ILLDet: Reliable Object Detection in Urban Underground Parking Using a Multithreaded Illumination Adapter

Bekhzod Norboev[1], Dmitry Churikov[2], Gulmira Pardaeva[1], Golib Berdiev[1], Alisher Rustamov[1] and Shamiljon Rustamov[1]

[1]*University of Information Technologies and Management, Saroy Str. 3, 180212 Qarshi, Uzbekistan*
[2]*MIREA - Russian Technological University, Vernadsky Ave. 78, 119454 Moscow, Russia*
norboyevbexzod98@gmail.com, cdv@ntcup.ru, mis.gulish@gmail.com, golibberdiev@gmail.com, alirus900@gmail.com, shamiljon1988@gmail.com

Keywords: Object Detection in Low Light, AVs' Navigation in Underground Parking, Computer Vision, YOLO.

Abstract: Reliable perception in underground parking remains difficult for autonomous vehicles, as rapid illumination changes degrade detector performance and increase safety risks. To mitigate this issue, we introduce the Urban Underground Parking (UUP) dataset, which contains more than 15,000 annotated images collected under dim, moderate, and bright lighting conditions across multiple regions and parking layouts. In addition, we develop ILLDet, a YOLO-based object detector enhanced with an illumination adapter that refines the input image before feature extraction. The adapter relies on learnable encoder–decoder blocks to form multi-exposure representations that reduce shadows, low-light regions, and glare while preserving spatial detail. Experimental evaluation on UUP indicates that ILLDet reaches 71.2% mAP@50, surpassing recent detection baselines in challenging lighting scenarios while keeping a balanced precision–recall trade-off. The detector shows more robust performance compared to baselines under illumination shifts, noise, and blur, suggesting good generalization to real underground environments. Although the enhancement module increases model complexity, the computational cost remains moderate and suitable for deployment. Overall, the proposed dataset and model establish a reproducible benchmark and provide a practical solution for vision-based perception in GPS-denied, safety-critical underground parking applications.

1 INTRODUCTION

Autonomous vehicles (AVs) are the next-generation technology that offers safe, efficient, and sustainable mobility to users. However, a significant challenge is the detection of various objects in an underground parking, which is essential for safe AV navigation. Real-world underground parking has several complications, such as low light, pillar occlusion, GPS-denied operation, etc., which cause a significant decrease in performance, such as about 40% lower mAP [1].

Previous research demonstrates that traditional detectors do not work well in low light [2]. This has triggered work on multi-sensor fusion and semantic SLAM to compensate [3]. Although the union of the two enhances geometric reasoning, the graphical front-end suffers especially in the underground areas due to noise, blurring, and color distortion. In such settings, the sensing range of LiDAR(r) is normally restricted to 50-100 m [4] and can only be aware of objects in the immediate vicinity. In addition, GPS signals disappear completely in enclosed structures, due to attenuation [5]. On the other hand, cameras have wider angular coverage, but their functionality declines drastically in low light, a long-established problem in driving in the underground and at night [6].

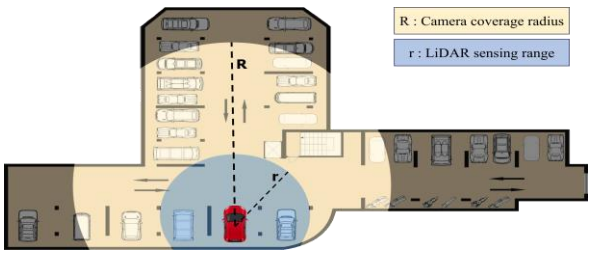

Figure 1: Sensor coverage in underground parking: LiDAR range *r* is limited, while camera coverage *R* extends further.

This is an essential gap and the core of our work. The model detects objects outside the LiDAR field of

view by taking advantage of the camera's wide field of view as it is designed to work in unfavourable light and visible conditions, as shown in Figure 1.

Recent progress in low-light enhancement, especially zero-reference depth-curve estimation with dehazing (ZDCE-DNet), has shown a high level of performance in underground mine images with the task of jointly addressing the illumination and haze elimination to enhance visual intelligibility [7]. However, direct application of Z-DCE-DNet in urban underground parking, including reflective surfaces, signage, heterogeneous vehicles, and dynamic occlusions, is inapplicable without architecture adjustment. We propose an end-to-end framework that first adapts illumination using a frequency-domain encoder, which is inspired by DarkIR [8], and then uses YOLOv11's backbone and FPN-PAN for object localization, extracting objects in the darkest underground corners, up to sunniest garage doors. Our contributions:

- We propose a new dataset called UUP, focusing on underground scenarios. UUP comprises more than 15,000 carefully collected images with various luminous intensities, such as low, medium, and high.
- We propose YOLOv11-based detector, which incorporates a specialized illumination adapter that enhances input images prior to feature extraction and feeds it to the detection block.
- We perform comprehensive benchmarking of state-of-the-art (SOTA) detectors along with ILLDet, establishing a reproducible baseline.

