

DarkDriving: A Real-World Day and Night Aligned Dataset for Autonomous Driving in the Dark Environment

Wuqi Wang^{1,*}, Haochen Yang^{2,*}, Baolu Li², Jiaqi Sun¹, Xiangmo Zhao¹, Zhigang Xu¹, Qing Guo³,
Haigen Min^{1,†}, Tianyun Zhang^{2,†}, Hongkai Yu^{2,†}

Abstract—The low-light conditions are challenging to the vision-centric perception systems for autonomous driving in the dark environment. In this paper, we propose a new benchmark dataset (named DarkDriving) to investigate the low-light enhancement for autonomous driving. The existing real-world low-light enhancement benchmark datasets can be collected by controlling various exposures only in small-ranges and static scenes. The dark images of the current nighttime driving datasets do not have the precisely aligned daytime counterparts. The extreme difficulty to collect a real-world day and night aligned dataset in the dynamic driving scenes significantly limited the research in this area. With a proposed automatic day-night Trajectory Tracking based Pose Matching (TTPM) method in a large real-world closed driving test field (area: 69 acres), we collected the first real-world day and night aligned dataset for autonomous driving in the dark environment. The DarkDriving dataset has 9,538 day and night image pairs precisely aligned in location and spatial contents, whose alignment error is in just several centimeters. For each pair, we also manually label the object 2D bounding boxes. DarkDriving introduces four perception related tasks, including low-light enhancement, generalized low-light enhancement, and low-light enhancement for 2D detection and 3D detection of autonomous driving in the dark environment. The experimental results show that our DarkDriving dataset provides a comprehensive benchmark for evaluating low-light enhancement for autonomous driving and it can also be generalized to enhance dark images and promote detection in some other low-light driving environment, such as nuScenes.

I. INTRODUCTION

The vision-centric perception systems are widely used on the intelligent vehicles due to the effective cost of camera sensors compared to the expensive cost of LiDAR sensors, such as the Tesla’s Full Self-Driving (FSD) system. The vision-centric perception systems are reliable to perceive the nearby environment in most of driving scenarios, however the low-light conditions in the dark environment might result in the significant challenges to the camera based vision-perception systems. For example, the nighttime fatal rate is 18% higher than the daytime fatal rate [9], and an Uber self-driving vehicle failed to detect a pedestrian and killed that pedestrian in Arizona at one dark night of 2018 [10]. These statistical data and driving fatalities during nighttime increase the demands to investigate autonomous driving in the dark environment.

One recent study [11] has demonstrated that advanced low-light enhancement methods can improve the robustness of vision-centric perception systems for autonomous



(a) DarkDriving Dataset



(b) LOL Dataset [1]

(c) Dark Zurich Dataset [2]

Fig. 1: Data examples of our DarkDriving and other representative low-light enhancement and nighttime driving datasets: (a) our DarkDriving with precisely aligned location and spatial contents (error: centimeters) of day and night, (b) low-light enhancement dataset by controlling exposures, and (c) nighttime driving dataset with roughly aligned daytime location (error: meters) and unaligned spatial contents using the GPS signal.

driving at night. However, there is currently no real-world dataset providing day and night paired data specifically for investigating low-light enhancement in autonomous driving. Existing real-world low-light enhancement datasets typically acquire paired images either by varying camera exposures or by capturing static scenes under different lighting conditions within small regions, *e.g.*, the LOL dataset [1]. Nighttime driving datasets can provide approximate daytime counterparts using GPS signals (error: meters), but they fail to ensure spatial content alignment due to moving objects and pose variations, *e.g.*, the Dark Zurich dataset [2]. Due to these limitations, some studies resort to synthetic datasets by artificially darkening daytime images [11], yet such methods do not realistically simulate dynamic nighttime conditions, including vehicle and street lighting variations. The lack of a real-world dataset with precisely aligned day and night image pairs presents a significant challenge in this field.

This paper introduces a real-world benchmark dataset, called DarkDriving, to advance research on low-light enhancement for autonomous driving. Leveraging the large-scale closed driving test field (69 acres) at Chang’an Uni-

¹Chang’an University. ²Cleveland State University. ³A*STAR.

*Equal contribution. [†]Co-corresponding authors: hgmin@chd.edu.cn, t.zhang85@csuohio.edu, h.yu19@csuohio.edu.

TABLE I: Comparison of our DarkDriving dataset and other low-light enhancement and nighttime driving datasets.

