

LED: Light Enhanced Depth Estimation at Night

Simon de Moreau^{1,2}

<https://simon.demoreau.fr/>

Yasser Almehio²

yasser.almehio@valeo.com

Andrei Bursuc³

<https://abursuc.github.io/>

Hafid El-Idrissi²

hafid.el-idrissi@valeo.com

Bogdan Stanciulescu¹

bogdan.stanciulescu@minesparis.psl.eu

Fabien Moutarde¹

<https://people.minesparis.psl.eu/fabien.moutarde/>

¹ Mines Paris - PSL University

Paris, France

² Valeo

Paris, France

³ Valeo AI

Paris, France

Abstract

Nighttime camera-based depth estimation is a highly challenging task, especially for autonomous driving applications, where accurate depth perception is essential for ensuring safe navigation. Models trained on daytime data often fail in the absence of precise but costly LiDAR. Even vision foundation models trained on large amounts of data are unreliable in low-light conditions. In this work, we aim to improve the reliability of perception systems at night time. To this end, we introduce Light Enhanced Depth (LED), a novel, cost-effective approach that significantly improves depth estimation in low-light environments by harnessing a pattern projected by high definition headlights available in modern vehicles. LED leads to significant performance boosts across multiple depth-estimation architectures (encoder-decoder, Adabins, DepthFormer, Depth Anything V2) both on synthetic and real datasets. Furthermore, increased performances beyond illuminated areas reveal a holistic enhancement in scene understanding. Finally, we release the Nighttime Synthetic Drive Dataset, a synthetic and photo-realistic nighttime dataset, which comprises 49,990 comprehensively annotated images. To facilitate further research, both synthetic dataset and code are publicly available at <https://simondemoreau.github.io/LED/>.

1 Introduction

Adverse conditions, such as harsh weather or nighttime, pose significant challenges to many computer vision applications. Despite impressive progress in perception systems for autonomous driving, enabled by powerful deep neural architectures and training techniques, the challenges of nighttime navigation persist. Accurate depth estimation is a crucial aspect

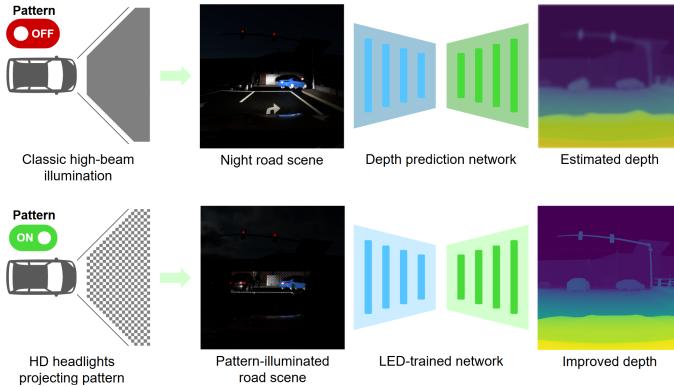


Figure 1: **Light Enhanced Depth (LED)** is a novel method that harnesses high-definition headlights’ pattern projected onto the scene to enhance nighttime depth estimation from RGB images. We release a synthetic nighttime dataset with high beam and pattern-illuminated images, along with comprehensive ground truth annotations, to advance research in nighttime perception.

of perception, profoundly impacting overall scene perception and comprehension, especially at night [8]. While LiDAR sensors offer high accuracy, their widespread adoption is impeded by their substantial cost. Additionally, in humid weather, LiDAR effectiveness can be significantly degraded due to beam reflection and increased noise levels [6, 7].

Cameras are powerful, cost-effective sensors that capture rich information about the environment and are deployed in most modern vehicles. Camera-based perception solutions display high reliability and accuracy [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], but mostly on clear daylight conditions: they struggle under distribution shift and low-light conditions. While foundation models [8, 9, 10, 11] show great improvement for depth estimation on multiple domains, they are still not robust to nighttime images that are long-tail in the training data distribution. The availability of labeled data necessary for training high-capacity deep neural networks is an additional challenge. For the task of depth estimation, some methods rely on supervised learning [1, 2, 3, 12], but most use self-supervised methods [8, 13, 14, 15, 16, 17, 18, 19, 20] since obtaining ground truth depth information is expensive. While dedicated approaches for specific conditions have emerged in recent years [21, 22, 23, 24, 25], they still only partially mitigate these issues. To the best of our knowledge, they remain limited to self-supervised strategies, due to the lack of publicly available, large-scale annotated datasets suitable for nighttime depth estimation. To address this gap, this paper releases a synthetic nighttime dataset annotated with dense depth maps, along with additional labels.

In this paper, we introduce Light Enhanced Depth (LED), a novel approach that significantly improves depth estimation in low-light environments, ensuring enhanced accuracy and reliability for autonomous vehicles. High-Definition (HD) headlights, commonly found in modern vehicles, have shown promising results in scene perception research [26, 27, 28]. Drawing inspiration from active stereovision [1, 2, 29, 30], we harness HD headlights to project a pattern into the scene, guiding the network and thereby improving performance.

