial data enhancement	70%-85%(CityPersons)
Multimodal integration	RODNet, MuFEM	Cross-modal complementary, attention-guided feature embeddings	85%-92%(KAIST)

3 Analysis of the impact of light

3.1 Classification system of lighting conditions

The lighting environment faced by the autonomous driving system can be cross-classified according to the time dimension (day and night), weather dimension (sunny/fog) and intensity dimension (light fog/heavy fog), forming six core scenarios.

Table 2 Six core scenarios under lighting conditions

Order of Scenarios	Name of Scenarios	Light of the Ambient	Visibility Conditions
1	Day and Clear Weather	Bright	Without limitation in the visibility
2	Day and Light Fog	Dark	Limited visibility due to darkness
3	Day and Heavy Fog	Bright	Limited visibility due to light fog(light fog's sight distance is about 250 meters)
4	Night and Clear Weather	Bright	Limited visibility due to heavy fog(heavy fog's sight distance is about 50 meters)
5	Night and Light Fog	Dark	Limited visibility due to both darkness and light fog(heavy fog's sight distance is about 250 meters)
6	Night and Heavy Fog	Dark	Limited visibility due to both darkness and heavy fog(heavy fog's sight distance is about 50 meters)

(1) Sunny daytime: bright light and unrestricted visibility (ideal conditions).
(2) Light fog during the day: The light is dim and visibility is limited to 250 meters
(3) Heavy fog during the day: bright light but visibility plummets to 50 meters (strong reflection interference).
(4) Clear nights: Light relies on artificial light sources, limiting visibility
(5) Light fog at night: dim light + visibility 250 meters (low contrast).
(6) Heavy fog at night: extremely dark light + visibility 50 meters (extreme challenge).

It is worth noting that there are certain differences in the distribution of various actual road data (such as the proportion of "heavy fog at night" in real driving); But for example, according to Waymo Open Dataset, heavy fog scenes at night account for only 0.7%, but contribute 12% of missed detection cases. Therefore, it can be seen that although the proportion of some scenarios is not high, the missed detection rate is extremely high and requires special attention.

3.2 Sensor layer: The degradation mechanism of imaging quality and detection accuracy

3.2.1 Vision sensors (e.g., cameras)

(1) Dynamic range crash: Strong light leads to overexposure (loss of detail), low light causes noise multiplication (signal-to-noise ratio ↓). For example, image entropy analysis shows that when the light is below 100 lux, the exposure time needs to be dynamically adjusted to maintain grayscale image quality.

(2) Physical degradation model: Illumination intensity (I) and exposure time (t) jointly determine the image quality Q, and the function curve is detailed in Figureure. 2

$$Q = k \cdot \frac{I \cdot t}{1 + \sigma_n^2(I)}$$

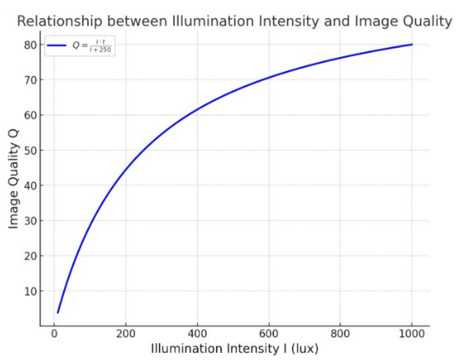

Figure.1 Curve of the relationship between light intensity (lux) and image quality Q

Among them, the light-related noise variance is. Experiments show that when I<250 lux and t>100ms, the Q value drops below the usable threshold.

(3) Color distortion: Foggy scattering leads to wavelength selectivity attenuation and RGB channel imbalance.

3.2.2 LiDAR/TOF sensor

(1) Ambient light interference: Sunlight or strong artificial light sources cause photon noise, resulting in ranging errors δ which is exponentially related to ambient light intensityI_{amb}

$\delta \propto e^{I_{amb}/I_0}$(I is the sensor threshold)

(2) Quantum noise: In low-light environments, the output of the photomultiplier tube follows the Poisson distribution, and the standard deviation of the ranging can reach 2 meters.

3.3 Decision-making link: system-level risk transmission caused by light

(1) Perceptual → prediction fault: The increase in the missed detection rate under low light leads to the expansion of the trajectory prediction error. The Reason2Drive test showed a 3.2-fold increase in the prediction error of pedestrian motion in dark environments.

(2) Planning delay: Brightlight glare causes an image processing delay of > 50 ms, resulting in a lag in emergency braking decisions [17].

(3) End-to-end impact: As shown in Figureure. 2, the MTADS index (86) was significantly lower than that of indoor light (92) under natural light, indicating that light indirectly constrained the generation of control instructions through perceived quality [20].