Dataset	Year	Real/ Simulated	Independent of Exposure Control	RGB Images	Scene	Day and Night Paired Images	Pair Alignment Error
LOL [1]	2018	Real	×	1,000	Static	✓	-
SCIE [3]	2018	Real	×	4,413	Static	✓	-
SID [4]	2018	Real	×	5,094	Static	✓	-
BBD100K-dark [5]	2020	Real	✓	14K	Driving	×	-
NuScenes-night [6]	2020	Real	✓	4.6K	Driving	×	-
Dark Zurich [2]	2020	Real	✓	8,779	Driving	✓	Meters
NightCity [7]	2021	Real	✓	4,297	Driving	×	-
SHIFT [8]	2022	Simulated	✓	2.5M	Driving	×	-
DarkDriving (Ours)	2026	Real	✓	19,076	Driving	✓	Centimeters

versity, we develop a novel Trajectory Tracking based Pose Matching (TTPM) method to automatically collect day and night image pairs under real-world driving conditions. The TTPM method ensures that the autonomous vehicle precisely follows the same trajectory at night as it did during the day, achieving fine-grained alignment in both location and spatial content, with an alignment error within centimeters. The proposed DarkDriving dataset consists of 9,538 precisely aligned day and night image pairs, with manually annotated objects. A comparative analysis with other representative datasets is presented in Fig. 1 and Table I. To further explore autonomous driving in low-light conditions, we extend our study to include four perception tasks for autonomous driving in dark environments, namely: low light enhancement, generalized low-light enhancement, and low-light enhancement for 2D detection and 3D detection. Our contributions are summarized as follows.

- We propose the first real-world benchmark dataset (DarkDriving) for studying low-light enhancement in autonomous driving, containing 9,538 day and night aligned image pairs collected from a large-scale closed driving test field (69 acres).
- We design a novel automatic Trajectory Tracking based Pose Matching (TTPM) method, which ensures precise location and spatial content alignment for each day and night image pair with an alignment error of just a few centimeters.
- Using the proposed DarkDriving dataset, we investigate four perception tasks, *i.e.*, low-light enhancement, generalized low-light enhancement, and low-light enhancement for 2D detection and 3D detection, to advance research on nighttime autonomous driving.

II. RELATED WORK

A. Datasets for Low-light Enhancement

The task of low-light image enhancement focuses on recovering obscured details in poorly illuminated regions, thereby improving the image visibility. In supervised learning approaches, paired datasets play a crucial role. Several publicly available datasets provide such paired images for low-light enhancement [12], [13], [14], [1], [3], [4]. The LOL dataset [1] was the first to introduce real-scene low/normal-light image pairs for enhancement tasks. SCIE [3] and SID [4] offer images captured at different exposure levels. However, these datasets only cover small static scenes not

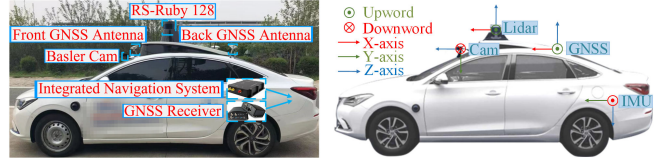


Fig. 2: Setup of sensors on the automated vehicle and their coordinate systems.

TABLE II: Specification of sensors on automated vehicle.

Sensors	Details
1×Camera	RGB, 2,448×2,048 resolution, 41° FOV
1×LiDAR	128 channels, 200m capturing range, -25° to +15° vertical FOV, Range Precision: ±2cm error
GPS & IMU	Xsens MTi-680G, 400 Hz SDI

designed for autonomous driving, so they are not suitable for investigating the dark driving problem.

B. Datasets for Nighttime Driving

Several datasets have been developed to support nighttime autonomous driving research. NuScenes-night [6] was the first dataset collected from an autonomous vehicle operating on public roads at night, including LiDAR, cameras, and radar data. BDD100K [5] contains a subset of nighttime driving scenes (BDD100K-dark) to benchmark image recognition algorithms in low-light conditions. Dark Zurich [2] offers a collection of images captured at different times, including roughly matched day-and-night image pairs aligned using GPS data. However, because of the coarse alignment and the lack of precisely paired images, these datasets are not ideal for supervised low-light enhancement training. Although simulation-based approaches [8] offer the advantage of generating paired day-night images, they cannot accurately capture the complexities of real-world nighttime driving. Despite these efforts, no existing dataset provides precisely aligned real-world day-and-night images for autonomous driving in dark.

III. DARKDRIVING DATASET

A. Automated Vehicle and Sensor Setup

We collect the DarkDriving dataset using one automated vehicle and multiple background vehicles. The Fig. 2 shows the automated vehicle (moving) used for our data collection. The background vehicles (static in the defined locations during day and night) consist of several normal vehicles, which

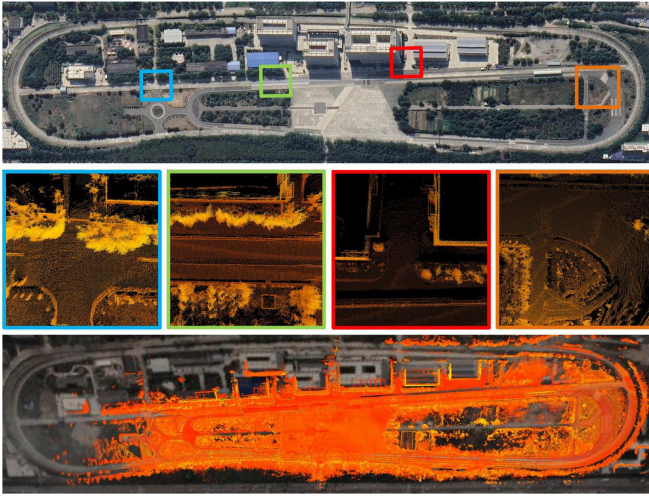


Fig. 3: **Details of the large real-world closed driving test field at Chang'an University.** From top to bottom: satellite image of driving test field, local and global demos of its High-Precision Point Cloud Map.

are employed to simulate diverse urban traffic scenarios. The automated vehicle is equipped with 128-channel LiDAR, an industrial front-facing camera, and an integrated navigation system. The layout of the automated vehicle sensors is also illustrated in Fig. 2, and the detailed sensor specification is provided in Table II. These sensor measurements use the same integrated navigation system for time synchronization, ensuring that the point cloud, camera data, and integrated navigation data within the same sequence are time-synchronized at the hardware level.