Our contributions can be summarized as follows:

- **Architecture-agnostic enhancement:** LED can be applied to any depth estimation architecture to improve nighttime performances (RMSE: -11% on encoder-decoder, -24.06% on Adabins [31], -8.00% on DepthFormer [26], -15.5% on Depth Anything V2 [28]).

- **Data-efficiency:** LED-trained models outperform the others with only 20% of training data.
- **Real prototype:** LED demonstrates great performances improvements on our in-house dataset, collected using a real car-mounted prototype.
- **Dataset:** We provide the Nighttime Synthetic Drive Dataset, a photorealistic synthetic dataset, comprising 49,995 comprehensively annotated nighttime images to foster future investigations in nighttime perception.

2 Related Work

2.1 Depth Estimation from Camera

Supervised Learning. Recent advances in supervised depth prediction leverage transformer architectures and attention mechanisms [1, 2, 3, 4] to improve performances. The use of bins [1, 2] improves depth prediction accuracy even with limited datasets available [5, 6, 7, 8] compared to self-supervised methods [9, 10, 11, 12, 13, 14, 15]. However, these methods are confined to daytime scenarios. Nighttime depth estimation approaches [16, 17] rely on self-supervised methods, due to the lack of large, annotated nighttime datasets. In response, we release the Nighttime Synthetic Drive Dataset to support the development of nighttime-specific methods.

Vision Foundation Models. Depth foundation models [18, 19, 20, 21] have achieved impressive generalization by leveraging extensive synthetic and real-world training data. However, we show in section 5.4 that these models perform poorly in zero-shot evaluations on our nighttime dataset. It highlights the need for models specialized in nighttime conditions and underscores the benefits of our method.

Nighttime Domain Adaptation. Nighttime depth estimation can be approached as a domain adaptation problem [22], focusing on aligning features between daytime and nighttime images [23, 24, 25, 26]. Our method specifically adapts to nighttime scenes features by leveraging the informative deformations of the light pattern emitted by a vehicle’s HD headlamp.

2.2 Light for Perception

Active Stereovision. Depth estimation based on active stereovision involves disparity measurement. Unlike traditional stereovision, disparity is computed between an image and a pattern projected onto the scene. Recent methods [27, 28, 29, 30] use deep learning models designed to take patterns and images as input. While active stereovision offers valuable scale information, it has demonstrated significant performance degradation when deployed outdoors, due to high ambient lighting and the low power of projectors [27, 28, 30]. We avoid this thanks to our nighttime environment and the utilization of high-powered HD headlights.

HD Lighting. A series of works [31, 32] highlight HD lighting potential applications and particularly its ability to design anti-glare systems while optimizing illumination for enhanced driver visibility, even in adverse conditions such as rain or snow. More recently, [33] propose reducing overall scene illumination to decrease power consumption while maintaining object detection capabilities. [34, 35, 36] also have proposed hardware-in-the-loop simulation for HD headlights. We introduce a novel application of HD headlights to improve understanding of overall scene geometry and enhance depth estimation. We achieve this by projecting a pattern onto the nighttime scene, providing guidance for the model.

3 Method

We improve depth estimation from monocular camera at nighttime by leveraging a pattern projected in front of the vehicle with HD headlights.

3.1 HD Pattern and Headlights

Networks designed for active stereovision exploit the disparity between the projected pattern and the camera’s view to estimate distances within the scene. With LED, a model identifies pattern’s areas that deviate from the implicitly learned reference pattern.

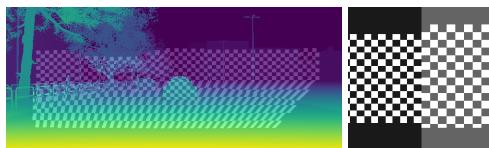


Figure 2: Pattern deformation in the scene: Left: For explanatory purpose, we project the pattern on a depth image. It demonstrates trapezoidal deformations on horizontal surfaces and undistorted squares on vertical surfaces. Complex deformations on the car reveal insights into its geometry. Right: Pattern projected on a wall at 10 m (darker) and 100 m (lighter). Square sizes increases with distance.

while vertical surfaces parallel to the image plane cause no deformation (see fig. 2). Non-planar surfaces, such as cars, result in more complex distortions. These deformations provide crucial geometric cues to the network, improving depth estimation accuracy. Additionally, the emitted light is not collimated: as the projection extends further, the light becomes more divergent, causing pixels to appear larger (see fig. 2), thus providing useful depth cues to the network.