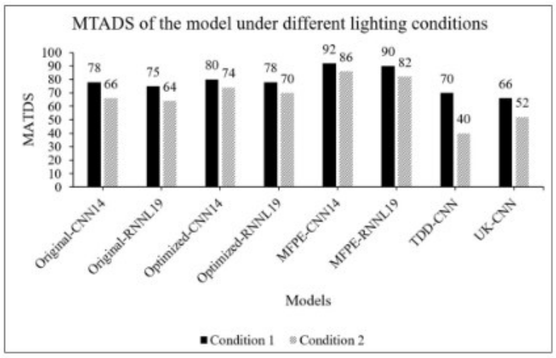

Figure 2 Comparison of MTADS metrics under two lighting conditions (black bar: Condition 1; gray bar: Condition 2). MTADS is the metric used in this article to evaluate the performance of the instrument, with higher values indicating better detection performance.

4 Optimization method: Multi-modal collaboration to solve the lighting bottleneck

4.1 Sensor fusion: heterogeneous complementarity and feature-level synergy

4.1.1 Convergence paradigm and typical architecture

Table 3 Comparison of different fusion levels and typical technologies

Integration level	Technical representatives	Core mechanism	Improved light robustness	limitations
High	YOLOv5+ radar projection	The radar detection frame is mapped to the image coordinates and fused with the YOLO detection results at the decision level	Foggy day small target recall ↑ 18.7%	Rely on single-sensor detection accuracy
Intermediate level	BEV spatial dynamic attention	LiDAR generates pseudo-depth maps→ optimizes the BEV features of the camera, and dynamically weights and fuses multimodal features	Night 3D Detection AP ↑ 12.3%	High computational complexity
Low level	Camera-dual light sensor	Lighting intensity data is directly involved in image pixel enhancement, improving edge contrast	The detection rate of lane lines under strong light ↑ 25.1%	Hardware synchronous calibration is required

4.1.2 Dedicated sensor improvement

(1) LiDAR anti-interference design

1550nm wavelength laser (vs traditional 905nm) is used to reduce the influence of fog scattering. Dynamic power adjustment: The transmitting power is increased to 30W in foggy days (5W in clear sky), and the detection distance is maintained at >100m.

(2)Dual optical sensing systemThe ambient light intensity and infrared spectrum are simultaneously collected to guide the adaptive adjustment of the camera's exposure parametersL_{amb}:

$t_{exp} = k \cdot \frac{L_{target}}{L_{amb}}$, k is the response coefficient of the camera

4.2 Algorithm improvement: lighting invariance and domain adaptation

4.2.1 Learning of light invariance characteristics

(1) Normalized Laplace operator：

Construct a lighting channel that eliminates shadow interference.$L_{inv} = \nabla^2 \left(\frac{L^*}{\backslash avg(L^*)}\right)$

(2) Divisive Normalization: Local normalization of luminance channels[17]：

$$I_{norm}(x,y) = \frac{I(x,y)}{\sqrt{\sum_{i,j \in \Omega} w_{ij} I^2(x+y, y+j) + \epsilon}}$$

The semantic segmentation accuracy of foggy days was significantly improved (mIoU ↑ 7.2%).

4.2.2 Multi-task joint optimization

(1) AllWeatherNet：
1)Hierarchical Enhancement Architecture:
Scene level: Global illumination correction
Object level: Target contrast enhanced
Texture level: Edge detail restoration
2)SIAM mechanismScaled Illumination-aware Attention dynamically adjusts the weights of each layer
(2) Combined light-detection training:
The LightDiff framework introduces Perceptual Scores (PS):

$$PS = E_{x\sim p_{data}}[f_{det}(G(x))]$$

Among them, G is the diffusion enhancement model, which is the detection network with frozen parameters, and the PS reward is maximized through reinforcement learning.f_{det}

4.3 Image enhancement: physics models merge with deep learning

4.3.1 Modernization of traditional models

HE-MSR-COM, SENA pipeline.

4.3.2 Generative augmentation breakthrough: lightdiff diffusion framework

Effect: nuScenes detected mAP ↑ 8.7% at night, and FID index was better than CycleGAN by 23.6%

5 Conclusion

This paper systematically reviews and compares the main methods of autonomous driving pedestrian detection under complex lighting conditions, and analyzes the mechanism of light degradation transmission to system performance through the sensor layer and algorithm layer. The main conclusions include:

(1) multimodal fusion (e.g., RGB-thermal imaging, LiDAR-camera) significantly improves recall and positioning accuracy in low-light and bad weather scenarios;

(2) Representation learning based on invariant lighting features and generative data enhancement partially make up for the shortcomings of data, but there are inevitable trade-offs with real-time.

(3) The physical limits of sensors (such as dynamic range and lack of point clouds) constitute the hard boundary of perception ability, which needs to be mitigated by sensor improvement and system redundancy.

(4) Data sharing, cross-domain generalization, and ethical supervision are key non-technical factors for long-term and secure deployment.

Based on the above conclusions, it is recommended that future research prioritize the development of low-cost and efficient fusion frameworks and lightweight robust algorithms, promote cross-agency data collaboration and federated learning schemes, and consider regulations and ethical constraints in the engineering deployment stage to promote the engineering realization of reliable pedestrian detection under all-climate conditions.

References

[1] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 1, 886–893.

[2] Chen, L., et al. (2020). Survey of pedestrian action recognition techniques for autonomous driving. Tsinghua Science and Technology, 25(4), 458–470.