1.1 Our Dataset: UUP

Although datasets such as nuScenes [9] and BDD100K [10] contain incidental parking scenes, none explicitly address severe occlusion and illumination changes in underground facilities, and this is a gap that UUP seeks to overcome. The UUP original real-world dataset was collected from different regions of the world and under various lighting conditions.

We obtained 15,798 frames through dashcam footage taken globally. We aimed at dashcams as they have a comparable height, motion, and field-of-view to autonomous vehicle front cameras, which have achieved high scalability in terms of realism [11] while implementing larger sensors is unfeasible in practice. We focused on the footage in high-density areas such as China, South Korea, and Western Europe, where underground parking is prevalent due to a lack of land [12]. We classified all videos in accordance with the heavy indicator (region) (America, Asia, Europe, Unknown) and the lighting condition: Dim (<5 lux), Modern (5-50 lux), and Bright (>50 lux), to allow conducting strict performance of illumination-robust detection. Figure 2 illustrates the geographic and lighting diversity of UUP. Each row and column corresponds to a lighting condition and a region, respectively. The variety in ceiling design, vehicle types, and illumination sources reflects real-world underground parking environments.

In order to guarantee privacy compliance, we obscured all identifiable personal data, such as faces or license plates. We have outlined six main categories, such as *Barrier Gate, Car, Cone, Motorcycle, Parking Spot* and *Pedestrian*. UUP will be released publicly with COCO-style annotations and train/val/test splits to support reproducible research.

2 THE PROPOSED FRAMEWORK: ILLDET

Our proposed architecture, ILLDet, boosts detection under diverse illumination conditions by integrating a multi-exposure illumination adapter with a YOLOv11-style detection backbone. Its design is a hybrid which includes (i) adaptive exposure fusion and (ii) residual integration together with backbone features and (iii) multi-scale detection heads. Figure 3 provides an overview of the model.

2.1 Multi-Exposure Generation

Given an RGB image $I \in \mathbb{R}^{\{H\times W\times 3\}}$, we generate three exposures: the original image I_{orig}, a low-light enhanced variant I_{low}, and a high-exposure enhanced variant I_{high}. The latter two are obtained through gamma correction:

$$I_\gamma(x) = I(x)^\gamma,\ \gamma \in \{1.5, 0.7\}, \qquad (1)$$

where x indexes pixel intensities normalized to [0,1]. Gamma correction has long been applied in low-light enhancement and HDR pipelines [13], [14], but here it is embedded directly within the detection pipeline to provide diverse illumination priors.

Figure 2: Some representative samples from the UUP dataset showing regional diversity and varying illumination levels.

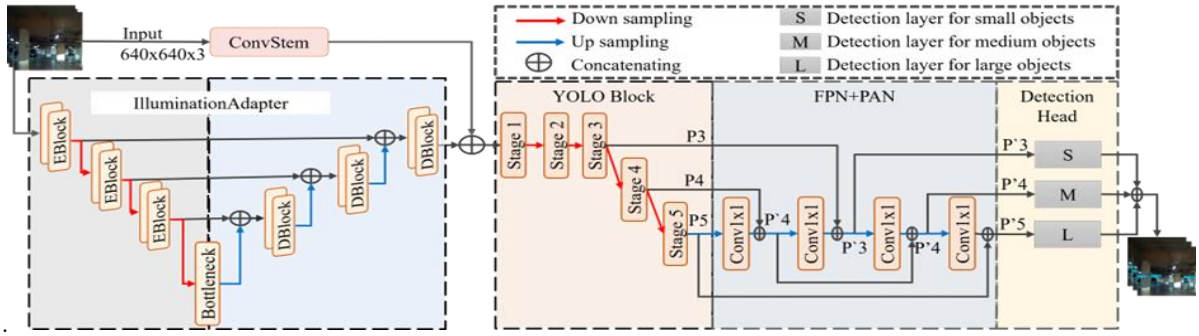

Figure 3: Architecture of ILLDet, integrating an Illumination Adapter with YOLO and FPN+PAN for feature enhancement under varying illumination and enabling robust detection of small, medium, and large objects in underground parking environments.