3.2 Architectures

Our method is architecture-agnostic. By relying on a single pattern, the model can implicitly learn it, eliminating the need for a dedicated network architecture. To demonstrate the benefits of projecting an HD pattern for depth estimation, we leverage a simple encoder-decoder model [56]. In spite of its straight-forward architecture, it achieves results on par with or superior to other state-of-the-art (SOTA) methods (see section 5.1). We also successfully apply our method to more complex SOTA architectures, such as DepthFormer [27] and Adabins [1], with similar performance gains (see section 5.1). Finally, we show that finetuning Depth Anything V2 [48] with LED, on a reduced amount of training samples, enable robust and precise depth estimation.

3.3 Implementation Details

We implement the encoder-decoder in PyTorch [33]. The model is trained from scratch during 70 epochs using the AdamW optimizer [22], with a batch size of 32 and a learning rate

Unlike traditional active stereovision, which relies on infrared lasers to project points, we harness an HD headlight. Similar to a projector, this headlight can dynamically project any image or shape onto the scene. To prevent pattern overlap, we use only the left headlight.

We employ a regular checkerboard pattern, because of its high contrast and sharp discontinuities. The dense concentration of corners and transitions makes it easily detectable by convolutional neural networks [8]. Upon projection, the checkerboard is deformed according to surface shapes (from the camera’s perspective). Horizontal planes stretch the pattern into a trapezoid,



Figure 3: Nighttime Synthetic Drive Dataset examples: (A) depicts HD pattern and (B) high-beam illumination. Ground truth annotations include dense depth maps, semantic segmentation, instance segmentation labels and bounding boxes.

of 10^{-3} . Input and output resolutions are set to 320×320 px. Similar to [20], our learning objective combines losses based on Log L1, gradient and normals. More information is available in the supplementary. The selection of the best epoch is based on the Root Mean Square Error (RMSE). The training process takes 4 hours on a single Nvidia RTX 4090.

We implement Adabins [2] using the official code and rely on the toolbox [25] for Depth-Former [26]. We apply Depth Anything V2 [28] official fine-tuning code to adapt the model for metric depth estimation. The only changes are limited to dataset format compatibility. The training setup follows the original papers' recommendations, excluding data augmentation to preserve the HD pattern from cropping.

4 Dataset

4.1 Nighttime Synthetic Drive Dataset (NSDD)

Due to the lack of public datasets containing HD pattern illumination, we create the Nighttime Synthetic Drive Dataset (NSDD) using the Nvidia Drive Sim (Drop 15) simulator [32], which generates road images with photorealism effort. To simulate realistic headlight projections, we adapt the vehicle's headlights based on photometric measurements. The dataset includes 24,995 images with pattern and 24,995 images with high beam illumination (see fig. 3). We simulate a real HD headlight with a resolution of 132×28 px and a field of view of $35^\circ \times 7^\circ$. The checkerboard cells are 0.5° , ensuring visibility with the camera resolution.

We choose High Beam (HB) as the comparison baseline, considering it the maximum normal lighting condition. Since high beam illumination exceeds that of the checkerboard pattern, performance improvements are attributed solely to the pattern's contribution. In each frame, vehicles, pedestrians, and traffic signs are randomized. To simulate light interference, we include other car headlights and randomize ambient light levels between 0 and 10 lux.

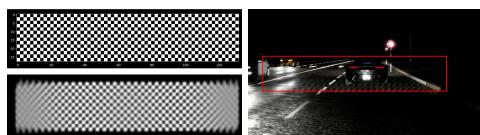


Figure 4: Simulation of HD pattern: Top-left: control matrix of the HD headlight. Bottom-left: photometry considering aberrations created by the headlight lens. Right: Resulting image using the photometry. The area outlined in red is the region of interest.

The dataset consists of 5 different maps (3 train / 1 val / 1 test), with 4,999 frames generated for both pattern and HB illuminated images in each. These maps represent major cities. Both parts of the dataset share the same randomization code, making the domains comparable.

To closely replicate reality, we account for the optical imperfections of headlights using real measured photometry data (see fig. 4). Aberrations observed in the pattern, particularly along the edges of the projec-

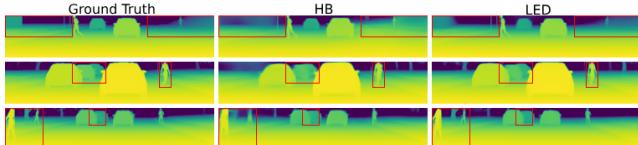


Figure 5: **Qualitative results on NSDD.** LED results exhibit higher precision, and more accurate object boundaries and shapes compared to HB. Red boxes indicate enhanced regions.