[3] Hemmati, M., Biglari-Abhari, M., & Niar, S. (2022). Adaptive real-time object detection for autonomous driving systems. Journal of Imaging, 8 (4), 106.

[4] Han, Y., et al. (2023). Low-illumination road image enhancement by fusing Retinex theory and histogram equalization. Electronics, 12(4), 990.
[5] Li, J., et al. (2024). Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. arXiv.
[6] Tian, D., Han, Y., Wang, B., Guan, T., & Wei, W. (2021). [Retracted] A review of intelligent driving pedestrian detection based on deep learning. Computational Intelligence and Neuroscience, 2021(1), 5410049.
[7] Schulz, D., & Perez, C. A. (2023). Two-stage pedestrian detection model using a new classification head for domain generalization. Sensors, 23 (23), 9380.
[8] Iftikhar, S., Zhang, Z., Asim, M., Muthanna, A., Koucheryavy, A., & Abd El-Latif, A. A. (2022). Deep learning-based pedestrian detection in autonomous vehicles: Substantial issues and challenges. Electronics, 11(21), 3551.1
[9] Dasgupta, K., Das, A., Das, S., Bhattacharya, U., & Yogamani, S. (2022). Spatio-contextual deep network-based multimodal pedestrian detection for autonomous driving. IEEE Transactions on Intelligent Transportation Systems, 23(9), 15940–15950.
[10] Geng, K., Dong, G., Yin, G., & Hu, J. (2020). Deep dual-modal traffic objects instance segmentation method using camera and LiDAR data for autonomous driving. Remote Sensing, 12(20), 3274.
[11] Qian, R., Lai, X., & Li, X. (2022). 3D object detection for autonomous driving: A survey Pattern Recognition, 130, 108796.
[12] Feng, D., et al. (2019). Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. arXiv.
[13] Wang, Y., Jiang, Z., Li, Y., Hwang, J. -N., Xing, G., & Liu, H. (2021). RODNet: A real-time radar object detection network cross-supervised by camera-radar fused object 3D localization. IEEE Journal of Selected Topics in Signal Processing, (Vol. 15, Issue 2).
[14] Zolali, M., Mirbaha, B., & Saffarzadeh, A. (2021). Drivers speed choice behavior considering drivers socio-demographic characteristics: A driving simulator study. IJTE, Online First.
[15] Meng, Q., Yang, X., & Liu, Y., et al. (2020). Visual navigation line extraction of agricultural mobile robot based on artificial bee colony algorithm in natural light environment Jiangsu Journal of Agricultural Sciences, (04), 919–929.
[16] Kurimo, E., Lepistö, L., Nikkanen, J., Grén, J., Kunttu, I., & Laaksonen, J. (2009). The effect of motion blur and signal noise on image quality in low light imaging. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 81–90.
[17] Zhang, Y., Carballo, A., Yang, H., & Takeda, K. (2023). Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. ISPRS Journal of Photogrammetry and Remote Sensing, 196, 146–177.
[18] Lange, R., Böhmer, S., & Buxbaum, B. (2020). CMOS-based optical time-of-flight 3D imaging and ranging. High Performance Silicon Imaging, 319–375.
[19] Nie, M., et al. (2023). Reason2Drive: Towards interpretable and chain-based reasoning for autonomous driving. arXiv.
[20] Li, Y., & Qu, J. (2022). MFPE: A loss function based on multi-task autonomous driving. ECTI-CIT Transactions, 16(4), 393–409. 4
[21] Ma, Y., Wang, M., Feng, Q., He, Z., & Tian, M. (2022). Current non-contact road surface condition detection schemes and technical challenges. Sensors, 22(24), 9583.
[22] Li, S., Luo, H., Shao, T., & Hishiyama, R. (2024). A mmWave sensor and camera fusion system for indoor occupancy detection and tracking. IEICE Transactions on Information and Systems, E107. D(9), 1192–1205.
[23] Qian, C., et al. (2024). AllWeatherNet: Unified image enhancement for autonomous driving under adverse weather and lowlight conditions. arXiv.
[24] Sabry, M., Schroeder, G., Varughese, J., & Olaverri-Monreal, C. (2025). Shadow erosion and nighttime adaptability for camera-based automated driving applications. arXiv.
[25] Kutila, M., Pyykonen, P., Ritter, W., Sawade, O., & Schaufele, B. (2016). Automotive LiDAR sensor development scenarios for harsh weather conditions. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 265–270.
[26] Ogunrinde, I., & Bernadin, S. (2023). Deep camera–radar fusion with an attention framework for autonomous vehicle vision in foggy weather conditions. Sensors, 23(14), 6255.
[27] Jin, Z. (2023). Research on 3D object detection and lane line segmentation for autonomous driving based on multi-sensor fusion. (Master's thesis, Chongqing Jiaotong University). Chongqing: Chongqing Jiaotong University.
[28] Lee, Y., Park, M., & Park, M. (2022). Improving lane detection performance for autonomous vehicle integrating camera with dual light sensors. Electronics, 11(9), 1474.