B. Large Real-world Closed Driving Test Field

To collect the day and night data of the consistent dynamic driving scenarios, an automated vehicle is used to collect the data in a large real-world closed driving test field. There are several advantages to use this driving test field for our DarkDriving data collection. The driving test field spans a large-scale area of 69 acres and includes over 40 typical vehicle driving scenarios in the real world, such as multi-lane road, curved road, intersection, streetlight and so on, which could well represent the real-world driving scenes in the dark environment. Because this driving test field is closed, we could avoid the inconsistency of moving objects in day and night and also pre-define and control the location and lighting conditions of the static background vehicles and streetlights to represent diverse nighttime driving scenarios. This driving test field provides us a High-Precision Point Cloud Map (constructed based on the LIO-SAM method [15] before) that built a digital-twin version of the real-world driving test field, whose mapping error is in just centimeters. Fig. 3 shows the satellite image of the test field and the local, global demonstrations of its High-Precision Point Cloud Map.

C. Data Acquisition by TTPM

The pipeline of our TTPM method is shown in Fig. 4. Let us introduce the details of each module in this section.

1) *High-Precision Map based Robust Localization:* Since GNSS is prone to the multipath effects in some complex driving environments, DarkDriving uses the available High-Precision Point Cloud Map (defined as \mathbf{M}) of the real-world closed driving test field for robust vehicle pose estimation. We provide continuous and accurate localization by associating the LiDAR-based point cloud data with the High-Precision Map using the Normal Distributions Transform (NDT) [16]. First of all, we uniformly divide the High-Precision Point Cloud Map \mathbf{M} into local grids, where each local grid can be represented as a Gaussian distribution \mathbf{G}_i . For each frame, the total m points $\mathbf{c} \in \mathbb{R}^{4 \times m}$ of the automated vehicle can be obtained via its LiDAR sensor, where each point is a 4D coordinate vector $[X, Y, Z, I]$. Then, NDT fits \mathbf{c} into the different local Gaussian distributions $\{\mathbf{G}_i \mid i = 1, 2, \dots, k\}$ via a Homogeneous Transformation Matrix (HTM) $\mathbf{h} \in \mathbb{R}^{4 \times 4}$ to find the Mahalanobis Distance based association error:

$$\mathbf{E}(\mathbf{h}) = \sum_{i=1}^k (\mathbf{h}\mathbf{c} - \mu_i)^T \Sigma_i^{-1} (\mathbf{h}\mathbf{c} - \mu_i), \quad (1)$$

where μ_i and Σ_i are the mean and covariance of the \mathbf{G}_i in the map \mathbf{M} . The objective is to minimize the association error to determine the optimal HTM for each frame:

$$\mathbf{h}^* = \arg \min_{\mathbf{h}} \mathbf{E}(\mathbf{h}). \quad (2)$$

After obtaining the optimal HTM \mathbf{h}^* for one frame, using \mathbf{M} as the coordinate system, the robust localization (vehicle pose) can be estimated by the homogeneous transformation. Finally, for that frame, let us re-define the estimated vehicle pose as $\mathbf{p} = [x, y, z, roll, yaw, pitch]$ after the optimal homogeneous transformation.

2) *Day and Night Trajectory Tracking:* To ensure the precisely aligned location and spatial contents of captured images during day and night, it is important to maintain the consistent moving trajectory (a time sequence of pose and velocity) of the automated vehicle during day and night. Because human driver cannot guarantee a perfectly uniform trajectory across multiple trips, substantial discrepancies in yaw and position are inevitable. Therefore, the automatic Trajectory Tracking method, widely used in robot control, is used to keep consistent trajectory of the automated vehicle during day and night.

First, we define a driving scenario in the closed driving test field and manually drive the automated vehicle for that driving scenario, where the vehicle's pose and velocity are recorded, leading to a desired driving trajectory \mathbf{T} . To facilitate the later processing and avoid motion blur, the desired driving trajectory \mathbf{T} has a rough velocity of 9 miles/hour in the manual human driving. Then, we make the automated vehicle autonomously drive and follow \mathbf{T} during daytime, resulting in an automatic daytime trajectory \mathbf{T}^d . Furthermore, we let the automated vehicle autonomously drive and follow \mathbf{T} during nighttime, generating an automatic nighttime trajectory \mathbf{T}^n .