2.2 Illumination Adapter

The Illumination Adapter encodes three exposures into a shared representation, applies exposure-adaptive fusion, and decodes the fused features to the spatial resolution of the stem. Each exposure is processed by a convolutional encoder $E(\cdot)$:

$$f_i = E(I_i), i \in \{orig, low, high\}. \quad (2)$$

Each encoder stage employs dilated depthwise convolutions and residual connections, which have proven effective in capturing multi-scale context in both CNNs and vision transformers [14] . We further apply a Spatial-Channel Attention (SCA) module to refine activations, similar in spirit to channel-attention designs such as SE-Net but adapted for illumination-variant features.

Adaptive Fusion. To combine exposure features, we concatenate and apply a global attention head over the global average pooling (GAP).

$$w = \text{Softmax}(\text{MLP}(\text{GAP}(b))), \quad (3)$$

$$f_{fused} = w1 f_{orig} + w2 f_{low} + w3 f_{high}. \quad (4)$$

This weighting resembles exposure-fusion methods in HDR imaging [15], but unlike traditional HDR, the weights are learned end-to-end to maximize detection accuracy under diverse lighting.

Decoder and Residual Integration: The fused features are decoded via PixelShuffle upsampling and lightweight residual blocks, producing f_{dec}. A residual merge with the initial stem features S_0 is then performed:

$$S_0' = S_0 + f_{dec} \quad (5)$$

This residual integration preserves spatial fidelity from the input while enriching the representation with illumination-aware corrections.

2.3 Backbone and Feature Fusion

The backbone follows a CSP-based design as in YOLOv11, producing pyramid features $\{P_3, P_4, P_5\}$ at resolutions 80×80, 40×40, and 20×20. To enhance cross-scale representation, we employ an FPN-PAN fusion strategy, originally introduced in FPN [16] and later improved by PANet [17]. Topdown and bottom-

up pathways propagate semantic and localization cues bidirectionally, where ϕ is a 3 × 3 convolution.

$$N_4 = \phi(\text{Upsample}(P_5) \oplus P_4), \quad (6)$$

$$N_3 = \phi(\text{Upsample}(N_4) \oplus P_3). \quad (7)$$

2.4 Detection Head

At each scale $s \in \{3,4,5\}$, we predict anchor-based bounding boxes and class probabilities that the output vector is:

$$y_{s,a} = [\hat{x}, \hat{y}, \hat{\omega}, \hat{h}, \hat{o}, \hat{C}_1, \ldots \hat{C}_i], \quad (8)$$

where $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ are bounding box parameters, $\hat{o}$ is objectness, and $\hat{c}_i$ is the probability of class i. Following prior YOLO variants [18], predictions across three scales capture both small and large objects effectively.

2.5 Loss Function

The total training loss integrates bounding box regression, objectness, and classification:

$$\mathcal{L} = \lambda_{\text{box}} \mathcal{L}_{\text{GIoU}} + \lambda_{\text{obj}} \mathcal{L}_{\text{BCE}}^{\text{obj}} + \lambda_{\text{cls}} \mathcal{L}_{\text{BCE}}^{\text{cls}}. \quad (9)$$

We adopt the Generalized IoU loss [19] for localization, as it is more robust than L1 loss or IoU losses, and binary cross-entropy for objectness and classification. Empirically, we set $\lambda_{box} = 1.0, \lambda_{obj} = 1.0, \lambda_{cls} = 0.5$.

3 EXPERIMENT

Setup: We conduct experiments on the UUP dataset comprising 15,798 images with 95,328 annotated objects across six categories: *Barrier Gate, Car, Cone, Motorcycle, Parking Spot* and *Pedestrian*. We split UUP dataset into 10,038 training, 3,870 validation, and 1,890 test images, ensuring no scene overlap. All models were implemented in PyTorch and trained on a server with 2 NVIDIA GeForce RTX 4060 (8GB) using the Adam optimizer with a learning rate of 2×10^{-3} and batch sizes ranging from 4 to 16. To address illumination challenges, we apply low-light simulation via multiexposure branches and gamma correction for high/low exposure variants. We also conduct experiments on the general object detection benchmarks, including YOLO family and RTDETR, and evaluated using mAP@50, mAP@0.5:095, precision, recall, and FPS. Our experimental results highlight that ILLDet achieves superior accuracy in low-light conditions while maintaining competitive efficiency.