Pattern	RMSE \downarrow	Abs Rel \downarrow	Log ₁₀ \downarrow	RMSE Log \downarrow	SILog \downarrow	Sq Rel \downarrow	$\delta^1 \uparrow$	$\delta^2 \uparrow$	$\delta^3 \uparrow$
<i>Encoder-decoder</i>									
HB	6.1204	0.0903	0.0233	0.1353	13.4847	3.3579	0.9489	0.9812	0.9908
LED	5.4259	0.1996	0.0188	0.1253	12.4900	16.1224	0.9603	0.9846	0.9927
<i>Adabins</i>									
HB	7.3520	0.1360	0.0320	0.1230	10.1880	4.2770	0.8790	0.9440	0.9740
LED	5.5830	0.0690	0.0240	0.0920	7.7290	2.5090	0.9210	0.9680	0.9840
<i>DepthFormer</i>									
HB	5.8528	0.0687	0.0281	0.1205	11.3841	0.8328	0.9432	0.9802	0.9913
LED	5.3845	0.1215	0.0276	0.1111	10.4146	3.4015	0.9497	0.9838	0.9950

Table 1: **Comparison of depth estimation performances on NSDD:** HB models are trained on high beam data and LED on HD light pattern. LED models outperform HB models across all metrics, exceptions are for Abs Rel and Sq Rel. Metrics are computed in the ROI.

tion, are due to these imperfections. All images have a resolution of 1920×1080 px, and we provide depth maps, annotations for 2D/3D object detection, normal estimation, and both semantic and instance segmentation. The provided depth values are not limited, though they are clipped at 100 m for all experiments. During training, a 640×640 px square is initially center-cropped and then resized to 320×320 px. To ensure reproducibility, code and synthetic dataset are publicly available. Releasing large-scale nighttime data, with precise and comprehensive annotations, will foster research in nighttime computer vision for autonomous driving.

4.2 Real-world Dataset

We validate LED on a real-world, in-house dataset collected using a car-mounted prototype. It comprises 50,000 images (70% train / 15% val / 15% test) from populated urban and rural roads, evenly split between Low Beam (LB) illumination and checkerboard pattern with 0.5° , 0.25° and 0.125° cells’ dimension. Ground truth is obtained from LiDAR data using the DOC-Depth method [42] and Exwayz software [43]. More details are available in supplementary materials.

5 Experiments

We evaluate our method through extensive experiments on the Nighttime Synthetic Drive Dataset. We validate the contribution of our light pattern in boosting performance, both inside and outside illuminated area. We apply our method to the encoder-decoder and other SOTA approaches: Adabins [44], DepthFormer [45] and Depth Anything V2 [46], showing its agnosticity. Finally, we evaluate LED robustness beyond its training domain and its capabilities on real-world scenarios. Metrics used align with prior works [42, 46, 47, 48, 49, 50].

Model	RMSE \downarrow	Abs Rel \downarrow	Log ₁₀ \downarrow	RMSE Log \downarrow	SILog \downarrow	Sq Rel \downarrow	$\delta^1 \uparrow$	$\delta^2 \uparrow$	$\delta^3 \uparrow$
(HB) O-ROI	9.2533	0.1598	0.0249	0.2047	20.4447	11.4068	0.9385	0.9714	0.9828
(LED) O-ROI	9.0070	0.1295	0.0221	0.2027	20.2348	8.7407	0.9427	0.9723	0.9828
(HB) Full	8.9702	0.1521	0.0247	0.1988	19.8572	10.5225	0.9396	0.9725	0.9836
(LED) Full	8.6963	0.1371	0.0217	0.1965	19.6137	9.5517	0.9446	0.9737	0.9839

Table 2: **Encoder-decoder performance beyond ROI on NSDD:** O-ROI stands for Outside ROI, where the evaluation mask is the inverse of the ROI. For Full, metrics are computed on the entire image. (LED) models trained on HD pattern outperform (HB) ones in all metrics.

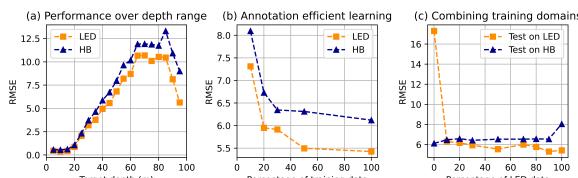


Figure 6: **Performance comparison of LED and HB in various settings and metrics on NSDD.** (a) Performance over distance; (b) Performance over training set size; (c) Robustness across domains: The amount of training data is fixed and composed of a given ratio of LED and HB data. Overall, LED achieves better long-range results while being more data-efficient. It also demonstrates improved performance across domains.

5.1 HD Pattern Impact

We assess LED impact on depth estimation by comparing results with and without pattern. As reference, we take models trained on high beam (HB), while the variants use our HD pattern (LED). Quantitative and qualitative results are presented in table 1 and fig. 5.

Quantitative Results. We evaluate our method on the encoder-decoder and two SOTA architectures commonly used in daytime scenarios, Adabins [2] and DepthFormer [26]. Their use of transformer architectures offers valuable insights into LED behavior on more recent architectures compared to the convolutional ones that are common in embedded devices.