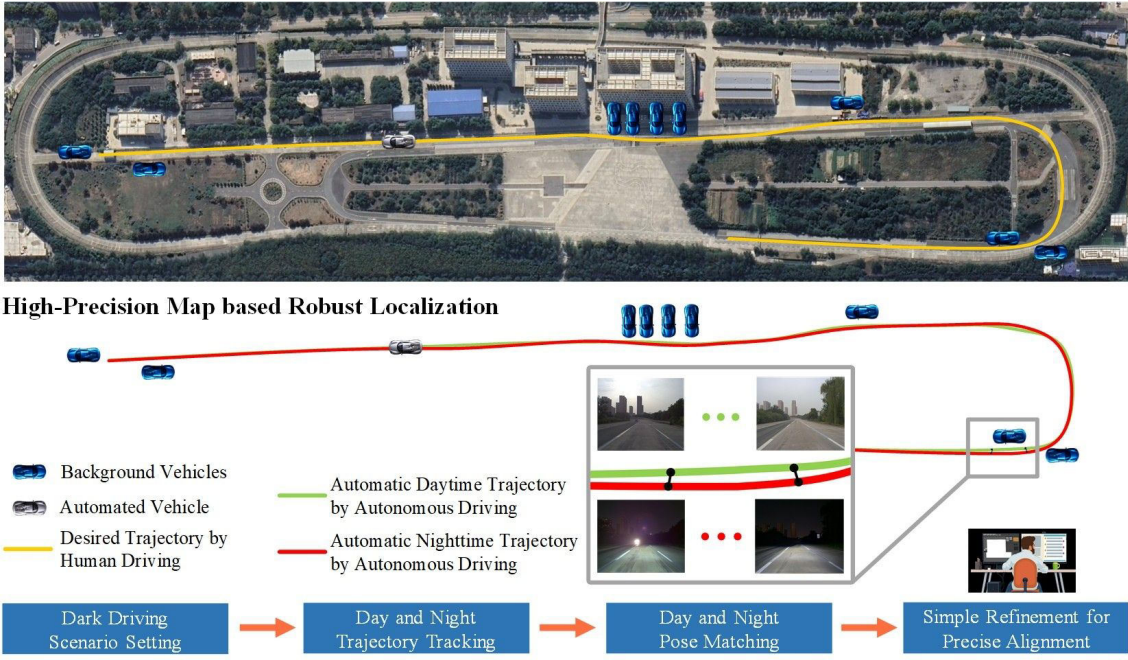


Fig. 4: **The pipeline of the automatic day-and-night Trajectory Tracking based Pose Matching (TTPM) method for data collection.** One example of the desired trajectories is shown here. Best view in color.

Our Trajectory Tracking during day and night is realized by the Pure Pursuit algorithm [17]. The Pure Pursuit algorithm Φ provides the motion planning for the next time steps, by computing the minimum Euclidean distance based on the vehicle's current pose, velocity, and the desired trajectory \mathbf{T} , which can be represented as:

$$[\mathbf{p}_{t+\Delta t}, \mathbf{v}_{t+\Delta t}] = \Phi(\mathbf{p}_t, \mathbf{v}_t, \mathbf{T}), \quad (3)$$

where Φ takes the current pose \mathbf{p}_t , current velocity \mathbf{v}_t at the time step t and the desired trajectory \mathbf{T} as the input and outputs the pose and velocity $[\mathbf{p}_{t+\Delta t}, \mathbf{v}_{t+\Delta t}]$ for motion planning in the future Δt time steps. These motion planning outputs are then fed to the PID (Proportional Integral Derivative) Control module to obtain the vehicle acceleration and steering angle for the automatic vehicle control. With this day and night Trajectory Tracking method, we could keep the daytime trajectory \mathbf{T}^d and nighttime trajectory \mathbf{T}^n consistent for the automated vehicle.

3) *Day and Night Pose Matching*: After day and night trajectory tracking, we get 23 day and night trajectory pairs and we denote each day and night trajectory pair as $[\mathbf{T}^d, \mathbf{T}^n]$. There are three challenges to find precisely aligned day and night image pairs. The trajectories \mathbf{T}^d and \mathbf{T}^n are highly consistent, but they might still have small control inconsistency due to the slight differences of vehicle weight and roadway surface in day and night. In our data collection, the collection time of different sensors is synchronized with same starting time, but they have different frequencies. The LiDAR based pose collection rate is high for robust localization (fps: 20), and the camera image collection rate during daytime is low for clear image details (fps: 10), while the camera image collection rate during nighttime is lower for more camera exposure time (fps: 6). Directly