3.1 Results and Analysis

We compare ILLDet with SOTA detectors including YOLO v5(lu) [20], YOLO v8l [21], YOLO v9m [22], YOLO v10l [23], YOLO v11l [24], YOLO v12l [25], and RF-DETR-S s [26]. As shown in Table 1, ILLDet has the highest mAP@0.5(71.23%), better than the best baseline, RF-DETR-S, by +2.4%. Although YOLOv11l had the best use of precision (85.20%), it had poor recall (56.62%), meaning that it could not identify minority classes. However, ILLDet strikes a balance between the precision (69.40%), and the highest recall (82.63%), resulting in the highest F1-score (75.44%). ILLDet contains 43.4 M parameters and 181 G FLOPS, which allows it to achieve multi-exposure and illumination correction. These results demonstrate that our design provides high detection reliability in challenging low-light conditions without compromising real-world performance for autonomous vehicles.

ILLDet demonstrates superior robustness under adverse conditions. various unwanted circumstances such as low-light, changes in brightness, blur and noise. Table 2 above demonstrates that ILLDet is the most successful at detection with all perturbation conditions, and most notably with extremely low-light and noisy conditions. In harsh low-light (S=5) ILLDet achieves 49.7% mAP under severe low-light (S=5), which is higher than RF-DETR-S by +5.9% and YOLOv9m by +13.9%. On the same note, ILLDet has a 45.8% mAP in noisy conditions (S=5), which is 40% and 20% higher than YOLOv11l(32.1) or YOLOv9m (36.8), respectively. These results provide further support for ILLDet that, besides operating with the strong baseline performance, it also provides solid generalization to difficult environments in real-world scenarios that are likely to occur in underground parking setups.

3.2 Ablation Study

In order to determine the role of each part in ILLDet, we conducted systematic ablations by freezing the encoder block (En), decoder block (De) or both. According to Table 3, freezing either block leads to notable drops in detection performance, with mAP@0.5 decreasing from 71.2% in the full ILLDet model to 66.5% and 59.6%, respectively. Similarly, the F1 score drops substantially, highlighting the synergistic role of both modules in robust object localization under challenging visual conditions.

Table 1: Comparison with SOTA detectors on the UUP dataset.

Model	Year	mAP@50	mAP@0.5:0.95	Precision	Recall	F1	Param (M)	FLOPs(G)
YOLOv5lu	2020	62.18	37.44	82.40	56.06	66.53	53.2	135.0
YOLOv8l	2023	61.01	38.28	79.27	55.55	65.27	43.7	165.2
YOLOv9m	2024	63.42	37.27	77.54	59.26	67.15	20.1	76.80
YOLOv10l	2024	43.04	24.27	64.02	47.16	54.33	24.4	120.3
YOLOv11l	2024	62.72	34.24	85.20	56.62	67.99	25.3	86.90
YOLOv12m	2025	50.29	29.13	67.88	51.17	58.34	20.2	67.50
RF-DETR-S	2025	69.53	36.15	69.53	59.56	64.17	32.1	45.01
ILLDet(Ours)	2025	71.23	35.69	69.40	82.63	75.44	43.4	181.0

Table 2: Robustness evaluation of detection models on visual distortions. Severity (S) denotes increasing distortion levels.

Effect	S	YOLOv11l	YOLOv9m	RF-DETR-S	ILLDet
Clean	–	62.7	63.4	69.5	71.2
Lowlight	1	52.3	54.1	55.6	57.3
	3	46.5	48.4	51.1	52.9
	5	41.5	43.6	46.9	49.7
Brightness	1	52.7	54.3	56.2	58.0
	3	48.2	50.2	52.5	54.5
	5	44.5	46.5	48.9	51.1
Blur	1	50.3	51.9	53.8	56.7
	3	44.0	46.2	49.3	51.7
	5	35.9	39.1	43.2	47.3
Noise	1	49.1	51.0	53.0	55.8
	3	41.0	44.2	47.7	50.8
	5	32.1	36.8	41.5	45.8

Table 3: Ablation study showing performance and parameter scaling across model variants (En is Encoder, De is Decoder).

Metric	w/o En+De	w/o En	w/o De	ILLDet
mAP@0.5	51.8	59.6	66.5	71.2
mAP@0.5:095	25.8	29.2	32.6	35.7
F1 Score	28.4	31.8	34.1	37.7

Note. That F1 in Table 3 is computed over IoU=0.5:0.95 and is therefore not directly comparable with Table 1.