Comparing the encoder-decoder LED with HB (table 1), we observe a significant improvement in metrics within the region-of-interest (ROI): LED yields -11% RMSE. Using LED, the model provides more accurate distance estimations. However, its errors are often related to pattern occlusions, which affect both near and far objects, resulting in a degradation of relative metrics (Abs Rel, Sq Rel). All metrics assessing global precision are improved.

Compared to the SOTA approaches [2, 26], LED-trained encoder-decoder either outperforms or matches the performance of their HB models. Given that the encoder-decoder is less tailored for depth estimation than more intricate SOTA designs, this result underscores the substantial improvements yield by LED, demonstrating that using the HD informative pattern even with a straightforward architecture is promising for challenging nighttime scenes.

With LED, Adabins demonstrates a substantial improvement compared to the HB model: -24.06% RMSE and -6.70% Abs Rel. Conversely, DepthFormer showcases -8.00% in RMSE but +5.28% in Abs Rel. Note that the most significant enhancement is observed in Adabins, which have the worst results by night on HB.

Improvements seen across diverse architectures confirm LED agnosticity to architectures. It implies potential effectiveness with future methods.

Performance Over Distance. Limited nighttime visibility makes depth estimation of distant objects challenging. In fig. 6 (a), we show the performance of the encoder-decoder over distances. LED maintains performance for close object and exhibits greater enhancement at longer ranges, thus addressing depth estimation of distant objects at night.

Domain	RMSE ↓	Abs Rel ↓	Log ₁₀ ↓	RMSE Log ↓	SILog ↓	Sq Rel ↓	δ ¹ ↑	δ ² ↑	δ ³ ↑
(HB → HB)	6.1204	0.0903	0.0233	0.1353	13.4847	3.3579	0.9489	0.9812	0.9908
(HB → LED)	17.3113	0.4479	0.1955	0.6817	52.5132	20.0263	0.4699	0.6652	0.7658
(LED → LED)	5.4259	0.1996	0.0188	0.1253	12.4900	16.1224	0.9603	0.9846	0.9927
(LED → HB)	8.0537	0.1511	0.0338	0.2026	20.1120	8.1999	0.9158	0.9626	0.9796

Table 4: **Encoder-decoder performance across domains of NSDD.** We denote (training domain → testing domain). The model trained on (HB) fails entirely when tested on (LED). Yet, (LED)-trained models are able to estimate depth when tested on the (HB) domain.

Pattern	# Images	RMSE ↓	Abs Rel ↓	Log ₁₀ ↓	RMSE Log ↓	SILog ↓	Sq Rel ↓	δ ¹ ↑	δ ² ↑	δ ³ ↑
HB	Zero-shot	19.058	0.430	0.243	0.587	27.004	8.134	0.086	0.186	0.785
LED	18.023	0.420	0.232	0.563	25.500	7.779	0.096	0.214	0.867	
HB	100	6.499	0.075	0.033	0.133	12.894	0.992	0.929	0.977	0.991
LED		6.198	0.073	0.032	0.131	12.182	0.898	0.933	0.977	0.991
HB	500	5.447	0.058	0.025	0.112	10.629	0.688	0.950	0.985	0.995
LED		5.225	0.057	0.025	0.108	10.038	0.648	0.953	0.986	0.995
HB	1000	5.124	0.058	0.026	0.107	10.038	0.607	0.953	0.985	0.995
LED		4.328	0.044	0.020	0.089	8.374	0.430	0.962	0.991	0.997

Table 5: **Depth Anything V2 performances on NSDD.** The model fails when used as zero-shot, but it correctly estimates depth with few-shot learning. LED always outperforms HB.

Pattern	RMSE↓	Abs Rel↓	SILog↓
LED	5.497	0.209	12.635
HB	6.443	0.095	13.723
HL	6.668	0.068	14.308
VL	7.360	0.172	15.774

Table 3: **Performances of the encoder-decoder using various patterns on NSDD.** VL and HL refers to vertical and horizontal lines. Results are obtained from 50% of NSDD training data.

Pattern Study. We generated additional data with various illumination patterns to compare the performance of the LED checkerboard against other common structured light patterns (horizontal and vertical lines). Table 3 shows that the checkerboard pattern outperforms the others across most metrics. While horizontal lines show slight improvement in relative metrics, they cause a significant decline in others.

5.2 Improving Global Scene Understanding

We investigate the impact of LED on depth estimation beyond the ROI, highlighting its contribution to a more holistic scene understanding. Results are presented in table 2.

Examining depth estimation both outside the ROI and across the entire image, we observe a consistent performance improvement: -2.66% RMSE outside the ROI and -3.05% RMSE on the entire image. We also report enhancement in relative metrics: Abs Rel (resp. -3.03% and -1.50%) and Sq Rel (resp. -23.27% and -9.23%). These results suggest that the HD headlight pattern provides valuable information, such as object size and scale, leading to enhanced overall scene understanding and more accurate depth estimation.