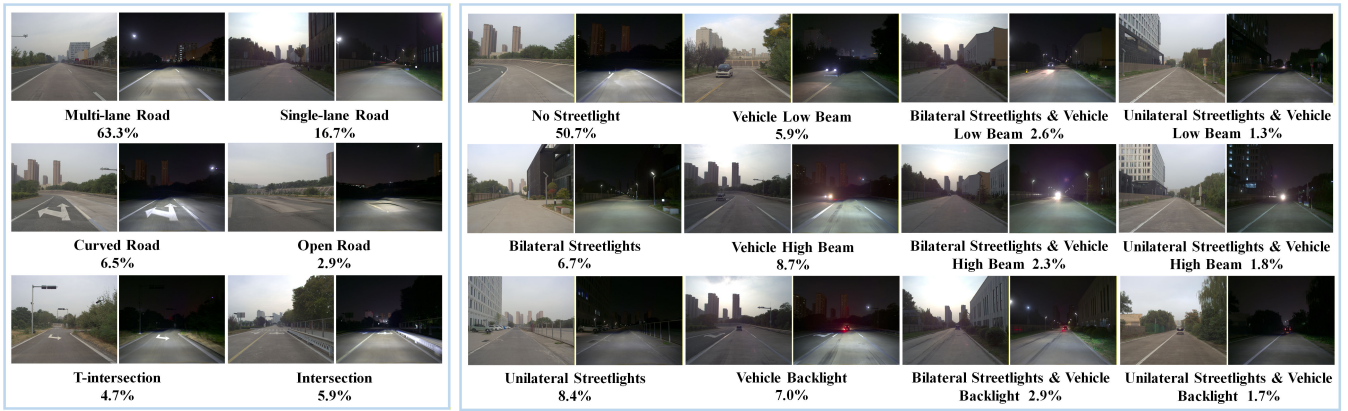
matching images in day and night is challenging because of significantly different image and lighting features in day and night.

To solve the matching challenges, we design a Day and Night Pose Matching method to find the precisely aligned day and night image pairs. Given a query camera image in daytime (camera image frame index: q), we can easily find its corresponding pose index in daytime by the nearest neighbor search in the time domain, denoted as a mapping function $\psi(\cdot)$, leading to the corresponding daytime pose $\mathbf{p}_{\psi(q)}^d$. Then we search for the minimum Euclidean distance over all the nighttime poses $\{\mathbf{p}_{\psi(f)}^n\}$ by solving the following optimization problem:

$$\begin{aligned} & \text{Minimize}_f \|\mathbf{p}_{\psi(q)}^d - \mathbf{p}_{\psi(f)}^n\|_2, \\ & \text{s.t. } \|\mathbf{p}_{\psi(q)}^d - \mathbf{p}_{\psi(f)}^n\|_2 \leq \delta \\ & \quad 1 \leq f \leq F \end{aligned} \quad (4)$$

where f is the camera image frame index and F is the total image frame number in the nighttime trajectory \mathbf{T}^n , and δ represents the maximum acceptable pose matching error, which is set to 5cm in this paper. In this way, after this pose matching based optimization, the day and night image pairs can be precisely aligned automatically.

4) *Simple Refinement for Precise Alignment*: Our data collection took 8 days in the large real-world driving test field, and the automatic Day and Night Trajectory Tracking generated 23 pairs of day-night image sequences, which was a total of over 60K images. To increase the data diversity, among them, 12 paired sequences have different desired driving trajectories \mathbf{T} . For the remaining paired sequences with shared \mathbf{T} , we set different locations and lighting conditions of background vehicles. The Day and Night Pose



(a) Diverse Types of Road Scenes

(b) Various Nighttime Lighting Conditions

Fig. 5: **Dark driving scenarios of our DarkDriving dataset.** One sample day-night image pair is shown for each dark driving scenario, followed by its percentage distribution in our dataset.

Matching automatically output 13,811 day-night image pairs with aligned location and spatial contents.

However, during our data collection, sometimes uncontrollable traffic participants might occasionally appeared in the camera. For example, uncontrollable staffs and third-party testing vehicles of the large driving test field might be occasionally captured by the camera in the daytime driving, while they were not recorded in the nighttime driving. In addition, the accumulated errors in the engineering implementation might accidentally result in the decimeter-level day and night alignment error. To further ensure the accuracy of the day and night correspondence, 4 human drivers were recruited, and each spent 5 hours to review the collected data to remove the day and night image pairs with different dynamic objects and the decimeter-level alignment errors. After this simple human refinement, we finally obtained 9,538 day and night image pairs with highly precise alignment of location and spatial contents, whose correspondence error is just in centimeter-level. The temporal consistency is significantly broken for the collected time-series image sequence after the Day and Night Pose Matching and the Simple Refinement, so we deal with each day-night image pair independently.

D. Dark Driving Scenario Setting and Analysis

To represent real nighttime driving scenarios, we collected data on 6 typical road scenarios in the large real-world driving test field: Multi-lane Road, Single-lane Road, Curved Road, Open Road, T-intersection, Intersection. By setting streetlights and beam conditions of background vehicles (static in the defined locations during day and night), we set 12 lighting conditions in nighttime driving: No Streetlight, Vehicle Low Beam, Bilateral Streetlight & Vehicle Low Beam, Unilateral Streetlight & Vehicle Low Beam, Bilateral Streetlight, Vehicle High Beam, Bilateral Streetlight & Vehicle High Beam, Unilateral Streetlight & Vehicle High Beam, Unilateral Streetlight, Vehicle Backlight, Bilateral Streetlight & Vehicle Backlight, Unilateral Streetlight & Vehicle Backlight.

We think setting diverse road scenarios and different lighting conditions of streetlights and background vehicles

is necessary to describe the various traffic scenes in the real-world nighttime driving, so we define the above multiple dark driving scenarios in this research. The Fig. 5 shows one sample day-night image pair and the percentage distribution of each dark driving scenario in our DarkDriving.