4 CONCLUSIONS

Reliable object perception in underground parking environments remains challenging for autonomous vehicles due to illumination variations, occlusions, and GPS-denied operation. In this work, we introduced the Urban Underground Parking (UUP) dataset, a dedicated benchmark containing diverse underground scenarios and lighting conditions for evaluating illumination-robust object detection. Unlike existing autonomous driving datasets, UUP specifically focuses on the visual challenges of enclosed parking environments and enables reproducible assessment of detection methods under difficult illumination conditions.

Based on UUP, we proposed ILLDet, a YOLO-based detection framework enhanced with a multi-exposure illumination adapter. The proposed module improves feature representation by adaptively combining exposure information before detection, allowing the model to better handle low-light regions, shadows, and brightness variations. Experimental results show that ILLDet achieves 71.23% mAP@50 with improved recall of 82.63%, outperforming representative state-of-the-art detectors on the proposed benchmark. Robustness and ablation studies further demonstrate the effectiveness of the illumination adapter in improving detection stability under challenging visual distortions.

Although the proposed framework demonstrates promising performance, further work is required to evaluate real-time deployment scenarios and extend the dataset with additional environments and dynamic conditions. Future research will focus on improving computational efficiency and integrating illumination-aware perception with higher-level autonomous driving modules, including collision avoidance and path planning. Overall, UUP and ILLDet provide a foundation for more reliable vision-

based perception in GPS-denied and safety-critical underground environments.

REFERENCES

[1] K. Haritha, P. B. Prakash, D. Pravallika, K. Venkatesh, and G. Venkatesh, "Enhancing object detection in autonomous vehicles under low-light conditions using federated learning and YOLOv5," International Journal of Modern Trends in Science and Technology, 2025.

[2] M. Ahmed et al., "Survey and performance analysis of deep learning-based object detection in challenging environments," Sensors, vol. 21, no. 15, art. 5116, 2021.

[3] J. S. O. Medina et al., "An overview of autonomous parking systems: Strategies, challenges, and future directions," Sensors, vol. 25, no. 14, art. 4328, 2025.

[4] J. Behley et al., "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 9297-9307.

[5] K. A. Fisher and J. F. Raquet, "Precision position, navigation, and timing without the global positioning system," Air Space Power Journal, 2011.

[6] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," Computer Vision and Image Understanding, vol. 178, pp. 30-42, 2019.

[7] Q. Du et al., "A hybrid zero-reference and dehazing network for joint low-light underground image enhancement," Scientific Reports, vol. 15, no. 1, art. 10135, 2025.

[8] D. Feijoo et al., "DarkIR: Robust low-light image restoration," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2025, pp. 10879-10889.

[9] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 11621-11631.

[10] F. Yu et al., "BDD100K: A diverse driving video database," Univ. of California, Berkeley, CA, USA, Tech. Rep., 2018.

[11] G. Yang et al., "DrivingStereo: A large-scale dataset for stereo matching in autonomous driving scenarios," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 899-908.

[12] Parkopedia, "Global parking index 2022," 2022, [Online]. Available: https://business.parkopedia.com/parking-data, [Accessed: Jan. 15, 2026].

[13] C. Chen et al., "Learning to see in the dark," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 3291-3300.

[14] Z. Liu et al., "A convnet for the 2020s," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 11976-11986.

[15] C. Li et al., "Learning a deep single-image contrast enhancer from multi-exposure images," IEEE Transactions on Image Processing, vol. 27, pp. 5634-5645, 2018.

[16] T.-Y. Lin et al., "Feature pyramid networks for object detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2117-2125.

[17] S. Liu et al., "Path aggregation network for instance segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 8759-8768.

[18] J. Redmon et al., "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779-788.

[19] H. Rezatofighi et al., "Generalized intersection over union: A metric and a loss for bounding box regression," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 658-666.

[20] G. Jocher, "Ultralytics YOLOv5," GitHub repository, 2020, [Online]. Available: https://github.com/ultralytics/yolov5.

[21] G. Jocher, A. Chaurasia, and J. Qiu, "YOLOv8," arXiv preprint arXiv:2305.09972, 2023.

[22] C.-Y. Wang and H.-Y. M. Liao, "YOLOv9," arXiv preprint, 2024.

[23] L. Liu et al., "YOLOv10: Real-time end-to-end object detection," arXiv preprint arXiv:2405.14458, 2024.

[24] G. Jocher and J. Qiu, "YOLOv11," Ultralytics Technical Report, 2024.

[25] Y. Tian et al., "YOLOv12: Attention-centric real-time object detectors," arXiv preprint, 2025.

[26] Robinson et al., "RF-DETR," GitHub repository, 2025, [Online]. Available: https://github.com/roboflow/rf-detr, [Accessed: Jan. 20, 2026].