5.3 Annotation-efficient Learning

Collecting annotated nighttime data is challenging and costly. We explore whether pattern extra guidance can reduce the need for training data (refer to fig. 6 (b)). Our results show that encoder-decoder model trained with less than 20% of the LED data outperforms one trained on the full HB dataset, reaching near-peak performance with just 50%. In addition, we observe that incorporating just 10% of LED data into the HB training set enable the network to learn relevant features and enhance its performances when used with pattern (see fig. 6 (c)). Therefore, any vehicle equipped with HD headlights can apply LED by adding only a few pattern images in their training set, thereby reducing the cost of specialized data.

Pattern	RMSE ↓	Abs Rel ↓	Log ₁₀ ↓	RMSE Log ↓	SILog ↓	Sq Rel ↓	δ ¹ ↑	δ ² ↑	δ ³ ↑
<i>Encoder-decoder</i>									
LB	14.199	0.187	0.072	0.292	28.938	5.620	0.772	0.909	0.956
LED 0.5°	11.089	0.119	0.050	0.213	21.251	2.304	0.864	0.950	0.977
LED 0.25°	8.695	0.109	0.040	0.190	18.635	2.128	0.899	0.963	0.984
LED 0.125°	10.154	0.096	0.040	0.190	18.862	2.132	0.907	0.969	0.984
<i>DepthFormer</i>									
LB	8.777	0.138	0.050	0.186	17.170	3.365	0.864	0.957	0.983
LED 0.5°	6.810	0.101	0.041	0.152	14.380	1.376	0.890	0.969	0.989
LED 0.25°	5.621	0.082	0.030	0.126	11.735	1.163	0.923	0.978	0.992
LED 0.125°	5.727	0.076	0.030	0.116	10.754	1.261	0.940	0.987	0.995
<i>DepthAnythingV2 - 1000 training images</i>									
LB	7.637	0.104	0.044	0.160	14.617	1.401	0.891	0.972	0.990
LED 0.5°	7.017	0.109	0.047	0.161	14.044	1.260	0.877	0.9697	0.989
LED 0.25°	6.304	0.102	0.044	0.146	12.300	1.079	0.885	0.973	0.993
LED 0.125°	5.785	0.082	0.037	0.126	10.673	0.840	0.921	0.981	0.992

Table 6: **Performance comparison on real-world data using various pattern resolutions.** LED models outperform LB on all metrics, showing LED suitability on real-world scenarios.

5.4 Few-shot Learning

Foundation models show strong zero-shot performance across various cameras and content types. However, when tested on NSDD, Depth Anything V2 [48] performs poorly, with an RMSE of ~ 19 m (see table 5), highlighting the complexity of nighttime data and the need for specialized models. It performs slightly better on LED images, possibly due to its ability to leverage the pattern’s geometric cues. Fine-tuning with as little as 100 images is sufficient for adapting the network to our distribution and produce consistent depth predictions on both LED and HB. LED-models demonstrate significant improvements across all metrics (-15.5% RMSE with 1,000 images). This underlines the effectiveness of our method in enhancing nighttime depth estimation even on foundation models. Results on our real-world dataset are provided in the supplementary material.

5.5 Robustness Across Domains

We assess model robustness when operating beyond the original training domain by testing a HB-trained model on pattern data and *vice versa*. We report results in table 4. The HB-trained model fails significantly on pattern data (+182.85% RMSE), while the LED-trained model performs reasonably on HB, with a +32.63% RMSE. The LED→HB performance is ~ 8 m RMSE, compared to ~ 17 m RMSE for HB→LED. Furthermore, we show that incorporating only 10% of HB data reduces performance drop (see fig. 6 (c)), limiting RMSE increase to just +17.11%. It enables a single network to perform well across both domains. This adaptability broadens the method’s applications, allowing selective use of the pattern for higher precision, its absence to avoid glare, or focused pattern projection on specific regions, such as objects of interest

5.6 Real-world Scenarios

We show the applicability of our method in real world scenarios using our in-house dataset. We report results in table 6. LED significantly boosts performance over LB across all metrics. DepthFormer (0.5°) shows -22.4% RMSE improvement. The qualitative results in fig. 7 show LED’s robustness under interfering light sources, *e.g.*, car headlights, street lights.

To account for objects passing through the pattern illumination, we investigate the impact of smaller checkerboard cells. Using a size of 0.25°, we observe great enhancement across all metrics (-17.4% RMSE, -18.8% Abs Rel against 0.5°). Further increasing the resolution to 0.125° does not lead to significant improvements.



Figure 7: Real-world qualitative results. DepthFormer LED produces accurate depth on complex scenes, even in presence of interfering lights, *e.g.*, car headlights, streetlights.