E. Data Processing and Annotation

Our dataset contains 9,538 pairs of precisely aligned day-night RGB images, where the image resolution is 2,448×2,048. Then, we randomly split each day-night sequence pair, which generates a training set of 5,906 image pairs and a testing set of 3,632 image pairs to train and test deep learning models for autonomous driving perception in the dark environment. The precisely aligned image pair data in our DarkDriving is excellent to train the deep learning models, especially for supervised learning.

For the task of Low-light Enhancement for 2D detection, we manually annotated the object 2D bounding boxes in each image of our DarkDriving dataset, leading to 13,184 annotated 2D bounding boxes. Because the data was collected in a closed driving test field without many traffic participants, we only labeled the major object class (Car) in the annotation.

IV. EXPERIMENTS

In our experiments, four autonomous driving related perception tasks are investigated.

A. Low-light Enhancement

Experiment Setup. For the low-light enhancement task, we evaluate five state-of-the-art methods on our DarkDriving dataset, including SNR-Aware [18], SNR-SKF [19], LLFormer [20], Retinexformer [21], and ControlNet [22]. For each enhancement method, we use publicly available code and train each supervised learning-based method until convergence. Specifically, for ControlNet, we finetune the pre-trained Stable Diffusion v1-5 model [23]. Before being fed into the network, all images are resized to 512×512 to maintain a consistent input format. We use three full-reference image quality metrics: PSNR, SSIM [24], and LPIPS [25] (Alex version), and four non-reference image

TABLE III: **Quantitative comparison of image quality for SOTA low-light enhancement methods on the testing set of DarkDriving.** The best and second best results are highlighted in **bold** and underline, respectively.

Methods	Full-reference Metrics			Non-reference Metrics			
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	MUSIQ \uparrow	NIQE \downarrow	HyperIQA \uparrow	CNNIQA \uparrow
Night Image	6.58	0.38	0.59	47.79	6.58	0.34	<u>0.54</u>
LLFormer	22.95	0.79	0.41	26.81	6.28	0.25	0.27
SNR-SKF	20.29	0.74	0.39	35.80	5.86	0.32	0.40
Retinexformer	<u>23.66</u>	<u>0.80</u>	<u>0.34</u>	33.18	6.06	0.31	0.34
ControlNet	13.17	0.58	0.50	65.16	<u>5.32</u>	0.66	0.67
SNR-Aware	28.55	0.81	0.15	<u>57.80</u>	4.44	<u>0.51</u>	0.51

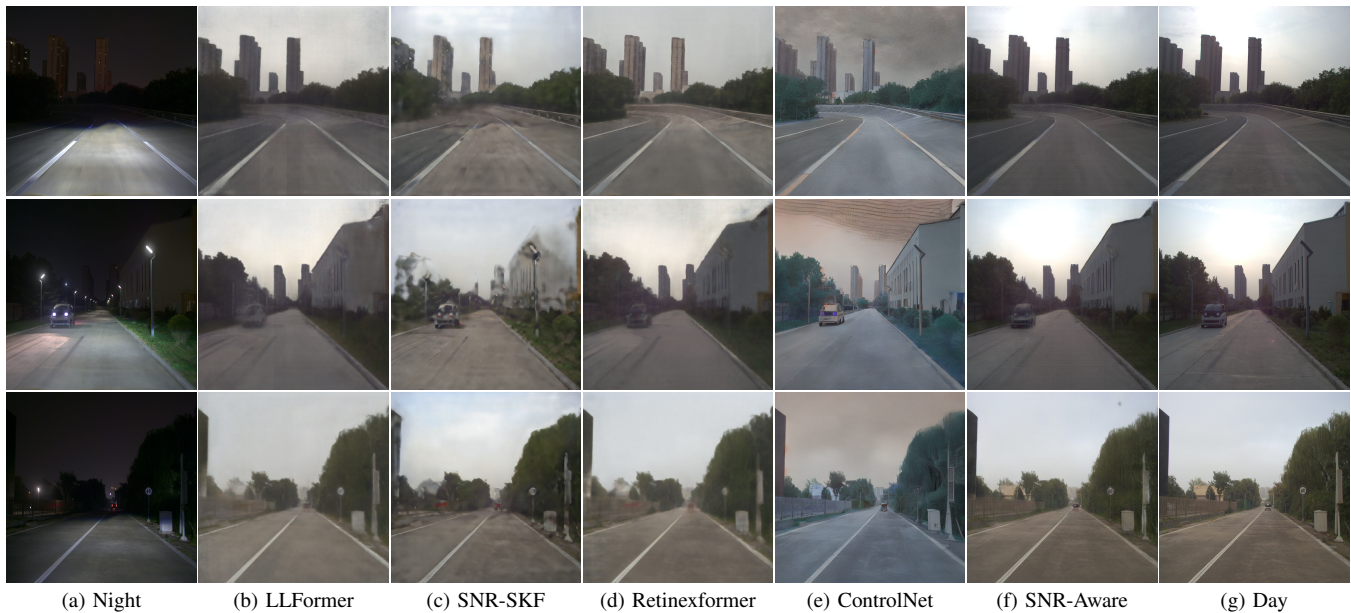


Fig. 6: **Visual comparison of nighttime image enhancement methods on the testing set of DarkDriving.** The input is the nighttime image (a) and the outputs (b–f) are to approximate the corresponding daytime image (g).

quality assessment metrics: MUSIQ [26], NIQE [27], HyperIQA [28], and CNNIQA [29] to evaluate quantitative results.