In contrast, Depth Anything V2 achieves its best performance with a 0.125° resolution, as it is better suited to capturing fine details. When fine-tuned with only 1,000 images, we observe -24.3% RMSE compared to the LB baseline. Performances shown by DepthFormer and Depth Anything V2 underpin the benefits of LED in complex real-world scenarios.

6 Limitations and Future Works

The method’s reliance on a single reference pattern ensures agnosticism but limits its flexibility. Designing specialized architectures that treat the pattern as an input could enable dynamic pattern optimization and improve generalization.

We tested varying projector-camera distances (real: 70 cm, NSDD: 150 cm), showing LED robustness on similar car models. Future work will assess its usage with larger variations.

LED patterns may cause glare for other road users. Leveraging [41] to mask them could mitigate this. Future work should assess both safety concerns and the impact on performance.

7 Conclusion

We introduce LED, a method for enhancing nighttime depth estimation by leveraging high-definition light patterns projected by modern vehicle headlights. Through extensive experiments, we demonstrate significant improvements in depth perception, both within and beyond illuminated areas. Our method’s versatility is highlighted by its successful integration with two state-of-the-art architectures, Adabins and DepthFormer, as well as the foundation model Depth Anything V2. Moreover, LED shows promising real-world performance. We also release the Nighttime Synthetic Drive Dataset, comprising 49,995 fully annotated images. We hope it will serve as a valuable resource for the research community, supporting exploration of various nighttime perception tasks.

Acknowledgment

We would like to thank Amandine Brunetto for her help and insights. We also thank Naimeric Villafruela from the Fablab of Mines Paris for his assistance in creating the real-world prototype. In addition, we extend our warm thanks to the Exwayz team for the excellent collaboration on the real-world ground truth processing. This work was granted access to the HPC resources of IDRIS under the allocation 2025-AD011015334 made by GENCI.

References

- [1] Ashutosh Agarwal and Chetan Arora. Attention attention everywhere: Monocular depth prediction with skip attention. In *WACV*, 2023.

- [2] Seung-Hwan Baek and Felix Heide. Polka Lines: Learning Structured Illumination and Reconstruction for Active Stereo. In *CVPR*, 2021.
- [3] Nicholas Baker, Hongjing Lu, Gennady Erlikhman, and Philip J Kellman. Local features and global shape information in object classification by deep convolutional neural networks. *VR*, 2020.
- [4] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. Adabins: Depth estimation using adaptive bins. In *CVPR*, 2021.
- [5] Jiawang Bian, Zhichao Li, Naiyan Wang, Huangying Zhan, Chunhua Shen, Ming-Ming Cheng, and Ian Reid. Unsupervised scale-consistent depth and ego-motion learning from monocular video. In *NeurIPS*, 2019.
- [6] Mario Bijelic, Tobias Gruber, and Werner Ritter. A benchmark for lidar sensors in fog: Is detection breaking down? In *IV*, 2018.
- [7] Mario Bijelic, Tobias Gruber, and Werner Ritter. Benchmarking image sensors under adverse weather conditions for autonomous driving. In *IV*, 2018.
- [8] Aleksei Bochkovskii, Amaël Delaunoy, Hugo Germain, Marcel Santos, Yichao Zhou, Stephan R. Richter, and Vladlen Koltun. Depth pro: Sharp monocular metric depth in less than a second. *arXiv*, 2024.
- [9] Gayan Brahmanage and Henry Leung. Outdoor RGB-D Mapping Using Intel-RealSense. In *IEEE Sensors*, 2019.
- [10] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liang, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, 2020.
- [11] Raoul De Charette, Robert Tamburo, Peter C. Barnum, Anthony Rowe, Takeo Kanade, and Srinivasa G. Narasimhan. Fast reactive control for illumination through rain and snow. In *ICCP*, 2012.
- [12] Simon de Moreau, Mathias Corsia, Hassan Bouchiba, Yasser Almehio, Andrei Bursuc, Hafid El-Idrissi, and Fabien Moutarde. Doc-depth: A novel approach for dense depth ground truth generation. *arXiv preprint arXiv:2502.02144*, 2025.
- [13] Exwayz. Exwayz 3d mapping: Create dense, accurate and georeferenced 3d point cloud using lidar slam. <https://www.exwayz.fr/>, 2024.
- [14] Sean Ryan Fanello, Christoph Rhemann, Vladimir Tankovich, Adarsh Kowdle, Sergio Orts Ecolano, David Kim, and Shahram Izadi. HyperDepth: Learning Depth from Structured Light without Matching. In *CVPR*, 2016.
- [15] Stefano Gasperini, Nils Morbitzer, HyunJun Jung, Nassir Navab, and Federico Tombari. Robust monocular depth estimation under challenging conditions. In *ICCV*, 2023.
- [16] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012.