Experimental Results and Findings. As shown in Table III and Fig. 6, SNR-Aware achieved high scores in both full-reference and no-reference metrics. ControlNet performs very well in non-reference metrics but sometimes shows some hallucination tradeoffs in full-reference metrics. Retinexformer also performs well. These results highlight two key findings: 1) Low-light enhancement methods could *significantly improve* the image quality of the original night image to approximate the corresponding daytime image (ground truth), leading to the promoted camera images for driving in the dark environment. 2) SNR-Aware shows the best balanced image quality after the low-light enhancement, whose strong SSIM and LPIPS results highlight its ability to enhance the dark images while maintaining structural integrity and perceptual quality.

B. Generalized Low-light Enhancement

Experimental Setup. To verify the generalization capability of our DarkDriving dataset, we test the trained enhancement models on a different night driving dataset nuScenes [6]. First, we pre-train the SNR-Aware model (best performance in the previous enhancement experiment) with LOL-v2 [1] and DarkDriving datasets, respectively. To bridge the domain gap between datasets, following [11], we generate synthetic corresponding nighttime counterpart images of nuScenes daytime training set via digital image processing, then fine-tune the pre-trained SNR-Aware models with the synthetic night-day image pairs of nuScenes. Finally, we test the fine-tuned models pre-trained with these two datasets on the nighttime validation set of nuScenes.

Experimental Results and Findings. Table IV shows that the model pre-trained on our DarkDriving dataset produces better image quality compared to the model pre-trained on the LOL-v2 dataset, which is consistent in the visualized

results of Fig. 7. These results confirm that our DarkDriving has better generalization than other datasets for low-light enhancement in the real-world dark driving environment.

TABLE IV: **Cross-dataset Enhancement.** Pre-trained on LOL-v2 or DarkDriving using SNR-Aware [18], Tested on the nighttime validation set of nuScenes.

Pre-trained Dataset	Non-reference Metrics			
	MUSIQ \uparrow	NIQE \downarrow	HyperIQA \uparrow	CNNIQA \uparrow
LOL-v2-real	28.23	4.63	0.30	0.51
LOL-v2-synthetic	28.45	4.44	0.27	0.54
DarkDriving	35.83	4.41	0.31	0.53



Fig. 7: **Visual Cross-dataset Enhancement on nuScenes.**

C. Low-light Enhancement for 2D Detection

Experimental Setup. We adopt YOLOv11 [30] as the 2D detection backbone and consider two configurations: (1) a model pretrained on the COCO dataset [31] without further adaptation, and (2) a variant finetuned on our DarkDriving daytime training set. Low-light enhancement methods are categorized into two groups: unsupervised and supervised. The raw nighttime image (“Night”) is used as the baseline. Evaluation is conducted on the dominant Car class following the YOLOv11 protocol, with AP₅₀ and AP₅₀₋₉₀ reported as the detection performance metrics.

TABLE V: **2D Detection on Darkdriving nighttime test set.** *: *unsupervised* enhancement approaches. Other enhancement methods: *supervised* and trained on the DarkDriving training set.

Method	YOLOv11-pretrained		YOLOv11-finetune	
	AP ₅₀ \uparrow	AP ₅₀₋₉₀ \uparrow	AP ₅₀ \uparrow	AP ₅₀₋₉₀ \uparrow
Night	0.316	0.178	0.166	0.083
CLIP-LIT* [32]	0.309	0.174	0.183	0.093
EnlightenGAN* [33]	0.291	0.165	0.093	0.048
LLFormer [20]	0.161	0.090	0.167	0.071
Retinexformer [21]	0.173	0.090	0.215	0.105
LightTheNight [11]	0.227	0.093	0.141	0.053
SNR-Aware [18]	0.363	0.187	0.550	0.265

Experimental Results and Findings. Table V shows that with the COCO-pretrained detector, only SNR-Aware improves nighttime detection, consistent with Table III where it achieves the best image quality. DarkDriving daytime data-finetuned detector does not perform well on the DarkDriving nighttime data (0.166 AP₅₀), while DarkDriving-trained SNR-Aware enhancement restores performance to



Fig. 8: **2D detection errors on nighttime images can be reduced after low-light enhancement.** (a) and (d) are night-day paired data of DarkDriving dataset.

0.550 AP₅₀. Figure 8 further confirms that missing detection in dark can be found after enhancement. These results demonstrate that paired enhancement on DarkDriving is effective in bridging the day–night gap for robust 2D detection.

D. Low-light Enhancement for 3D Detection

Experimental Setup. We train the supervised SNR-Aware model on the paired DarkDriving dataset and then evaluate it on the nuScenes 3D detection benchmark. Our experimental setup and comparison methods follow the protocol in LightTheNight [11]. During testing, we applied low-light enhancement exclusively to the front camera images of the nuScenes nighttime validation set, while keeping the other five camera views unchanged. The evaluation metrics are Average Precision (AP), Average Translation Error (ATE), and Average Orientation Error (AOE).