- [17] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, 2017.
- [18] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. Digging into self-supervised monocular depth estimation. In *ICCV*, 2019.
- [19] Vitor Guzilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. 3d packing for self-supervised monocular depth estimation. In *CVPR*, 2020.
- [20] Mohit Gupta, Qi Yin, and Shree K. Nayar. Structured Light in Sunlight. In *ICCV*, 2013.
- [21] Junjie Hu, Mete Ozay, Yan Zhang, and Takayuki Okatani. Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries. In *WACV*, 2019.
- [22] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *CVPR*, 2024.
- [23] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. *arXiv preprint arXiv:1907.10326*, 2019.
- [24] Chunyu Li, Yusuke Monno, and Masatoshi Okutomi. Deep Hyperspectral-Depth Reconstruction Using Single Color-Dot Projection. In *CVPR*, 2022.
- [25] Zhenyu Li. Monocular depth estimation toolbox. <https://github.com/zhyever/Monocular-Depth-Estimation-Toolbox>, 2022.
- [26] Zhenyu Li, Zehui Chen, Xianming Liu, and Junjun Jiang. Depthformer: Exploiting long-range correlation and local information for accurate monocular depth estimation. *MIR*, 2023.
- [27] Lina Liu, Xibin Song, Mengmeng Wang, Yong Liu, and Liangjun Zhang. Self-supervised monocular depth estimation for all day images using domain separation. In *ICCV*, 2021.
- [28] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2018.
- [29] Will Maddern, Geoff Pascoe, Chris Linegar, and Paul Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *IJRR*, 2017.
- [30] Christoph Mertz, Sanjeev J. Koppal, Solomon Sia, and Srinivasa Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In *CVPR Workshops*, 2012.
- [31] Pushmeet Kohli, Nathan Silberman, Derek Hoiem, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.
- [32] NVIDIA. Nvidia drive sim. <https://www.nvidia.com/en-us/self-driving-cars/simulation/>. Accessed: 2024-01-18.

- [33] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.
- [34] Matteo Poggi, Filippo Aleotti, Fabio Tosi, and Stefano Mattoccia. Towards real-time unsupervised monocular depth estimation on cpu. In *IROS*, 2018.
- [35] Gernot Riegler, Yiyi Liao, Simon Donne, Vladlen Koltun, and Andreas Geiger. Connecting the Dots: Learning Representations for Active Monocular Depth Estimation. In *CVPR*, 2019.
- [36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [37] Chang Shu, Kun Yu, Zhixiang Duan, and Kuiyuan Yang. Feature-metric loss for self-supervised learning of depth and egomotion. In *ECCV*, 2020.
- [38] Nathan Silberman and Rob Fergus. Indoor scene segmentation using a structured light sensor. In *ICCV Workshops*, 2011.
- [39] Jaime Spencer, Richard Bowden, and Simon Hadfield. Defeat-net: General monocular depth via simultaneous unsupervised representation learning. In *CVPR*, 2020.
- [40] Robert Tamburo, Eriko Nurvitadhi, Abhishek Chugh, Mei Chen, Anthony Rowe, Takeo Kanade, and Srinivasa G. Narasimhan. Programmable Automotive Headlights. In *ECCV*, 2014.
- [41] Madhu Vankadari, Sourav Garg, Anima Majumder, Swagat Kumar, and Ardhendu Behera. Unsupervised monocular depth estimation for night-time images using adversarial domain feature adaptation. In *ECCV*, 2020.
- [42] Mirko Waldner and Torsten Bertram. Optimal Real-time Digitization of Matrix-Headlights. In *AIM*, 2021.
- [43] Mirko Waldner, Maximilian Kramer, and Torsten Bertram. Hardware-in-the-Loop-Simulation of the light distribution of automotive Matrix-LED-Headlights. In *AIM*, 2019.
- [44] Mirko Waldner, Maximilian Kramer, and Torsten Bertram. Digitization of Matrix-Headlights That Move as in the Real Test Drive. In *AIM*, 2020.
- [45] Mirko Waldner, Nathalie Müller, and Torsten Bertram. Energy-Efficient Illumination by Matrix Headlamps for Nighttime Automated Object Detection. *IJECER*, 2022.
- [46] Kun Wang, Zhenyu Zhang, Zhiqiang Yan, Xiang Li, Baobei Xu, Jun Li, and Jian Yang. Regularizing nighttime weirdness: Efficient self-supervised monocular depth estimation in the dark. In *CVPR*, 2021.

- [47] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, 2024.
- [48] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv:2406.09414*, 2024.
- [49] Zhichao Yin and Jianping Shi. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In *CVPR*, 2018.
- [50] Yupeng Zheng, Chengliang Zhong, Pengfei Li, Huan-ang Gao, Yuhang Zheng, Bu Jin, Ling Wang, Hao Zhao, Guyue Zhou, Qichao Zhang, et al. Steps: Joint self-supervised nighttime image enhancement and depth estimation. In *ICRA*, 2023.
- [51] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In *CVPR*, 2017.