TABLE VI: **3D Detection on nuScenes nighttime validation set.** BEVDepth (RGB-Only) [34] and CRN (RGB+Radar) [35] are trained on nuScenes daytime training set. *: *unsupervised* enhancement approaches. LightTheNight: trained on nuScenes. SNR-Aware: pre-trained on DarkDriving training set (same as that in Table IV).

Method	BEVDepth (RGB)			CRN (RGB+Radar)		
	AP \uparrow	ATE \downarrow	AOE \downarrow	AP \uparrow	ATE \downarrow	AOE \downarrow
Night	0.134	0.787	0.957	0.435	0.474	0.597
CLIP-LIT* [32]	0.131	0.791	0.972	0.430	0.475	0.607
EnlightenGAN* [33]	0.138	0.786	0.948	0.436	0.473	0.585
LightTheNight [11]	0.176	0.774	1.108	0.485	0.478	0.464
SNR-Aware [18]	0.148	0.786	0.911	0.481	0.468	0.575

Experimental Results and Findings. As shown in Table VI, compared to directly using nighttime images, the low-light enhancement model SNR-Aware simply pre-trained with our DarkDriving dataset could improve 3D detection on nuScenes and get comparable performance with the state-of-the-art method LightTheNight [11] intensively trained on nuScenes, which indicates that our DarkDriving can be generalized to promote detection in other low-light driving environment, such as nuScenes.

V. CONCLUSION

This paper introduces the first real-world benchmark dataset, DarkDriving, to advance research on low-light enhancement for autonomous driving in nighttime. By designing a new Trajectory Tracking based Pose Matching (TTPM) method for the large closed driving test field, the day-night image pairs with precise location and spatial-content alignments (error: centimeters) are collected automatically. Using DarkDriving, we study four perception tasks: low-light enhancement, generalized low-light enhancement, and low-light enhancement for 2D and 3D detection. Our DarkDriving work established a comprehensive benchmark of evaluating low-light enhancement for autonomous driving, and it can be generalized to benefit low-light enhancement and enhance detection on other datasets, such as nuScenes.

REFERENCES

- [1] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv:1808.04560*, 2018.
- [2] C. Sakaridis, D. Dai, and L. Van Gool, "Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3139–3153, 2020.
- [3] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.
- [4] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.
- [5] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2636–2645.
- [6] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 621–11 631.
- [7] X. Tan, K. Xu, Y. Cao, Y. Zhang, L. Ma, and R. W. Lau, "Night-time scene parsing with a large real dataset," *IEEE Transactions on Image Processing*, vol. 30, pp. 9085–9098, 2021.
- [8] T. Sun, M. Segu, J. Postels, Y. Wang, L. Van Gool, B. Schiele, F. Tombari, and F. Yu, "Shift: a synthetic driving dataset for continuous multi-task domain adaptation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 371–21 382.
- [9] I. Ashraf, S. Hur, M. Shafiq, and Y. Park, "Catastrophic factors involved in road accidents: Underlying causes and descriptive analysis," *PLoS one*, vol. 14, no. 10, p. e0223473, 2019.
- [10] National Transportation Safety Board, "Inadequate safety culture contributed to uber automated test vehicle crash," <https://www.nts.gov/news/press-releases/Pages/NR20191119c.aspx>, 2019.
- [11] J. Li, B. Li, Z. Tu, X. Liu, Q. Guo, F. Juefei-Xu, R. Xu, and H. Yu, "Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 205–15 215.
- [12] M. Afifi, K. G. Derpanis, B. Ommer, and M. S. Brown, "Learning multi-scale photo exposure correction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9157–9167.
- [13] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2011, pp. 97–104.
- [14] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6849–6857.
- [15] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping," in *IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2020, pp. 5135–5142.
- [16] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, 2003, pp. 2743–2748.
- [17] S. Macenski, S. Singh, F. Martín, and J. Ginés, "Regulated pure pursuit for robot path tracking," *Autonomous Robots*, vol. 47, no. 6, pp. 685–694, 2023.
- [18] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 714–17 724.
- [19] Y. Wu, C. Pan, G. Wang, Y. Yang, J. Wei, C. Li, and H. T. Shen, "Learning semantic-aware knowledge guidance for low-light image enhancement," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1662–1671.
- [20] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 2654–2662.
- [21] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 504–12 513.
- [22] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3836–3847.
- [23] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2022, pp. 10 684–10 695.
- [24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [25] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.
- [26] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.
- [27] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.
- [28] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3667–3676.
- [29] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.
- [30] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.
- [31] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755.
- [32] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8094–8103.
- [33] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
- [34] Y. Li, Z. Ge, G. Yu, J. Yang, Z. Wang, Y. Shi, J. Sun, and Z. Li, "Bevdepth: Acquisition of reliable depth for multi-view 3d object detection," in *AAAI conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 1477–1485.
- [35] Y. Kim, J. Shin, S. Kim, I.-J. Lee, J. W. Choi, and D. Kum, "Crn: Camera radar net for accurate, robust, efficient 3d perception," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 615–17 626.